UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Fournier, L., Heeren, A., Baggio, S., Clark, L., Verdejo-García, A., Perales, J. C., & Billieux, J. (2026). Reducing assessment time without sacrificing content validity in short-form development: an illustration through the revised 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20-R). *Collabra: Psychology*, *12*(1). https://doi.org/10.1525/collabra.150350

Methodology and Research Practice

Reducing Assessment Time Without Sacrificing Content Validity in Short-Form Development: An Illustration Through the Revised 20-Item Short Version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20-R)

Loïs Fournier[1][a], Alexandre Heeren[2], Stéphanie Baggio[1,3], Luke Clark[4,5], Antonio Verdejo-García[6], José C. Perales[7], Joël Billieux[1,8]

[1] Institute of Psychology, University of Lausanne, Lausanne, Switzerland, [2] Psychological Sciences Research Institute, Catholic University of Louvain, Louvain-la-Neuve, Belgium, [3] Institute of Primary Health Care, University of Bern, Bern, Switzerland, [4] Department of Psychology, Center for Gambling Research, University of British Columbia, Vancouver, Canada, [5] Djavad Mowafaghian Center for Brain Health, University of British Columbia, Vancouver, Canada, [6] School of Psychological Sciences and Turner Institute for Brain and Mental Health, Monash University, Melbourne, Australia, [7] Department of Experimental Psychology, Mind, Brain, and Behavior Research Center, University of Granada, Granada, Spain, [8] Center for Excessive Gambling, Department of Addiction Medicine, Lausanne University Hospital, Lausanne, Switzerland

Keywords: content validity, impulsivity, short-form development, shortr, UPPS-P Impulsive Behavior Scale, UPPS-P-20-R

https://doi.org/10.1525/collabra.150350

Collabra: Psychology

Vol. 12, Issue 1, 2026

The UPPS-P Impulsive Behavior Scale is a well-established psychometric instrument for assessing impulsivity, a key multidimensional psychological construct encompassing distinct facets that are differentially associated with numerous psychopathological and neuropathological symptoms across various disorders. To facilitate its integration in clinical and research settings, Billieux et al. (2012) developed and evaluated the psychometric properties of a 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20), reducing assessment time to approximately one-third of that of the original version of the psychometric instrument. Despite its widespread popularity, legitimate concerns were raised about the content validity of this short version of the UPPS-P Impulsive Behavior Scale. Indeed, the inevitable loss of construct-level content coverage in short-form development, although often unaddressed by researchers, warrants in-depth examination. Therefore, the present registered report aimed to develop and evaluate the psychometric properties of a revised short version of the UPPS-P Impulsive Behavior Scale through an expert-driven and data-driven methodological approach to short-form development emphasizing item-level and construct-level content validity. Taken together, the present registered report provides professionals in clinical and research settings (1) a revised short English-language psychometric instrument for assessing impulsivity – i.e., the revised 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20-R) – that presents strong psychometric properties, notably addressing the legitimate concerns critically highlighted about the content validity of the UPPS-P-20, and (2) a data-driven methodological approach to construct-level content validity in short-form development – i.e., the *R* package *shortr* – that can be adopted to develop short versions of any psychometric instrument.

1. Introduction

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 distinct facets that 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 premedita-*

a Corresponding author at: lois.fournier@unil.ch

tion (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).

Originally, the UPPS-P Impulsive Behavior Model was developed based on four trait-like impulsivity-related facets of the Revised NEO Personality Inventory (i.e., impulsiveness, excitement seeking, self-discipline, and deliberation) (Costa & McCrae, 1992) and seventeen classic scales or subscales measuring impulsivity (Whiteside & Lynam, 2001). By federating twenty-one coexisting conceptualizations of impulsivity and thereby correcting the jingle (i.e., distinct constructs designated by one same label) and jangle (i.e., distinct labels designating one same construct) fallacies that characterized the research field of impulsivity, the UPPS-P Impulsive Behavior Model and its corresponding original assessment tool – i.e., the 59-item original version of the UPPS-P Impulsive Behavior Scale (UPPS-P-59; Cyders et al., 2007; Whiteside & Lynam, 2001) – have received considerable interest and have exerted a substantial impact on following impulsivity research over the past decades. 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).

To facilitate the integration of the UPPS-P Impulsive Behavior Scale in clinical and research settings, Billieux et al. (2012) developed and evaluated the psychometric properties of a short version of the corresponding original assessment tool, i.e., the 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20). In reducing assessment time to approximately one-third of that of the original version of the psychometric instrument, this short version has gained considerable popularity. Moreover, given its well-established psychometric properties and relevance for various problematic behaviors and mental disorders, the UPPS-P-20 (Billieux et al., 2012) was adapted to diverse linguistic, developmental, and clinical populations, such as individuals across various languages (Cándido et al., 2012; d'Orta et al., 2015; Fournier, Bőthe, et al., 2025; Zsila et al., 2020), children (Geurten et al., 2021), and patients with substance use disorders (Calzada et al., 2017; Kempeneers et al., 2023; Sánchez-Domínguez et al., 2023).

The methodological approach to short-form development with respect to the UPPS-P-20 (Billieux et al., 2012) consisted of retaining the four items per factor that presented the highest factor loadings. Indeed, it is common practice in short-form development to retain items presenting the highest factor loadings or total-item correlations for a psychometric instrument's given factor(s), as it presumably maximizes internal consistency reliability and minimizes measurement error (Smith et al., 2000). However, considering the attenuation paradox in psychometrics (Boyle, 1991; Loevinger, 1954), internal consistency reliability countervails with construct-level content validity and, consequently, such common practice in short-form development tends to narrow the breadth of construct-level content coverage of psychometric instruments (Clark & Watson, 2019). Reflecting this, Cyders et al. (2014) criticized the methodological approach to short-form development with respect to the UPPS-P-20 (Billieux et al., 2012), raising legitimate concerns about the content validity of this short version of the UPPS-P Impulsive Behavior Scale.

As short-form development in psychometrics consists of retaining a subset of items included in the original version of a psychometric instrument, loss of construct-level content coverage is inevitable, therefore warranting in-depth examination (Smith et al., 2000). To address the content validity issue critically highlighted by Cyders et al. (2014), the present registered report aims to develop and evaluate the psychometric properties of a revised short version of the UPPS-P Impulsive Behavior Scale through an expert-driven and data-driven methodological approach to short-form development emphasizing item-level and construct-level content validity.

The present registered report consists of a three-phase protocol, the overarching objectives of which are to develop and evaluate the psychometric properties of the revised 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20-R). In the first phase (i.e., development phase I), the primary objectives are to collect participant data with respect to an initial item pool derived from the original form and to ensure it presents adequate psychometric properties (i.e., content validity, internal structure validity, internal consistency reliability) prior to proceeding with short-form development. In the second phase (i.e., development phase II), the primary objectives are to develop the short form from the initial item pool derived from the original form and to ensure it presents adequate psychometric properties (i.e., content validity, internal structure validity, internal consistency reliability) prior to proceeding with short-form evaluation. Lastly, in the third phase (i.e., evaluation phase), the primary objectives are to collect participant data with respect to the short form – along with forms assessing related psychological constructs – and to evaluate its psychometric properties (i.e., internal structure validity, internal consistency reliability, test-retest reliability, convergent validity, criterion validity).

2. Methods

2.1. Development phase I

In the first phase (i.e., development phase I), the primary objectives are to collect participant data with respect to an initial item pool derived from the original form (UPPS-P-n_i)

and to ensure it presents adequate psychometric properties (i.e., content validity, internal structure validity, internal consistency reliability) prior to proceeding with short-form development (see Table 2.1.).

Therefore, the research questions pertaining to the first phase of the present registered report are as follows:

- **RQ1.** What is the internal structure validity of the UPPS-P-n_i?
- **RQ2.** What is the internal consistency reliability of the UPPS-P-n_i?

Box 2.1. An expert-driven methodological approach to item-level content validity in short-form development

The *59-item original version of the UPPS-P Impulsive Behavior Scale* (UPPS-P-59; Cyders et al., 2007; Whiteside & Lynam, 2001) is a self-administered English-language psychometric instrument that assesses the applicability of 59 statements related to the five different dimensions of the UPPS-P Impulsive Behavior Model, namely (1) *lack of premeditation* (e.g., item 1: "I have a reserved and cautious attitude toward life."), (2) *negative urgency* (e.g., reverse-scored item 2*: "I have trouble controlling my impulses."), (3) *sensation seeking* (e.g., reverse-scored item 3*: "I generally seek new and exciting experiences and sensations."), (4) *lack of perseverance* (e.g., item 4: "I generally like to see things through to the end."), and (5) *positive urgency* (e.g., reverse-scored item 5*: "When I am very happy, I can't seem to stop myself from doing things that can have bad consequences."). Each of the five instrument's dimensions includes 10 to 14 items that are scored (or reverse-scored) on a four-point Likert-type scale (from 1 = "Strongly agree" to 4 = "Strongly disagree") and that provide arithmetic mean scores likewise ranging from 1 (i.e., the lowest level of endorsement of the corresponding impulsive behavior dimension) to 4 (i.e., the highest level of endorsement of the corresponding impulsive behavior dimension).

In examining the individual items comprised in the UPPS-P-59 (Cyders et al., 2007; Whiteside & Lynam, 2001), we, the authors, consider that some items fail to provide relevant information about one's endorsement of the impulsive behavior dimensions. Therefore, we developed an initial item pool derived from the original form (UPPS-P-n_i) to ensure that all items provide relevant information about one's endorsement of the impulsive behavior dimensions. To develop the said initial item pool, we proceeded as follows. Three authors – Loïs Fournier, José C. Perales, and Joël Billieux – were instructed to independently review all 59 items of the UPPS-P-59 (Cyders et al., 2007; Whiteside & Lynam, 2001), and, while adopting a conservative approach, to evaluate their item-level content validity by annotating them with "keep", "revise", or "discard". Evaluations were then compared, and discrepancies were discussed until a consensus was reached. Overall, evaluations of the item-level content validity of the items of the UPPS-P-59 (Cyders et al., 2007; Whiteside & Lynam, 2001) can be summarized in three different categories.

The first category regards items that fail to provide relevant information about one's attraction to excitement and openness 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). Therefore, we argue that the seven items listed below should be discarded:

- Sensation seeking reverse-scored item 13*: "I like sports and games in which you have to choose your next move very quickly.".
- Sensation seeking reverse-scored item 18*: "I would enjoy water skiing.".
- Sensation seeking reverse-scored item 26*: "I would enjoy parachute jumping.".
- Sensation seeking reverse-scored item 36*: "I would like to learn to fly an airplane.".
- Sensation seeking reverse-scored item 46*: "I would enjoy the sensation of skiing very fast down a high mountain slope.".
- Sensation seeking reverse-scored item 51*: "I would like to go scuba diving.".
- Sensation seeking reverse-scored item 56*: "I would enjoy fast driving.".

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). Therefore, we argue that the two items listed below should be discarded:

- Negative urgency reverse-scored item 7*: "I have trouble resisting my cravings (for food, cigarettes, etc.).".
- Positive urgency reverse-scored item 57*: "When I am very happy, I feel like it is ok to give in to cravings or overindulge.".

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 intense negative or positive emotional states in which sudden actions are emitted is relevant, as negative urgency and positive urgency are conceptualized as two distinct facets that are differentially associated with symptoms across various disorders (Cyders & Smith, 2007). Yet, we consider that this is straightforwardly addressable by appending intense negative emotional states to such items. To this end, we generated appendices incorporating different intense negative emotional states (e.g., "distress", "bad mood", "frustration"). Therefore, we argue that the five items listed below should be revised by appending what reads in italics:

- Negative urgency reverse-scored revised item 2*: "*When I feel distressed,* I have trouble controlling my impulses.".
- Negative urgency reverse-scored revised item 12*: "*When I feel bad,* I often get involved in things I later wish I could get out of.".
- Negative urgency reverse-scored revised item 39*: "It is hard for me to resist acting on my *negative* feelings.".
- Negative urgency revised item 53: "*Even when I am in a bad mood,* I always keep my feelings under control.".
- Negative urgency reverse-scored revised item 58*: "*When I feel frustrated,* I sometimes do impulsive things that I later regret.".

In keeping 45 items, revising five items, and discarding nine items with respect to the UPPS-P-59 (Cyders et al., 2007; Whiteside & Lynam, 2001), we, the authors, consider that the UPPS-P-n_i provides relevant information about one's endorsement of the impulsive behavior dimensions.

2.1.1. Materials

We collected participant data with respect to their sociodemographic information (i.e., level of English proficiency, age, gender identity, and country of residence) and to their answers on the *initial item pool derived from the original version of the UPPS-P Impulsive Behavior Scale* (UPPS-P-n_i).

The *initial item pool derived from the original version of the UPPS-P Impulsive Behavior Scale* (UPPS-P-n_i) is a self-administered English-language psychometric instrument that assesses the applicability of n_i statements related to the five different dimensions of the UPPS-P Impulsive Behavior Model, namely (1) *lack of premeditation* (e.g., item 1: "I have a reserved and cautious attitude toward life."), (2) *negative urgency* (e.g., reverse-scored revised item 2*: "*When I feel distressed,* I have trouble controlling my impulses."), (3) *sensation seeking* (e.g., reverse-scored item 3*: "I generally seek new and exciting experiences and sensations."), (4) *lack of perseverance* (e.g., item 4: "I generally like to see things through to the end."), and (5) *positive urgency* (e.g., reverse-scored item 5*: "When I am very happy, I can't seem to stop myself from doing things that can have bad consequences."). Each of the five instrument's dimensions includes 5 to 13 items that are scored (or reverse-scored) on a four-point Likert-type scale (from 1 = "Strongly agree" to 4 = "Strongly disagree") and that provide arithmetic mean scores likewise ranging from 1 (i.e., the lowest level of endorsement of the corresponding impulsive behavior dimension) to 4 (i.e., the highest level of endorsement of the corresponding impulsive behavior dimension).

2.1.2. Data

Participants were recruited using *Prolific* (https://www.prolific.com/), a third-party crowdsourcing platform for recruiting and compensating participants for online research, and were redirected to *Qualtrics* (https://www.qualtrics.com/), a third-party crowdsourcing platform for implementing online surveys.

We collected participant data with respect to their sociodemographic information and to their answers on the UPPS-P-n_i. Prior to proceeding with participant data collection, N = 32 "pilot" participants (Perneger et al., 2015) were recruited to complete all statements implemented in

Table 2.1. Development phase I

Development phase I			
Materials	**Data**	**Analyses**	**Results**
Sociodemographic information	Participant data collection Development phase I (N = 948)	Descriptive analyses	Sociodemographic information
UPPS-P-n_i		Confirmatory factor analyses	Internal structure validity (RQ1)
			Internal consistency reliability (RQ2)

UPPS-P-n_i

UPPS-P-n_i = initial item pool derived from the original version of the UPPS-P Impulsive Behavior Scale.

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 were paid for their full completion of the online survey. Of note, all "pilot" participant data was 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 10.0 minutes, "pilot" participants were paid $\lceil 10.0 \rceil$*(9/60) = 1.50 GBP for their full completion of the online survey to equate to a "good" hourly rate of 9.00 GBP, in accordance with *Prolific* (https://www.prolific.com/). Let T be the median time (in minutes) it took "pilot" participants to complete all statements implemented in the pilot full online survey; participants were paid $\lceil T \rceil$*(9/60) = $\lceil 6.2 \rceil$*(9/60) = 1.05 GBP for their full completion of the online survey to equate to a "good" hourly rate of 9.00 GBP, in accordance with *Prolific* (https://www.prolific.com/).

The inclusion criteria for data collection are as follows: (1) being fluent in English and (2) being 18 years of age or older. Endorsement of all of the aforementioned criteria establishes inclusion.

The exclusion criteria for data collection are as follows: (1) failing to complete the full online survey, (2) failing the attention check implemented in the UPPS-P-n_i (i.e., "This is an attention check. Please select "Agree strongly" as an answer to the present statement."), and (3) completing the full online survey in less time than three standard deviations below the mean time it took participants to complete the full online survey. Endorsement of any of the aforementioned criteria establishes exclusion.

With respect to data collection, we required that participants provide answers to all statements implemented in the full online survey and adopted a listwise deletion approach to missing data handling by excluding data from participants who failed to complete the full online survey.

To estimate the minimum number of participants to be recruited for data collection, Monte Carlo simulation analyses were performed with respect to the pre-established five-factor structure of the 59-item original version of the UPPS-P Impulsive Behavior Scale (UPPS-P-59; Cyders et al., 2007; Whiteside & Lynam, 2001). In summary, using estimates from prior research conducted on the structural equation model of interest, the Monte Carlo simulation method is employed to iteratively generate samples of varying sample sizes, fit them to the said structural equation model, assess their quality of adjustment, and estimate the minimum sample size that consistently yields adequate quality of adjustment (Muthén & Muthén, 2002). In the present registered report, Monte Carlo simulation analyses were performed using the *R* packages *lavaan* version 0.6-17 (Rosseel et al., 2023) and *simsem* version 0.5-16 (Pornprasertmanit et al., 2021). The model parameter values employed in generating the data were derived from the model-implied standardized estimates provided by Fournier, Bőthe, et al. (2025) in their confirmatory factor analyses performed with respect to the pre-established five-factor structure of the 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20; Billieux et al., 2012). The mean model parameter values derived from the model-implied standardized estimates provided by Fournier, Bőthe, et al. (2025) were then employed in specifying the data generation model parameter values by assuming equivalence across the UPPS-P-20 (Billieux et al., 2012) and the UPPS-P-59 (Cyders et al., 2007; Whiteside & Lynam, 2001). Using the latter data generation model, 2,000 samples were generated across twenty incremental sample sizes (i.e., 100 samples per sample size of interest) varying from N = 59*1 = 59 to N = 59*20 = 1,180. To fit the 2,000 samples to the structural equation model, weighted least squares mean-and-variance-adjusted robust estimation methods were employed (Finney & di Stefano, 2013). Of note, in fitting the structural equation model, the ordered categorical nature of the observed variables was accounted for. To assess the quality of adjustment of the 2,000 samples to the structural equation model, three conventional model-implied fit indices were employed: the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). Good fit was determined by a CFI $\geq$ 0.950, a TLI $\geq$ 0.950, and an RMSEA $\leq$ 0.050. As a result, the Monte Carlo simulation analyses suggested that the minimum sample size that consistently yielded good quality of adjustment – as determined by the 95th percentiles of the distributions of each of the three conventional model-implied fit indices – was equal to $N = 59*14 = 826$.

All in all, with respect to the minimum number of participants to be recruited for data collection, which equals $N = 826$, we terminated data collection upon reaching a 20% increment, i.e., upon reaching $N = \lceil 826*(120/100) \rceil = 992$.

Of the 992 recruited participants, 11 (1.109%) did not complete any statement implemented in the full online survey, 23 (2.319%) were not included for not being fluent in English, and 10 (1.008%) were excluded for failing the attention check implemented in the UPPS-P-n_i (i.e., "This is an attention check. Please select "Agree strongly" as an answer to the present statement."), yielding a sample size equal to $N = 948$.

2.1.3. Analyses

Descriptive analyses of the participant sociodemographic information were performed with respect to the total sample. Subsequently, participant sociodemographic information is reported in the "Results" section of the manuscript through (1) sample size, (2) mean, standard deviation, and range values of their age, (3) percentage frequency values of their gender identity, and (4) percentage frequency values of their country of residence.

Confirmatory factor analyses of the UPPS-P-n_i were performed with respect to its pre-established five-factor structure using the *R* package *lavaan* version 0.6-19 (Rosseel et al., 2024). To fit the structural equation model, weighted least squares mean-and-variance-adjusted robust estimation methods were employed (Finney & di Stefano, 2013). Of note, in fitting the structural equation model, the ordered categorical nature of the observed variables was accounted for. Adequate model-implied standardized estimates were determined by model-implied non-null λ standardized estimates $\geqslant 0.500$. If the latter decision rule is not met, alternative structural equation models were iteratively fitted by omitting the 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, global exact, global approximate, and local fit were examined. To examine global exact fit, a global exact fit hypothesis test was 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 equal to $\alpha = 0.050$, 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 global exact fit was determined by a corresponding probability value equal to or greater than a significance level equal to $\alpha = 0.050$. To examine global approximate fit, three model-implied fit indices were employed: the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR) (Kline, 2023). To estimate dynamic threshold values for the three aforementioned model-implied global approximate fit indices, Monte Carlo simulation analyses were performed with respect to the structural equation model using the *R* package *dynamic* version 1.1.0 (Wolf & McNeish, 2025). 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), as determined by the 95th percentiles of the distributions of each of the three model-implied fit indices (McNeish, 2023; McNeish & Wolf, 2023). Adequate global approximate fit was determined by a CFI greater than or equal to the dynamic lower-bound value, an RMSEA lower than or equal to the dynamic upper-bound value, and an SRMR 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 was inspected (Kline, 2023). If the aforementioned decision rules for adequate global exact and global approximate fit are not met, alternative structural equation models were fitted based on local fit until meeting the said decision rules. Subsequently, to highlight internal structure validity and internal consistency reliability evidence, confirmatory factor analysis results for all structural equation models that were relevant to this investigation are 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 global approximate fit indices (i.e., comparative fit indices (CFI), root mean square errors of approximation (RMSEA) along with their corresponding 90% confidence intervals, and standardized root mean square residuals (SRMR)) 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. All highlighted internal structure validity and internal consistency reliability evidence was employed to decide which structural equation model corresponded to the UPPS-P-n_i.

2.2. Development phase II

In the second phase (i.e., development phase II), the primary objectives are to develop the short form (UPPS-P-20-R) from the initial item pool derived from the original form (UPPS-P-n_i) and to ensure it presents adequate psychometric properties (i.e., content validity, internal structure validity, internal consistency reliability) prior to proceeding with short-form evaluation (see Table 2.2.).

Therefore, the research questions pertaining to the second phase of the present registered report are as follows:

- **RQ3.** What is the content validity of the pre-established UPPS-P-20-R?
- **RQ4.** What is the internal structure validity of the pre-established UPPS-P-20-R?
- **RQ5.** What is the internal consistency reliability of the pre-established UPPS-P-20-R?

Box 2.2. A data-driven methodological approach to construct-level content validity in short-form development

In psychometrics, the network approach refers to a set of methods employed to model and analyze the relationships among psychological variables. Unlike traditional psychometric approaches, such as the structural equation approach focusing on latent variables, the network approach emphasizes the interconnections among observed variables. Due to the latter emphasis, modeling and analyzing network models to complement structural equation models when developing psychometric instruments offers several advantages. Most notably, in undirected weighted network models, a subtype of network models, observed variables are represented by nodes, and associations between observed variables, each assigned a weight that represents the magnitude of the association, are represented by edges. In this perspective, undirected weighted network models provide estimates of the magnitude of associations (i.e., the shared variance) among items of psychometric instruments that structural equation models cannot, providing critical insight into the construct-level content validity of subsets of items of psychometric instruments. To illustrate, if an undirected weighted network model suggests that a subset of items of a psychometric instrument presents a high magnitude of associations with another subset of items, the shared variance of the said subsets of items is therefore high: the content they assess and the information they provide is highly similar. From the standpoint of the latter illustration, undirected weighted network modeling allows for the estimation of whether a subset of items assesses a “narrow” or a “broad” proportion of the construct-level content of a psychometric instrument. Hence, identifying an optimal subset of a desired number of items that assesses the “broadest” proportion of the construct-level content of a psychometric instrument consists of a combinatorial optimization problem.

Consider an undirected weighted network model $G = (V, E)$, where V denotes the set of nodes and E the set of edges. Each edge e_{ij} is associated with a positive or negative weight w_{ij}. Let $S \subset V$ be a subset of nodes of fixed size k, and let $\bar{S}$ denote its complement in V. The objective is to identify the optimal subset S of size k such that the sum of the absolute values of the edge weights connecting S to its complement $\bar{S}$ of size $n - k$ is maximized. Formally, the combinatorial optimization problem can be expressed as:

$$\max_{S \subset V,\, |S|=k} \left(\sum_{i \in S,\, j \in \bar{S}} |w_{ij}| \right) \qquad (2.2.)$$

Solving the combinatorial optimization problem expressed in Equation 2.2. allows identifying what optimal subset of a desired number of items presents the highest magnitude of associations (i.e., the highest shared variance) within the set of items.

To this end, we developed and released the *R* package *shortr* (Fournier, Heeren, et al., 2025), which implements combinatorial search algorithms (i.e., brute force combinatorial search algorithm, simulated annealing combinatorial search algorithm) to identify what optimal subset of a desired number of items should be retained in a short version of a psychometric instrument to assess the “broadest” proportion of the construct-level content of the set of items included in the original version of the said psychometric instrument.

The *R* package *shortr* (Fournier, Heeren, et al., 2025) is shared publicly under a *GNU General Public License* on the *Comprehensive R Archive Network* (https://doi.org/10.32614/CRAN.package.shortr).

2.2.1. Materials and data

We employed participant data that was collected in the prior development phase (i.e., development phase I) with respect to their answers on the *initial item pool derived from the original version of the UPPS-P Impulsive Behavior Scale* (UPPS-P-n_{i}).

2.2.2. Analyses

Network analyses of the UPPS-P-n_{i} were performed with respect to its n_{i} items and to the n_{j} items comprised in each of its five factors using the *R* packages *bootnet* version

Table 2.2. Development phase II

<table>
<tr><th colspan="4">Development phase II</th><th rowspan="6">Pre-established UPPS-P-20-R</th></tr>
<tr><th>Materials</th><th>Data</th><th>Analyses</th><th>Results</th></tr>
<tr><td rowspan="3">UPPS-P-n_i</td><td rowspan="3">Participant data Development phase I (N = 948)</td><td>Network analyses</td><td>Content validity (RQ3)</td></tr>
<tr><td rowspan="2">Confirmatory factor analyses</td><td>Internal structure validity (RQ4)</td></tr>
<tr><td>Internal consistency reliability (RQ5)</td></tr>
</table>

UPPS-P-n_i = initial item pool derived from the original version of the UPPS-P Impulsive Behavior Scale. UPPS-P-20-R = revised 20-item short version of the UPPS-P Impulsive Behavior Scale.

1.6 (Epskamp, 2024), *glasso* version 1.11 (Friedman et al., 2019), *huge* version 1.3.5 (Jiang et al., 2021), *igraph* version 2.1.4 (Csárdi et al., 2025), and *qgraph* version 1.9.8 (Epskamp et al., 2023). A prominent (undirected weighted) network model in psychometrics is the Gaussian graphical model (Borsboom et al., 2021; Epskamp et al., 2018; Lauritzen, 1996). In Gaussian graphical models, observed variables are represented by nodes, and associations between observed variables are represented by edges, each assigned a weight – derived from the partial correlation coefficient value of its corresponding nodes – that represents the pairwise conditional (in-)dependence between observed variables after conditioning on all other observed variables included in the model. Prior to fitting each of the n_i-item and the five n_j-item Gaussian graphical models, as all items of the UPPS-P-n_i on which network analyses were performed are ordered categorical, data was subjected to a non-paranormal transformation, a semi-parametric transformation designed to relax the assumption of multivariate normality inherent to Gaussian graphical models (Isvoranu & Epskamp, 2023; Liu et al., 2018). To fit each of the n_i-item and the five n_j-item Gaussian graphical models, the *ggmModSelect* algorithm was employed to search for an optimal Gaussian graphical model by (1) fitting a hundred Gaussian graphical models using the graphical lasso regularization algorithm with varying tuning parameter values, (2) refitting all models without regularization, (3) selecting the model with the lowest Bayesian information criterion value, and (4) iteratively adding and removing edges to the latter model upon reaching minimal Bayesian information criterion value (Isvoranu & Epskamp, 2023). To estimate the community structure of each of the n_i-item and the five n_j-item Gaussian graphical models, 1,000 executions of the *spin glass* community detection algorithm were employed to search for an optimal quality of partitioning by selecting the community structure with the highest modularity Q value (Newman & Girvan, 2004; Reichardt & Bornholdt, 2006; Traag & Bruggeman, 2009). To identify what five optimal subsets of four items should be retained in the pre-established UPPS-P-20-R to assess the "broadest" proportion of the construct-level content of the set of items included in the UPPS-P-n_i, a brute force combinatorial search algorithm was implemented using the *R* package *shortr* version 1.0.1 (Fournier, Heeren, et al., 2025) to solve the combinatorial optimization problem expressed in Equation 2.2. with respect to each of the five n_j-item Gaussian graphical models. Subsequently, to highlight construct-level content validity evidence, network analysis results for all Gaussian graphical models that were relevant to this investigation are reported in the "Results" section of the manuscript through (1) model-implied graph drawings, (2) model-implied community structures and modularity Q values, and (3) model-implied optimal subsets.

Confirmatory factor analyses of the five model-implied optimal subsets of four items that should be retained in the pre-established UPPS-P-20-R were performed with respect to its pre-established five-factor structure using the *R* package *lavaan* version 0.6-19 (Rosseel et al., 2024). To fit the structural equation model, weighted least squares mean-and-variance-adjusted robust estimation methods were employed (Finney & di Stefano, 2013). Of note, in fitting the structural equation model, the ordered categorical nature of the observed variables was accounted for. Adequate model-implied standardized estimates were determined by model-implied non-null λ standardized estimates $\geqslant 0.500$. If the latter decision is not met, alternative Gaussian graphical models and structural equation models were iteratively fitted by omitting the 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, global exact, global approximate, and local fit were examined. To examine global exact fit, a global exact fit hypothesis test was 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 equal to $\alpha = 0.050$, 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 global exact fit was determined by a corresponding probability value equal to or greater than a significance level equal to α = 0.050. To examine global approximate fit, three model-implied fit indices were employed: the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR) (Kline, 2023). To estimate dynamic threshold values for the three aforementioned model-implied global approximate fit indices, Monte Carlo simulation analyses were performed with respect to the structural equation model using the *R* package *dynamic* version 1.1.0 (Wolf & McNeish, 2025). 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), as determined by the 95th percentiles of the distributions of each of the three model-implied fit indices (McNeish, 2023; McNeish & Wolf, 2023). Adequate global approximate fit was determined by a CFI greater than or equal to the dynamic lower-bound value, an RMSEA lower than or equal to the dynamic upper-bound value, and an SRMR 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 was inspected (Kline, 2023). If the aforementioned decision rules for adequate global exact and global approximate fit are not met, alternative structural equation models were fitted based on local fit until meeting the said decision rules. Subsequently, to highlight internal structure validity and internal consistency reliability evidence, confirmatory factor analysis results for all structural equation models that were relevant to this investigation are 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 global approximate fit indices (i.e., comparative fit indices (CFI), root mean square errors of approximation (RMSEA) along with their corresponding 90% confidence intervals, and standardized root mean square residuals (SRMR)) 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. All highlighted internal structure validity and internal consistency reliability evidence was employed to decide which structural equation model corresponded to the pre-established UPPS-P-20-R.

2.3. Evaluation phase

Lastly, in the third phase (i.e., evaluation phase), the primary objectives are to collect participant data with respect to the short form (UPPS-P-20-R) – along with forms assessing related psychological constructs – and to evaluate its psychometric properties (i.e., internal structure validity, internal consistency reliability, test-retest reliability, convergent validity, criterion validity) (see Table 2.3.).

Therefore, the research questions pertaining to the third phase of the present registered report are as follows:

- **RQ6.** What is the internal structure validity of the UPPS-P-20-R?
- **RQ7.** What is the internal consistency reliability of the UPPS-P-20-R?
- **RQ8.** What is the test-retest reliability of the UPPS-P-20-R?
- **RQ9.** What is the convergent validity of the UPPS-P-20-R?
- **RQ10.** What is the criterion validity of the UPPS-P-20-R?

2.3.1. Materials

We collected participant data with respect to their sociodemographic information (i.e., level of English proficiency, age, gender identity, and country of residence) and to their answers on five different psychometric instruments: (1) the *revised 20-item short version of the UPPS-P Impulsive Behavior Scale* (UPPS-P-20-R), (2) the *13-item Brief Self-Control Scale* (BSCS-13; Tangney et al., 2004), (3) the *7-item Generalized Anxiety Disorder Scale* (GAD-7; Spitzer et al., 2006), (4) the *9-item Patient Health Questionnaire* (PHQ-9; Kroenke et al., 2001), and (5) the *15-item short version of the Problematic Mobile Phone Use Questionnaire* (PMPUQ-SV-15; López-Fernández et al., 2018).

The *revised 20-item short version of the UPPS-P Impulsive Behavior Scale* (UPPS-P-20-R) (see Appendix) is a self-administered English-language psychometric instrument that assesses the applicability of 20 statements related to the five different dimensions of the UPPS-P Impulsive Behavior Model, namely (1) *lack of premeditation* (e.g., item 1: "My thinking is usually careful and purposeful."), (2) *positive urgency* (e.g., reverse-scored item 2*: "When I am in a great mood, I tend to get into situations that could cause me problems."), (3) *sensation seeking* (e.g., reverse-scored item 3*: "I generally seek new and exciting experiences and sensations."), (4) *negative urgency* (e.g., reverse-scored item 4*: "When I feel bad, I will often do things I later regret in order to make myself feel better now."), and (5) *lack of perseverance* (e.g., item 5: "I generally like to see things through to the end."). Each of the five instrument's dimensions includes 4 items that are scored (or reverse-scored) on a four-point Likert-type scale (from 1 = "Strongly agree" to 4 = "Strongly disagree") and that provide arithmetic mean scores likewise ranging from 1 (i.e., the lowest level of endorsement of the corresponding impulsive behavior dimension) to 4 (i.e., the highest level of endorsement of the corresponding impulsive behavior dimension).

The *13-item Brief Self-Control Scale* (BSCS-13; Tangney et al., 2004) is a self-administered English-language psychometric instrument that assesses the applicability of 13 statements related to *self-control* (e.g., reverse-scored item 1*: "I am good at resisting temptation."). The instrument

Table 2.3. Evaluation phase

<table>
<tr><th colspan="5">Evaluation phase</th><th rowspan="10">Established UPPS-P-20-R</th></tr>
<tr><th>Materials</th><th colspan="2">Data</th><th>Analyses</th><th>Results</th></tr>
<tr><td>Sociodemographic information</td><td rowspan="8">Participant data collection Evaluation phase (N = 398)</td><td rowspan="3"></td><td>Descriptive analyses</td><td>Sociodemographic information</td></tr>
<tr><td rowspan="3">Pre-established UPPS-P-20-R</td><td rowspan="2">Confirmatory factor analyses</td><td>Internal structure validity (RQ6)</td></tr>
<tr><td>Internal consistency reliability (RQ7)</td></tr>
<tr><td>Participant data collection Evaluation phase (N = 67)</td><td rowspan="5">Correlation analyses</td><td>Test-retest reliability (RQ8)</td></tr>
<tr><td>BSCS-13</td><td rowspan="4"></td><td>Convergent validity (RQ9)</td></tr>
<tr><td>GAD-7</td><td rowspan="3">Criterion validity (RQ10)</td></tr>
<tr><td>PHQ-9</td></tr>
<tr><td>PMPUQ-SV-15</td></tr>
</table>

UPPS-P-20-R = revised 20-item short version of the UPPS-P Impulsive Behavior Scale. BSCS-13 = 13-item Brief Self-Control Scale (Tangney et al., 2004). GAD-7 = 7-item Generalized Anxiety Disorder Scale (Spitzer et al., 2006). PHQ-9 = 9-item Patient Health Questionnaire (Kroenke et al., 2001). PMPUQ-SV-15 = 15-item short version of the Problematic Mobile Phone Use Questionnaire (López-Fernández et al., 2018).

includes 13 items that are scored (or reverse-scored) on a five-point Likert-type scale (from 1 = "Not at all like me" to 5 = "Very much like me") and that provide an arithmetic mean score likewise ranging from 1 (i.e., the lowest level of endorsement of self-control) to 5 (i.e., the highest level of endorsement of self-control). Participant data collected with respect to the present instrument was employed to evaluate convergent validity evidence, as previous original research articles have suggested that self-control presents differential associations with impulsive behavior dimensions (Brevers et al., 2017; Hasegawa et al., 2020).

The *7-item Generalized Anxiety Disorder Scale* (GAD-7; Spitzer et al., 2006) is a self-administered English-language psychometric instrument that assesses the applicability of 7 statements related to *anxiety* (e.g., item 1: "Feeling nervous, anxious or on edge."). The instrument includes 7 items that are scored on a four-point Likert-type scale (from 0 = "Not at all" to 3 = "Nearly every day") and that provide an arithmetic mean score likewise ranging from 0 (i.e., the lowest level of endorsement of anxiety) to 3 (i.e., the highest level of endorsement of anxiety). Participant data collected with respect to the present instrument was employed to evaluate criterion validity evidence, as previous original research articles have suggested that anxiety presents differential associations with impulsive behavior dimensions (Billieux et al., 2012; Claréus et al., 2017; Lim & Kim, 2018).

The *9-item Patient Health Questionnaire* (PHQ-9; Kroenke et al., 2001) is a self-administered English-language psychometric instrument that assesses the applicability of 9 statements related to *depression* (e.g., item 1: "Little interest or pleasure in doing things."). The instrument includes 9 items that are scored on a four-point Likert-type scale (from 0 = "Not at all" to 3 = "Nearly every day") and that provide an arithmetic mean score likewise ranging from 0 (i.e., the lowest level of endorsement of depression) to 3 (i.e., the highest level of endorsement of depression). Participant data collected with respect to the present instrument was employed to evaluate criterion validity evidence, as previous original research articles have suggested that depression presents differential associations with impulsive behavior dimensions (Billieux et al., 2012; Claréus et al., 2017; Lim & Kim, 2018).

The *15-item short version of the Problematic Mobile Phone Use Questionnaire* (PMPUQ-SV-15; López-Fernández et al., 2018) is a self-administered English-language psychometric instrument that assesses the applicability of 15 statements related to three different dimensions of problematic mobile phone use, namely (1) *dependent mobile phone use* (e.g., item 1: "It is easy for me to spend all day not using my mobile phone."), (2) *dangerous mobile phone use* (e.g., reverse-scored item 2*: "I use my mobile phone while driving."), and (3) *prohibited mobile phone use* (e.g., item 3: "I don't use my mobile phone when it is completely forbidden to use it."). Each of the three instrument's dimensions includes 5 items that are scored (or reverse-scored) on a four-point Likert-type scale (from 1 = "Strongly agree" to 4 = "Strongly disagree") and that provide arithmetic mean scores likewise ranging from 1 (i.e., the lowest level of endorsement of the corresponding problematic mobile phone use dimension) to 4 (i.e., the highest level of endorsement of the corresponding problematic mobile phone use dimension). Participant data collected with respect to the present instrument was employed to evaluate criterion validity evidence, as previous original research articles have suggested that dependent mobile phone use, dangerous mobile phone use, and prohibited mobile phone use present differential associations with impulsive behavior dimensions (Billieux et al., 2008; Canale et al., 2021).

2.3.2. Data

Participants were recruited using *Prolific* (https://www.prolific.com/), a third-party crowdsourcing platform for recruiting and compensating participants for online research, and were redirected to *Qualtrics* (https://www.qualtrics.com/), a third-party crowdsourcing platform for implementing online surveys.

We collected participant data with respect to their sociodemographic information and to their answers on the UPPS-P-20-R, the 13-item Brief Self-Control Scale (BSCS-13; Tangney et al., 2004), the 7-item Generalized Anxiety Disorder Scale (GAD-7; Spitzer et al., 2006), the 9-item Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001), and the 15-item short version of the Problematic Mobile Phone Use Questionnaire (PMPUQ-SV-15; López-Fernández et al., 2018). Prior to proceeding with participant data collection, N = 32 "pilot" participants (Perneger et al., 2015) were 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 were paid for their full completion of the online survey. Of note, all "pilot" participant data was 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 11.8 minutes, "pilot" participants were paid $\lceil 11.8 \rceil$*(9/60) = 1.80 GBP for their full completion of the online survey to equate to a "good" hourly rate of 9.00 GBP, in accordance with *Prolific* (https://www.prolific.com/). Let T be the median time (in minutes) it took "pilot" participants to complete all statements implemented in the pilot full online survey; participants were paid $\lceil T \rceil$*(9/60) = $\lceil 6.0 \rceil$*(9/60) = 0.90 GBP for their full completion of the online survey to equate to a "good" hourly rate of 9.00 GBP, in accordance with *Prolific* (https://www.prolific.com/). Furthermore, in order to evaluate the test-retest reliability of the UPPS-P-20-R, previously recruited participants were invited to re-complete the UPPS-P-20-R following a four-week interval and were paid $\lceil T \rceil$*(9/60) = $\lceil 6.0 \rceil$*(9/60) = 0.90 GBP for their full re-completion of the psychometric instrument.

The inclusion criteria for data collection are as follows: (1) being fluent in English and (2) being 18 years of age or older. Endorsement of all of the aforementioned criteria establishes inclusion.

The exclusion criteria for data collection are as follows: (1) failing to complete the full online survey, (2) failing the attention check implemented in the UPPS-P-20-R or – if applicable – failing two to five of the five attention checks implemented in the five psychometric instruments (e.g., "This is an attention check. Please select "Agree strongly" as an answer to the present statement."), (3) completing the full online survey in less time than three standard deviations below the mean time it took participants to complete the full online survey, and (4) having completed the full online survey in the prior development phase (i.e., development phase I). Endorsement of any of the aforementioned criteria establishes exclusion.

With respect to data collection, we required that participants provide answers to all statements implemented in the full online survey and adopted a listwise deletion approach to missing data handling by excluding data from participants who failed to complete the full online survey.

To estimate the minimum number of participants to be recruited for data collection, Monte Carlo simulation analyses were performed with respect to the pre-established five-factor structure of the 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20; Billieux et al., 2012). In summary, using estimates from prior research conducted on the structural equation model of interest, the Monte Carlo simulation method is employed to iteratively generate samples of varying sample sizes, fit them to the said structural equation model, assess their quality of adjustment, and estimate the minimum sample size that consistently yields adequate quality of adjustment (Muthén & Muthén, 2002). In the present registered report, Monte Carlo simulation analyses were performed using the *R* packages *lavaan* version 0.6-17 (Rosseel et al., 2023) and *simsem* version 0.5-16 (Pornprasertmanit et al., 2021). The model parameter values employed in generating the data were derived from the model-implied standardized estimates provided by Fournier, Bőthe, et al. (2025) in their confirmatory factor analyses performed with respect to the pre-established five-factor structure of the UPPS-P-20 (Billieux et al., 2012). The mean model parameter values derived from the model-implied standardized estimates provided by Fournier, Bőthe, et al. (2025) were then employed in specifying the data generation model parameter values. Using the latter data generation model, 2,000 samples were generated across twenty incremental sample sizes (i.e., 100 samples per sample size of interest) varying from N = 20*1 = 20 to N = 20*20 = 400. To fit the 2,000 samples to the structural equation model, weighted least squares mean-and-variance-adjusted robust estimation methods were employed (Finney & di Stefano, 2013). Of note, in fitting the structural equation model, the ordered categorical nature of the observed variables was accounted for. To assess the quality of adjustment of the 2,000 samples to the structural equation model, three conventional model-implied fit indices were employed: the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). Good fit was determined by a CFI $\geq$ 0.950, a TLI $\geq$ 0.950, and an RMSEA $\leq$ 0.050. As a result, the Monte Carlo simulation analyses suggested that the minimum sample size that consistently yielded good quality of adjustment – as determined by the 95th percentiles of the distributions of each of the three conventional model-implied fit indices – was equal to N = 20*17 = 340.

To estimate the minimum number of participants to be recruited to re-complete the UPPS-P-20-R following a four-week interval, two-sided correlation test power analyses were performed. In summary, specifying all but the sample size N parameter values (i.e., correlation coefficient ρ, significance level α, power β) involved in the two-sided correlation test equation, the two-sided correlation test power method is employed to iteratively input varying sample size values to the said equation and estimate the minimum sample size that yields an adequate effect size (Cohen, 1988). In the present registered report, two-sided correlation test power analyses were performed using the *R* package *pwr* version 1.3-0 (Champely, 2020). The parameter values employed in estimating the minimum sample size that yields an adequate effect size, which was determined by ρ^2 = 0.500 (ρ = 0.707), were equal to α = 0.001 and β = 0.999. As a result, the two-sided correlation test power analyses suggested that the minimum sample size that yielded an adequate effect size was equal to N = 56.

All in all, with respect to the minimum number of participants to be recruited for data collection, which equals N = 340 and N = 56, respectively, we terminated data collection upon reaching 20% increments, i.e., upon reaching $N = \lceil 340*(120/100) \rceil = 408$ and $N = \lceil 56*(120/100) \rceil = 68$, respectively.

Of the 408 recruited participants, 4 (0.980%) did not complete any statement implemented in the full online survey, and 6 (1.471%) were not included for not being fluent in English, yielding a sample size equal to N = 398. Of the 68/398 previously recruited participants recruited to re-complete the UPPS-P-20-R following a four-week interval, 1 (1.471%) did not complete any statement implemented in the full online survey, yielding a sample size equal to N = 67.

2.3.3. Analyses

Descriptive analyses of the participant sociodemographic information were performed with respect to the total sample. Subsequently, participant sociodemographic information is reported in the “Results” section of the manuscript through (1) sample size, (2) mean, standard deviation, and range values of their age, (3) percentage frequency values of their gender identity, and (4) percentage frequency values of their country of residence.

Confirmatory factor analyses of the UPPS-P-20-R were performed with respect to its pre-established five-factor structure using the *R* package *lavaan* version 0.6-19 (Rosseel et al., 2024). To fit the structural equation model, weighted least squares mean-and-variance-adjusted robust estimation methods were employed (Finney & di Stefano, 2013). Of note, in fitting the structural equation model, the ordered categorical nature of the observed variables was accounted for. To assess the quality of adjustment to the data of the structural equation model, global exact, global approximate, and local fit were examined. To examine global exact fit, a global exact fit hypothesis test was 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 equal to α = 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 global exact fit was determined by a corresponding probability value equal to or greater than a significance level equal to α = 0.050. To examine global approximate fit, three model-implied fit indices were employed: the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR) (Kline, 2023). To estimate dynamic threshold values for the three aforementioned model-implied global approximate fit indices, Monte Carlo simulation analyses were performed with respect to the structural equation model using the *R* package *dynamic* version 1.1.0 (Wolf & McNeish, 2025). 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), as determined by the 95th percentiles of the distributions of each of the three model-implied fit indices (McNeish, 2023; McNeish & Wolf, 2023). Adequate global approximate fit was determined by a CFI greater than or equal to the dynamic lower-bound value, an RMSEA lower than or equal to the dynamic upper-bound value, and an SRMR 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 was inspected (Kline, 2023). If the aforementioned decision rules for adequate global exact and global approximate fit are not met, alternative structural equation models were 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 internal structure validity and internal consistency reliability evidence, confirmatory factor analysis results for all structural equation models that were relevant to this investigation are 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 global approximate fit indices (i.e., comparative fit indices (CFI), root mean square errors of approximation (RMSEA) along with their corresponding 90% confidence intervals, and standardized root mean square residuals (SRMR)) 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.

Using the participant data collected prior to and posterior to the four-week interval with respect to the UPPS-P-20-R, correlation analyses were performed using two-sided Spearman's ρ rank correlation tests on the psychometric instrument's arithmetic mean scores. Subsequently, to highlight test-retest reliability evidence, correlation analysis results for the psychometric instrument's dimensions are reported in the "Results" section of the manuscript through Spearman's ρ^2 rank correlation coefficient values. Test-retest reliability evidence was labeled "strong" if $\rho^2 \geqslant 0.500$, "adequate" if $\rho^2 \in [0.250, 0.500[$, "weak" if $\rho^2 \in [0.100, 0.250[$, and "absent" if $\rho^2 < 0.100$.

Using the participant data collected with respect to the UPPS-P-20-R and the 13-item Brief Self-Control Scale (BSCS-13; Tangney et al., 2004), correlation analyses were performed using two-sided Spearman's ρ rank correlation tests on the psychometric instruments' arithmetic mean scores. Subsequently, to highlight convergent validity evidence, correlation analysis results for the psychometric instruments' dimensions are reported in the "Results" section of the manuscript through Spearman's ρ rank correlation coefficient values. Convergent validity evidence was labeled "strong" if $|\rho| \geqslant 0.500$, "adequate" if $|\rho| \in [0.250, 0.500[$, "weak" if $|\rho| \in [0.100, 0.250[$, and "absent" if $|\rho| < 0.100$.

Using the participant data collected with respect to the UPPS-P-20-R, the 7-item Generalized Anxiety Disorder Scale (GAD-7; Spitzer et al., 2006), the 9-item Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001), and the 15-item short version of the Problematic Mobile Phone Use Questionnaire (PMPUQ-SV-15; López-Fernández et al., 2018), correlation analyses were performed using two-sided Spearman's ρ rank correlation tests on the psychometric instruments' arithmetic mean scores. Subsequently, to highlight criterion validity evidence, correlation analysis results for the psychometric instruments' dimensions are reported in the "Results" section of the manuscript through Spearman's ρ rank correlation coefficient values. Criterion validity evidence was labeled "strong" if $|\rho| \geqslant 0.500$, "adequate" if $|\rho| \in [0.250, 0.500[$, "weak" if $|\rho| \in [0.100, 0.250[$, and "absent" if $|\rho| < 0.100$.

3. Results

3.1. Development phase I

Participant data collection yielded a sample size equal to $N = 948$. With respect to their age, participants were between 18 and 83 years of age ($M = 37.825$, $SD = 13.303$). With respect to their gender identity, 484 participants (51.055%) identified as men, 455 (47.996%) as women, and 9 (0.949%) as belonging to other gender identities. With respect to their country of residence, 409 participants (43.143%) resided in the United Kingdom of Great Britain and Northern Ireland, 170 (17.932%) in the United States of America, 93 (9.810%) in Canada, and 276 (29.114%) across 35 other countries of residence.

- **RQ1.** What is the internal structure validity of the UPPS-P-n_i?

An original structural equation model corresponding to the initial item pool derived from the original version of the UPPS-P Impulsive Behavior Scale (UPPS-P-n_i) – describing five latent variables and 50 observed variables – was fitted.

With respect to global exact fit examination, model-implied χ^2 test statistics along with their corresponding degrees of freedom and probability values are as follows:

> χ^2 (1,165) = 6,271.321 ($p < 0.001$).

The null hypothesis that the difference between the population covariance matrix and the model-implied covariance matrix is null can be rejected at a significance level equal to $\alpha = 0.050$. In this light, decision rules for adequate global exact fit are not met.

With respect to global approximate fit examination, model-implied global approximate fit indices are as follows:

> CFI = 0.928, RMSEA = 0.068 [0.066, 0.070], SRMR = 0.078.

The CFI is not greater than or equal to the level four dynamic lower-bound value (CFI $\geqslant$ 0.977), the RMSEA is not lower than or equal to the level four dynamic upper-bound value (RMSEA $\leqslant$ 0.039), and the SRMR is not lower than or equal to the level four dynamic upper-bound value (SRMR $\leqslant$ 0.044), all of which are suggested by the results of the Monte Carlo simulation analyses (see Supplementary materials, Table S–3.1.I.). In this light, decision rules for adequate global approximate fit are not met.

With respect to local fit examination, upon inspection of the difference between the population covariance matrix and the model-implied covariance matrix, nine instances of local misfit were observed (see Supplementary materials, Table S–3.1.II.).

In light of the highlighted internal structure validity evidence, we decided to reject the original structural equation model corresponding to the UPPS-P-n_i.

An alternative structural equation model corresponding to the UPPS-P-n_i – describing five latent variables and 40 observed variables – was fitted by omitting four observed variables based on instances of model-implied non-null λ standardized estimates < 0.500 and six observed variables based on instances of local misfit.

With respect to global exact fit examination, model-implied χ^2 test statistics along with their corresponding degrees of freedom and probability values are as follows:

> χ^2 (730) = 2,285.704 ($p < 0.001$).

The null hypothesis that the difference between the population covariance matrix and the model-implied covariance matrix is null can be rejected at a significance level equal to $\alpha = 0.050$. In this light, decision rules for adequate global exact fit are not met.

With respect to global approximate fit examination, model-implied global approximate fit indices are as follows:

> CFI = 0.976, RMSEA = 0.047 [0.045, 0.050], SRMR = 0.048.

The CFI is greater than or equal to the level three dynamic lower-bound value (CFI ⩾ 0.977), the RMSEA is lower than or equal to the level three dynamic upper-bound value (RMSEA ⩽ 0.061), and the SRMR is lower than or equal to the level three dynamic upper-bound value (SRMR ⩽ 0.056), all of which are suggested by the results of the Monte Carlo simulation analyses (see Supplementary materials, Table S–3.1.III.). In this light, decision rules for adequate global approximate fit are met.

With respect to local fit examination, upon inspection of the difference between the population covariance matrix and the model-implied covariance matrix, no instance of local misfit was observed (see Supplementary materials, Table S–3.1.IV.).

In light of the highlighted internal structure validity evidence, we decided not to reject the alternative structural equation model corresponding to the UPPS-P-n_i.

- **RQ2.** What is the internal consistency reliability of the UPPS-P-n_i?

With respect to the original structural equation model corresponding to the UPPS-P-n_i, model-implied McDonald's ω internal consistency values are as follows:

> $\omega_{\text{lack of premeditation}} = 0.859$, $\omega_{\text{negative urgency}} = 0.931$, $\omega_{\text{sensation seeking}} = 0.839$, $\omega_{\text{lack of perseverance}} = 0.910$, $\omega_{\text{positive urgency}} = 0.956$.

With respect to the alternative structural equation model corresponding to the UPPS-P-n_i, model-implied McDonald's ω internal consistency values are as follows:

> $\omega_{\text{lack of premeditation}} = 0.870$, $\omega_{\text{negative urgency}} = 0.934$, $\omega_{\text{sensation seeking}} = 0.807$, $\omega_{\text{lack of perseverance}} = 0.911$, $\omega_{\text{positive urgency}} = 0.962$.

All in all, all highlighted internal structure validity and internal consistency reliability evidence was employed to decide that the alternative structural equation model corresponded to the UPPS-P-n_i.

3.2. Development phase II

- **RQ3.** What is the content validity of the pre-established UPPS-P-20-R?

A n_i-item Gaussian graphical model corresponding to the initial item pool derived from the original version of the UPPS-P Impulsive Behavior Scale (UPPS-P-n_i) – describing 40 observed variables – and five n_j-item Gaussian graphical models corresponding to each of its five factors – describing four to 12 observed variables – were fitted.

With respect to the n_i-item Gaussian graphical model corresponding to the UPPS-P-n_i describing 40 observed variables, the model-implied graph drawing is reported below (see Figure 3.2.). The model-implied community structure (modularity Q value = 0.561) (see Figure 3.2.) corroborates that of the alternative structural equation model corresponding to the UPPS-P-n_i. The five model-implied optimal subsets (see Figure 3.2.) are suggested by the results of the brute force combinatorial search algorithm implemented to solve the combinatorial optimization problem expressed in Equation 2.2. with respect to each of the five n_j-item Gaussian graphical models (see Supplementary materials, Figure S–3.2.I. to Figure S–3.2.V.).

In Gaussian graphical models, observed variables are represented by nodes, and associations between observed variables are represented by edges, each assigned a weight – derived from the partial correlation coefficient value of its corresponding nodes – that represents the pairwise conditional (in-)dependence between observed variables after conditioning on all other observed variables included in the model. In the present figure, nodes (i.e., observed variables) are depicted as circles, and edges (i.e., associations between observed variables) are depicted as lines. Edge color reflects the sign of the edges, with green reflecting positive edges and red reflecting negative edges. Edge opacity/width reflects the absolute values of the edge weights. Node color reflects the model-implied community structure (modularity Q value = 0.561). Of note, the model-implied community structure corroborates that of the alternative structural equation model corresponding to the UPPS-P-n_i, with orange nodes reflecting lack of premeditation, sky blue nodes reflecting negative urgency, bluish green nodes reflecting sensation seeking, yellow nodes reflecting lack of perseverance, and reddish purple nodes reflecting positive urgency. Node opacity/width reflects the five model-implied optimal subsets. Of note, the five model-implied optimal subsets are suggested by the results of the brute force combinatorial search algorithm implemented to solve the combinatorial optimization problem expressed in Equation 2.2. with respect to each of the five n_j-item Gaussian graphical models (see Supplementary materials, Figure S–3.2.I. to Figure S–3.2.V.).

All in all, all highlighted content validity evidence was employed to decide that the five model-implied optimal subsets corresponded to the pre-established revised 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20-R).

- **RQ4.** What is the internal structure validity of the pre-established UPPS-P-20-R?

An original structural equation model corresponding to the pre-established UPPS-P-20-R – describing five latent variables and 20 observed variables – was fitted.

With respect to global exact fit examination, model-implied χ^2 test statistics along with their corresponding degrees of freedom and probability values are as follows:

> χ^2 (160) = 606.230 ($p < 0.001$).

The null hypothesis that the difference between the population covariance matrix and the model-implied covariance matrix is null can be rejected at a significance level

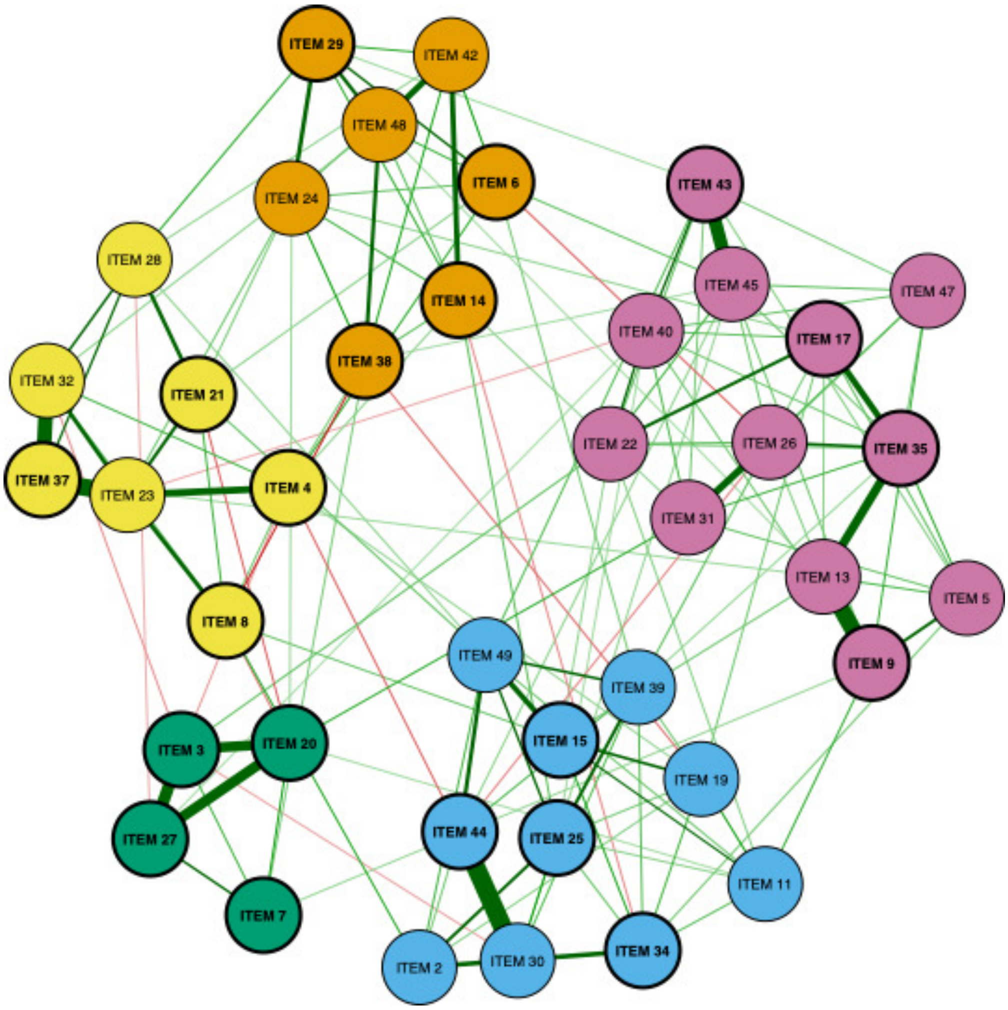

Figure 3.2. Model-implied graph drawing for the n_i-item Gaussian graphical model corresponding to the initial item pool derived from the original version of the UPPS-P Impulsive Behavior Scale (UPPS-P-n_i)

In Gaussian graphical models, observed variables are represented by nodes, and associations between observed variables are represented by edges, each assigned a weight – derived from the partial correlation coefficient value of its corresponding nodes – that represents the pairwise conditional (in-)dependence between observed variables after conditioning on all other observed variables included in the model. In the present figure, nodes (i.e., observed variables) are depicted as circles, and edges (i.e., associations between observed variables) are depicted as lines. Edge color reflects the sign of the edges, with green reflecting positive edges and red reflecting negative edges. Edge opacity/width reflects the absolute values of the edge weights. Node color reflects the model-implied community structure (modularity Q value = 0.561). Of note, the model-implied community structure corroborates that of the alternative structural equation model corresponding to the UPPS-P-n_i, with orange nodes reflecting lack of premeditation, sky blue nodes reflecting negative urgency, bluish green nodes reflecting sensation seeking, yellow nodes reflecting lack of perseverance, and reddish purple nodes reflecting positive urgency. Node opacity/width reflects the five model-implied optimal subsets. Of note, the five model-implied optimal subsets are suggested by the results of the brute force combinatorial search algorithm implemented to solve the combinatorial optimization problem expressed in Equation 2.2. with respect to each of the five n_j-item Gaussian graphical models (see Supplementary materials, Figure S–3.2.I. to Figure S–3.2.V.).

equal to α = 0.050. In this light, decision rules for adequate global exact fit are not met.

With respect to global approximate fit examination, model-implied global approximate fit indices are as follows:

> CFI = 0.978, RMSEA = 0.054 [0.050, 0.059], SRMR = 0.051.

The CFI is greater than or equal to the level two dynamic lower-bound value (CFI ⩾ 0.959), the RMSEA is lower than or equal to the level two dynamic upper-bound value (RMSEA ⩽ 0.074), and the SRMR is lower than or equal to the level two dynamic upper-bound value (SRMR ⩽ 0.058), all of which are suggested by the results of the Monte Carlo simulation analyses (see Supplementary materials, Table S–3.2.VI.). In this light, decision rules for adequate global approximate fit are met.

With respect to local fit examination, upon inspection of the difference between the population covariance matrix and the model-implied covariance matrix, no instance of local misfit was observed (see Supplementary materials, Table S–3.2.VII.).

In light of the highlighted internal structure validity evidence, we decided not to reject the original structural equation model corresponding to the pre-established UPPS-P-20-R.

- **RQ5.** What is the internal consistency reliability of the pre-established UPPS-P-20-R?

With respect to the original structural equation model corresponding to the pre-established UPPS-P-20-R, model-implied McDonald's ω internal consistency values are as follows:

> $\omega_{\text{lack of premeditation}} = 0.764$, $\omega_{\text{negative urgency}} = 0.829$, $\omega_{\text{sensation seeking}} = 0.798$, $\omega_{\text{lack of perseverance}} = 0.774$, $\omega_{\text{positive urgency}} = 0.896$.

All in all, all highlighted internal structure validity and internal consistency reliability evidence was employed to decide that the original structural equation model corresponded to the pre-established UPPS-P-20-R.

3.3. Evaluation phase

Participant data collection yielded a sample size equal to $N = 398$. With respect to their age, participants were between 18 and 77 years of age ($M = 37.503$, $SD = 12.882$). With respect to their gender identity, 211 participants (53.015%) identified as women, 183 (45.980%) as men, and 4 (1.005%) as belonging to other gender identities. With respect to their country of residence, 204 participants (51.256%) resided in the United Kingdom of Great Britain and Northern Ireland, 69 (17.337%) in the United States of America, 24 (6.030%) in Canada, and 101 (25.377%) across 25 other countries of residence.

- **RQ6.** What is the internal structure validity of the UPPS-P-20-R?

An original structural equation model corresponding to the revised 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20-R) – describing five latent variables and 20 observed variables – was fitted.

With respect to global exact fit examination, model-implied χ^2 test statistics along with their corresponding degrees of freedom and probability values are as follows:

> χ^2 (160) = 404.325 ($p < 0.001$).

The null hypothesis that the difference between the population covariance matrix and the model-implied covariance matrix is null can be rejected at a significance level equal to $\alpha = 0.050$. In this light, decision rules for adequate global exact fit are not met.

With respect to global approximate fit examination, model-implied global approximate fit indices are as follows:

> CFI = 0.963, RMSEA = 0.062 [0.055, 0.070], SRMR = 0.063.

The CFI is greater than or equal to the level four dynamic lower-bound value (CFI ⩾ 0.959), the RMSEA is lower than or equal to the level four dynamic upper-bound value (RMSEA ⩽ 0.077), and the SRMR is lower than or equal to the level four dynamic upper-bound value (SRMR ⩽ 0.065), all of which are suggested by the results of the Monte Carlo simulation analyses (see Supplementary materials, Table S–3.3.I.). In this light, decision rules for adequate global approximate fit are met.

With respect to local fit examination, upon inspection of the difference between the population covariance matrix and the model-implied covariance matrix, two instances of local misfit were observed (see Supplementary materials, Table S–3.3.II.).

The first instance of local misfit is as follows:

- Lack of premeditation item 19: "Before I get into a new situation, I like to find out what to expect from it.".

Differences between the population covariance matrix and the model-implied covariance matrix suggest that the item listed above may also assess the impulsive behavior dimension of sensation seeking. Yet, we, the authors, consider that the item listed above does not provide relevant information about one's attraction to excitement and openness to new experiences (i.e., sensation seeking), therefore neither prompting *post hoc* model modification nor rejection of the original structural equation model corresponding to the UPPS-P-20-R.

The second instance of local misfit is as follows:

- Sensation seeking reverse-scored item 14*: "I quite enjoy taking risks.".

Differences between the population covariance matrix and the model-implied covariance matrix suggest that the item listed above may also assess the impulsive behavior dimension of lack of premeditation. Yet, we, the authors, consider that the item listed above does not provide relevant information about one's lack of reflection on the potential consequences of actions preceding their emission (i.e., lack of premeditation), therefore neither prompting *post hoc* model modification nor rejection of the original structural equation model corresponding to the UPPS-P-20-R.

In light of the highlighted internal structure validity evidence, we decided not to reject the original structural equation model corresponding to the UPPS-P-20-R.

- **RQ7.** What is the internal consistency reliability of the UPPS-P-20-R?

With respect to the original structural equation model corresponding to the UPPS-P-20-R, model-implied McDonald's ω internal consistency values are as follows:

> $\omega_{\text{lack of premeditation}} = 0.722$, $\omega_{\text{positive urgency}} = 0.870$, $\omega_{\text{sensation seeking}} = 0.769$, $\omega_{\text{negative urgency}} = 0.823$, $\omega_{\text{lack of perseverance}} = 0.812$.

All in all, all highlighted internal structure validity and internal consistency reliability evidence was employed to decide that the original structural equation model corresponded to the UPPS-P-20-R.

The model-implied structural equation drawing for the original structural equation model corresponding to the UPPS-P-20-R is reported below (see Figure 3.3.).

- **RQ8.** What is the test-retest reliability of the UPPS-P-20-R?

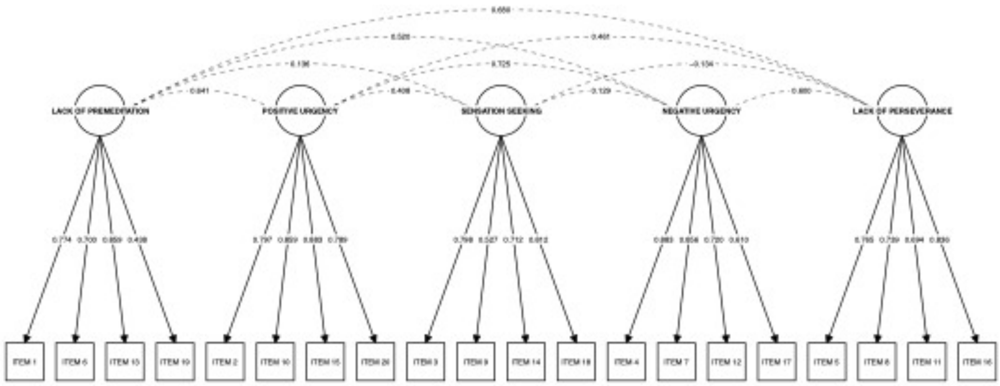

Figure 3.3. Model-implied structural equation drawing for the original structural equation model corresponding to the revised 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20-R)

In the present figure, factors (i.e., latent variables) are depicted as circles, and items (i.e., observed variables) are depicted as squares. Lines connecting factors and items reflect non-null λ parameters (i.e., factor loadings) along with their corresponding model-implied non-null λ standardized estimates. Lines connecting factors reflect off-diagonal ψ parameters (i.e., factor covariances) along with their corresponding model-implied off-diagonal ψ standardized estimates. Of note, diagonal ψ parameters (i.e., factor variances) were set to unity.

Table 3.3.1. Convergent validity evidence

		UPPS-P-20-R				
		Lack of premeditation	Positive urgency	Sensation seeking	Negative urgency	Lack of perseverance
BCSC-13	Self-control	-0.481***	-0.541***	-0.122*	-0.645***	-0.687***

UPPS-P-20-R = revised 20-item short version of the UPPS-P Impulsive Behavior Scale. BSCS-13 = 13-item Brief Self-Control Scale (Tangney et al., 2004). In the present table, Spearman's ρ rank correlation coefficient values are reported along with their corresponding probability values, with no asterisk reflecting $p \geqslant 0.050$, one asterisk reflecting $p \in]0.050, 0.010]$, two asterisks reflecting $p \in]0.010, 0.001]$, and three asterisks reflecting $p < 0.001$. Convergent validity evidence is labeled "strong" if $|\rho| \geqslant 0.500$, "adequate" if $|\rho| \in [0.250, 0.500[$, "weak" if $|\rho| \in [0.100, 0.250[$, and "absent" if $|\rho| < 0.100$.

With respect to test-retest reliability evidence, the results of the two-sided Spearman's ρ rank correlation tests on the psychometric instrument's arithmetic mean scores are as follows:

> $\rho^2_{\text{lack of premeditation}} = 0.367$ ($\rho = 0.606$, $p < 0.001$), $\rho^2_{\text{positive urgency}} = 0.508$ ($\rho = 0.713$, $p < 0.001$), $\rho^2_{\text{sensation seeking}} = 0.438$ ($\rho = 0.662$, $p < 0.001$), $\rho^2_{\text{negative urgency}} = 0.576$ ($\rho = 0.759$, $p < 0.001$), $\rho^2_{\text{lack of perseverance}} = 0.532$ ($\rho = 0.729$, $p < 0.001$).

Strong test-retest reliability evidence ($\rho^2 \geqslant 0.500$) is highlighted with respect to the impulsive behavior dimensions of positive urgency, negative urgency, and lack of perseverance. Adequate test-retest reliability evidence ($\rho^2 \in [0.250, 0.500[$) is highlighted with respect to the impulsive behavior dimensions of lack of premeditation and sensation seeking.

- **RQ9.** What is the convergent validity of the UPPS-P-20-R?

With respect to convergent validity evidence, the results of the two-sided Spearman's ρ rank correlation tests on the psychometric instruments' arithmetic mean scores are reported below (see Table 3.3.1.).

Strong convergent validity evidence ($|\rho| \geqslant 0.500$) is highlighted with respect to the impulsive behavior dimensions of positive urgency, negative urgency, and lack of perseverance. Adequate convergent validity evidence ($|\rho| \in [0.250, 0.500[$) is highlighted with respect to the impulsive behavior dimension of lack of premeditation. Weak convergent validity evidence ($|\rho| \in [0.100, 0.250[$) is highlighted with respect to the impulsive behavior dimension of sensation seeking.

- **RQ10.** What is the criterion validity of the UPPS-P-20-R?

With respect to criterion validity evidence, the results of the two-sided Spearman's ρ rank correlation tests on the psychometric instruments' arithmetic mean scores are reported below (see Table 3.3.2.).

Adequate criterion validity evidence ($|\rho| \in [0.250, 0.500[$) is highlighted with respect to the impulsive behavior dimensions of lack of premeditation, positive urgency, negative urgency, and lack of perseverance. Weak criterion validity evidence ($|\rho| \in [0.100, 0.250[$) is highlighted with respect to the impulsive behavior dimension of sensation seeking.

4. Discussion

The UPPS-P Impulsive Behavior Scale is a well-established psychometric instrument for assessing impulsivity, a key multidimensional psychological construct encompassing distinct facets that are differentially associated with numerous psychopathological and neuropathological symptoms across various disorders (Berg et al., 2015; Rochat et

Table 3.3.2. Criterion validity evidence

		UPPS-P-20-R				
		Lack of premeditation	Positive urgency	Sensation seeking	Negative urgency	Lack of perseverance
GAD-7	Anxiety	0.166***	0.331***	-0.029	0.452***	0.429***
PHQ-9	Depression	0.216***	0.361***	0.050	0.458***	0.498***
PMPUQ-SV-15	Dependent mobile phone use	0.057	0.112*	-0.034	0.129**	0.183***
	Dangerous mobile phone use	0.104*	0.336***	0.154**	0.120*	0.020
	Prohibited mobile phone use	0.260***	0.359***	0.196***	0.231***	0.252***

UPPS-P-20-R = revised 20-item short version of the UPPS-P Impulsive Behavior Scale. GAD-7 = 7-item Generalized Anxiety Disorder Scale (Spitzer et al., 2006). PHQ-9 = 9-item Patient Health Questionnaire (Kroenke et al., 2001). PMPUQ-SV-15 = 15-item short version of the Problematic Mobile Phone Use Questionnaire (López-Fernández et al., 2018). In the present table, Spearman's ρ rank correlation coefficient values are reported along with their corresponding probability values, with no asterisk reflecting $p \geqslant 0.050$, one asterisk reflecting $p \in]0.050, 0.010]$, two asterisks reflecting $p \in]0.010, 0.001]$, and three asterisks reflecting $p < 0.001$. Criterion validity evidence is labeled "strong" if $|\rho| \geqslant 0.500$, "adequate" if $|\rho| \in [0.250, 0.500[$, "weak" if $|\rho| \in [0.100, 0.250[$, and "absent" if $|\rho| < 0.100$.

al., 2018; Smith et al., 2007). To facilitate its integration in clinical and research settings, Billieux et al. (2012) developed and evaluated the psychometric properties of a 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20), reducing assessment time to approximately one-third of that of the original version of the psychometric instrument. Despite its widespread popularity, legitimate concerns were raised about the content validity of this short version of the UPPS-P Impulsive Behavior Scale (Cyders et al., 2014). Indeed, the inevitable loss of construct-level content coverage in short-form development, although often unaddressed by researchers, warrants in-depth examination (Smith et al., 2000). Therefore, the present registered report aimed to develop and evaluate the psychometric properties of a revised short version of the UPPS-P Impulsive Behavior Scale through an expert-driven and data-driven methodological approach to short-form development emphasizing item-level and construct-level content validity.

Our results suggest that the psychometric properties of the revised 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20-R) are stronger than – or in line with – those reported in previous original research articles with respect to the UPPS-P-20 (Billieux et al., 2012) and to other versions of the said psychometric instrument.

First, all highlighted internal structure validity evidence is adequate. The model-implied global approximate fit indices with respect to the UPPS-P-20-R (N = 398, χ^2 (160) = 404.325 (p < 0.001), CFI = 0.963, RMSEA = 0.062 [0.055, 0.070], SRMR = 0.063) are in line with those reported by Fournier, Bőthe, et al. (2025) (N = 82,243, χ^2 (160) = 52,867.560 (p < 0.001), CFI = 0.957, RMSEA = 0.063 [0.063, 0.064], SRMR = NA). Although two instances of local misfit were observed, neither prompted *post hoc* model modification nor rejection of the original structural equation model corresponding to the UPPS-P-20-R.

Second, all highlighted internal consistency reliability evidence is adequate. The model-implied McDonald's ω internal consistency values with respect to the UPPS-P-20-R (N = 398, $\omega \in [0.722, 0.870]$) are in line with those reported by Fournier, Bőthe, et al. (2025) (N = 82,243, $\omega \in [0.736, 0.841]$).

Third, all highlighted test-retest reliability evidence is strong to adequate. Strong test-retest reliability evidence ($\rho^2 \geqslant 0.500$) is highlighted with respect to the impulsive behavior dimensions of positive urgency, negative urgency, and lack of perseverance. Adequate test-retest reliability evidence ($\rho^2 \in [0.250, 0.500[$) is highlighted with respect to the impulsive behavior dimensions of lack of premeditation and sensation seeking.

Fourth, all highlighted convergent validity evidence is strong to weak. Strong convergent validity evidence ($|\rho| \geqslant 0.500$) is highlighted with respect to the impulsive behavior dimensions of positive urgency, negative urgency, and lack of perseverance. Adequate convergent validity evidence ($|\rho| \in [0.250, 0.500[$) is highlighted with respect to the impulsive behavior dimension of lack of premeditation. Weak convergent validity evidence ($|\rho| \in [0.100, 0.250[$) is highlighted with respect to the impulsive behavior dimension of sensation seeking. Compared to those reported by previous original research articles that have suggested that self-control presents differential associations with impulsive behavior dimensions (Brevers et al., 2017; Hasegawa et al., 2020), the Spearman's ρ^2 rank correlation coefficient values with respect to the UPPS-P-20-R are stronger than those in 3 out of 9 instances (33.333%) and in line with those in 6 out of 9 instances (66.666%) (see Table 4.1.).

Fifth, all highlighted criterion validity evidence is adequate to weak. Adequate criterion validity evidence ($|\rho| \in [0.250, 0.500[$) is highlighted with respect to the impulsive behavior dimensions of lack of premeditation, positive urgency, negative urgency, and lack of perseverance. Weak criterion validity evidence ($|\rho| \in [0.100, 0.250[$) is highlighted with respect to the impulsive behavior dimension of sensation seeking. Compared to those reported by previous original research articles that have suggested that anxiety, depression, dependent mobile phone use, dangerous mobile phone use, and prohibited mobile phone use present differential associations with impulsive behavior dimensions (Billieux et al., 2008, 2012; Canale et al., 2021; Claréus et al., 2017; Lim & Kim, 2018), the Spearman's

Table 4.1. Convergent validity evidence

			UPPS-P-20-R / UPPS-P-20 / Other versions of the said psychometric instrument				
			Lack of premeditation	Positive urgency	Sensation seeking	Negative urgency	Lack of perseverance
BCSC-13	Self-control	Present registered report	Adequate	Strong	Weak	Strong	Strong
		Brevers et al. (2017)	Adequate	NA	Absent	Adequate	Strong
		Hasegawa et al. (2020)	Adequate	Strong	Weak	Strong	Adequate

UPPS-P-20-R = revised 20-item short version of the UPPS-P Impulsive Behavior Scale. UPPS-P-20 = 20-item short version of the UPPS-P Impulsive Behavior Scale (Billieux et al., 2012). BSCS-13 = 13-item Brief Self-Control Scale (Tangney et al., 2004). Convergent validity evidence is labeled “strong” if $|\rho| \geqslant 0.500$, “adequate” if $|\rho| \in [0.250, 0.500[$, “weak” if $|\rho| \in [0.100, 0.250[$, and “absent” if $|\rho| < 0.100$.

ρ^2 rank correlation coefficient values with respect to the UPPS-P-20-R are stronger than those in 17 out of 54 instances (31.481%), in line with those in 32 out of 54 instances (59.259%), and weaker than those in 5 out of 54 instances (9.259%) (see Table 4.2.).

Last, in examining the individual items comprised in the UPPS-P-20-R, we, the authors, consider that all highlighted content validity evidence is strong. More specifically, posterior to our expert-driven and data-driven methodological approach to short-form development emphasizing item-level and construct-level content validity, we consider that all individual items comprised in the UPPS-P-20-R (1) provide relevant information about one’s endorsement of the impulsive behavior dimensions (item-level content validity) and (2) assess a “broad” proportion of the construct-level content of the set of items included in the original version of the said psychometric instrument (construct-level content validity), therefore addressing the legitimate concerns critically highlighted by Cyders et al. (2014) about the content validity of the UPPS-P-20 (Billieux et al., 2012).

Nevertheless, several limitations with respect to the present registered report ought to be acknowledged.

In the first phase (i.e., development phase I), as decision rules for adequate global exact and global approximate fit were not met with respect to the original structural equation model corresponding to the initial item pool derived from the original version of the UPPS-P Impulsive Behavior Scale (UPPS-P-n_i), *post hoc* model modification based on local fit was carried out, shifting short-form development analyses from a “confirmatory mode” to an “exploratory mode” (Kline, 2023). Yet, such a shift does not compromise the overarching objectives of the three-phase short-form development and evaluation protocol: adopting an “exploratory mode” to develop the pre-established UPPS-P-20-R was adequate, inasmuch as a “confirmatory mode” was subsequently adopted on independent participant data to evaluate the UPPS-P-20-R.

In the second phase (i.e., development phase II), as the five model-implied optimal subsets of four items were identified from five sets of four to 12 items (i.e., $n_{\text{lack of premeditation}} = 7$, $n_{\text{negative urgency}} = 10$, $n_{\text{sensation seeking}} = 4$, $n_{\text{lack of perseverance}} = 7$, $n_{\text{positive urgency}} = 12$), the cardinality of the combinatorial search space (i.e., the number of candidate subsets), calculated using the binomial coefficient, varied extensively. On the one hand, with respect to the impulsive behavior dimension of positive urgency, as the model-implied optimal subset of four items was identified from a set of 12 items, the cardinality of the combinatorial search space was the highest (i.e., $C_4^{12} = 495$). On the other hand, with respect to the impulsive behavior dimension of sensation seeking, as the model-implied optimal subset of four items was identified from a set of four items, the cardinality of the combinatorial search space was the lowest (i.e., $C_4^4 = 1$). As such, the impulsive behavior dimensions benefited from our data-driven methodological approach to construct-level content validity in short-form development with varying extents of optimization: the impulsive behavior dimensions of lack of premeditation, negative urgency, lack of perseverance, and positive urgency underwent substantial optimization, whereas the impulsive behavior dimension of sensation seeking underwent no optimization. Yet, in the first phase (i.e., development phase I), as the primary objectives were to collect participant data with respect to an initial item pool derived from the original form (UPPS-P-n_i) and to ensure it presented adequate psychometric properties (i.e., content validity, internal structure validity, internal consistency reliability) prior to proceeding with short-form development, the impulsive behavior dimension of sensation seeking nonetheless benefited from our data-driven methodological approach to item-level content validity in short-form development.

In the third phase (i.e., evaluation phase), weak convergent validity evidence ($|\rho| \in [0.100, 0.250[$) and weak criterion validity evidence ($|\rho| \in [0.100, 0.250[$) is highlighted with respect to the impulsive behavior dimension of sensation seeking. Yet, weak convergent validity evidence ($|\rho| \in [0.100, 0.250[$) and weak criterion validity evidence ($|\rho| \in [0.100, 0.250[$) was also reported by previous original research articles that have suggested that self-control, anxiety, depression, dependent mobile phone use, dangerous mobile phone use, and prohibited mobile phone use present differential associations with the impulsive behavior dimension of sensation seeking (Billieux et al., 2008, 2012; Brevers et al., 2017; Canale et al., 2021; Claréus et al., 2017; Hasegawa et al., 2020; Lim & Kim, 2018).

Taken together, the output of the present registered report is twofold.

Table 4.2. Criterion validity evidence

			UPPS-P-20-R / UPPS-P-20 / Other versions of the said psychometric instrument				
			Lack of premeditation	Positive urgency	Sensation seeking	Negative urgency	Lack of perseverance
GAD-7	Anxiety	Present registered report	Weak	Adequate	Absent	Adequate	Adequate
		Billieux et al. (2012)	Absent	Absent	Absent	Adequate	Adequate
		Claréus et al. (2017)	Absent	Adequate	Absent	Adequate	Absent
		Lim and Kim (2018)	Weak	Adequate	Weak	Adequate	Adequate
PHQ-9	Depression	Present registered report	Weak	Adequate	Absent	Adequate	Adequate
		Billieux et al. (2012)	Absent	Absent	Absent	Adequate	Adequate
		Claréus et al. (2017)	Absent	Adequate	Weak	Adequate	Absent
		Lim and Kim (2018)	Weak	Adequate	Absent	Adequate	Weak
PMPUQ-SV-15	Dependent mobile phone use	Present registered report	Absent	Weak	Absent	Weak	Weak
		Billieux et al. (2008)	Absent	NA	Absent	Adequate	Absent
		Canale et al. (2021)	Absent	Weak	Absent	Adequate	NA
	Dangerous mobile phone use	Present registered report	Weak	Adequate	Weak	Weak	Absent
		Billieux et al. (2008)	Absent	NA	Weak	Weak	Absent
		Canale et al. (2021)	Adequate	Weak	Weak	Weak	NA
	Prohibited mobile phone use	Present registered report	Adequate	Adequate	Weak	Weak	Adequate
		Billieux et al. (2008)	Weak	NA	Absent	Absent	Absent
		Canale et al. (2021)	Adequate	Weak	Weak	Weak	NA

UPPS-P-20-R = revised 20-item short version of the UPPS-P Impulsive Behavior Scale. UPPS-P-20 = 20-item short version of the UPPS-P Impulsive Behavior Scale (Billieux et al., 2012). GAD-7 = 7-item Generalized Anxiety Disorder Scale (Spitzer et al., 2006). PHQ-9 = 9-item Patient Health Questionnaire (Kroenke et al., 2001). PMPUQ-SV-15 = 15-item short version of the Problematic Mobile Phone Use Questionnaire (López-Fernández et al., 2018). Criterion validity evidence is labeled "strong" if $|\rho| \geqslant 0.500$, "adequate" if $|\rho| \in [0.250, 0.500[$, "weak" if $|\rho| \in [0.100, 0.250[$, and "absent" if $|\rho| < 0.100$.

First, the present registered report provides professionals in clinical and research settings a revised short English-language psychometric instrument for assessing impulsivity – i.e., the revised 20-item short version of the UPPS-P Impulsive Behavior Scale (UPPS-P-20-R) – that presents strong psychometric properties, notably addressing the legitimate concerns critically highlighted by Cyders et al. (2014) about the content validity of the UPPS-P-20 (Billieux et al., 2012). The UPPS-P-20-R is stored on European servers based in Frankfurt (Germany) and shared publicly under a *Creative Commons Attribution 4.0 International Public License* on the *Open Science Framework* (https://doi.org/10.17605/osf.io/s84k7) (see Appendix). In reducing assessment time to approximately one-third of that of the original version of the psychometric instrument without sacrificing content validity, the UPPS-P-20-R offers promising potential in clinical and research settings for assessing impulsivity concisely yet comprehensively. Future research is warranted to investigate its adaptability to diverse linguistic, developmental, and clinical populations.

Second, the present registered report provides professionals in research settings a data-driven methodological approach to construct-level content validity in short-form development – i.e., the *R* package *shortr* (Fournier, Heeren, et al., 2025) – that can be adopted to develop short versions of any psychometric instrument. Whereas it is common practice in short-form development to retain items presenting the highest factor loadings or total-item correlations for a psychometric instrument's given factor(s), such common practice tends to narrow the breadth of construct-level content coverage of psychometric instruments (Clark & Watson, 2019). Yet, as short-form development in psychometrics consists of retaining a subset of items included in the original version of a psychometric instrument, loss of construct-level content coverage is inevitable, therefore warranting in-depth examination (Smith et al., 2000). To this end, we developed and released the *R* package *shortr* (Fournier, Heeren, et al., 2025), which implements combinatorial search algorithms (i.e., brute force combinatorial search algorithm, simulated annealing combinatorial search algorithm) to identify what optimal subset of a desired number of items should be retained in a short version of a psychometric instrument to assess the "broadest" proportion of the construct-level content of the set of items included in the original version of the said psychometric instrument. The *R* package *shortr* (Fournier, Heeren, et al., 2025) is shared publicly under a *GNU General Public License* on the *Comprehensive R Archive Network* (https://doi.org/10.32614/CRAN.package.shortr). Concise psychometric instruments are essential assets in clinical and research settings, as they save valuable financial and temporal resources. Yet, such conciseness inevitably entails a trade-off with comprehensiveness. In mitigating the latter trade-off, the *R* package *shortr* (Fournier, Heeren, et al., 2025) offers promising potential in future research for developing short versions of any psychometric instrument without sacrificing content validity. As such, it not only addresses a long-standing limitation in short-form development but also paves the way for transforming methods and practices in psychometrics.

Acknowledgments

Loïs Fournier was supported by the *Swiss National Science Foundation* under a "Doc.CH" Doctoral Research Fellowship [Grant ID: P000PS_211887]. The *Swiss National Science Foundation* had no role in registered report design, data collection and analysis, publication decision, or manuscript preparation. The authors thank *Peer Community In Registered Reports* for recommending and reviewing the present registered report: Veli-Matti Karhulahti (recommender), Yuqi Huang (reviewer), Ivan Ropovik (reviewer), and Orestis Zavlis (reviewer).

Author contributions

Loïs Fournier: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original draft, Writing – Review & editing. **Alexandre Heeren:** Conceptualization, Methodology, Writing – Review & editing. **Stéphanie Baggio:** Writing – Review & editing. **Luke Clark:** Writing – Review & editing. **Antonio Verdejo-García:** Writing – Review & editing. **José C. Perales:** Conceptualization, Methodology, Writing – Review & editing. **Joël Billieux:** Conceptualization, Methodology, Project Administration, Supervision, Writing – Review & editing.

Competing interests

The authors declare no competing interests.

Ethics statement

The present registered report protocol describing research with human participants complies with the Declaration of Helsinki. The present registered report protocol received ethical approval from the Research Ethics Commission of the Faculty of Social and Political Sciences of the University of Lausanne, Lausanne, Switzerland [Project ID: C_SSP_102022_00013] on December 13, 2022. Informed consent was provided by all human participants. Human participant compensation equated to an hourly rate of 9.00 GBP.

Code availability

All code is stored on European servers based in Frankfurt (Germany) and shared publicly under a *Creative Commons Attribution 4.0 International Public License* on the *Open Science Framework* (https://doi.org/10.17605/osf.io/3nztw). The *R* package *shortr* (Fournier, Heeren, et al., 2025) is shared publicly under a *GNU General Public License* on the *Comprehensive R Archive Network* (https://doi.org/10.32614/CRAN.package.shortr).

Data availability

All data is stored on European servers based in Frankfurt (Germany) and shared publicly under a *Creative Commons Attribution 4.0 International Public License* on the *Open Science Framework* (https://doi.org/10.17605/osf.io/6v5qb).

Materials availability

All materials are stored on European servers based in Frankfurt (Germany) and shared publicly under a *Creative Commons Attribution 4.0 International Public License* on the *Open Science Framework* (https://doi.org/10.17605/osf.io/s84k7).

Preprints availability

All preprints are stored on European servers based in Frankfurt (Germany) and shared publicly under a *Creative Commons Attribution 4.0 International Public License* on the *Open Science Framework* (https://doi.org/10.17605/osf.io/fxpdn).

Supplementary information availability

All supplementary information is stored on European servers based in Frankfurt (Germany) and shared publicly under a *Creative Commons Attribution 4.0 International Public License* on the *Open Science Framework* (https://doi.org/10.17605/osf.io/wc2vy).

Peer Community In Registered Reports information

The authors thank *Peer Community In Registered Reports* for recommending and reviewing the present registered report: Veli-Matti Karhulahti (recommender), Yuqi Huang (reviewer), Ivan Ropovik (reviewer), and Orestis Zavlis (reviewer). The first stage of the present registered report (https://osf.io/8je2w) was recommended by *Peer Community In Registered Reports* on January 10, 2025 (https://rr.peercommunityin.org/articles/rec?id=862). The second stage of the present registered report (https://osf.io/th8px) was recommended by *Peer Community In Registered Reports* on November 4, 2025 (https://rr.peercommunityin.org/articles/rec?id=1118; https://doi.org/10.24072/pci.rr.101118). The preprint of the present registered report, as recommended by *Peer Community In Registered Reports* on November 4, 2025 (https://rr.peercommunityin.org/articles/rec?id=1118; https://doi.org/10.24072/pci.rr.101118), is stored on European servers based in Frankfurt (Germany) and shared publicly under a *Creative Commons Attribution 4.0 International Public License* on the *Open Science Framework* (https://osf.io/wtbgr).

Editors: Don van Ravenzwaaij (Editor-in-Chief)

Submitted: November 10, 2025 PST. Accepted: November 19, 2025 PST. Published: January 05, 2026 PST.

References

American Psychiatric Association (Ed.). (2022). *Diagnostic and statistical manual of mental disorders (DSM-5-TR)*. https://doi.org/10.1176/appi.books.9780890425787

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

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

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

Billieux, J., Van Der Linden, M., & Rochat, L. (2008). The role of impulsivity in actual and problematic use of the mobile phone. *Applied Cognitive Psychology*, *22*(9), 1195–1210. https://doi.org/10.1002/acp.1429

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G., Isvoranu, A.-M., Wysocki, A. C., Van Borkulo, C. D., Van Bork, R., & Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, *1*(1), 58. https://doi.org/10.1038/s43586-021-00055-w

Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences*, *12*(3), 291–294. https://doi.org/10.1016/0191-8869(91)90115-R

Brevers, D., Foucart, J., Verbanck, P., & Turel, O. (2017). Examination of the validity and reliability of the French version of the Brief Self-Control Scale. *Canadian Journal of Behavioural Science*, *49*(4), 243–250. https://doi.org/10.1037/cbs0000086

Calzada, G., Rothen, S., Radziejewska, D., Martins, D., Aranda, L., Bassini, L., Zuka, H., Thorens, G., Khazaal, Y., & Zullino, D. (2017). Validation of a short French UPPS-P Impulsive Behavior Scale in patients with substance use disorder. *International Journal of Mental Health and Addiction*, *15*(5), 1096–1102. https://doi.org/10.1007/s11469-017-9763-x

Canale, N., Moretta, T., Pancani, L., Buodo, G., Vieno, A., Dalmaso, M., & Billieux, J. (2021). A test of the pathway model of problematic smartphone use. *Journal of Behavioral Addictions*, *10*(1), 181–193. https://doi.org/10.1556/2006.2020.00103

Cándido, A., Orduña, E., Perales, J. C., Verdejo-García, A., & Billieux, J. (2012). Validation of a short Spanish version of the UPPS-P Impulsive Behavior Scale. *Trastornos Adictivos*, *14*(3), 73–78. https://doi.org/10.1016/S1575-0973(12)70048-X

Champely, S. (2020). *pwr: basic functions for power analysis* (1.3-0) [Computer software]. https://doi.org/10.32614/CRAN.package.pwr

Claréus, B., Daukantaitė, D., Wångby-Lundh, M., & Lundh, L.-G. (2017). Validation of a Swedish version of the short UPPS-P Impulsive Behavior Scale among young adults. *Addictive Behaviors Reports*, *6*, 118–122. https://doi.org/10.1016/j.abrep.2017.10.001

Clark, L. A., & Watson, D. (2019). Constructing validity: new developments in creating objective measuring instruments. *Psychological Assessment*, *31*(12), 1412–1427. https://doi.org/10.1037/pas0000626

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI)*.

Csárdi, G., Nepusz, T., Traag, V., Horvát, S., Zanini, F., Noom, D., & Müller, K. (2025). *igraph: network analysis and visualization* (2.1.4) [Computer software]. https://doi.org/10.32614/CRAN.package.igraph

Cyders, M. A., Littlefield, A. K., Coffey, S., & Karyadi, K. A. (2014). Examination of a short English version of the UPPS-P Impulsive Behavior Scale. *Addictive Behaviors*, *39*(9), 1372–1376. https://doi.org/10.1016/j.addbeh.2014.02.013

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

d'Orta, I., Burnay, J., Aiello, D., Niolu, C., Siracusano, A., Timpanaro, L., Khazaal, Y., & Billieux, J. (2015). Development and validation of a short Italian UPPS-P Impulsive Behavior Scale. *Addictive Behaviors Reports*, *2*, 19–22. https://doi.org/10.1016/j.abrep.2015.04.003

Epskamp, S. (2024). *bootnet: bootstrap methods for various network estimation routines* (1.6) [Computer software]. https://doi.org/10.32614/CRAN.package.bootnet

Epskamp, S., Costantini, G., Haslbeck, J., & Isvoranu, A. (2023). *qgraph: graph plotting methods, psychometric data visualization, and graphical model estimation* (1.9.8) [Computer software]. https://doi.org/10.32614/CRAN.package.qgraph

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *53*(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823

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

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

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., … Billieux, J. (2025). 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*, *32*(5), 635–653. https://doi.org/10.1177/10731911241259560

Fournier, L., Heeren, A., Baggio, S., Clark, L., Verdejo-García, A., Perales, J. C., & Billieux, J. (2025). *shortr: Optimal Subset Identification in Undirected Weighted Network Models (Version 1.0.1)* [Computer software]. https://doi.org/10.32614/CRAN.package.shortr

Friedman, J., Hastie, T., & Tibshirani, R. (2019). *glasso: graphical lasso estimation of Gaussian graphical models* (1.11) [Computer software]. https://doi.org/10.32614/CRAN.package.glasso

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

Geurten, M., Catale, C., Gay, P., Deplus, S., & Billieux, J. (2021). Measuring impulsivity in children: adaptation and validation of a short version of the UPPS-P Impulsive Behavior Scale in children and investigation of its links with ADHD. *Journal of Attention Disorders*, *25*(1), 105–114. https://doi.org/10.1177/1087054718775831

Hasegawa, T., Kawahashi, I., Fukuda, K., Imada, S., & Tomita, Y. (2020). Reliability and validity of a short Japanese version of the UPPS-P Impulsive Behavior Scale. *Addictive Behaviors Reports*, *12*, 100305. https://doi.org/10.1016/j.abrep.2020.100305

Isvoranu, A.-M., & Epskamp, S. (2023). Which estimation method to choose in network psychometrics? Deriving guidelines for applied researchers. *Psychological Methods*, *28*(4), 925–946. https://doi.org/10.1037/met0000439

Jiang, H., Fei, X., Liu, H., Roeder, K., Lafferty, J., Wasserman, L., Li, X., & Zhao, T. (2021). *huge: high-dimensional undirected graph estimation* (1.3.5) [Computer software]. https://doi.org/10.32614/CRAN.package.huge

Kempeneers, P., Mreyen, K., Pallincourt, R., Remacle, F., Wildemeersch, G., & Simon, J. (2023). Validation of the UPPS-P Impulsive Behavior Scale and clinical correlates of its scores in French-speaking patients starting a residential detoxification program. *Indian Journal of Psychological Medicine*, *45*(5), 503–510. https://doi.org/10.1177/02537176231157411

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

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lauritzen, S. L. (1996). *Graphical models*. https://doi.org/10.1093/oso/9780198522195.001.0001

Lim, S. Y., & Kim, S. J. (2018). Validation of a short Korean version of the UPPS-P Impulsive Behavior Scale. *Asia-Pacific Psychiatry*, *10*(3), e12318. https://doi.org/10.1111/appy.12318

Liu, H., Lafferty, J., & Wasserman, L. (2018). *The nonparanormal: semiparametric estimation of high-dimensional undirected graphs*. https://doi.org/10.1184/R1/6610712

Loevinger, J. (1954). The attenuation paradox in test theory. *Psychological Bulletin*, *51*(5), 493–504. https://doi.org/10.1037/h0058543

López-Fernández, O., Kuss, D. J., Pontes, H. M., Griffiths, M. D., Dawes, C., Justice, L. V., Männikkö, N., Kääriäinen, M., Rumpf, H.-J., Bischof, A., Gässler, A.-K., Romo, L., Kern, L., Morvan, Y., Rousseau, A., Graziani, P., Demetrovics, Z., Király, O., Schimmenti, A., … Billieux, J. (2018). Measurement invariance of the short version of the Problematic Mobile Phone Use Questionnaire (PMPUQ-SV) across eight languages. *International Journal of Environmental Research and Public Health*, *15*(6), 1213. https://doi.org/10.3390/ijerph15061213

López-Guerrero, J., Navas, J. F., Perales, J. C., Rivero, F. J., & Muela, I. (2023). The interrelation between emotional impulsivity, craving, and symptom 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

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

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

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

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*(2), 026113. https://doi.org/10.1103/PhysRevE.69.026113

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

Pornprasertmanit, S., Miller, P., Schoemann, A., & Jorgensen, T. D. (2021). *simsem: simulated structural equation modeling* (0.5-16) [Computer software]. https://doi.org/10.32614/CRAN.package.simsem

Reichardt, J., & Bornholdt, S. (2006). Statistical mechanics of community detection. *Physical Review E*, *74*(1), 016110. https://doi.org/10.1103/PhysRevE.74.016110

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

Rosseel, Y., Jorgensen, T. D., & De Wilde, L. (2023). *lavaan: latent variable analysis* (0.6-17) [Computer software]. https://doi.org/10.32614/CRAN.package.lavaan

Rosseel, Y., Jorgensen, T. D., & De Wilde, L. (2024). *lavaan: latent variable analysis* (0.6-19) [Computer software]. https://doi.org/10.32614/CRAN.package.lavaan

Sánchez-Domínguez, R., Benjet, C., Marín-Navarrete, R., & Nicolini, H. (2023). Validity and reliability of the short version of the UPPS-P Impulsive Behavior Scale in patients with substance use disorders. *Journal of Substance Use*, *28*(5), 721–729. https://doi.org/10.1080/14659891.2022.2087777

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

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

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

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*(1), 102–111. https://doi.org/10.1037/1040-3590.12.1.102

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder. *Archives of Internal Medicine*, *166*(10), 1092. https://doi.org/10.1001/archinte.166.10.1092

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, *72*(2), 271–324. https://doi.org/10.1111/j.0022-3506.2004.00263.x

Traag, V. A., & Bruggeman, J. (2009). Community detection in networks with positive and negative links. *Physical Review E*, *80*(3), 036115. https://doi.org/10.1103/PhysRevE.80.036115

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

Wolf, M. G., & McNeish, D. (2025). *dynamic: dynamic fit indices cutoffs for latent variable models (Version 1.1.0)* [Computer software]. https://github.com/melissagwolf/dynamic

World Health Organization (Ed.). (2024). *Clinical descriptions and diagnostic requirements for ICD-11 mental, behavioral, and neurodevelopmental disorders*.

Zsila, Á., Bőthe, B., Demetrovics, Z., Billieux, J., & Orosz, G. (2020). Further exploration of the S-UPPS-P Impulsive Behavior Scale's factor structure: evidence from a large Hungarian sample. *Current Psychology*, *39*(1), 378–388. https://doi.org/10.1007/s12144-017-9773-7

Supplementary Materials

Peer review communication

Download: https://collabra.scholasticahq.com/article/150350-reducing-assessment-time-without-sacrificing-content-validity-in-short-form-development-an-illustration-through-the-revised-20-item-short-version-of/attachment/314427.docx?auth_token=p9l1b-equOmLV5pQkrCe

Supplementary materials

Download: https://collabra.scholasticahq.com/article/150350-reducing-assessment-time-without-sacrificing-content-validity-in-short-form-development-an-illustration-through-the-revised-20-item-short-version-of/attachment/314428.docx?auth_token=p9l1b-equOmLV5pQkrCe

Appendix

Download: https://collabra.scholasticahq.com/article/150350-reducing-assessment-time-without-sacrificing-content-validity-in-short-form-development-an-illustration-through-the-revised-20-item-short-version-of/attachment/314429.docx?auth_token=p9l1b-equOmLV5pQkrCe