From: Loïs Fournier (corresponding author), on behalf of all the authors involved (Alexandre Heeren, Stéphanie Baggio, Luke Clark, Antonio Verdejo-García, José C. Perales, and Joël Billieux)

To: Veli-Matti Karhulahti (recommender), Ivan Ropovik (reviewer), and Orestis Zavlis (reviewer)

Dear Veli-Matti Karhulahti, Ivan Ropovik, and Orestis Zavlis,

We were pleased to receive the decision with respect to the original version of our stage one registered report manuscript, *Development and evaluation of a revised 20-item short version of the UPPS-P Impulsive Behavior Scale* (Manuscript ID: 862), in your communication of August 27, 2024.

We thank you, Veli-Matti Karhulahti, for your willingness to consider a revised version of our manuscript. We also thank you, Veli-Matti Karhulahti, Ivan Ropovik, and Orestis Zavlis, for your positive evaluation of the original version of our manuscript. We certainly appreciate your insightful comments and consider that all evaluations contributed to improving the original version of our manuscript.

To facilitate the review of the revised version of our manuscript, comments are numbered and presented in table cells, below which are our responses. All revisions to the original version of our manuscript have been highlighted in green in the (.pdf) present document. All revisions to the original version of our manuscript have been highlighted using the "track changes" function in *Microsoft Word* in the (.docx) revised version of our manuscript.

I, Loïs Fournier, act as the corresponding author of the present revised version of our manuscript and remain at your disposal for any further information.

Sincerely,

Loïs Fournier (corresponding author), on behalf of all the authors involved (Alexandre Heeren, Stéphanie Baggio, Luke Clark, Antonio Verdejo-García, José C. Perales, and Joël Billieux)

Loïs Fournier, M.Sc. (lois.fournier@unil.ch)
Institute of Psychology, University of Lausanne, Lausanne, Switzerland

From: Loïs Fournier (corresponding author), on behalf of all the authors involved (Alexandre Heeren, Stéphanie Baggio, Luke Clark, Antonio Verdejo-García, José C. Perales, and Joël Billieux)

To: Veli-Matti Karhulahti (recommender)

Comment VMK-00

Dear Loïs Fournier and colleagues,

Thank you again for submitting to *PCI RR*, and my apologies for the small delay in this stage one review. Due to the summer holidays and some coincidences, it took me longer than usual to find suitable reviewers, and, unfortunately, one of the tentatively consenting reviewers with topic expertise had to turn down the task in the end. Considering the months that have already passed since your submission, I decided to move onwards, exceptionally, with only two reviews. To compensate for the gap in construct-specific feedback, I do my best to leave related comments myself, having followed the impulsivity literature to some limited extent.

I have not been involved in research projects explicitly on impulsivity, so the comments are mainly based on my understanding of the theoretical literature. You're naturally free to rebut any of the suggestions if they involve misunderstandings. Please also carefully consider the methodological concerns expressed by the second reviewer. I hope you find the feedback useful overall, and you're naturally welcome to contact me during the process if any related questions occur. By the way, this was among the most carefully written initial stage one submissions I've handled so far.

Best wishes

Veli-Matti Karhulahti

Dear Veli-Matti Karhulahti,

Thank you for agreeing to be the recommender for our stage one registered report. We believe that the review process was timely, especially during the summer holidays, and that all comments, both in terms of quality and quantity, greatly contributed to improving the original version of our manuscript. To facilitate the review of the revised version of our manuscript, comments are numbered and presented in table cells, below which are our responses. All revisions to the original version of our manuscript have been highlighted in green in the (.pdf) present document. All revisions to the original version of our manuscript have been highlighted using the "track changes" function in *Microsoft Word* in the (.docx) revised version of our manuscript. Thank you for your positive message recognizing the care with which we prepared our submission!

Loïs Fournier (corresponding author), on behalf of all the authors involved (Alexandre Heeren, Stéphanie Baggio, Luke Clark, Antonio Verdejo-García, José C. Perales, and Joël Billieux)

Loïs Fournier, M.Sc. (lois.fournier@unil.ch)
Institute of Psychology, University of Lausanne, Lausanne, Switzerland

Comment VMK-01

> Although impulsivity has been a useful term for communicating certain entities related to human psychology and psychopathology, there has also been active discussion concerning its conceptual and theoretical assumptions as a construct (see Fried, 2020; Zavlis & Fried, 2024). I understand this goes mostly beyond scale development, and I would not expect the paper to address such questions in detail. However, it would make the work stronger if the assumptions behind the chosen theory of impulsivity were more explicit. I know the authors have worked on this topic for a long time and are well aware of various conceptual and theoretical viewpoints. It would likely be a relatively small effort to address this briefly in the introduction, even though building on the existing UPPS-P work sets limits to considering it in practice.

We acknowledge the present comment. Indeed, we are familiar with the references listed in the present comment. In line with the work of Fried (2020) and of Zavlis and Fried (2024), we share the scientific view that a single, unitary construct of impulsivity (e.g., an "*i*-factor") is untenable. This echoes criticisms that a single, unitary construct of psychopathology (e.g., a "*p*-factor") would be no adequate or useful summary of psychopathology variation (Watts et al., 2023). Critically, the UPPS-P Impulsive Behavior Model (Cyders et al., 2007; Whiteside & Lynam, 2001) accounts for the non-unitary nature of impulsivity (Evenden, 1999; Sharma et al., 2014) by conceptualizing it as a multidimensional psychological construct encompassing five facets. Moreover, distinguishing among these five facets is crucial: substantial evidence from the scientific literature has highlighted differential transdiagnostic associations between the facets and symptoms of various disorders (Berg et al., 2015; Rochat et al., 2018; Smith et al., 2007). Therefore, the first paragraph of the "Introduction" section of the revised version of our manuscript now reads as follows:

"*Impulsivity* is a key psychological construct integrated into most major personality models (Whiteside & Lynam, 2001) and stands as one of the most prevalent diagnostic criteria in foremost nosography manuals (American Psychiatric Association, 2022; World Health Organization, 2024). Among prevailing models of impulsivity, the UPPS-P Impulsive Behavior Model (Cyders et al., 2007; Whiteside & Lynam, 2001) accounts for the non-unitary nature of impulsivity (Evenden, 1999; Sharma et al., 2014) by conceptualizing it as a multidimensional psychological construct encompassing five distinct facets which are differentially associated with numerous psychopathological and neuropathological symptoms across various disorders (Berg et al., 2015; Rochat et al., 2018; Smith et al., 2007): (1) *lack of premeditation* (i.e., lack of reflection on the potential consequences of actions preceding their emission), (2) *positive urgency* (i.e., emission of sudden actions in intense positive emotional states), (3) *sensation seeking* (i.e., attraction to excitement and openness to new experiences), (4) *negative urgency* (i.e., emission of sudden actions in intense negative emotional states), and (5) *lack of perseverance* (i.e., difficulty sustaining focus on demanding or monotonous tasks).".

Berg, J. M., Latzman, R. D., Bliwise, N. G., & Lilienfeld, S. O. (2015). Parsing the heterogeneity of impulsivity: a meta-analytic review of the behavioral implications of the UPPS for psychopathology. *Psychological Assessment*, *27*(4), 1129–1146. https://doi.org/10.1037/pas0000111

Cyders, M. A., Smith, G. T., Spillane, N. S., Fischer, S., Annus, A. M., & Peterson, C. (2007). Integration of impulsivity and positive mood to predict risky behavior: development and validation of a measure of positive urgency. *Psychological Assessment*, *19*(1), 107–118. https://doi.org/10.1037/1040-3590.19.1.107

Evenden, J. L. (1999). Varieties of impulsivity. *Psychopharmacology*, *146*(4), 348–361. https://doi.org/10.1007/PL00005481

Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry*, *31*(4), 271–288. https://doi.org/10.1080/1047840X.2020.1853461

Rochat, L., Billieux, J., Gagnon, J., & Van Der Linden, M. (2018). A multifactorial and integrative approach to impulsivity in neuropsychology: insights from the UPPS model of impulsivity. *Journal of Clinical and Experimental Neuropsychology*, *40*(1), 45–61. https://doi.org/10.1080/13803395.2017.1313393

Sharma, L., Markon, K. E., & Clark, L. A. (2014). Toward a theory of distinct types of impulsive behaviors: a meta-analysis of self-report and behavioral measures. *Psychological Bulletin*, *140*(2), 374–408. https://doi.org/10.1037/a0034418

Smith, G. T., Fischer, S., Cyders, M. A., Annus, A. M., Spillane, N. S., & McCarthy, D. M. (2007). On the validity and utility of discriminating among impulsivity-like traits. *Assessment*, *14*(2), 155–170. https://doi.org/10.1177/1073191106295527

Watts, A. L., Greene, A. L., Bonifay, W., & Fried, E. I. (2023). A critical evaluation of the p-factor literature. *Nature Reviews Psychology*, *3*(2), 108–122. https://doi.org/10.1038/s44159-023-00260-2

Whiteside, S. P., & Lynam, D. R. (2001). The five-factor model and impulsivity: using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, *30*(4), 669–689. https://doi.org/10.1016/S0191-8869(00)00064-7

Zavlis, O., & Fried, E. I. (2024). *No evidence for a psychological trait of impulsivity*. https://doi.org/10.31234/osf.io/m2xdj

Comment VMK-02

> As an example, I've personally found the conceptual distinction between 'stopping' and 'waiting' impulsivity helpful when interpreting clinical data, as in the inability to stop using social media (after starting) tends to manifest differently than the inability to resist starting (e.g., Dalley et al., 2011; Dalley & Ersche, 2018). At the same time, dimensions like novelty/sensation seeking have their own literature, and various parallel theories have been proposed (e.g., Mestre-Bach et al., 2020). I am noting this due to the increasing concern over multiple overlapping measures (e.g., Elson et al., 2023) – again, your team is clearly aware of this (and addressing the jingle-jangle fallacy on p. 2), but further reasoning for the superiority or utility of the chosen model would be helpful in the introduction.

We acknowledge the present comment. Yet, the cognitive, affective, behavioral, and motivational aspects underlying the impulsive behavior dimensions described in the UPPS-P Impulsive Behavior Model (Cyders et al., 2007; Whiteside & Lynam, 2001) are beyond the scope of the present registered report. Still, we provided a brief overview of these aspects in the "Introduction" section by adding a statement which reads as follows:

"Extensive research has investigated the cognitive, affective, behavioral, and motivational mechanisms underlying the impulsive behavior dimensions described in the UPPS-P Impulsive Behavior Model, including decision-making processes in complex conditions (mainly associated with lack of

premeditation), prepotent response inhibition in intense emotional states (mainly associated with negative and positive urgency), approach-avoidance system functioning (mainly associated with sensation seeking), and resistance to proactive interference in working memory (mainly associated with lack of perseverance) (Bechara & Van Der Linden, 2005; Gay et al., 2008; Rochat et al., 2018).".

Moreover, as per the present comment and per "Comment VMK-03", in the "Box 2.1. An expert-driven methodological approach to construct-level content validity in short-form development and evaluation" section, we provided additional information regarding the aspects supporting our evaluations of the item-level content validity of the items of the 59-item original version of the UPPS-P Impulsive Behavior Scale (UPPS-P-59; Cyders et al., 2007; Whiteside & Lynam, 2001), which is reported in our response to "Comment VMK-03".

With respect to the second part of the present comment, in fact, similar reasoning to that reported in the work of Elson et al. (2023) – and in the present comment – led us to prefer to develop and evaluate the psychometric properties of a revised 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20-R) rather than a novel psychometric instrument. Given the considerable popularity of the 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20; Billieux et al., 2012) in clinical and research settings, we strongly believe it best to provide clinicians and researchers with a revised version of a psychometric instrument that they are familiar with. Therefore, in line with the metaphor of Elson et al. (2023), we strongly believe that the revised 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20-R), unlike toothbrushes, will be reused by others.

Bechara, A., & Van Der Linden, M. (2005). Decision-making and impulse control after frontal lobe injuries. *Current Opinion in Neurology*, *18*(6), 734–739. https://doi.org/10.1097/01.wco.0000194141.56429.3c

Billieux, J., Rochat, L., Ceschi, G., Carré, A., Offerlin-Meyer, I., Defeldre, A.-C., Khazaal, Y., Besche-Richard, C., & Van Der Linden, M. (2012). Validation of a short French version of the UPPS-P Impulsive Behavior Scale. *Comprehensive Psychiatry*, *53*(5), 609–615. https://doi.org/10.1016/j.comppsych.2011.09.001

Cyders, M. A., Smith, G. T., Spillane, N. S., Fischer, S., Annus, A. M., & Peterson, C. (2007). Integration of impulsivity and positive mood to predict risky behavior: development and validation of a measure of positive urgency. *Psychological Assessment*, *19*(1), 107–118. https://doi.org/10.1037/1040-3590.19.1.107

Elson, M., Hussey, I., Alsalti, T., & Arslan, R. C. (2023). Psychological measures aren't toothbrushes. *Communications Psychology*, *1*(1), 25. https://doi.org/10.1038/s44271-023-00026-9

Gay, P., Rochat, L., Billieux, J., d'Acremont, M., & Van Der Linden, M. (2008). Heterogeneous inhibition processes involved in different facets of self-reported impulsivity: evidence from a community sample. *Acta Psychologica*, *129*(3), 332–339. https://doi.org/10.1016/j.actpsy.2008.08.010

Rochat, L., Billieux, J., Gagnon, J., & Van Der Linden, M. (2018). A multifactorial and integrative approach to impulsivity in neuropsychology: insights from the UPPS model of impulsivity. *Journal of Clinical and Experimental Neuropsychology*, *40*(1), 45–61. https://doi.org/10.1080/13803395.2017.1313393

Whiteside, S. P., & Lynam, D. R. (2001). The five-factor model and impulsivity: using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, *30*(4), 669–689. https://doi.org/10.1016/S0191-8869(00)00064-7

Comment VMK-03

> The above could also aid further in explaining item exclusions/edits (pp. 6-7), which have been outlined clearly but partially lack theoretical justification. For example, when three authors evaluated item-level content validity (p. 6), the ontology against which the assessment took place is not mentioned. This becomes relevant, e.g., when modifying items to correctly measure 'negative urgency' (p. 7), which involves the distinction between negative/positive urgency as a premise. The rationale appears to be that the assessment was done against the original five-dimensional model but justifying that model (over its alternatives) could make the work even more convincing. One reviewer suggests spelling out RQs for each phase. This would be an excellent opportunity to clarify the goals related to UPPS-P and impulsivity construct(s) in general.

We acknowledge the present comment. As outlined in the "Introduction" section of our manuscript, the original development of the UPPS-P Impulsive Behavior Model and its corresponding assessment tool (Cyders et al., 2007; Whiteside & Lynam, 2001) unified twenty-one coexisting conceptualizations of impulsivity, thereby addressing the jingle and jangle fallacies across the various theoretical models and assessment tools that characterized the research field. Moreover, as per the "Introduction" section of our manuscript and our response to "Comment VMK-01", substantial evidence from the scientific literature has replicated the five-factor structure of the UPPS-P Impulsive Behavior Scale across numerous adaptations and highlighted the importance of distinguishing among these five facets due to their differential transdiagnostic associations with symptoms of various disorders (Berg et al., 2015; Rochat et al., 2018; Smith et al., 2007). In this light, the well-established five-factor UPPS-P Impulsive Behavior Model served as the framework and ontology for our expert-driven methodological approach to construct-level content validity.

As per the present comment and per "Comment VMK-02", which we acknowledge, in the "Box 2.1. An expert-driven methodological approach to construct-level content validity in short-form development and evaluation" section, we provided additional information regarding the aspects supporting our evaluations of the item-level content validity of the items of the 59-item original version of the UPPS-P Impulsive Behavior Scale (UPPS-P-59; Cyders et al., 2007; Whiteside & Lynam, 2001), which reads as follows:

"The first category regards items that fail to provide relevant information about one's tendency to seek excitement and to be open to new experiences (i.e., sensation seeking) due to their specificity and dependence on external factors: one might present a high level of endorsement of sensation seeking, but might not like or enjoy the specific activities listed in such items, or might not be able to access such activities due to environmental, financial, mental, or physical factors. Moreover, one might like or enjoy the specific activities listed in such items independently of their level of endorsement of sensation seeking (Maples-Keller et al., 2016).".

"The second category regards items that fail to provide relevant information about one's emission of sudden actions in intense negative or positive emotional states (i.e., negative urgency, positive urgency) due to their specificity and explicit focus on "cravings": one might present a high level of endorsement of negative urgency or positive urgency but might not experience the "cravings" listed in such items. Moreover, the explicit focus on "cravings" listed in such items conflates impulsive behavior

with a feature that is specific to disorders due to substance use and addictive behaviors: whereas the associations between craving and symptoms across such disorders are indeed well-established, the associations between craving, negative urgency, and positive urgency are inconsistent across original research articles (López-Guerrero et al., 2023).".

"The third and last category regards items that fail to provide relevant information about one's emission of sudden actions in intense negative emotional states (i.e., negative urgency), for they do not incorporate the intense negative emotional states of the impulsive behavior dimension they purportedly assess. Moreover, the incorporation of the negative or positive intense emotional states in which sudden actions are emitted is relevant, as negative urgency and positive urgency are conceptualized as two distinct facets which are differentially associated with symptoms across various disorders (Cyders & Smith, 2007).".

With respect to the outlining of the research questions pertaining to our development and evaluation protocol, we acknowledge and agree with the corresponding comments of Orestis Zavlis (i.e., "Comment OZ-02", "Comment OZ-04"), which you emphasized in the present comment. All the research questions now outlined are listed in our response to "Comment OZ-02", yet we suggest referring to the revised version of our manuscript for a contextual presentation of the research questions outlined.

Berg, J. M., Latzman, R. D., Bliwise, N. G., & Lilienfeld, S. O. (2015). Parsing the heterogeneity of impulsivity: a meta-analytic review of the behavioral implications of the UPPS for psychopathology. *Psychological Assessment*, *27*(4), 1129–1146. https://doi.org/10.1037/pas0000111

Cyders, M. A., & Smith, G. T. (2007). Mood-based rash action and its components: positive and negative urgency. *Personality and Individual Differences*, *43*(4), 839–850. https://doi.org/10.1016/j.paid.2007.02.008

Cyders, M. A., Smith, G. T., Spillane, N. S., Fischer, S., Annus, A. M., & Peterson, C. (2007). Integration of impulsivity and positive mood to predict risky behavior: development and validation of a measure of positive urgency. *Psychological Assessment*, *19*(1), 107–118. https://doi.org/10.1037/1040-3590.19.1.107

López-Guerrero, J., Navas, J. F., Perales, J. C., Rivero, F. J., & Muela, I. (2023). The interrelation between emotional impulsivity, craving, and symptoms severity in behavioral addictions and related conditions: a theory-driven systematic review. *Current Addiction Reports*, *10*(4), 718–736. https://doi.org/10.1007/s40429-023-00512-4

Maples-Keller, J. L., Berke, D. S., Few, L. R., & Miller, J. D. (2016). A review of sensation seeking and its empirical correlates: dark, bright, and neutral hues. In V. Zeigler-Hill & D. K. Marcus (Eds.), *The dark side of personality: science and practice in social, personality, and clinical psychology* (pp. 137–156). American Psychological Association. https://doi.org/10.1037/14854-008

Rochat, L., Billieux, J., Gagnon, J., & Van Der Linden, M. (2018). A multifactorial and integrative approach to impulsivity in neuropsychology: insights from the UPPS model of impulsivity. *Journal of Clinical and Experimental Neuropsychology*, *40*(1), 45–61. https://doi.org/10.1080/13803395.2017.1313393

Smith, G. T., Fischer, S., Cyders, M. A., Annus, A. M., Spillane, N. S., & McCarthy, D. M. (2007). On the validity and utility of discriminating among impulsivity-like traits. *Assessment*, *14*(2), 155–170. https://doi.org/10.1177/1073191106295527

Whiteside, S. P., & Lynam, D. R. (2001). The five-factor model and impulsivity: using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, *30*(4), 669–689. https://doi.org/10.1016/S0191-8869(00)00064-7

From: Loïs Fournier (corresponding author), on behalf of all the authors involved (Alexandre Heeren, Stéphanie Baggio, Luke Clark, Antonio Verdejo-García, José C. Perales, and Joël Billieux)

To: Ivan Ropovik (reviewer)

Comment IR-00

Thanks to the authors for the opportunity to read their manuscript. Overall, I think that the proposed study would be informative. It is a well-designed and thought-through validation study. I especially liked the approach to item subset selection by means of network models, which I see as conceptually strong. The authors plan to collect and examine several types of validity evidence and make them an integral part of short-form development.

That said, I also have some critical takes and suggestions for improvement. An acknowledgment upfront: I don't have much expert knowledge about the substantive aspects of the constructs measured by the present validation study. In my review, I will mainly focus on the measurement, design, and analysis side of things. As my role as a reviewer is mainly to provide critical feedback, I provide it in the form of comments below, not in order by importance but rather chronologically as I read the paper. I leave it to the authors' discretion which suggestions they find sensible and choose to incorporate. I hope that the authors find at least some of the suggestions below helpful.

Best wishes,

Ivan Ropovik

Dear Ivan Ropovik,

Thank you for agreeing to review our stage one registered report. We have carefully considered each of your suggestions, which, from a methodological point of view, have sparked many stimulating discussions between us. To facilitate the review of the revised version of our manuscript, comments are numbered and presented in table cells, below which are our responses. All revisions to the original version of our manuscript have been highlighted in green in the (.pdf) present document. All revisions to the original version of our manuscript have been highlighted using the "track changes" function in *Microsoft Word* in the (.docx) revised version of our manuscript. Thank you again for your time and expertise. We truly appreciate your positive and encouraging feedback on our development and evaluation protocol.

Sincerely,

Loïs Fournier (corresponding author), on behalf of all the authors involved (Alexandre Heeren, Stéphanie Baggio, Luke Clark, Antonio Verdejo-García, José C. Perales, and Joël Billieux)

Loïs Fournier, M.Sc. (lois.fournier@unil.ch)
Institute of Psychology, University of Lausanne, Lausanne, Switzerland

Comment IR-01

> I think the authors outline the risks of retaining items based on factor loadings well. However, narrowing down the construct breadth is not the only risk. It is also the fact that a "good" item among "poor" items tends to get low loading, and being a naive empiricist in this sense runs the risk that the highly correlating poor items (that measure the given construct poorly) hijack the construct validity of any scale's score.

We agree with the present comment. In fact, similar reasoning to that which you report in the present comment led us to supplement our data-driven methodological approach (described in "Box 2.2." in the original version of our manuscript) with our expert-driven methodological approach to construct-level content validity in short-form development and evaluation (described in "Box 2.1." in the original version of our manuscript).

Comment IR-02

> I suppose that each of the three exclusion criteria is sufficient for a participant exclusion. If this is the case, I suggest it is clearly stated. Also, especially given that it is an online-based data collection with limited fidelity control, it would make sense to try to screen out careless responders (e.g., based on long string detection or some sort of insufficient variance in responding pattern or being a multivariate outlier indicating random response pattern). For example, the *careless R* package (Yentes & Wilhelm, 2023) might prove useful.

With respect to participant data exclusion criteria, we acknowledge the present comment and have added a statement which reads as follows:

"Endorsement of any of the aforementioned criteria establishes exclusion.".

For consistency, with respect to participant data inclusion criteria, we have added a statement which reads as follows:

"Endorsement of all of the aforementioned criteria establishes inclusion.".

Two participant data exclusion criteria will be implemented with respect to "careless" respondents: attention checks and lower limits on the time it will take participants to complete the full online survey.

Attention checks (e.g., "This is an attention check. Please select "Agree strongly" as an answer to the present statement.") will be implemented in each participant data collection and in each psychometric instrument. In each participant data collection, failing the attention check implemented in the first psychometric instrument or – if applicable – failing two or more attention checks implemented in all psychometric instruments will result in participant data exclusion.

Lower limits on the time it will take participants to complete the full online survey will be implemented in each participant data collection. In each participant data collection, completing the full online survey in less time than three standard deviations below the mean time it will take participants to complete the full online survey will result in participant data exclusion.

Of note, endorsement of any of the two abovementioned criteria (i.e., attention checks, lower limits on the time it will take participants to complete the full online survey) establishes exclusion.

Comment IR-03

"We will exclude data from participants who will have failed to complete the full online survey". Why? If the responses do not show patterns of carelessness or exclusion criteria are not met, then certainly, some data is better than none. As you will be using SEM, you can leverage FIML to impute the missing data. This is conceptually a much sounder approach than using listwise deletion. Solving the missing data issue by using forcing responses comes with several – both methodological and ethical – problems. A better solution would be to use a single reminder that the participant missed some questions on the given page. It can be easily done in *Qualtrics* (https://www.qualtrics.com/).

The present comment sparked important exchanges between us with respect to our approach to participant data exclusion criteria and missing data handling.

In fact, following acknowledgment of the present comment, we first agreed that a full information maximum likelihood (FIML) approach would be the most suitable option for missing data handling. However, when Loïs Fournier (corresponding author) tested a full information maximum likelihood (FIML) approach on data collected on the 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20; Billieux et al., 2012) in the context of a previous original research article (Fournier et al., 2024)[*], the output consisted in an error message which reads as follows:

"Error: lavaan -> lav_options_set(): missing = "ml" not available in the categorical setting".

Indeed, with respect to the missing data handling argument (i.e., "missing"), the documentation accompanying the *R* package *lavaan* version 0.6-18 (Rosseel et al., 2024) reads as follows:

"If (part of) the data is categorical, and the estimator is from the (weighted) least squares family, the only option (besides listwise deletion) is pairwise deletion.".

Therefore, in accounting for the categorical nature of the raw data we will collect on all psychometric instruments, a full information maximum likelihood (FIML) approach cannot be adopted.

Additionally, in light of "Comment IR-04", we thank you for emphasizing that a multiple imputation (MI) approach to missing data handling would be impractical, which we fully agree with. Performing confirmatory factor, network, and correlation analyses on multiple imputed datasets would indeed prove to be a "major headache", and we strongly believe it best to avoid such a situation.

All in all, considering that a full information maximum likelihood (FIML) approach cannot be adopted and a multiple imputation (MI) approach would be impractical, we strongly believe that the best approach consists of requiring that participants provide answers to all statements implemented in the full online survey and to exclude data from participants who will have failed to complete the full online survey, as reported in the original version of the manuscript. Such an approach will trivially guarantee

---

[*] Employing the *R* packages *bootnet* version 1.6 (Epskamp, 2024) and *lavaan* version 0.6-18 (Rosseel et al., 2024) that are required to perform the planned confirmatory factor and network analyses, using categorical data as an input, specifying weighted least squares mean-and-variance-adjusted for the estimator argument, and specifying full information maximum likelihood (FIML) for the missing data handling argument.

the absence of missing data. Therefore, with respect to missing data, we added a statement which reads as follows:

"With respect to data collection, as we will require that participants provide answers to all statements implemented in the full online survey and as we will exclude data from participants who will have failed to complete the full online survey, no missing data will arise.".

Billieux, J., Rochat, L., Ceschi, G., Carré, A., Offerlin-Meyer, I., Defeldre, A.-C., Khazaal, Y., Besche-Richard, C., & Van Der Linden, M. (2012). Validation of a short French version of the UPPS-P Impulsive Behavior Scale. *Comprehensive Psychiatry*, *53*(5), 609–615. https://doi.org/10.1016/j.comppsych.2011.09.001

Epskamp, S. (2024). *bootnet: bootstrap methods for various network estimation routines* (1.6) [Computer software]. https://CRAN.R-project.org/package=bootnet

Fournier, L., Bőthe, B., Demetrovics, Z., Koós, M., Kraus, S. W., Nagy, L., Potenza, M. N., Ballester-Arnal, R., Batthyány, D., Bergeron, S., Briken, P., Burkauskas, J., Cárdenas-López, G., Carvalho, J., Castro-Calvo, J., Chen, L., Ciocca, G., Corazza, O., Csako, R. I., Fernandez, D. P., Fujiwara, H., Fernandez, E. F., Fuss, J., Gabrhelík, R., Gewirtz-Meydan, A., Gjoneska, B., Gola, M., Grubbs, J. B., Hashim, H. T., Islam, M. S., Ismail, M., Jiménez-Martínez, M. C., Jurin, T., Kalina, O., Klein, V., Költő, A., Lee, S.-K., Lewczuk, K., Lin, C.-Y., Lochner, C., López-Alvarado, S., Lukavská, K., Mayta-Tristán, P., Miller, D. J., Orosová, O., Orosz, G., Ponce, F. P., Quintana, G. R., Quintero Garzola, G. C., Ramos-Diaz, J., Rigaud, K., Rousseau, A., Scanavino, M. D. T., Schulmeyer, M. K., Sharan, P., Shibata, M., Shoib, S., Sigre-Leirós, V., Sniewski, L., Spasovski, O., Steibliene, V., Stein, D. J., Strizek, J., Sungkyunkwan University Research Team, Tsai, M.-C., Ünsal, B. C., Vaillancourt-Morel, M.-P., Van Hout, M. C., & Billieux, J. (2024). Evaluating the factor structure and measurement invariance of the 20-item short version of the UPPS-P Impulsive Behavior Scale across multiple countries, languages, and gender identities. *Assessment*, 10731911241259560. https://doi.org/10.1177/10731911241259560

Rosseel, Y., Jorgensen, T. D., & De Wilde, L. (2024). *lavaan: latent variable analysis* (0.6-18) [Computer software]. https://CRAN.R-project.org/package=lavaan

Comment IR-04

The suggested approach to handling missing data (which, according to the current rules, is super unlikely, as you plan to exclude anyone not finishing the survey) would turn out to be a major headache. Imagine needing to implement the entire analysis within a multiple imputation approach due to, say, less than 1% of missing data. I guess literally no one uses MI when doing latent modeling, where FIML is the natural candidate for missing data treatment.

With respect to the present comment, we suggest referring to our response to "Comment IR-03" for detailed information regarding missing data. Yet again, as per our response to "Comment IR-03", we thank you for emphasizing that a multiple imputation (MI) approach to missing data handling would be impractical, which we fully agree with. Performing confirmatory factor, network, and correlation analyses on multiple imputed datasets would indeed prove to be a "major headache", and we strongly believe it best to avoid such a situation.

Comment IR-05

> The "Data" part has rather weird structuring. I'd split the information presented there into "Participants" and "Analysis" (and possibly "Sample size determination").

We acknowledge the present comment. As per "Comment OZ-02", we outlined the research questions pertaining to our development and evaluation protocol and believe that the revised version of our manuscript now benefits from an improved presentation. All the research questions now outlined are listed in our response to "Comment OZ-02", yet we suggest referring to the revised version of our manuscript for a contextual presentation of the research questions outlined.

Comment IR-06

> The sample size determination is well done. You've arrived at a relatively tight boundary between the minimum sample size and the stopping rule. Maybe requiring good quality of adjustment (I have never heard that term, though, in this context) determined by the 95th percentiles is unnecessarily strict. 90% would do the job well enough, IMO, given your validation goals.

We thank you for your positive evaluation of our sample size determination protocol. With respect to our Monte Carlo simulation analyses, as we are inclined towards a conservative approach, we decided to rely on the 95th percentiles of the distributions of model-implied approximate fit indices, which is indeed more stringent than relying on the 90th percentiles.

Comment IR-07

> As you plan to use the WLSMV estimator, I presume that you plan to model the latent indicators as ordinal. If this is the case (as it should be), it needs to be stated explicitly.

We acknowledge the present comment. Indeed, we will account for the ordered categorical nature of the raw data we will collect on all psychometric instruments despite such an approach being implicit in the original version of our manuscript. Therefore, in the corresponding sections of the revised version of our manuscript, we added a statement which reads as follows:

"Of note, in fitting the structural equation model, the ordered categorical nature of the observed variables will be accounted for.".

Comment IR-08

> I think the primary assessment of model fit should be based on the chi^2 test, not AFIs. Chi^2 is the only formal test of model fit and the most powerful indication of model fit. Of course, "dichotomania" based on whether the associated p-value is significant or not doesn't help here. But if the chi^2 test is significant, it should be the impetus to assess local fit. Only if there is no evidence of severe local misspecification should the model be regarded as a good enough representation of the data. Looking solely at global fit is just not enough (regardless of whether it is chi^2 or AFIs). I understand you want the RR to be as script-based as possible, but I would sacrifice some of the easiness of looking at global AFI indices and build in a procedure of assessing local fit.

We acknowledge the present comment. With respect to the assessment of the quality of adjustment to the data of the structural equation models that we will fit in the context of the present registered report, we augmented and revised our approach. As per the present comment, in addition to approximate fit, exact and local fit will now be examined. As per "Comment IR-09", rather than using fixed threshold values, approximate fit will now be examined using dynamic threshold values. Therefore, in the corresponding sections of the revised version of our manuscript, the assessment of the quality of adjustment to the data of the structural equation models now read as follows:

"Confirmatory factor analyses of the 50-item version of the UPPS-P Impulsive Behavior Scale (UPPS-P-50) will be performed with respect to its pre-established five-factor structure using the *R* package *lavaan* version 0.6-17 or later (Rosseel et al., 2023). To fit the structural equation model, weighted least squares mean-and-variance-adjusted robust estimation methods will be employed (Finney & di Stefano, 2013). Of note, in fitting the structural equation model, the ordered categorical nature of the observed variables will be accounted for. Adequate model-implied standardized estimates will be determined by model-implied non-null λ standardized estimates ≥ 0.500. If the latter decision rule is not met, alternative structural equation models will be iteratively fitted by omitting observed variables based on model-implied non-null λ standardized estimates until meeting the said decision rule. To assess the quality of adjustment to the data of the structural equation model, exact, approximate, and local fit will be examined. To examine exact fit, an exact fit hypothesis test will be performed under the null hypothesis that the difference between the population covariance matrix and the model-implied covariance matrix is null, against the alternative hypothesis that the difference between the population covariance matrix and the model-implied covariance matrix is non-null (Kline, 2023). If the corresponding probability value is lower than a significance level equaling $\alpha = 0.050$, then the null hypothesis can be rejected at the said significance level, else the null hypothesis cannot be rejected at the said significance level. Adequate exact fit will be determined by a corresponding probability value equal to or greater than a significance level equaling $\alpha = 0.050$. To examine approximate fit, three model-implied fit indices will be employed: the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA) (Kline, 2023). To estimate dynamic threshold values for the three aforementioned model-implied approximate fit indices, Monte Carlo simulation analyses will be performed with respect to the structural equation model using the *R* package *dynamic* version 1.1.0 or later (Wolf & McNeish, 2022). In summary, using estimates from the "correctly specified" structural equation model of interest, the Monte Carlo simulation method is employed to iteratively generate samples under varying magnitudes of misspecification applied to the null λ parameters of the "correctly specified" structural equation model of interest, fit them to the said "correctly specified" structural equation model, assess their quality of adjustment, and estimate the dynamic threshold values that consistently reject "incorrectly specified" models (i.e., misspecification applied to the null λ parameters of the "correctly specified" structural equation model of interest ≥ 0.500), as determined by the 95th percentiles of the distributions of each of the three model-implied fit indices (McNeish, 2023; McNeish & Wolf, 2023). Adequate approximate fit will be determined by a CFI greater than or equal to the dynamic lower-bound value, a TLI greater than or equal to the dynamic lower-bound value, and an RMSEA lower than or equal to the dynamic upper-bound value, all of which are suggested by the results of the Monte Carlo simulation analyses. To examine local fit, the difference between the population covariance matrix and the model-implied covariance matrix will be inspected (Kline, 2023). If the aforementioned decision rules for adequate exact and approximate fit are not met, alternative structural equation models will be fitted based on local fit until meeting the said decision rules, if and only if such *post hoc* model modification can be carried out sparingly and be theoretically justified (MacCallum, 1986; MacCallum et al., 1992; Silvia & MacCallum, 1988). Subsequently, to highlight construct validity and internal consistency reliability evidence, confirmatory factor analysis results for all structural equation models that will have been relevant to this investigation will be reported

Finney, S. J., & di Stefano, C. (2013). *Structural equation modeling: a second course*.

Kline, R. B. (2023). *Principles and practice of structural equation modeling*.

MacCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, *100*(1), 107–120. https://doi.org/10.1037/0033-2909.100.1.107

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychological Bulletin*, *111*(3), 490–504. https://doi.org/10.1037/0033-2909.111.3.490

McNeish, D. (2023). Dynamic fit index cutoffs for categorical factor analysis with Likert-type, ordinal, or binary responses. *American Psychologist*, *78*(9), 1061–1075. https://doi.org/10.1037/amp0001213

McNeish, D., & Wolf, M. G. (2023). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*, *28*(1), 61–88. https://doi.org/10.1037/met0000425

Rosseel, Y., Jorgensen, T. D., & De Wilde, L. (2023). *lavaan: latent variable analysis* (0.6-17) [Computer software]. https://CRAN.R-project.org/package=lavaan

Silvia, E. S. M., & MacCallum, R. C. (1988). Some factors affecting the success of specification searches in covariance structure modeling. *Multivariate Behavioral Research*, *23*(3), 297–326. https://doi.org/10.1207/s15327906mbr2303_2

Wolf, M. G., & McNeish, D. (2022). *dynamic: dynamic fit indices cutoffs for latent variable models* (1.1.0) [Computer software]. https://cran.r-project.org/package=dynamic

Comment IR-09

> Also, the proposed AFI cutoffs are too lenient. Even Hu and Bentler (1999) suggest higher thresholds. With your sort of (ordinal) data, model misfit is easier to miss, so you have to be stricter compared to a situation where the data are continuous. I would go for these stricter thresholds, or alternatively, you may want to employ dynamic fit index cutoffs, which is, IMO, the best option here (see McNeish, 2023).

We acknowledge the present comment. As per "Comment IR-08", with respect to the assessment of the quality of adjustment to the data of the structural equation models that we will fit in the context of the present registered report, we augmented and revised our approach. Notably, rather than using fixed threshold values, approximate fit will now be examined using dynamic threshold values. Therefore, with respect to the present comment, we suggest referring to our response to "Comment IR-08" for detailed information regarding approximate fit examination.

Comment IR-10

"The structural equation model presenting the highest quality of adjustment to the data will correspond to the 50-item version of the UPPS-P Impulsive Behavior Scale". How are you going to do the quality of adjustment assessment if you have (1) model-implied standardized estimates, (2) model-implied fit indices, and (3) model-implied McDonald's internal consistency values? I guess it cannot be a purely algorithmic RR-style decision, which is fine by me. Psychometric validation isn't easily implemented in an RR workflow.

We acknowledge and agree with the present comment. The structural equation models that will correspond to the 50-item version of the UPPS-P Impulsive Behavior Scale (UPPS-P-50) and to the revised 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20-R) should be selected based on the extensive evidence of construct validity and internal consistency reliability, rather than based on an "algorithm-like" rationale.

As per our response to "Comment IR-08", with respect to the construct validity and internal consistency reliability evidence of the structural equation models of interest, the corresponding sections of the revised version of our manuscript now read as follows:

"Subsequently, to highlight construct validity and internal consistency reliability evidence, confirmatory factor analysis results for all structural equation models that will have been relevant to this investigation will be reported in the "Results" section of the manuscript through (1) model-implied χ^2 test statistics along with their corresponding degrees of freedom and probability values, (2) model-implied approximate fit indices (i.e., comparative fit indices (CFI), Tucker-Lewis indices (TLI), and root mean square errors of approximation (RMSEA) along with their corresponding 90% confidence intervals) along with their corresponding dynamic threshold values, (3) differences between the population covariance matrices and the model-implied covariance matrices, and (4) model-implied McDonald's ω internal consistency values.".

In the "2.1. Development phase I" section, what reads as follows will be appended to the abovementioned paragraph:

"All highlighted construct validity and internal consistency reliability evidence will be employed to decide which structural equation model will correspond to the 50-item version of the UPPS-P Impulsive Behavior Scale (UPPS-P-50).".

In the "2.2. Development phase II" section, what reads as follows will be appended to the abovementioned paragraph:

"All highlighted construct validity and internal consistency reliability evidence will be employed to decide which structural equation model will correspond to the revised 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20-R) to be evaluated in posterior phases.".

Comment IR-11

I lacked more details about how the validity evidence will be assessed. I would personally consider convergent and criterion validity as part of an overwhelming concept of construct validity (see

> Cronbach & Meehl, 1955). I think it would be a conceptually stronger solution to establish and test a nomological network than such rather piecewise evidence. But I leave that up to the author's discretion.

We agree that a "nomological network" approach (Cronbach & Meehl, 1955) would be the most suitable option for validity evidence highlighting. However, accounting for issues related to statistical power, a "nomological network" approach cannot be adopted. As we cannot use estimates from prior research conducted on the structural equation model corresponding to the "nomological network" of interest, we cannot perform Monte Carlo simulation analyses to estimate the minimum number of participants to be recruited for data collection to yield adequate statistical power. In addition, given the complexity of the structural equation model corresponding to the "nomological network" of interest, which would involve 11 latent variables and 64 observed variables, the minimum number of participants to be recruited for data collection to yield adequate statistical power would require a substantial increase. However, we strongly believe that such a substantial increase would not be worth the corresponding substantial increase in the financial resources we would need to secure to compensate participants.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

From: Loïs Fournier (corresponding author), on behalf of all the authors involved (Alexandre Heeren, Stéphanie Baggio, Luke Clark, Antonio Verdejo-García, José C. Perales, and Joël Billieux)

To: Orestis Zavlis (reviewer)

Comment OZ-00

Thank you for the opportunity to review this registration. I have found the authors' research outline to be both well-written and well-argued throughout. Below, you can find more specific comments pertaining to the Stage 1 review.

Dear Orestis Zavlis,

Thank you for agreeing to review our stage one registered report. We believe that outlining the research questions pertaining to our development and evaluation protocol in accordance with your comments greatly contributed to improving the original version of our manuscript. To facilitate the review of the revised version of our manuscript, comments are numbered and presented in table cells, below which are our responses. All revisions to the original version of our manuscript have been highlighted in green in the (.pdf) present document. All revisions to the original version of our manuscript have been highlighted using the "track changes" function in *Microsoft Word* in the (.docx) revised version of our manuscript. Thank you again for your time and expertise. We truly appreciate your positive and encouraging feedback on our development and evaluation protocol.

Sincerely,

Loïs Fournier (corresponding author), on behalf of all the authors involved (Alexandre Heeren, Stéphanie Baggio, Luke Clark, Antonio Verdejo-García, José C. Perales, and Joël Billieux)

Loïs Fournier, M.Sc. (lois.fournier@unil.ch)
Institute of Psychology, University of Lausanne, Lausanne, Switzerland

Comment OZ-01

> 1A. The scientific validity of the research question(s).
>
> The questions the authors wish to address are scientifically valid since they aim to address the relatively well-known problem of whether short scales can adequately capture a phenomenon of interest (that is usually measured with larger scales).

We thank you for the present positive and encouraging comment.

Comment OZ-02

> 1B. The logic, rationale, and plausibility of the proposed hypotheses, as applicable.
>
> The authors do not appear to have explicit hypotheses but instead outline a number of different research questions, each associated with specific (development) phases of their project. Although I have found those research questions applicable in terms of addressing the authors' aims (and I particularly enjoyed the authors' use of network psychometrics), I think the preregistration will benefit by outlining clearly these research questions at the start of each section (i.e., outline the research questions of Development Phase 1, 2, and 3 in a numbered way).

We thank you for the present comment, which we acknowledge and agree with. In outlining the research questions pertaining to our development and evaluation protocol, we believe that the revised version of our manuscript now benefits from an improved presentation. All the research questions now outlined are listed below, yet we suggest referring to the revised version of our manuscript for a contextual presentation of the research questions outlined.

In "2.1. Development phase I", the research questions read as follows:

"RQ1. What is the construct validity of the established 50-item version of the UPPS-P Impulsive Behavior Scale (UPPS-P-50)?"

"RQ2. What is the internal consistency reliability of the established 50-item version of the UPPS-P Impulsive Behavior Scale (UPPS-P-50)?"

In "2.2. Development phase II", the research questions read as follows:

"RQ1. What is the content validity of the pre-established revised 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20-R)?"

"RQ2. What is the construct validity of the pre-established revised 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20-R)?"

"RQ3. What is the internal consistency reliability of the pre-established revised 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20-R)?"

In "2.3. Evaluation phase", the research questions read as follows:

Comment OZ-03

1C. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable).

The statistical analyses are comprehensively outlined, directly correspond to the research questions, and appear sound and realistic.

We thank you for your positive evaluation of our development and evaluation protocol.

Comment OZ-04

1D. Whether the clarity and degree of methodological detail are sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses.

As briefly alluded to earlier, I have found the authors' outline of statistical analyses to be very comprehensive and detailed. At the same time, however, I believe that the registration will be strengthened with the inclusion of detailed and numbered research questions (and a brief parenthetical note of the statistical analyses aimed at addressing them).

All the research questions now outlined are listed in our response to "Comment OZ-02", yet we still suggest referring to the revised version of our manuscript for a contextual presentation of the research questions outlined.

Comment OZ-05

1E. Whether the authors have considered sufficient outcome-neutral conditions (e.g., absence of floor or ceiling effects, positive controls, other quality checks) for ensuring that the obtained results are able to test the stated hypotheses or answer the stated research question(s).

The authors have proposed a number of simulation, sensitivity, and power analyses to ensure that their data can address the authors' research questions.

We thank you again for your positive evaluation of our development and evaluation protocol.

Supplementary revisions

Supplementary revision 1

We thank you, Veli-Matti Karhulahti, Ivan Ropovik, and Orestis Zavlis, for agreeing to be recommender and reviewer with respect to the present registered report. In the "Acknowledgments" section of the revised version of our manuscript, we added a statement which reads as follows:

"The authors thank *Peer Community In Registered Reports* for recommending and reviewing the present registered report: Veli-Matti Karhulahti (recommender), Ivan Ropovik (reviewer), and Orestis Zavlis (reviewer).".

Feel free to reach out to Loïs Fournier (corresponding author; lois.fournier@unil.ch) if you would like your name to be removed.

Supplementary revision 2

In recent exchanges between us, we have come to realize two important points with respect to participant data collection. First, in our experience, the estimates provided by *Qualtrics* (https://www.qualtrics.com/) regarding the time required to complete all statements implemented in a full online survey are severely overestimated. Therefore, new estimates will be provided through a "pilot" approach. Second, the location from which financial resources – with respect to the present registered report – will be transferred to *Prolific* (https://www.prolific.com/) requires the use of GBP rather than USD. Therefore, in light of these two important points with respect to participant data collection, we added what reads as follows:

"We will collect participant data with respect to their sociodemographic information and to their answers on the 50-item version of the UPPS-P Impulsive Behavior Scale (UPPS-P-50). Prior to proceeding with participant data collection, N = 32 "pilot" participants (see Perneger et al., 2015) will be recruited to complete all statements implemented in the full online survey – strictly following the protocol of participant data collection – in order to (1) ensure the absence of technical and "pilot"-participant-reported issues in the full online survey, (2) estimate the time required to complete all statements implemented in the full online survey, and (3) determine how much participants will be paid for their full completion of the online survey. Of note, all "pilot" participant data will be permanently deleted prior to proceeding with participant data collection. As *Qualtrics* (https://www.qualtrics.com/) estimates that the time required to complete all statements implemented in the full online survey equals 13.8 minutes, "pilot" participants will be paid $\lceil 13.8 \rceil$*(9/60) = 2.1 GBP for their full completion of the online survey to equate to a "good" hourly rate of 9 GBP, in accordance with *Prolific* (https://www.prolific.com/). Let T be the median time (in minutes) it will take "pilot" participants to complete all statements implemented in the pilot full online survey; participants will be paid $\lceil T \rceil$*(9/60) GBP for their full completion of the online survey to equate to a "good" hourly rate of 9 GBP, in accordance with *Prolific* (https://www.prolific.com/).".

Perneger, T. V., Courvoisier, D. S., Hudelson, P. M., & Gayet-Ageron, A. (2015). Sample size for pre-tests of questionnaires. *Quality of Life Research*, *24*(1), 147–151. https://doi.org/10.1007/s11136-014-0752-2

For consistency, as participant data collection will occur in two phases, revisions similar to those mentioned above have been applied to all corresponding sections of the revised version of our manuscript.

Supplementary revision 3

In accordance with the seventh edition of the *Publication Manual of the American Psychological Association* (2020), the "References" section title now reads "Reference list".

Publication Manual of the American Psychological Association (7th ed.). (2020). American Psychological Association. https://doi.org/10.1037/0000165-000

Supplementary revision 4

Following online publication on July 26, 2024 (i.e., 29 days after submitting the original version of our manuscript), of an original research article referenced in the original version of our manuscript, the in-text citation and reference list entry for the said original research article were updated and now respectively read as follows:

"(Fournier et al., 2024)"

"Fournier, L., Bőthe, B., Demetrovics, Z., Koós, M., Kraus, S. W., Nagy, L., Potenza, M. N., Ballester-Arnal, R., Batthyány, D., Bergeron, S., Briken, P., Burkauskas, J., Cárdenas-López, G., Carvalho, J., Castro-Calvo, J., Chen, L., Ciocca, G., Corazza, O., Csako, R. I., Fernandez, D. P., Fujiwara, H., Fernandez, E. F., Fuss, J., Gabrhelík, R., Gewirtz-Meydan, A., Gjoneska, B., Gola, M., Grubbs, J. B., Hashim, H. T., Islam, M. S., Ismail, M., Jiménez-Martínez, M. C., Jurin, T., Kalina, O., Klein, V., Költő, A., Lee, S.-K., Lewczuk, K., Lin, C.-Y., Lochner, C., López-Alvarado, S., Lukavská, K., Mayta-Tristán, P., Miller, D. J., Orosová, O., Orosz, G., Ponce, F. P., Quintana, G. R., Quintero Garzola, G. C., Ramos-Diaz, J., Rigaud, K., Rousseau, A., Scanavino, M. D. T., Schulmeyer, M. K., Sharan, P., Shibata, M., Shoib, S., Sigre-Leirós, V., Sniewski, L., Spasovski, O., Steibliene, V., Stein, D. J., Strizek, J., Sungkyunkwan University Research Team, Tsai, M.-C., Ünsal, B. C., Vaillancourt-Morel, M.-P., Van Hout, M. C., & Billieux, J. (2024). Evaluating the factor structure and measurement invariance of the 20-item short version of the UPPS-P Impulsive Behavior Scale across multiple countries, languages, and gender identities. *Assessment*, 10731911241259560. https://doi.org/10.1177/10731911241259560"