Peer Review and Communication History

Ms Title: Leisure Activities as a Driver of Personality Development? A Random-Intercept Cross-lagged Panel Model Across 13 Years in Adulthood

Author Namees: Julia Sander, Paul Schumann, David Richter, Jule Specht

Submitted: Apr 14, 2020

Editor First Decision—Revise & Resubmit

June 29, 2020

Dear Julia Sander,

Thank you for submitting your work to Collabra: Psychology. I was able to secure reviews from two outstanding scholars who are experts about personality trait development. Both of these individuals are extraordinarily well qualified to review this paper and I thank them for their service to this journal. I independently read the paper and then consulted their comments.

As you will read below, the reviewers had mixed reactions to the paper. I believe they liked the sample and motivating question. I agree that these are strengths of the work. However, I think they raised concerns about the specific theorizing behind any expected associations between social participation and personality trait development, the measurement of social participation, and the analytic approach used in the paper. All of these issues then influence whether the Discussion is appropriate. These issues might be addressable, but it will take considerable revision and likely some supplemental analyses.

In short, I will extend a resubmit decision, but you should consider this a “high risk” R&R decision. I will send any revision to back the reviewers given the scope of their concerns. Thus, I cannot make any firm predictions about the ultimate fate of this work. You might decide that the current package will face an easier road to publication at another outlet.

I think both reviewers raised excellent points and I suspect both of their views are representative of the reaction of other readers. I know I agreed with their reactions. You should address each of their concerns either in the revised text or in the response letter. Addressing their 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. Both Reviewers commented on the Introduction and questioned whether it best framed the current research. Thus, I think the Introduction will need a substantial revision. I agree that greater attention to why social participation is an important variable to consider is needed as well as a strong justification of the theoretical model and how that ties to the analytic approach in the paper.
2. The Reviewers noted concerns with the analytic model. There is on-going debate about the standard cross-lagged model. I think it is important to clarify why the model used in the paper is appropriate and you should consider the CPLP-RI and correlated growth model variously mentioned by the Reviewers as alternatives. To be clear, the model used in the paper might be the one that is best suited to test your hypotheses, but that rationale should be articulated for readers.
3. I think the point raised by Reviewer #2 about age is important. Stability coefficients seem to change with age and thus age would seem to be a viable moderator of selection and socialization effects. This issue should be considered both theoretically/conceptually and analytically.
4. I think the points about the nature of the social participation variable raised by Reviewer 1 (and to some extent, Reviewer #2) are important to address. I think a stronger defense of the validity of the measure is needed and I also think it would be important to consider some specificity in terms of the nature of the activities. For example, “artistic activities” seem more relevant to Openness than other trait (Extraversion and “fun activities” seem like another intuitive pair).
5. I did not understand the theoretical rationale for controlling for income. This should be expanded in the paper.
6. I think reporting degrees of freedom and chi-square for all models in the primary text is important.

Those were the major issues that I think require attention. I think clarifying the theory motivating the work, the measurement strategy, and analytic model will strongly enhance the work. I realize this is a significant revision of the current paper, but I think this is what is required to bring the work in alignment with the expectations of readers of papers published in this outlet. I will also reiterate that the reviewers made strong points and that you should also address any of their concerns that I did not summarize in this letter. My goal was to provide “big picture” guidance about how to approach any revision in this letter.

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](mailto:editorialoffice@collabra.org).

Thank you for trusting us with your paper. I want to reiterate that this is a high-risk revision to be transparent about the nature of the task as I see it. If you have any questions or difficulties during the revision process, please do contact us. We hope you can submit your revision within the next three months. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Brent Donnellan

**Reviewer 1**

**Open response questions**

Your name will not be revealed unless you wish to sign your review. If you chose to show your name, please type it in the text field below. If you prefer your comments to be anonymous, leave the field blank.

In this study, researchers examined cross-lagged co-development between the Big Five personality traits and social participation in a large, representative sample of Germany. They found significant co-developmental effects for the traits of O, E, and N, and less consistent effects for A and C. I believe that this paper has potential to be a well-cited work of personality development. However, this initial paper has major room for refinement that I would require before I recommend this study for publication. Beyond these revisions, I also believe that it would be valuable to disambiguate the social participation items and examining them separately. This is not required, but I believe it would substantially improve this work and up its impact in the field.

Major issues:

Intro

* The first eight paragraphs of the introduction are rather meandering and go into detail about some foundational issues that can be presented much more concisely (personality is important, there are normative changes in personality across the lifespan, there are substantial individual differences in change). The major idea of this paper, that social participant may be linked to personality change, isn’t even mentioned until five paragraphs in. This part of the introduction should be shortened.
* The authors review research on specific, discrete life events (e.g. retirement, having children) that is only somewhat related to the present research, which focuses more on the gradually unfolding effect of daily experiences on personality development. I think that these references can be culled.
* On the other hand, theoretical aspects of the introduction should be expanded. In particular, the mechanisms that may underlie development are only discussed in passing (with respect to the sociogenomic model). A reader not well-versed in personality development may be rather confused as to how changes in one’s daily behaviors may affect personality. The TESSERA model (Wrzus & Roberts, 2017) lays out specific pathways that may lead personality states to congeal into personality trait change. Conceptualizing co-development in terms of each component of this framework would substantially improve the paper’s theoretical focus and provide evidence by which to evaluate a recently-proposed model of change.
* Socialization and selection effects are discussed throughout the paper but never given sufficient description or definition.

Methods

* The authors do not provide evidence for the internal reliability or factor structure of this social participation scale.
* The authors also do not provide evidence for measurement invariance of the social participation scale. Establishing weak invariance is critically important so that we know whether scale scores are comparable across measurement occasions.
* I’m not sure what happened with Table 2, but it’s there five times. This might have been caused by the manuscript upload system?
* The authors should be aware of the debate and distinction between CLPMs and Random Intercept CLPMS (Hamaker et al., 2015). Because CLPMs do not adjust for individual differences in baseline levels of social participation and personality, they conflate selection and socialization effects to some extent. I would suggest that the authors drop the CPLM analyses and only run RI-CLPMs (as they seem more germane to the inferential goals of the present study), or estimate these models in addition to the already-reported CLPMs.

Results

* Standardizing personality and social participation to a mean of 0 and a SD of 1 would allow the authors to evaluate effect sizes in absolute terms (comparing to other personality research) and in relative terms (comparing effect sizes across traits and cross-lag directionalities).

Minor issues:

* The authors frequently use the phrase “go along with,” which is rather nonscientific. Perhaps the phrases “co-occur with” or “are associated with” could be used in their place.
* Similarly, the authors frequently use the terms “supposed” and “assume” when they mean “theorized” (these terms are technically grammatically correct but have the everyday connotation of “in theory it does, but in actuality it doesn’t”).
* I think the p-value sign is flipped in the second to last paragraph of the openness results.
* The stability of social participation does not need to be re-reported in each model’s results.
* Figure 1 would be much more informative if it included parameter point estimates.
* Table 1 can include age as well.
* In the discussion, the authors mention that this study provides evidence that change in social participation was associated with change… beyond change in openness to actions. I think this should be openness to ideas, instead, as that is the NEO-PI-R facet most associated with intellect.

Finally, I would like to outline reasons why I think the authors should disambiguate the social participation activities. One of the most interesting parts of this paper is the multifaceted nature of social participation. Although the authors are justified in summing all activities into a formative index, we could learn a lot more about how social participation is associated with personality development if these activities are broken apart in additional analyses. The authors even allude to this in the introduction, when they make points about the multifaceted nature of social participation.

Benefits of separating activities:

* The introduction would be much more well-structured if different types of social participation were reviewed in turn. For example, the authors could first discuss social life and personality development, and then cultural activity and personality development, and finally civic life (voluntary and political activities) and personality development. Each of these activities may have different links to personality development, and past studies have typically examined these everyday behaviors in isolation.
* Scores on the social participation scale would be more interpretable. As Table 1 shows, these variables are not all highly associated with one another, and likely would not load highly onto a single factor. It is okay to have a formative index of behavior, where it is not assumed that a latent level of social participation causes all activities equally, but the scale is currently only interpretable in terms of “how much stuff the participant does in their free time” because specific content is averaged out.
* Analyzing these scales separately would allow the researchers to examine how particular kinds of social participation are associated with personality development, which is theoretically quite interesting. Looking at table 1, it seems that O and E are linked to change in most activities, but, if cross-sectional associations translate longitudinally, change in A may only be associated with interpersonal activities and increases in C may predict decreases in fun activities (as alluded to in the discussion). These specific links are both interesting for theory and make for a more interesting paper.
* The major drawback from separating activities is that individual items may not be highly reliable. Given the sample size, I’m not too worried about the decrease in signal that would come from lowering reliability, but extremely low test-retest correlations, for example, might preclude the authors from these extra analyses.

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. Psychological methods, 20(1), 102.

Wrzus, C., & Roberts, B. W. (2017). Processes of personality development in adulthood: The TESSERA framework. Personality and Social Psychology Review, 21(3), 253-277.

**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**

Your name will not be revealed unless you wish to sign your review. If you chose to show your name, please type it in the text field below. If you prefer your comments to be anonymous, leave the field blank.

Jenn Lodi-Smith

**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

Nov 29, 2020

Dear Brent Donnellan,

Thank you very much for your letter of June 29, 2020 concerning our manuscript titled “Co-development of Personality and Social Participation across the Lifespan”, submitted for publication in *Collabra: Psychology*.

We were very pleased to hear that both reviewers evaluated the manuscript a potentially well-cited work of personality development. We also greatly appreciate the thoughtfulness of the constructive comments that you and the two reviewers provided. And we are very grateful for the chance to revise and resubmit our work.

We have substantially revised the manuscript, now entitled *“Leisure Activities as a Driver of Personality Development? – A Random-Intercept Cross-lagged Panel Model Across 13 Years in Adulthood”*, according to your recommendations. We have carefully considered and responded in detail to each of the points made by you and the reviewers, paying particular attention to the points that you highlighted. We believe that the paper is now much stronger. In the following sections, we describe in detail how we have addressed each point.

# Your editorial comment

|  |
| --- |
| I think both reviewers raised excellent points and I suspect both of their views are representative of the reaction of other readers. I know I agreed with their reactions. You should address each of their concerns either in the revised text or in the response letter. Addressing their 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. Both Reviewers commented on the **Introduction** and questioned whether it best framed the current research. Thus, I think the Introduction will need a substantial revision. I agree that greater attention to why social participation is an important variable to consider is needed as well as a strong justification of the theoretical model and how that ties to the analytic approach in the paper. |
| **Our response:**  Thank you very much for your appreciative comments and suggestions.  We agree that we need a better framed Introduction with stronger justification of the theoretical model and the analytic approach. Thus, we substantially revised the Introduction.  As suggested, we have reconsidered our analytic approach and decided to follow the advice of Reviewer#1 to additionally disambiguate the overall participation score and examine the leisure activities separately. In line with Reviewer #1 we believe that this extension significantly improves the work and increases its potential impact to the field of personality development. Thus, the focus of the paper shifted, and we adapted the title and Introduction accordingly.  We now, arrive more quickly at the point that participation in leisure activities may be linked to personality change. We shortened the paragraphs on the importance of personality development research and state the aim of this paper right at the beginning:  “Personality has very concrete implications on many life domains such as health, mortality, relationship success, educational and occupational attainment, income, and job satisfaction (for an overview, see Soto (2019)). Because personality is also subject to change, there is great interest in understanding developmental patterns and change mechanisms of personality (Aschwanden & Allemand, 2020; Bleidorn et al. 2019).  In line with recent theories on personality change and stability (Roberts, 2018; Wagner, Orth, Bleidorn, Hopwood, & Kandler, 2020; Wrzus & Roberts, 2017, for an overview see also Specht et al, 2014), we assume a gradually unfolding effect of leisure activities on personality and vice versa. That is, changes in leisure participation, in the sense of carrying out daily activities and repetitive experiences may be a source of individual differences in personality development across the lifespan.  In the present study, we investigate the transactions of overall participation in leisure time and six specific leisure activities with change in the Big Five personality traits (John & Srivastava, 1999) over time in adulthood. For this purpose, we use data from 55,790 individuals who provide data across a period of 13 years.” (p.3)  In a following paragraph on “What drives personality development?”, we now describe recent theoretical frameworks on personality development that may explain this association in more detail (p.4). In particular, we focus on the TESSERA framework (Wrzus & Roberts, 2017) to derive hypotheses on the associations of participation in leisure activities and personality trait change.  “In terms of recent theories on personality stability and change (Roberts, 2018; Wagner, Orth, Bleidorn, Hopwood, & Kandler, 2020; Wrzus & Roberts, 2017) personality change may be stimulated through the bottom-up process of a person’s daily behaviors (leisure activities) and experiences. On the other hand, an individual’s personality traits will impact its’ situation selection, i.e. the type and frequency of leisure activities (Wrzus & Roberts, 2017).  Applying the TESSERA framework, personality traits may change if the following short-term process is repeated: A certain leisure activity, e.g. political engagement [Triggering situation] is perceived as relevant to a trait, e.g. extraversion [Expectancy] and thus elicits a relevant state, e.g. speaking in front of a crowd [State/State expression] and this state level does not correspond to the actual trait level, e.g. low extraversion, but the reaction is positive, e.g.: political talk is a success, people are convinced [Reaction]. In the long-run, reflective processes such as self-narration, or associative processes such as habit formation may lead to higher extraversion in this example (Wrzus & Roberts, 2017).  According to this theory, being active and having a high-volume participation in leisure activities may enhance the number of potentially triggering situations and thus enhance the probability of personality trait change.” (pp. 4-5).  Further, in the Introduction we now also describe the results of earlier research investigating associations of leisure activities and personality traits in a paragraph on “What personality − leisure activity − transactions to expect?” (please see pages 5-8). By this we now review literature that is directly related to the present research only. However, we find that:  “[…] a systematic understanding of whether and how participation in leisure activities contributes to personality change and vice versa is still lacking. To get closer to the underlying mechanisms of personality change, we must investigate how changes in an individual's behaviour and experiences (e.g. leisure activities), affect its future personality. Comparing people who are more active in their leisure time with those who are less active, in terms of personality change, would not allow such causal conclusions. Thus, to get valid insights about prospective effects between personality and leisure activities we must distinguish between-person and within-person variance in longitudinal data.” (p. 8) |
| 1. The Reviewers noted concerns with the **analytic model**. There is on-going debate about the standard cross-lagged model. I think it is important to clarify why the model used in the paper is appropriate and you should consider the RI-CLPM and correlated growth model variously mentioned by the Reviewers as alternatives. To be clear, the model used in the paper (ARCL) might be the one that is best suited to test your hypotheses, but that rationale should be articulated for readers. |
| **Our response:**  Thank you for pointing out the alternative analytic approaches. We have reconsidered our analytic model and have decided that the random-intercept model (RI-CLPM) fits best to our research question, since we are interested in the within-person directional processes of leisure activities and personality. We articulate that as follows to the reader:  “The RI-CLPM model offers some advantages over the traditional cross-lagged panel model (Finkel, 1995, see e.g. Mund & Nestler, 2019 or Orth, Clark, Donnellan & Robins, 2020 for more statistical reasoning and model comparison). Most importantly, the model allows to distinguish between within-person and between-person effects (Allison, 2009). […] Thus, the RI-CLPM controls for stable differences due to e.g. gender or other stable characteristics, that make individuals systematically different over time. Therefore, estimated auto-regressive (AR) and cross-lagged (CL) terms are no longer confounded by stable between-person differences.” (p. 12)  “By using a random-intercept cross-lagged panel model (RI-CLPM), we will disentangle between-person and within-person effects, making the within-person longitudinal relationship between participation in leisure activities and personality development visible. Thus, we will provide evidence by which to evaluate recently proposed models of personality stability and change.” (p. 10) |
| 1. I think the point raised by Reviewer #2 about age is important. Stability coefficients seem to change with **age** and thus age would seem to be a viable moderator of selection and socialization effects. This issue should be considered both theoretically/conceptually and analytically. |
| **Our response:**  We now consider potential age differences in prospective effects of leisure activities on personality traits and vice versa, more explicitly. From the theoretical point of view, we added that  “The authors of the TESSERA framework explain age differences in personality development by varying components and processes due to physical, cognitive, social, and societal changes related to age” (p.4)  However, we were not able to derive age-group specific hypothesis from the literature regarding the within-person effects of participation in leisure activities and personality change. Nevertheless, we expanded our analysis by adding age groups  “For additional analysis, we generate groups, based on age in 2005: old adults (born 1930 −1954) middle-aged adults (born 1955 − 1974), and young adults (born 1975 − 1987).” (p.11)  “Subsequently we expanded the RI-CPLM by adding an age group variable to the model (Mulder & Hamaker, 2020)” (p.14)  Indeed, we find that some within-person cross-lagged effects differ between age groups and discussed these findings in detail, please see Discussion section on pages 41-45. Therefore, we are very obliged for the advice regarding age from you and the reviewers. |
| 1. I think the points about the **nature of the social participation variable** raised by Reviewer #1 (and to some extent, Reviewer #2) are important to address. I think a stronger defense of the validity of the measure is needed and I also think it would be important to consider some specificity in terms of the nature of the activities. For example, “artistic activities” seem more relevant to Openness than other trait (Extraversion and “fun activities” seem like another intuitive pair). |
| **Our response:**  In the course of the new analytical approach described above, we have moved away from the designation of social participation. Instead, we now speak of "overall participation" to describe the sum score of the six leisure activities. We think that this term is more precise, because not every leisure activity studied is necessarily social. Additional to overall participation, we now investigate the effects of six leisure activities (physical activity, socializing, volunteering, political activity, artistic and musical activity, going out [previously called” fun activities”) separately. Based on the literature available we derived specific hypotheses e.g. the evidence from cross-sectional data suggested that changes in frequency of socializing may result in changes in extraversion and agreeableness and vice versa. However, since (to our knowledge) this is the first time that prospective within-person effects of leisure activities and personality traits are investigated, this study is rather explorative and therefore we report the results for each combination of personality trait and leisure activity. Interestingly, we found that the intuitive pairs like openness and artistic and musical activities are associated at between-person level but not at within-person level (please see Results section on pages 14-40) |
| 1. I did not understand the theoretical rationale for **controlling for income.** This should be expanded in the paper. |
| **Our response:**  In the first version of the manuscript we controlled for income, because we assumed that net income determines how often certain leisure activities can be carried out. In the current version of the manuscript, we estimate a RI-CLPM where the between-person variance captures possible confounders. Therefore, we did not explicitly estimate the effect of income in this revision. Please see paragraph on statistical model on pages 11-14 for details. |
| 1. I think reporting **degrees of freedom and chi-square for all models** in the primary text is important. |
| **Our response:**  Due to the large number (7 x 5 plus Granger-Sims causality tests) of models reported in the manuscript we have once again refrained from reporting the degrees of freedom and chi-square for all modelsin the main text. In this way we want to increase the readability of the Results section. However, all model fit parameters can be seen in the Tables 2-8 and the supplemental Tables S3-S9. We hope that this approach is in your interest.  By following the helpful suggestions of you and the reviewers, we believe that our manuscript is now more properly anchored in theory and analytic approach. We are optimistic that due to the extended scope (specific leisure activities, age groups) and the more detailed analysis (between-person and with-person level) the manuscript now has considerably more significance.  Further, we addressed all minor issues stated by the reviewers. Please see below. |

# Reviewer #1

|  |
| --- |
| In this study, researchers examined cross-lagged co-development between the Big Five personality traits and social participation in a large, representative sample of Germany. They found significant co-developmental effects for the traits of O, E, and N, and less consistent effects for A and C. I believe that this paper has potential to be a well-cited work of personality development. |
| **Our response:**  Thank you for your positive comments! It is very motivating that you believe this paper has potential to be a well-cited work of personality development. |
| However, this initial paper has major room for refinement that I would require before I recommend this study for publication. Beyond these revisions, I also believe that it would be valuable to disambiguate the social participation items and examining them separately. This is not required, but I believe it would substantially improve this work and up its impact in the field. |
| **Our response:**  We thank you very much for this helpful suggestion. We agree that extending the work by examining the leisure activities separately gives it much more strength, especially as the relationship between leisure activities and personality development has not been studied in depth before. As a result, we have incorporated your suggestion and completely restructured the manuscript accordingly. Please see the manuscript for details. |
| Intro:   * The first eight paragraphs of the introduction are rather meandering and go into detail about some foundational issues that can be presented much more concisely (personality is important, there are normative changes in personality across the lifespan, there are substantial individual differences in change). The major idea of this paper, that social participant may be linked to personality change, isn’t even mentioned until five paragraphs in. This part of the introduction should be shortened. * The authors review research on specific, discrete life events (e.g. retirement, having children) that is only somewhat related to the present research, which focuses more on the gradually unfolding effect of daily experiences on personality development. I think that these references can be culled. * On the other hand, theoretical aspects of the introduction should be expanded. In particular, the mechanisms that may underlie development are only discussed in passing (with respect to the sociogenomic model). A reader not well-versed in personality development may be rather confused as to how changes in one’s daily behaviors may affect personality. The TESSERA model (Wrzus & Roberts, 2017) lays out specific pathways that may lead personality states to congeal into personality trait change. Conceptualizing co-development in terms of each component of this framework would substantially improve the paper’s theoretical focus and provide evidence by which to evaluate a recently proposed model of change. * Socialization and selection effects are discussed throughout the paper but never given sufficient description or definition. |
| **Our response:**   * Thank you for pointing to this inadequacy. We now omitted this needless part of the Introduction and, arrive more quickly to the point that change in participation in leisure activities may be linked to personality change. We severely curtailed the first paragraphs and state the aim of this paper right at the beginning. Please see page 3. * As part of the restructuring of the Introduction, we have also omitted the content on personality and life events. Instead, we now focus on the literature on leisure activities and personality traits and review on “What personality − leisure activity − transactions to expect?” from earlier research. (pp. 5-8).   “[…] Up to now, literature on personality-leisure activity-relationships is rare. In line with Stephan and colleagues (2014), we review earlier studies broadly ordered by physical, social, and cognitive leisure activities. However, some leisure activities may not be uniquely assigned to one of these domains. We summarize the evidence on this in a paragraph on “Other activities”. […]” (p.5)   * Thanks to your recommendation we revised the theoretical anchoring of the study and newly drafted a paragraph on “What drives personality development?” (p.4). We now describe recent theoretical frameworks on personality development that may explain this association of overall participation, leisure activities and personality change in more detail.   “Most theoretical frameworks agree that personality can be shaped by biological factors, e.g. genetic influences, or health issues, as well as environmental factors, e.g. social roles, normative life transitions or major individual life events. There is empirical evidence for the influence of both, genes and life experiences in shaping personality stability and change (Bleidorn, Kandler, & Caspi, 2014; Briley & Tucker-Drob, 2014). Also, time, role scripts, and active involvement of the person themself appear to be sources of change of personality traits (cf. Specht et al., 2014).  Despite differences in details, recent theoretical frameworks (Roberts, 2018; Wagner, Orth, Bleidorn, Hopwood, & Kandler, 2020; Wrzus & Roberts, 2017) broadly describe personality change as a bottom-up process: Recurrent and enduring short-term changes in behavior and daily life experiences drive long-term changes in personality traits. Further, these theories propose transactional processes: Personality traits evoke a characteristic pattern of behavior which increases the probability of exposure to specific environments and to life experiences (i.e., selection effects). Similarly, environmental factors act via situational processes, through the filter of individual experiences, on personality traits (i.e., socialization effects).” (p.4)  In particular, we now center the Introduction on the TESSERA framework (Wrzus & Roberts, 2017) to derive hypotheses on the associations of participation in leisure activities and personality trait change. There is now a separate section entitled “Participation in leisure activities – a driver of personality development?” (pp. 4-5). By this, we hope to improve the paper’s theoretical focus and provide evidence by which to evaluate the recently proposed model of change.  “The recently published TESSERA framework on personality development in adulthood postulates that personality changing short-term processes can be generalized as “recursive sequences of Triggering situations, Expectancy, States/State expressions, and Reactions” (Wrzus & Roberts, 2017). Internal reflective or associative processes transform repeated TESSERA sequences into long-term personality development. Internal reflective processes can be self-reflection, accommodation, assimilation, life reflection, or self-narration. Internal associative processes may be implicit learning, reinforcement learning, or habit formation. The authors of the TESSERA framework explain age differences in personality development by varying components and processes due to physical, cognitive, social, and societal changes related to age. […]  Applying the TESSERA framework, personality traits may change if the following short-term process is repeated: A certain leisure activity, e.g. political engagement [Triggering situation] is perceived as relevant to a trait, e.g. extraversion [Expectancy] and thus elicits a relevant state, e.g. speaking in front of a crowd [State/State expression] and this state level does not correspond to the actual trait level, e.g. low extraversion, but the reaction is positive, e.g.: political talk is a success, people are convinced [Reaction]. In the long-run, reflective processes such as self-narration, or associative processes such as habit formation may lead to higher extraversion in this example (Wrzus & Roberts, 2017).  According to this theory, being active and having a high-volume participation in leisure activities may enhance the number of potentially triggering situations and thus enhance the probability of personality trait change.” (pp. 4-5).   * We now shortly mention selection and socialization effects and give examples in the Introduction:   “Personality traits evoke a characteristic pattern of behavior which increases the probability of exposure to specific environments and thus experiences (i.e. selection effects). In turn, environmental factors act via situational processes, through the filter of the individual experiences, on personality traits (i.e. socialization effects)” (p.4)  However, we do not discuss socialization and selection effects throughout the paper anymore. Because of the new approach to analysis (RI-CLPM), we now speak of within-person cross-lagged effects from change in frequency of a certain leisure activity to personality trait change or from personality trait change to change in frequency of a certain leisure activities. |
| Methods:   * The authors do not provide evidence for the internal reliability or factor structure of this social participation scale. * The authors also do not provide evidence for measurement invariance of the social participation scale. Establishing weak invariance is critically important so that we know whether scale scores are comparable across measurement occasions. * I’m not sure what happened with Table 2, but it’s there five times. This might have been caused by the manuscript upload system? * The authors should be aware of the debate and distinction between CLPMs and Random Intercept CLPMS (Hamaker et al., 2015). Because CLPMs do not adjust for individual differences in baseline levels of social participation and personality, they conflate selection and socialization effects to some extent. I would suggest that the authors drop the CPLM analyses and only run RI-CLPMs (as they seem more germane to the inferential goals of the present study), or estimate these models in addition to the already-reported CLPMs. |
| **Our response:**   * In fact, the initial social participation construct was not a factor. We have only calculated the sum of the frequency of the six different leisure activities − which do not load on one factor − to measure a person’s overall participation in leisure activities. In this revised version of the manuscript we no longer speak of “social participation” but of “overall participation”. We believe this is more precise, since the leisure activities investigated (physical activity, socializing, volunteering, political activity, artistic and musical activity, going out activities [previously called “fun activities”]) do not necessarily always have a social component. In mathematical terms, however, it is still the same sum score. The idea behind this cumulative score is that we wanted to know not only what the interaction between certain leisure activities and personality traits is like, but also what the influence of the general level of activity is (expressed by the sum score “overall participation”). * Unfortunately, the SOEP does not provide a social or overall participation scale (see above). So, we could not estimate a corresponding latent variable. Therefore, it is also not possible to report the measurement invariance. * We are sorry that the first submission was so difficult to read due to the multiple insertion of the Table. This was a technical error, which hopefully will not happen again in this version. However, due to the extension of analysis we provide eight Tables (Tables 1-8) in the manuscript and nine more Tables in the supplements now (Tables S1-S9). * Thanks to your advice, we once again took a close look at the analytical aspects of our study and finally decided to switch to an adapted version of the random- intercept cross-lagged panel model (RI-CLPM). Please see Method section on statistical model for details (pp. 12-14). We have weighed up the pros and cons thoroughly, also considering the findings of Orth, Clark, Donnellan and Robins (2020). We concluded that our research question − How does a change in the frequency of a particular leisure activity affect the development of a person’s personality trait and vice versa? − can best be answered with the RI-CLPM (extended by age grouping, see Mulder & Hamaker, 2020). Orth and colleagues (2020) compared seven cross-lagged model and recommend using either the CLPM or the RI-CLPM, depending on whether the research question concerns between- or within-person effects. When the research question concerns prospective effects of within-person deviations from trait levels, as in our case, the RI-CLPM should be selected. Ultimately, we are interested in the mechanisms of personality change within a person. |
| Results:   * Standardizing personality and social participation to a mean of 0 and a SD of 1 would allow the authors to evaluate effect sizes in absolute terms (comparing to other personality research) and in relative terms (comparing effect sizes across traits and cross-lag directionalities). |
| * According to your suggestion we standardized all study variables to a mean of 0 and standard deviation of 1, before further analysis. By this we now allow to compare effect sizes within this study and across personality research. |
| Minor issues:   * The authors frequently use the phrase “go along with,” which is rather nonscientific. Perhaps the phrases “co-occur with” or “are associated with” could be used in their place. * Similarly, the authors frequently use the terms “supposed” and “assume” when they mean “theorized” (these terms are technically grammatically correct but have the everyday connotation of “in theory it does, but in actuality it doesn’t”). * I think the p-value sign is flipped in the second to last paragraph of the openness results. * The stability of social participation does not need to be re-reported in each model’s results. * Figure 1 would be much more informative if it included parameter point estimates. * Table 1 can include age as well. * In the discussion, the authors mention that this study provides evidence that change in social participation was associated with change… beyond change in openness to actions. I think this should be openness to ideas, instead, as that is the NEO-PI-R facet most associated with intellect |
| **Our response:**   * As non-native speakers we are very grateful for these helpful comments and have adapted the wording. We have replaced the terms " go along with " and " supposed/ assume" accordingly. * Thank you for this remark. The p-value sign was incorrect. However, the Results section has been completely renewed. * Due to the new statistical models the reporting of results in the tables changed (see Tables 2-8). We hope that the presentation is well comprehensible now, without containing unnecessary details. * Thank you for pointing to this mistake in the discussion. However, we fully reworked the discussion section, due to new analysis and results. |
| Finally, I would like to outline reasons why I think the authors should disambiguate the social participation activities. One of the most interesting parts of this paper is the multifaceted nature of social participation. Although the authors are justified in summing all activities into a formative index, we could learn a lot more about how social participation is associated with personality development if these activities are broken apart in additional analyses. The authors even allude to this in the introduction when they make points about the multifaceted nature of social participation.  Benefits of separating activities:   * The introduction would be much more well-structured if different types of social participation were reviewed in turn. For example, the authors could first discuss social life and personality development, and then cultural activity and personality development, and finally civic life (voluntary and political activities) and personality development. Each of these activities may have different links to personality development, and past studies have typically examined these everyday behaviors in isolation. * Scores on the social participation scale would be more interpretable. As Table 1 shows, these variables are not all highly associated with one another, and likely would not load highly onto a single factor. It is okay to have a formative index of behavior, where it is not assumed that a latent level of social participation causes all activities equally, but the scale is currently only interpretable in terms of “how much stuff the participant does in their free time” because specific content is averaged out. * Analyzing these scales separately would allow the researchers to examine how particular kinds of social participation are associated with personality development, which is theoretically quite interesting. Looking at table 1, it seems that O and E are linked to change in most activities, but, if cross-sectional associations translate longitudinally, change in A may only be associated with interpersonal activities and increases in C may predict decreases in fun activities (as alluded to in the discussion). These specific links are both interesting for theory and make for a more interesting paper. * The major drawback from separating activities is that individual items may not be highly reliable. Given the sample size, I’m not too worried about the decrease in signal that would come from lowering reliability, but extremely low test-retest correlations, for example, might preclude the authors from these extra analyses.   Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. Psychological methods, 20(1), 102.  Wrzus, C., & Roberts, B. W. (2017). Processes of personality development in adulthood: The TESSERA framework. Personality and Social Psychology Review, 21(3), 253-277. |

**Our response:**

Thanks again for your outstandingly valuable feedback. We were able to follow your reasoning very well and have restructured the paper in the exactly corresponding way.

* In line with previous research (Stephan, Boiché, Canada, & Terracciano, 2014), we now review the literature on certain leisure activities and their relationship to personality traits broadly ordered by physical, social, and cognitive activities in the Introduction. However, some leisure activities may not be uniquely assigned to one of these domains. Please see pages 5-10 for literature review and resulting hypotheses.
* Earlier research shows that personality traits not only correlate with the frequency of certain leisure activities but also with the variety of activities (Stephan, Boiché, Canada, & Terracciano, 2014). Thus, we were still interested in the effect of “how much stuff the participant does in their free time” on personality traits and vice versa. So, in this revision we examine both the individual leisure activity and the sum score, now called “overall participation”. Please see the paragraph on “Current study” (pp-9-10) for details on this revised paper’s concept.
* As proposed, we now discuss the results for each combination of leisure activity and personality trait separately. Indeed, we found highly specific results. There were even differences in within-person effects across age groups. Due to the new analyses we came to the following new conclusion:

“Taken together, by applying RI-CLPM to four waves of 13-year longitudinal data from the socioeconomic panel study (SOEP), the present study has shown that at the between-person level, the investigated leisure activities as well as the overall participation are most strongly associated to openness. However, at within-person level, prospective feedback effects are found for extraversion with socializing and overall participation only. It was also found that there are some within-person cross-lagged effects of specific leisure activities with certain personality traits in certain age groups. For example, increase in volunteering behavior leads to prospective increase in extraversion in young and old adults.

Overall, this study strengthens the idea of the TESERRA framework, that repetitive trait triggering situations can lead to personality change. However, for openness and neuroticism traits, we could not identify triggering leisure activities, in the current study. To develop a broader picture of personality trait triggering leisure situations, additional studies will be needed that investigate more leisure experiences in connection with personality change at a shorter time scale.” (pp.48-49)

* Most test-retest-correlations vary around r=.50 (see supplement S2 – intercorrelations). Thus, we were not prevented from the suggested analyses.

# Reviewer #2

|  |
| --- |
| This manuscript presents a test of the developmental relationship of everyday social behavior and personality traits over time in the German Socio-Economic Panel. This is an important question given that the majority of research testing the role of social engagement on personality development has focused on role status rather than actual social behavior. A number of refinements can be made to better deliver this valuable research question in this robust dataset. |
| **Our response:**  Thank you for your positive comments and the appreciation of our efforts! |
| At the outset, the introduction was somewhat confusing. The very broad opening of the importance of personality in general obfuscates the value of the important process questions being asked in the paper. A focus on setting up the developmental question would help clarify the scope of the paper given that life outcomes are not incorporated in the analyses. |
| **Our response:**  Thank you very much for this suggestion. We agree and now focus the Introduction more on the developmental research question and anchor it in recent theoretical frameworks. As suggested, we now omitted the literature review on life events. Instead we included earlier research on different leisure activities and personality traits (as proposed by Reviewer #1). Thus, the focus of the paper shifted, and we adapted the title and Introduction accordingly.  In the Introduction we now, arrive more quickly at the point that participation in leisure activities may be linked to personality change. We shortened the paragraphs on the importance of personality development research and state the aim of this paper right at the beginning (see page 3). This is followed by a section reviewing recent theoretical frameworks on personality stability and change (see “What drives personality development?”, see p. 4). Thereafter we now apply the TESSERA framework (Wrusz & Roberts, 2017) to the current research question, setting out how and why participation in leisure activities may be a driver of personality development (see pp. 4-5). We now close the introduction reviewing earlier research on personality and overall participation in leisure activities and specific leisure activities (see pp. 5-8, paragraph “What personality − leisure activity − transactions to expect?”)  We hope to have erased the appearance of being “all over the place” and to have clarified the focus of the paper in the Introduction. |
| Further, articulating the developmental literature, precision is needed in differentiating stability and change – at times these terms are confounded with findings of each intermingled throughout p. 4 & 5. For example, “rank-order changes” (p. 5), is ambiguous and the introduction currently overlooks the important differences between rank-order stability and mean-level change and the very real occurrence of both simultaneously in the personality development literature. In general, a careful proofread would improve the readability of the manuscript. |
| **Our response:**  We have thoroughly revised the Introduction and think that there is no longer a confusing usage of the terms between rank-order stability and mean-level change. Thank you for this remark. |
| Clarity and precision are also needed on analytic choices. Please justify the choice of the ARCL over other modeling possibilities such as the latent growth model. Articulation of modeling choices and potentially changes regarding demographic variables is also warranted. The manuscript states that gender was not controlled for because it did not relate to ROS or MLC in another sample. However, that study was in a very different sample. These choices should be based in the dataset itself and analyses modified as needed. |
| **Our response:**  Thank you for this remark. We once again took a close look at the analytical aspects of our study. Correlating intercepts and slopes, as done by the Bivariate Latent Growth Curve Models (LGC) would have measured correlated mean-level change. We think that our research question − How does change in frequency of a particular leisure activity affect the development of a person’s personality trait and vice versa? − can best be answered with a cross-lagged panel model. Orth and colleagues (2020) compared seven cross-lagged models and recommend using either the CLPM or the RI-CLPM, depending on whether the research question concerns between- or within-person effects. When the research question concerns prospective effects of within-person deviations from trait levels, as in our case, the RI-CLPM should be selected. Ultimately, we are interested in the mechanisms of personality change within a person. We have weighed up the pros and cons thoroughly, also considering the findings of Orth, Clark, Donnellan and Robins (2020) and finally decided to switch from the ARCL to an adapted version of the random-intercept cross-lagged panel model (RI-CLPM, see Hamaker, Kuiper, & Grasman, 2015). Please see Method section on statistical model for details (pp. 11-14). The RI-CLPM controls for stable differences due to e.g. gender or other stable characteristics, that make individuals systematically different over time. Therefore, estimated auto-regressive and cross-lagged terms are no longer confounded by stable between-person differences and gender has not to be controlled for. |
| On a related and potentially important note, justification is needed regarding the treatment of age. Much of the literature review was focused on personality in late life. However, the sample is a lifespan sample. Simply controlling for age and age2 may not get at age differences in developmental patterns. In the code provided on OSF, it appears demographic age groups may have been created but results are not reported within age groups. Alternatively, this may be an opportunity to apply local structural equation modeling. This emerging technique in personality psych (Olaru et al., 2019; Wagner et al., 2019) could be particularly interesting with this research question. |
| **Our response:**  Due to the extension of the research questions to the single leisure activities and the resulting fundamental revision of the Introduction the literature review is not focused on personality in late life anymore.  We agree that age differences in personality change patterns can be expected from a theoretical point of view (Roberts, 2018; Wagner, Orth, Bleidorn, Hopwood, & Kandler, 2020; Wrzus & Roberts, 2017). However, we were not able to derive age-group specific hypothesis from the literature regarding the within-person effects of participation in leisure activities and personality change, since literature on personality-leisure-relationships is scar (see p.5).  Nevertheless, for additional analysis, we generated groups, based on age in 2005: old adults (born 1930 −1954), middle-aged adults (born 1955 − 1974), and young adults (born 1975 − 1987 (see p.10). We expanded the RI-CPLM by adding an age group variable to the model (Mulder & Hamaker, 2020) (p.14)  Indeed, we find that some within-person cross-lagged effects differ between age groups and discussed these findings in detail, please see Discussion section on pages 41-45. Therefore, we are very obliged for the advice regarding age. |
| Finally, please state whether or not the hypotheses were preregistered. The authors are encouraged to explore preregistration (https://osf.io/x4gzt/) and other options for open science approaches to secondary data analysis (Weston et al., 2019) in their future research. |
| **Our response:**  We now state that hypotheses were not preregistered in the paper. Thank you for pointing to the recommendations of Weston and colleagues (2019) that we will take heed in our future research. |

This summarizes the ways in which we have responded to your and the reviewers’ constructive comments. We believe that we have addressed all issues that were raised in a satisfactory manner and hope you will find this revised manuscript acceptable for publication in *Collabra: Psychology – Personality Psychology.* We appreciate your continued consideration of this manuscript.

Sincerely yours,

Julia Sander

(on behalf of all authors)

**Editor Second Decision—Revise & Resubmit**

Jan 28, 2021

Dear Julia Sander,

Thank you for submitting your your revised manuscript, “Leisure Activities as a Driver of Personality Development? – A Random-Intercept Cross-lagged Panel Model Across 13 Years in Adulthood” to Collabra:Psychology. I sent the paper back to both reviewers and their comments appear at the end of this letter. I thank both experts for providing helpful feedback. As with their reviews last time, they did outstanding work in identifying the strengths and weaknesses of the paper.

The bottom line is that this paper is moving in the right direction but there are still several issues that should be resolved before I am comfortable with publication. The reviewers raised important points and I will ask you to address all of them either in a letter of response or in the revised paper. I think taking their suggestions seriously will enhance this paper. I will highlight some issues that came up as I read your paper and I will try to synthesize the editorial feedback across reviewers. I do not plan to send the revised version back to the reviewers but reserve that right if anything changes dramatically.

1. Both reviewers raise questions about the age groupings. I understand that categorical variables are easier to deal with when testing moderation in the SEM framework. My recommendation is to provide some greater justification for the specific groupings used in the paper and refer to them by age (see Reviewer #1) rather than birth year. Likewise, additional details will help address concerns raised by Reviewer #2.
2. The use of the RI-CLPM was appreciated by both reviewers. At the same time, there were questions. The second issue raised by reviewer #2 about how the “trait” factor was identified was something that crossed my mind. I think this reviewer offered useful suggestions for handling the issues. I think this specification probably does have some important implications for interpreting the results that should be explained to readers. I will add that the loadings might be worth reporting and interpreting if this is the final model that ends up in the paper. My understanding is that AR stability coefficients should be interpreted in light of how much variance is accounted for by the “trait” factor. If the trait factor accounts for a large chunk of variance at each wave, then this has implications for what variance at each wave represents and for comparing the AR stability coefficients. (I might be wrong so correct me in the response letter).
3. I also think Reviewer #2’s point #1 about interpreting more parameters from the RI-CLPM model was useful.
4. Both reviewers commented on the terminology in the paper and the value in having consistency of the language and tenses throughout. Please attend to these issues. On a related point, I found the “causation” reference to particular parameters to raise some reservations. For example, the phrase “overall participation -> extraversion causation (beta = .024, …)” appeared on page 17. I think it would be easy enough to swap out causation here with something more neutral such as prospective effect and just use that throughout.
5. More generally, it would be good to give the paper a final edit for clarity and typos as I noted some of those when I read the paper (e.g., scrabbles should likely be Scrabble, on p. 7; “in for the young adults” on page 33 could probably just be “in the young adult sample”).
6. When describing the RI-CLPM on page 12, I think it would be useful to have a sub-heading to set-up that section. I would then call out to Figure 1 right away so readers have a reference point for the parameters. I do not know that it is 100% necessary to use the Greek letters/old-school LISREL notation (that’s what I call it) in the figure but that is probably just stylistic. Likewise, the triangles could probably be deleted to enhance clarity but I will leave that up to your team.
7. It might be useful to give readers some additional pieces of descriptive information such as how the leisure activities correlated with each other as well as the zero-order stabilities of those variables.
8. I liked the summary table idea as I think readers will benefit from a structured section with take away messages given the large number of analyses. However, I found Table 9 a bit hard to apprehend at first so I encourage you to consider alternative ways to display the information. For example, I did not find the italics to be helpful and it made the table harder to read. You might consider doing away with the Xs and just using plusses and minuses and not having anything in italicized font. You could delete the political activities row and just add to the note that nothing was p < .05. These are just ideas and provided in the spirit of making the paper maximally accessible.
9. In thinking through all of the results, I was a bit concerned about the sheer number of analyses. I don’t have strong opinion here but you might want to adopt a more stringent p-value for identifying a discovery or at least acknowledge this issue a bit more in the paper.

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](mailto: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. Thank you for considering Collabra as an outlet for your work.

Sincerely,

Brent Donnellan

# 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.

The current manuscript has a number of notable improvements on the original manuscript and presents quite interesting data. The reframing of the research to focus on the specific leisure activities as well as overall activity level and the refining focus to this goal greatly enhances the delivery of this interesting research. The additional details regarding the model and modeling choices is also valuable. While the choices in analysis and granularity make for a more complex set of findings, Table 9 helps organize the multiple results well. The addition of age groups is also a substantial add both from the perspective of interest and analytic insights. I have a few additional suggestions that I hope will further enhance and refine this work.

My primary suggestion is to more deeply address the one caveat mentioned that “age is a proxy for unknown confounders for development” (p. 43). This can be paired with additional limitations and future directions around the measurement of leisure activities. Specifically, more can be done both in the discussion and in future work to explore why the patterns may be present in the data. Age and developmental context could potentially inform this but so could other and/or related factors. Right now the measure only addresses frequency but what about other aspects of the leisure activities such as how rewarding they are, how voluntary they are, and how enjoyable they are? The example in the intro using the TESSRA model is helpful – perhaps one or two could help these points in the discussion as well.

On a smaller note, please include the mean of the age groups not just the year of birth to help be more clear in what the age groups mean. Additionally, perhaps it has just been a while since I sat down with the RI-CLPM but I am confused by the use of “unexpected change” in describing the results. “Short-term boost” also seems a bit confusing given the 4-year intervals of the study. A general read for clarity on terms like this would help deliver the take home message of the findings.

In sum, these are interesting findings from a large, respected panel study that have the potential to catalyze future work in the area.

##### 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.

These revisions were well done and, though they required quite a lot of changes, have vastly improved the paper. I have four comments that concern changes made in revision.

First, the authors should report and interpret correlated change effects (in terms of the authors’ model: ψ(lvi, lui)) alongside their cross-lagged effects. These parameters test whether temporary deviations from a person’s personality trait score are associated with simultaneous temporary deviations from their leisure activities score. They’re an essential piece of the within-person developmental puzzle.

Second, as I have not encountered time-varying unit effects before, I am unsure if it is appropriate to use them here. Typically, unit effects are modeled as time-invariant, making them function as random intercepts that account solely for individual differences at baseline. When unit effects are modeled as time-variant, however, I am unsure if they then function as a combination of random intercepts and random slopes, modeling individual differences at baseline and also change across the study period. If they do (which, after reading the Zyphur et al., 2020 reference, might be the case), these time-varying unit effects may account for meaningful within-person developmental trends that the authors likely do not wish to remove from the cross-lagged portion of their model. In less jargony terms, personality development research usually treats sustained trends (e.g. a steady increase in extraversion) and deviations from that trend (e.g. an especially big increase in extraversion at t2) as important variance to explain. The time-varying unit effect may remove the sustained trend from the model. Because this may have a drastic effect on the results (or no effect at all!), I request that the authors estimate additional variants of their basic models that include time invariant unit-effects (i.e. all factor loadings on the unit effect are set to 1). If the authors’ cross-lagged and correlated change findings are quite different across time-varying and time-invariant unit effect models, I would suggest that they primarily interpret the time-invariant unit effect models, as it is much more clear how to interpret those findings.

Third, I would appreciate clarification in the manuscript about how age effects were estimated. Were they estimated by adding a single covariate? How were age groups compared with one another? Are differences across age groups robust enough to merit fine-grained interpretation in the discussion?

Fourth, the tense of language in the revised parts of the introduction and discussion sections is somewhat imprecise. The authors toggle between past-tense and present-tense in both the introduction and the results. I suggest that they frame results of both past research and this study in the past tense, as psychological research generally describes what has been found (e.g. people who did X were Y (in Z et al.)) not what is universally true (e.g. people who X are Y).

##### 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

Apr 21, 2021

Dear Professor Brent Donnellan,

Thank you very much for your letter of January 28, 2021, concerning our manuscript entitled “Leisure Activities as a Driver of Personality Development? A Random-Intercept Cross-lagged Panel Model Across 13 Years in Adulthood”, submitted for publication in *Collabra: Psychology*.

We were very pleased to hear that you and both reviewers evaluated the revised manuscript as “moving in the right direction” and that it has the potential to catalyze future work in the area. We once more greatly appreciate the thoughtfulness of the constructive comments that you and the two reviewers provided.

Again, we have carefully considered and responded in detail to each of the points made by you and the reviewers. In the following sections, we describe in detail how we have addressed each point.

# Your editorial comment

|  |
| --- |
| […] The bottom line is that this paper is moving in the right direction but there are still several issues that should be resolved before I am comfortable with publication. The reviewers raised important points and I will ask you to address all of them either in a letter of response or in the revised paper. I think taking their suggestions seriously will enhance this paper. I will highlight some issues that came up as I read your paper and I will try to synthesize the editorial feedback across reviewers. *I do not plan to send the revised version back to the reviewers but reserve that right if anything changes dramatically.*   1. Both reviewers raise questions about the **age groupings**. I understand that categorical variables are easier to deal with when testing moderation in the SEM framework. My recommendation is to provide some greater justification for the specific groupings used in the paper and refer to them by age (see Reviewer #1) rather than birth year. Likewise, additional details will help address concerns raised by Reviewer #2. |
| **Our response:**  Thank you very much for your appreciative comments and suggestions.  As suggested, we now refer to groups by age range at t1 throughout the manuscript. Further, we make our age group justifications explicit and give additional details on how age effects were estimated. We added a new paragraph in the Methods section:  “Age Groups  For additional analysis, we generated age groups of approximately equal size: Group 1, young adults (M = 24.1, min = 18, max = 30 years old at T1, i.e. in 2005; born 1975–1987; N = 17151); Group 2, middle-aged adults (M = 39.6, min = 31, max = 50 years old at T1; born 1955–1974; N = 24546), and Group 3, older adults (M = 61.3, min = 51, max = 75 years old at T1; born 1930–1954; N = 14093). For descriptive statistics based on these age groups, in 2005, please see Supplement S1.” (p.11)  To answer questions raised by Reviewer #2, such as: “Were age effects estimated by adding a single covariate and how were age groups compared with one another”, we now describe the age effects test procedure in the Methods section in more detail:  “*Moderation Effects of Age Groups in RI-CLPM*  Subsequently we expanded the RI-CPLM by adding the categorical age grouping variable to the model. Age group differences were investigated by comparing a multiple group version of the RI-CLPM in which there are no constraints across the age groups with a model in which the lagged regression coefficients are constrained to be identical across the groups (see Mulder & Hamaker, 2020). We therefore know from chi-square difference tests whether there is a moderation effect of age group or not. […]“(p. 15)  We agree with Reviewer #2 that model fit testing with and without constraints across age groups only gives a first impression on age effects. That is why we now interpret results a bit more carefully and recommend investigating the cross-lagged effects of personality traits and leisure activities in the different age groups in more detail in further studies.  “Our results on age effects should also be interpreted with caution. We obtained a first impression of the impact of age through model fit testing with and without constraints across age groups, but the mechanisms in the different age groups should be investigated in more detail in further studies.” (p.49) |
| 1. The use of the RI-CLPM was appreciated by both reviewers. At the same time, there were questions. The second issue raised by reviewer #2 about how the **“trait” factor** was identified was something that crossed my mind. I think this reviewer offered useful suggestions for handling the issues. I think this specification probably does have some important implications for interpreting the results that should be explained to readers. I will add that the loadings might be worth reporting and interpreting if this is the final model that ends up in the paper. My understanding is that **AR stability coefficients** should be interpreted in light of how much variance is accounted for by the “trait” factor. If the trait factor accounts for a large chunk of variance at each wave, then this has implications for what variance at each wave represents and for comparing the AR stability coefficients. (I might be wrong so correct me in the response letter). |
| **Our response:**  We thank you and Reviewer #2 for this comment, and we can certainly see the confusion that might have come with our choice of uncommon model estimation. However, we think that restraining the factor loadings of the unit effect latent factor over such a long period of time would be an unreasonable assumption.  If the variance accounted for by the trait factor/unit effect had been very different for each wave, the interpretation of the AR coefficients might have needed an overhaul. However, as can be seen in the respective outputs, the difference between the various factor loadings was quite marginal. Thus, we do not think that the AR stability coefficients must be interpreted any differently.  However, as we were eager to test whether the results would change with restricted factor loadings, we calculated the full models once again with time-invariant unit effects for every respective trait/leisure activity combination. Results can now be found in the Appendix online. While the AR terms tend to be slightly smaller when the unit effects are allowed to differ across time, these differences had no influence on the significance and ranged, if at all, around 2 to 3 hundredths.  Please see <https://osf.io/fdxzp/?view_only=eaffa966ffe14e7c8a4ccb5ace76ace6> for R outputs. |
| 1. I also think Reviewer #2’s point #1 about interpreting more parameters from the RI-CLPM model was useful. |
| **Our response:**  We thank you and Reviewer #2 for the comment. We now pay more attention to correlated change effects throughout the manuscript. That is, we now describe this parameter and its interpretation in the Method section in more detail:  “The covariance of the phantom variables at each measurement occasion (ψ(lvi, lui)) shows whether temporary deviations from a person’s leisure activities score are associated with simultaneous temporary deviations from their personality trait score. We call this covariance “within-person correlated change effects” in the following.” (p. 14)  We have further added the correlated change values to the Results tables in the manuscript (Tables 2−8, pages 17–39). In addition, we now report on them in the results section with a few sentences each:  “At the within-person level, temporary deviations from a person’s **overall participation** score are positively associated with simultaneous temporary deviations from their openness (*rs* of .029−.054, *p*< .001) and extraversion trait scores (*rs* of .020−.054, p< .001). Temporary deviations from a person’s overall participation score were negatively associated with simultaneous temporary deviations from their neuroticism trait score at first and last measurement occasions only (*rs* of -.039 − -.018, *p*< .001). There were no simulations temporary deviations of overall participation in leisure activities with conscientiousness and agreeableness trait scores.” (p. 19)  “At the within-person level, temporary deviations from a person’s **physical activity** score are positively associated with simultaneous temporary deviations from their openness (*r*s of .029 and .018, *p*< .01), conscientiousness (*r*s of .013 and .014, *p*< .05), and extraversions (*r*s of .025 and .014, *p*< .001) trait scores at two measurement occasions. Temporary deviations from a person’s physical activity score were negatively associated with simultaneous temporary deviations from their neuroticism trait at the first and last measurement occasion only (*r*s of -.023 and-.016, *p*< .001). There were no simultaneous temporary deviations of physical activity and agreeableness trait scores.” (p. 21)  “At the within-person level, temporary deviations from a person’s **socializing** score are positively associated with simultaneous temporary deviations from their openness (*rs* of .015−.036, *p*< .05), extraversion (*rs* of .044−.052, *p*< .01), and agreeableness (*rs* of .016−.022, *p*< .05) trait scores, at least at three of four measurement occasions. Temporary deviations from a person’s socializing score were negatively associated with simultaneous temporary deviations from their neuroticism trait at three measurement occasions (*rs* of -.035, -.023 and -.016, *p*< .001). For simulations temporary deviations of socializing and conscientiousness trait scores, there were associations in both directions at different measurement occasions (*rs* of -.014, .017, and .034, *p*< .05.” (p. 24)  “At the within-person level, temporary deviations from a person’s **volunteering** score are positively associated with simultaneous temporary deviations from their openness (*rs* of .015 − .026, *p*< .05) and extraversion (*rs* of .013 − .035, *p*< .01) trait scores at three measurement occasions. Temporary deviations from a person’s volunteering score were negatively associated with simultaneous temporary deviations from their agreeableness (*rs* of -.018 and -.013, *p*< .05) and neuroticism (*rs* of -.016 and -.011, *p*<.05) trait scores at first and last measurement occasion only. There were no simulations temporary deviations of volunteering and conscientiousness trait scores.” (p. 28)  “At the within-person level, temporary deviations from a person’s **political activity** score are positively associated with simultaneous temporary deviations from their openness (*r* = .016, *p*< .05), conscientiousness (*r* = .026, *p*< .001), and extraversions (*r* = .015, *p*< .05) trait scores at the first measurement occasions. Temporary deviations from a person’s political activity score were negatively associated with simultaneous temporary deviations from their neuroticism trait score (*r* = -.026, *p*< .001) at the first measurement occasion and from their conscientiousness trait score (*rs* of -.011 and -.018*, p*< .05) at the third and fourth measurement occasions.” (p.31)  “At the within-person level, temporary deviations from a person’s **artistic and musical activity** score are positively associated with simultaneous temporary deviations from their openness trait score (rs of .027 −.068, p< .001). There are further simultaneous temporary deviations of artistic and musical activity from their extraversion trait scores (rs of .017 and .031, p< .01) at the first and third measurement occasions. There were no simulations temporary deviations of artistic and musical activity and conscientiousness, agreeableness, and neuroticism trait scores.” (p. 34)  “At the within-person level, temporary deviations from a person’s **going-out activity** score are positively associated with simultaneous temporary deviations from their openness (*rs* of .016−.025, *p*< .001) and their extraversion (*rs* of .012−.038, *p*< .05) trait scores. Further, at first measurement occasions only, there were negative simultaneous temporary deviations of going-out activities with conscientiousness (*r* = -0.048, *p*< .001) and agreeableness (*r* = -0.013, *p*< .05). There were no simulations temporary deviations of going-out activities and neuroticism trait scores.” (p.38)  “**Summary of Results** […] At the within-person level, temporary deviations from a person’s openness trait score are positively associated with simultaneous temporary deviations from their overall participation in leisure activities, and from their socializing, artistic, and musical activities at all measurement occasions. The same applies to the relationship between fluctuations in extraversion with simultaneous temporary deviations in overall participation in leisure activities, socializing activities, and going-out activities. Temporary deviations from a person’s physical activities, volunteering, and political activities are not unambiguously associated with simultaneous deviations in personality traits. That is, there are only a few significant correlations at some measurement points for the aforementioned leisure activities with personality traits; see Tables 2−8.” (p.41)  Furthermore, we now include correlated change effects in the discussion of the results, e.g.:  “We expected that changes in the frequency of artistic and musical activities may trigger changes in openness and vice versa. […] We found no effect, however, of changes in artistic and musical activities on an individual’s openness. At first glance, this appears somewhat contradictory to earlier findings by Schwaba et al. (2018), […] The measurement occasions in the current study were four years apart, whereas participants in the research of Schwaba et al. (2018) completed the survey every year or every second year. In the present study, we also found simultaneous temporary deviation in the same direction of openness and artistic and musical activity within persons. This points to shorter cause-effect relationships.” (p. 46-47). |
| 1. Both reviewers commented on the terminology in the paper and the value in having consistency of the language and tenses throughout. Please attend to these issues. On a related point, I found the “causation” reference to particular parameters to raise some reservations. For example, the phrase “overall participation -> extraversion causation (beta = .024, …)” appeared on page 17. I think it would be easy enough to swap out causation here with something more neutral such as prospective effect and just use that throughout. |
| **Our response:**  We were pleased to follow the reviewers' recommendations to use consistent, precise terminology and the correct tenses.  We have used the term "causation" because the underlying statistical test, the Granger-Sims causality test, claims to investigate it. However, we also understand the reservations about such statements, which is why we were happy to implement your suggestion and now speak of “prospective effects” instead throughout Results and Discussion sections. |
| 1. More generally, it would be good to give the paper a final **edit for clarity and typos** as I noted some of those when I read the paper (e.g., scrabbles should likely be Scrabble, on p. 7; “in for the young adults” on page 33 could probably just be “in the young adult sample”). |
| **Our response:**  Thank you for the careful reading. We now had a scientific translator who is a native English speaker proofread the manuscript and we hope that style is improved and that the typos are solved now. |
| 1. When **describing the RI-CLPM** on page 12, I think it would be useful to have a sub-heading to set-up that section. I would then call out to **Figure 1** right away so readers have a reference point for the parameters. I do not know that it is 100% necessary to use the Greek letters/old-school LISREL notation (that’s what I call it) in the figure but that is probably just stylistic. Likewise, the triangles could probably be deleted to enhance clarity, but I will leave that up to your team. |
| **Our response:**  Thank you for this remark. We have been happy to implement the following subheadings to improve the clarity of the section on Statistical Model:   * “*RI-CLPM: Distinguishing Between-Person From Within-Person Effects.* * *Interpretation of RI-CLPM* * *Granger-Sims Causality Tests* * *Moderation Effects of Age Groups in RI-CLPM”* (pp. 13-15)   Further we now call out to Figure1 right away at the beginning of section on page 12.  While we can see that the used Greek letters notation might be confusing for some readers, we personally still prefer this notation as it is a more interdisciplinary approach. In addition, it facilitates comparisons with the work we cite, see e.g., Zyphur et al., 2020, that uses this notation as well. |
| 1. It might be useful to give readers some additional pieces of **descriptive information** such as how the leisure activities correlated with each other as well as the zero-order stabilities of those variables. |
| **Our response:**  As suggested, we now provide more information on descriptive statistics in the Results section:  “[…] We present means, standard deviations, minimums, and maximums of all study variables across waves in Table 1 and by age groups in Supplement 1 (S1). Zero-order stabilities of leisure activities, that is, the correlation of one measurement occasion with the following, vary around *r* = .50, ranging between *r* = .42 for socializing across waves and *r* = .60 for physical activity across waves. For comparison: The zero-order stability of the Big Five personality traits ranges between *r* = .53 and *r* = .69. The strongest correlation between leisure activities was observed between physical activities and going-out activities (*r* = .40 at t1). The weakest association was found for political activities and socializing (*r* = 0.03 at t1). Intercorrelations among all leisure activities and Big Five personality traits over four waves can be found in the supplemental materials (S2). Of the Big Five trait domains, openness had the strongest concurrent associations with overall participation in leisure activities (*rs* of .22 −.28).” (p. 15) |
| 1. I liked the **summary table** idea as I think readers will benefit from a structured section with take away messages given the large number of analyses. However, I found **Table 9** a bit hard to apprehend at first, so I encourage you to consider alternative ways to display the information. For example, I did not find the italics to be helpful and it made the table harder to read. You might consider doing away with the Xs and just using plusses and minuses and not having anything in italicized font. You could delete the political activities row and just add to the note that nothing was p < .05. These are just ideas and provided in the spirit of making the paper maximally accessible. |
| **Our response:**  Thank you for this remark. We have revised Table 9 summarizing the cross-lagged effects accordingly. We think that the results are now even more easily visible immediately. However, we have kept the line on political activities so that it is clear at first glance that there were no significant effects here. We now hope that Table 9 can be understood more easily. |
| 1. In thinking through all of the results, I was a bit concerned about the sheer number of analyses. I don’t have strong opinion here, but you might want to adopt a **more stringent p-value** for identifying a discovery or at least acknowledge this issue a bit more in the paper. |
| **Our response:**  We understand the concerns and why one would consider adopting a more stringent p-value, given the large number of tests. However, we remain with the larger probability of error. Due to the study's limitation of measuring only every four years, the potential effects are very small. For future studies, we would recommend a narrower measurement interval and a more stringent p-value. We acknowledge this issue in the limitations section:  “It is also important to bear in mind the possible alpha inflation (type-I error) due to multiple testing in this study. It may be that we are reporting effects that do not actually exist. However, we decided not to use a stricter significance threshold than *p*<.05. Due to the long measurement intervals and the small effects that can therefore be expected, the risk of overlooking an effect (type-II error) that can be examined more thoroughly in future studies would have been high.” (p.49) |

# Reviewer #1

|  |
| --- |
| The current manuscript has a number of notable improvements on the original manuscript and presents quite interesting data. The reframing of the research to focus on the specific leisure activities as well as overall activity level and the refining focus to this goal greatly enhances the delivery of this interesting research. The additional details regarding the model and modeling choices is also valuable. While the choices in analysis and granularity make for a more complex set of findings, Table 9 helps organize the multiple results well. The addition of age groups is also a substantial add both from the perspective of interest and analytic insights. I have a few additional suggestions that I hope will further enhance and refine this work. |
| **Our response:**  Thank you for your positive comments! |
| My primary suggestion is to more deeply address the one caveat mentioned that “age is a proxy for unknown confounders for development” (p. 43). This can be paired with additional limitations and future directions around the measurement of leisure activities. Specifically, more can be done both in the **discussion and** in **future work** to explore why the patterns may be present in the data. Age and developmental context could potentially inform this but so could other and/or related factors. Right now the measure only addresses frequency but what about other aspects of the leisure activities such as how rewarding they are, how voluntary they are, and how enjoyable they are? The example in the intro using the TESSRA model is helpful – perhaps one or two could help these points in the discussion as well. |
| **Our response:**  Thank you very much for this valuable reflection. To address your comment, we have extended the discussion of the findings in several places:  In the section on “How do our findings relate to the TESSERA framework of personality development?” on page 48, we now provide an example to illustrate what influence aspects of leisure activities might have beyond frequency:  “Applying the TESSERA framework on personality development in adulthood (Wrzus & Roberts, 2017) to our results, overall participation in leisure activities and the activities of socializing (i.e., meeting friends, family, acquaintances etc.) and volunteering may be triggering situations for extraversion development. According to the theory, extraversion may change because a triggering situation is repeatedly perceived as trait-relevant and thus elicits a relevant state—e.g., talking at length about one’s own experiences—that does not correspond to the actual low extraversion, but nevertheless elicits a positive reaction: e.g., people are grateful or interested in meeting again. The same mechanism can be applied to other traits and situations: Socializing and going-out activities may be triggering situations for conscientiousness development. And a change in agreeableness may be triggered by a change in volunteering, according to our results.  According to this framework, internal reflective or associative processes transform repeated TESSERA sequences into long-term personality development. A note of caution should be added here since we did not assess these cognitive processes. Our measure of leisure activity only addresses frequency. However, cognitive and emotional aspects of leisure activities such as how rewarding they are, how voluntary they are, or how enjoyable they are could have an influence on related personality changes as well. In the above-described relationship between socializing and extraversion, for example, it would be decisive that the reaction of the environment is perceived as positive.” (p.48)  Further, we now discuss the findings in more depth and against the background of additional leisure activity characteristics, e.g., regarding socializing:  The hypothesis on agreeableness was not confirmed. After estimating between-person differences, we still found simultaneous temporary deviations between socializing and agreeableness in the same directions but found no cross-lagged effects for socializing and agreeableness at the within-person level. It may be that the frequency of socializing by itself is not a trait-triggering situation for agreeableness. Hence, it could conceivably be hypothesized that only meetings with emotionally secure attachment figures are state-relevant in the sense of the TESSERA framework. However, we did not capture the valence or other characteristics of leisure activities in this study.” (p. 44)  Additionally, in the Limitations and Future Directions section, we now elaborate more on the importance of other aspects of leisure activities:  “Age and developmental contexts could potentially inform the relationship between leisure activities and personality change, but so could other factors. To better understand the mechanisms underlying the link between change in participation in leisure activities and change in personality, future research might utilize in-vivo data on thoughts and feelings centered on the environmental changes (e.g., Wrzus & Roberts, 2017). This would enable them to examine the relevance of leisure activity characteristics beyond their frequency, such as voluntariness, enjoyment, and consequences and reactions. “(p.49) |
| On a smaller note, please include the mean of the **age groups** not just the year of birth to help be more clear in what the age groups mean. Additionally, perhaps it has just been a while since I sat down with the RI-CLPM but I am confused by the use of “unexpected change” in describing the results. “Short-term boost” also seems a bit confusing given the 4-year intervals of the study. A general read for clarity on terms like this would help deliver the take home message of the findings. |
| **Our response:**  Thank you for pointing out these unclarities. As there was some confusion about the age groups, we have now added a short paragraph on this in the Methods section (see pages 10-11). We now also include the corresponding age range at T1.  “Unexpected change” in this instance simply means that this change was not predicted by the underlying statistical model and thus occurs as a “shock” or “disturbance” to the system. We now explain that term in more detail in the Method section (p. 14).  It is important to note that in the RI-CLPM, cross-lagged effects capture only temporary effects of one construct on the other. In this paper we refer to those as “short-term boosts”, which might be misleading due to the four-year interval between the respective timepoints. However, as the RI-CLPM cannot detect sustained prospective effects over multiple timepoints, which can be seen as a flaw of the model (see Orth, Clark, Donnellan & Robins, 2020), this description seemed like an acceptable solution. We have supplemented the explanation of RI-CLPM terms in the manuscript at page 14 accordingly. |
| In sum, these are interesting findings from a large, respected panel study that have the potential to catalyze future work in the area. |
| **Our response:**  Thank you for appreciating our efforts! |

# Reviewer #2

|  |
| --- |
| These revisions were well done and, though they required quite a lot of changes, have vastly improved the paper. I have four comments that concern changes made in revision. |
| **Our response:**  Thank you for your positive evaluation and the appreciation of our efforts! |
| First, the authors should report and interpret **correlated change effects** (in terms of the authors’ model: ψ(lvi, lui)) alongside their cross-lagged effects. These parameters test whether temporary deviations from a person’s personality trait score are associated with simultaneous temporary deviations from their leisure activities score. They’re an essential piece of the within-person developmental puzzle. |
| **Our response:**  We thank you for the comment. We now pay more attention to correlated change effects throughout the manuscript. That is, we now describe this parameter and its interpretation in the Method section in more detail:  “The covariance of the phantom variables at each measurement occasion (ψ(lvi, lui)) shows whether temporary deviations from a person’s leisure activities score are associated with simultaneous temporary deviations from their personality trait score. We call this covariance “within-person correlated change effects” in the following.” (p. 14)  We have further added the correlated change values to the results tables in the manuscript. In addition, we now report on them in the Results section with a few sentences each. Please see pages 18 – 42.  Furthermore, we now include correlated change effects in the discussion of the results, e.g.:  “We expected that changes in the frequency of artistic and musical activities may trigger changes in openness and vice versa. […] We found no effect, however, of changes in artistic and musical activities on an individual’s openness. At first glance, this appears somewhat contradictory to earlier findings by Schwaba et al. (2018), […] The measurement occasions in the current study were four years apart, whereas participants in the research of Schwaba et al. (2018) completed the survey every year or every second year. In the present study, we also found simultaneous temporary deviation in the same direction of openness and artistic and musical activity within persons. This points to shorter cause-effect relationships.” (p. 46-47). |
| Second, as I have not encountered **time-varying unit effects** before, I am unsure if it is appropriate to use them here. Typically, unit effects are modeled as time-invariant, making them function as random intercepts that account solely for individual differences at baseline. When unit effects are modeled as time-variant, however, I am unsure if they then function as a combination of random intercepts and random slopes, modeling individual differences at baseline and also change across the study period. If they do (which, after reading the Zyphur et al., 2020 reference, might be the case), these time-varying unit effects may account for meaningful within-person developmental trends that the authors likely do not wish to remove from the cross-lagged portion of their model. In less jargony terms, personality development research usually treats sustained trends (e.g. a steady increase in extraversion) and deviations from that trend (e.g. an especially big increase in extraversion at t2) as important variance to explain. The time-varying unit effect may remove the sustained trend from the model. Because this may have a drastic effect on the results (or no effect at all!), I request that the authors estimate additional variants of their basic models that include time invariant unit-effects (i.e. all factor loadings on the unit effect are set to 1). If the authors’ cross-lagged and correlated change findings are quite different across time-varying and time-invariant unit effect models, I would suggest that they primarily interpret the time-invariant unit effect models, as it is much more clear how to interpret those findings. |
| **Our response:**  Thank you very much for the in-depth examination of our methods—we appreciate that! We assume that even “stable” between-person differences do not remain exactly the same over 13 years. This would be too strong an assumption to make over such a long period of time. We therefore estimated the unit effects separately and did not equate them across all four measurement points. While we certainly understand the argument that was made here and think that this might actually be a reasonable assumption to make, we do not think that time-varying unit effects may account for meaningful within-person developmental trends. An additional latent construct would be needed to properly capture those (see also Zyphur et al., 2020, Appendix, for further reasoning).  However, we were very eager to do the check. We therefore calculated the full models once again with time-invariant unit effects, that is, with all factor loadings fixed to 1, and compared them to the non-restricted models (time-varying unit effects). Within-person auto-regressive terms tend to be a bit smaller if unit effects are allowed to differ across time, compared to time-invariant unit effects. But the differences in within-person auto-regressive effects had no influence on the significance and ranged, if at all, around 2 to 3 hundredths. We found no noteworthy differences of the cross-lagged effects and the correlated change effects.  An exception was the relationship between some leisure activities and openness. At the between-person level, there was the greatest positive correlation of leisure activities compared too other personality traits. Here, the restriction of the factor loadings of the unit effect led to a few hundredths larger within-person auto-regressive, cross-lagged terms and correlated change effects compared to the time-varying unit effect models.  Due to these minor changes caused by equating the factor loadings of the unit effects, we decided to stay with our original, time-varying unit effect models. They have a better model fit to data and, in our opinion, also represent the real world more accurately over this long measurement period.  In the manuscript, we now briefly refer to these findings in footnote in the Methods section on page 13:  “4The comparison of time-invariant and time-varying unit effect models showed that the within-person cross-lagged terms and correlated change effects generally do not differ. The within-person auto-regressive terms are slightly larger in the more restricted models. But there was no relevant difference in the significance of the parameters. For full results on restricted versus unrestricted factor loadings of unit effects, please see R scripts and outputs on OSF: <https://osf.io/fdxzp/?view_only=eaffa966ffe14e7c8a4ccb5ace76ace6> “ |
| Third, I would appreciate clarification in the manuscript about **how age effects were estimated**. Were they estimated by adding a single covariate? How were age groups compared with one another? Are differences across age groups robust enough to merit fine-grained interpretation in the discussion? |
| **Our response:**  Age effects were estimated by adding a categorical grouping variable in a multiple group analysis (please see Mulder & Hamaker, 2020). This approach implies that not only the means can differ across the age groups, but also the lagged regression coefficients, the (residual) variances, and the (residual) covariances. Age group differences were investigated by comparing a multiple group version of the RI-CLPM in which there are no constraints across the age groups with a model in which the lagged regression coefficients are constrained to be identical across the groups. If the chi-square difference test indicated that this constraint cannot be imposed, this implies that (some of) the lagged coefficients differ across the groups: The lagged effects of the variables on each other depend on the level of the grouping variable. In contrast, when the equality constraints on the lagged parameters across the groups hold, this implies there is no moderation effect. Thus, by chi-square difference test we know whether there are age group differences, but not between which age groups these occur. We now describe the procedure in more detail in the Methods section:  “*Moderation Effects of Age Groups in RI-CLPM*  Subsequently we expanded the RI-CPLM by adding the categorical age grouping variable to the model. Age group differences were investigated by comparing a multiple group version of the RI-CLPM in which there are no constraints across the age groups with a model in which the lagged regression coefficients are constrained to be identical across the groups (see Mulder & Hamaker, 2020). We therefore know from chi-square difference tests whether there is a moderation effect of age group or not. To keep the number of models at a presentable and parsimonious level, we decided to test only whether the full model would show differences between the age groups and present the results in the supplemental materials S1 and S3-S9 only. […]” (p.15)  Further, we double-checked whether we are using descriptive vocabulary to report on age group effects throughout Results and Discussion section.  Additionally, we now refer to this issue again in the Limitations section:  “Our results on age effects should also be interpreted with caution. We obtained a first impression of the impact of age through model fit testing with and without constraints across age groups, but the mechanisms in the different age groups should be investigated in more detail in further studies. […]” (p.49) |
| Fourth, the **tense of language** in the revised parts of the introduction and discussion sections is somewhat imprecise. The authors toggle between past-tense and present-tense in both the introduction and the results. I suggest that they frame results of both past research and this study in the past tense, as psychological research generally describes what has been found (e.g. people who did X were Y (in Z et al.)) not what is universally true (e.g. people who X are Y). |
| **Our response:**  Thank you very much for pointing this out. We have revised both the introduction and the discussion so that both are now in past tense. |

This summarizes the ways in which we have responded to your and the reviewers’ constructive comments. We believe that we have addressed all issues that were raised in a satisfactory manner and hope you will find this once again revised manuscript acceptable for publication in *Collabra: Psychology – Personality Psychology.* We appreciate your continued consideration of this manuscript.

Sincerely yours,

Julia Sander

(on behalf of all authors)

**Editor Final Decision—Accept**

Apr 23, 2021

Dear Julia Sander,

I have now had a chance to read over your manuscript “Leisure Activities as a Driver of Personality Development? A Random-Intercept Cross-lagged Panel Model Across 13 Years in Adulthood”, 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. Our managing editor will contact you in case there are any pre-prodution 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.

Sincerely, Brent Donnellan