Journal of Research in Personality 85 (2020) 103914

Contents lists available at ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

Full Length Article

State-level impulsivity, affect, and alcohol: A psychometric evaluation of the momentary impulsivity scale across two intensive longitudinal samples ☆

Angela K. Stevens [a,b,*], Brittany E. Blanchard [a], Amelia E. Talley [a], Jennifer L. Brown [c], Max A. Halvorson [d], Tim Janssen [b], Kevin M. King [d], Andrew K. Littlefield [a]

[a] Texas Tech University Department of Psychological Sciences, Texas Tech University, Psychology Building, Box 42051, Lubbock, TX 79409, USA
[b] Center for Alcohol and Addiction Studies, Brown University School of Public Health, Providence, RI 02912, USA
[c] Addiction Sciences Division, Department of Psychiatry & Behavioral Neuroscience, University of Cincinnati College of Medicine, 3131 Harvey Ave, Suite 104, Cincinnati, OH 45229, USA
[d] Department of Psychology, University of Washington, USA

ARTICLE INFO

Article history:
Received 6 August 2019
Revised 13 November 2019
Accepted 13 January 2020
Available online 15 January 2020

Keywords:
State-level impulsivity
Trait-level impulsivity
Multi-trait, multi-method
Daily diary
Ecological momentary assessment
Psychometrics
Alcohol
Affect

ABSTRACT

We reexamined the psychometric properties of the Momentary Impulsivity Scale (MIS) in two young adult samples using daily diary (N = 77) and ecological momentary assessment (N = 147). A one-factor between- and within-person structure was supported, though "I felt impatient" loaded poorly within-person. MIS scores consistently related to emotion-driven trait impulsivity; however, MSSDs of MIS scores were unrelated to outcomes after accounting for aggregate MIS scores. We observed positive, within-person correlations with negative, but not positive, affect. Between-person MIS scores correlated with alcohol problems, though within-person MIS-alcohol relations were inconsistent. MIS scores were unrelated to laboratory-based impulsivity tasks. Findings inform the assessment of state-level impulsivity in young adults. Future research should prioritize expanding the MIS to capture the potential multi-dimensionality of state-level impulsivity.

© 2020 Elsevier Inc. All rights reserved.

1. Introduction[1]

Impulsivity, the most ubiquitous transdiagnostic symptom criterion in the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition (*DSM-5*; American Psychiatric Association, 2013), can be conceptualized in a variety of ways (e.g., risk-taking, distractibility, motor impulsivity, behavioral activation; Cyders, 2015; see Evenden, 1999).[2] Many of these impulsivity-related constructs are related to different risky behaviors (e.g., Smith et al., 2007), including specific types of substance use (e.g., Blanchard, Stevens, Littlefield, Talley, & Brown, 2017). However, impulsivity facets change over time (e.g., Littlefield, Stevens, Ellingson, King, & Jackson, 2016), and individuals may not behave consistently across situations (Fleeson, 2004). Because trait-level impulsivity assessments assume the propensity for impulsive responses is consistent across situations, intraindividual variability in impulsivity is often neglected. This is potentially problematic, as evidence suggests links between personality and expressive behavior can vary widely across

☆ Manuscript preparation was supported in part by grant number T32 DA016184 from the National Institute on Drug Abuse at the National Institutes of Health.

* Corresponding author at: Center for Alcohol and Addiction Studies, Brown University School of Public Health, G-S121-4 121 S. Main St., Providence, RI 02912, USA.

E-mail address: angela_stevens@brown.edu (A.K. Stevens).

[1] The two reported studies in this manuscript were not pre-registered before submission. The authors' contributions are as follows: AKS, AET, JLB, and AKL designed, implemented, and collected daily diary data; BEB assisted with daily diary and lab-based data collection for Sample 1; MAH and KMK designed, implemented, and collected data for the EMA study. AKS conducted all statistical analyses with assistance from AKL. AKS drafted the initial version of the manuscript, which was edited by BEB and AKL, who also made significant writing contributions, as well as AET, JLB, MAH, TJ, and KMK. All authors read and approved the final manuscript. The universities' Institutional Review Boards preclude authors from disseminating raw data, but variance-covariance matrices and syntax are available from the authors by request.

[2] Trait "impulsivity" is a broad phenotype characterized by different behavioral tendencies. Although trait "impulsivity" is most accurately operationalized by separate, but related, constructs and was coined a misnomer in recent literature (see Cyders, 2015), this has yet to be established for state-level impulsivity. Thus, we use the term "impulsivity" throughout, unless more nuanced language is appropriate.

https://doi.org/10.1016/j.jrp.2020.103914
0092-6566/© 2020 Elsevier Inc. All rights reserved.

situations (DeYoung, 2015) and relatively brief time windows (Fleeson & Jayawickreme, 2001, 2007; see Fleeson, 2017; see also Vazire & Sherman, 2017).

The conceptualization of personality as a "person"-related variable (i.e., trait-level personality) versus a "situation"-related variable (i.e., state-level personality) has sparked contentious debates (e.g., Kenrick & Funder, 1988; Roberts, 2009). Nevertheless, research in this area has progressed suggesting that personality traits are robust predictors of psychopathology (e.g., impulsivity and alcohol-related outcomes; Littlefield & Sher, 2016), and that the expression of personality may be context-dependent (see Fleeson & Jayawickreme, 2015; see also Wilson, Thompson, & Vazire, 2017). Despite these advancements, research examining within-person fluctuations in personality has been hindered by a paucity of assessment tools designed for this purpose. As clinical and research endeavors increase focus on facets of impulsivity based, in part, on its value as a transdiagnostic feature (Barlow, Sauer-Zavala, Carl, Bullis, & Ellard, 2014; Carver, Johnson, & Timpano, 2017), identifying psychometrically sound state-level impulsivity assessments is an important next step. Some researchers utilize laboratory-based impulsivity tasks to assess state-level impulsivity (see Cyders & Coskunpinar, 2011); however, extant psychometric evidence does not support this interpretation of these tasks, as their construct validity has been called into question (King et al., 2019; Sharma, Markon, & Clark, 2014; Stevens, Blanchard, & Littlefield, 2018).

In an attempt to better capture state-level impulsivity, Tomko et al. (2014) created the Momentary Impulsivity Scale (MIS), which is a brief, 4-item scale. Using ecological momentary assessment (EMA; six assessments/day across 28 days) in an adult clinical sample seeking outpatient treatment for either borderline personality disorder (BPD) or a depressive disorder, their findings demonstrated a single MIS factor structure at between- and within-person levels. There was also support for the MIS's content and convergent validity using the UPPS Impulsive Behavior Scale (Whiteside & Lynam, 2001) and Barratt Impulsiveness Scale-11 (BIS-11; Patton, Stanford, & Barratt, 1995). Specifically, Tomko et al. (2014) found positive and large-to-very large correlations between UPPS urgency, UPPS lack of perseverance, BIS motor, and BIS attention. Tomko et al. (2014) also found positive and medium-to-large correlations with UPPS lack of planning and BIS non-planning and a positive and small-to-medium correlation with UPPS sensation seeking. Tomko et al. (2014) also included an index of the temporal instability of MIS responses in a time series – the mean squared successive difference (MSSD) – which reflects the MIS's state-like properties in a manner that aggregate MIS scores (i.e., averaged across EMAs) purportedly cannot. Interestingly, patterns of UPPS and BIS-11 correlations were similar to correlations with MSSD of MIS scores (Tomko et al., 2014). Less is known about MIS scores' relations to criterion outcomes. Trull, Wycoff, Lane, Carpenter, and Brown (2016) recently examined MIS scores' relations to alcohol using an adult clinical sample similar to Tomko et al. (2014) and found binary alcohol use endorsement was positively related, within-individuals, to MIS scores on the same assessment occasion, as well as at the daily level. Nevertheless, they did not find this relation at the between-person level (Trull et al., 2016). However, to our knowledge, the reliability and validity of the MIS has not been reexamined outside of its original study sample and relations between MIS scores and criterion outcomes have yet to be tested among young adults.

Despite its recommendation for use in the extant literature (e.g., Ansell, Laws, Roche, & Sinha, 2015; Barker, Romaniuk, Cardinal, Pope, Nicol, & Hall, 2015; Trull et al., 2016) and its widespread use, the MIS has not undergone a psychometric evaluation outside of its original sample comprised of adults in outpatient mental health treatment. Thus, the purpose of the current study was to reevaluate the psychometric properties of the MIS in two young adult samples using daily diary and EMA methodologies and a multi-trait, multi-method approach ([MTMM]; Campbell & Fiske, 1959). This study consisted of two samples: (1) young adults recruited for a daily diary study (Sample 1), and (2) young adults recruited for an EMA study (Sample 2). Across both samples, we sought to replicate findings of Tomko et al. (2014) by examining: (a) the MIS factor structure using multilevel confirmatory factor analysis, (b) between- and within-person reliabilities, (c) temporal instability of the MIS via the MSSD of MIS scores, and (d) content validity using the UPPS-P Impulsive Behavior Scale (Lynam, Smith, Cyders, Fischer, & Whiteside, 2007). We also sought to extend the work by Tomko et al. (2014) by including additional psychometric tests not examined in the original MIS study. Importantly, we tested criterion validity of the MIS at between- and within-person levels using multiple alcohol indices, positive affect, and negative affect. To our knowledge, this is the first examination of criterion validity for MIS scores outside of the original adult clinical sample. In Sample 1, we also examined associations between MIS scores and three laboratory-based impulsivity tasks that purportedly assess state-level impulsivity. We are the first to examine convergent and discriminant validity of MIS scores using laboratory-based behavioral tasks. In both samples, we also included between- and within-person Omega estimates for internal consistency, which is a novel extension of Tomko et al. (2014).

Based on limited extant research, we hypothesized that we would find the following: (1) support for a one-factor structure of the MIS at within- and between-person levels; (2) higher between-person reliabilities than within-person reliabilities consistent with Tomko et al. (2014); (3) at least small-to-medium between-person bivariate correlations between UPPS-P facets and aggregate (between-person) MIS scores, with higher magnitude correlations detected for urgency facets, similar to Tomko et al. (2014); (4) at least small within-person correlations among MIS scores, positive and negative affect ratings, and reported alcohol consumption; (5) at least small between-person correlations between aggregate MIS scores and alcohol indices (Sample 2 only); and (6) small associations (r < 0.15) between laboratory-based impulsivity tasks and aggregate MIS scores, consistent with research noting robust method effects (Sample 1 only; Sharma et al., 2014). In general, we anticipated comparable effect sizes to Tomko et al. (2014). However, the reproducibility of effect sizes in psychological sciences is equivocal, thus reduced effect sizes in the present study (at least compared to Tomko et al., 2014) are possible (Open Science Collaboration, 2015). Identical data analytic procedures were used across samples when possible, thus abbreviated analytic procedures are provided for Sample 2.

2. Methods (Sample 1: Daily Diary)

2.1. Participants

Seventy-seven[3] participants (M age = 20.8, SD = 1.9) were recruited for a daily diary study from a large, southwestern city

[3] The Level-2 sample size recruited for Sample 1 is comparable to that of similar work using multilevel-confirmatory factor analyses and between-person bivariate correlations (i.e., Level-2 n = 77; see Tomko et al., 2014). There are limited documented procedures for power analyses involving complex multilevel models, therefore a within-factors repeated measures ANOVA power analysis was computed using G*Power 3.1.9.2 assuming small, medium, and large effect sizes (corresponding to partial eta squared of 0.02, 0.06, and 0.14) and an average correlation of repeated assessments (r = 0.33). Assuming the largest number of groups (n = 76) and 10 repeated measurements, there appears to be adequate power to detect small (0.94), medium (0.99) and large (1.00) effects. Assuming the smallest possible number of groups (n = 55), adequate power remains: small (0.82), medium (0.99), large (1.00).

via flyers, advertisements, and social media for a broader study examining behavior, substance use, and relationships among young adults endorsing high-risk behavior. Many participants self-identified as female (60.5%) and White (76.3%), with 26% of participants self-identifying as Hispanic/Latinx. Ninety-three percent of participants reported current school (i.e., undergraduate- or graduate-level) enrollment. Eligible participants were required to (a) be between ages of 18–25, (b) endorse at least one binge-drinking episode in the past month, and (c) report at least one unprotected sexual encounter in the past month. Pregnancy was an exclusion criterion (see Stevens, Littlefield, Talley, & Brown, 2017, for additional recruitment details).

2.2. Study protocol

Eligible participants completed baseline and follow-up measures in the laboratory. All behavioral tasks were counterbalanced across participants, as well as counterbalanced across baseline and follow-up visits, such that one individual did not receive the same ordering of tasks at both visits. Before completing self-report measures, participants received training on standard alcoholic drink equivalencies outlined by the National Institute on Alcohol Abuse and Alcoholism (NIAAA, 2005). Following the baseline visit, participants entered a 10-day daily diary phase where reports were completed online (i.e., starting Friday until the following Sunday). Assessment links were distributed daily at 7:30 a.m. via email, and the survey closed at 2:00 p.m. each day. An average of 8.1 (*SD* = 2.4; range = 0–10) daily reports were completed. Completion rates for the daily diary study (80%) were comparable to similar daily diary studies (e.g., Simons, Dvorak, Batien, & Wray, 2010).

2.3. Measures

2.3.1. Demographics

Eligible participants completed a baseline measure of demographic questions including age, gender, race, and ethnicity.

2.3.2. Self-reported impulsivity-like facets

We assessed the following impulsivity-like facets via the 59-item UPPS-P Impulsive Behavior Scale (Lynam et al., 2007) using a 4-point Likert-type scale: (1) negative urgency, or a tendency to act rashly under extreme negative mood; (2) positive urgency, or a tendency to act rashly under extreme positive mood; (3) sensation seeking, or the tendency to seek out novel or thrilling experiences; (4) lack of planning, or the tendency to act without thinking; and (5) lack of perseverance, or the inability to remain focused on a task. Higher subscale sum scores reflected higher impulsivity (α = 0.81–0.93 across subscales).

2.3.3. Laboratory-based impulsivity assessments

(1) The GoStop Impulsivity Paradigm was used to measure prepotent response inhibition (Dougherty, Mathias, Marsh, & Jagar, 2005). Participants were presented 5-digit numbers on the screen and instructed to click the mouse if the subsequent trial was a matching trial but to inhibit the click if the number turned red. Specifically, this 11.67-minute task included two blocks of trials, with each block including randomly generated 5-digit numbers presented on the screen for 500 ms at 1500 ms intervals. Half of the target trials included a target-stop trial, where the color of the numbers changed from black to red at 50, 150, 250, and 250 ms after being presented on the computer screen. Target and stop-target trials occurred 25% of the time. Percentage of inhibited responses (e.g., Bagge, Littlefield, Rosellini, & Coffey, 2013) was the outcome of interest. Test-retest reliability from baseline to follow-up was moderate-to-high (*r* = 0.47).

(2) The Immediate Memory Task (IMT) was used to measure response initiation (Dougherty, Marsh, & Mathias, 2002; Dougherty, Marsh-Richard, Hatzis, Nouvion, & Mathias, 2008). During a 10.5-min testing session, participants were presented 5-digit numbers on the computer screen and were instructed to click the mouse when the 5-digit number span matched the set immediately preceding it. Each number was presented for 500 ms at 500 ms intervals. Stimuli could also be nonmatching, including catch stimuli (i.e., a 5-digit number which differed from the previous set by only one digit) and filler stimuli (i.e., a novel 5-digit number). For each trial, target and catch stimuli occurred 33% of the time, whereas novel stimuli occurred 34% of the time. The IMT ratio was calculated as the proportion of commission errors divided by the proportion of correct detections, which reflects errors in response initiation (e.g., Bagge et al., 2013). Test-retest reliability from baseline to follow-up was high (*r* = 0.75).

(3) The Two Choice Impulsivity Paradigm (TCIP) was used to assess delay discounting, or an individual's preference for smaller-sooner rewards compared to larger-later rewards, via a discrete-choice task (Dougherty et al., 2005). Participants completed four training trials before the testing session to associate the two shapes (i.e., circles and squares) with their respective delays and rewards. Participants were then presented with 50 trials of a black circle and square side-by-side on a computer screen, with the presentation order of the two shapes being randomly determined for each trial. Participants chose between clicking a circle to earn five points after a 5-second delay or clicking a square to earn 15 points after a 15-second delay. The outcome of interest was the number of immediate choices selected. Test-retest reliability from baseline to follow-up was moderate-to-high (*r* = 0.46).

2.3.4. Daily momentary impulsivity

We assessed momentary impulsivity at each daily report in reference to the prior day via the 4-item MIS (Tomko et al., 2014). Participants read each item and selected which number on a 5-point Likert scale best described their experiences. Items included: "I said things without thinking"; "I have felt impatient"; "I spent more money than I meant to"; and "I made a 'spur of the moment' decision." Higher sum scores reflected higher momentary impulsivity. Reliabilities are reported and discussed in detail in *Results*.

2.3.5. Daily alcohol consumption

We assessed alcohol consumption at each daily diary report for yesterday's alcohol consumption (i.e., "How many drinks did you consume yesterday?").

2.3.6. Daily affect

Within each daily diary report, we assessed yesterday's daily positive (6 items) and negative affect (5 items) using an 11-item modification of the Positive and Negative Affect Scale (PANAS; Watson, Clark, & Tellegen, 1998). The following positive affect items were administered: "interested," "strong," "enthusiastic," "alert," "determined," and "active." The following negative affect items were administered: "upset," "scared," "irritable," "ashamed," and "jittery." Between- and within-person reliabilities, respectively, were high for positive affect (R_{KF} = 0.98; R_{CN} = 0.79) and negative affect (R_{KF} = 0.99; R_{CN} = 0.86; Shrout & Lane, 2012).

2.4. Data analytic plan

Various analytic procedures were conducted using Mplus Version 8.2 (Muthén & Muthén, 1998–2018), SAS 9.4™ software (SAS Institute Inc., Cary, NC, USA),[4] and R Version 3.6.0 (R Core Team, 2013). Using recently published guidelines by Funder and Ozer (2019), effect sizes for within- and between-person correlations are interpreted as follows: r = 0.05 (very small), r = 0.10 (small), r = 0.20 (medium), r = 0.30 (large), and r = 0.40 (very large). Ninety-five percent confidence intervals (CIs) are included for all effects.

2.4.1. Factor structure, instability, and reliability

To examine the internal structure of the MIS at within- and between-person levels, we conducted a multilevel confirmatory factor analysis (ML-CFA) in Mplus using weighted least squares mean- and variance- (WLSMV) adjusted estimation, which employs pairwise deletion for missing data. Given the use of categorical endogenous indicators, WLSMV is recommended because it provides consistent parameter estimates and unbiased standard errors with non-normal variables (see Asparouhov & Muthén, 2010). The following model fit indices are reported: χ^2, comparative fit index (CFI), root-mean-square error of approximation (RMSEA), and standardized root mean square residual (SRMR). Model fit indices were inspected, such that a CFI approaching one and RMSEA and SRMR approaching zero indicated good fit to the data. To estimate the temporal instability of MIS scores across time, we computed an average daily MSSD of MIS scores as recommended by Jahng, Wood, and Trull (2008; see Eq. (1)).

$$MSSD = \frac{1}{N-1}\sum_{i=1}^{N-1}(x_{i+1}-x_i)^2 \tag{1}$$

Indeed, the MSSD also is a function of the first-order autocorrelation (iACORR(1)) and the variance (iVAR) of a time series (Eq. (2); Jahng, 2008):

$$MSSD = 2(iVAR)(1 - iACORR(1)) \tag{2}$$

High MSSD values reflect higher variability and lower temporal dependency, which, in this case, suggests temporal instability in impulsivity across days. Jahng et al. (2008) determined the MSSD is preferred over other indices of temporal instability in a time series, at least for affective variables.[5]

Internal consistency of MIS scores was examined using two approaches. Consistent with Tomko et al. (2014), we estimated variance components in SAS™ PROC MIXED using restricted maximum likelihood (REML) estimation as recommended by Shrout and Lane (2012). We then computed between- and within-person reliabilities using Eqs. (3) and (4), respectively (Shrout & Lane, 2012).

$$R_{KF} = \frac{\sigma^2 person + (\sigma^2 person * item/m)}{[\sigma^2 person + (\sigma^2 person * item/m) + (\sigma^2 error/km)]} \tag{3}$$

$$R_{CN} = \frac{\sigma^2 person * item}{[\sigma^2 person * item + (\sigma^2 error/m)]} \tag{4}$$

Between- and within-person internal consistencies were also examined using Coefficient Omega consistent with procedures outlined by Geldhof et al. (2014).

2.4.2. Content, criterion, and convergent validity

We examined content validity by computing between-person bivariate correlations among UPPS-P subscales and aggregate MIS scores (averaged across time points) in SAS™ using Pearson's r. For concurrent validity, we tested within-person correlations of MIS scores, alcohol consumption, and positive and negative affect in R using the 'rmcorr' package (Bakdash & Marusich, 2018). We examined convergent validity using between-person bivariate correlations between aggregate MIS scores and UPPS-P impulsivity-like facets computed in SAS™ using Pearson's r. We also computed between-person bivariate correlations among aggregate MIS scores and three laboratory-based impulsivity task outcomes, which purport to measure state-level impulsivity. Consistent with recommendations by Dejonckheere et al. (2019), correlations between the MSSD of MIS scores and UPPS-P impulsivity facets and laboratory-based impulsivity task outcomes reflect partial correlations after accounting for aggregate MIS scores.

3. Results (Sample 1: Daily Diary)

3.1. Factor structure, instability, and reliability

Consistent with Hypothesis 1, the one-factor solution of the MIS at between- and within-person levels exhibited excellent fit to the data ($\chi^2(4)$ = 12.04, p = .02; CFI = 0.98, RMSEA = 0.06, within-person SRMR = 0.03, between-person SRMR = 0.02). Variance attributed to between-person differences ranged from 38.1% to 55.6%. At the within-person level, the "impatient" item exhibited a notably low standardized factor loading (λ = 0.38). See Table 1 for standardized factor loadings and 95% CIs. See Table 2 for descriptive statistics of aggregate MIS scores, MSSD of MIS scores, laboratory-based tasks, and UPPS-P trait impulsivity facets. Rates of alcohol consumption and binge drinking episodes across days are also provided for both samples (see Supplementary Table 1).

Using a generalizability theory (GT) approach, the between-person reliability in Sample 1 was high (R_{KF} = 0.98), and the within-person reliability was also high (R_{CN} = 0.79), though comparably lower, consistent with Hypothesis 2. When using Coefficient Omega as the index of internal consistency, between-person reliability remained high in Sample 1 (ω = 0.90) whereas within-person reliability was low Sample 1 (ω = 0.61).

3.2. Content, criterion, and convergent validity

See Table 3 for between-person bivariate correlations among aggregate MIS scores, MSSD of MIS scores, and UPPS-P impulsivity facets. At the aggregate level and consistent with Hypothesis 3, MIS scores exhibited very large, positive correlations with positive and negative urgency facets. Aggregate MIS scores also evinced medium-to-large correlations with lack of planning and sensation seeking, whereas the correlation between MIS scores and lack of perseverance, across individuals, was small. MSSD of MIS scores were also positively correlated with aggregate MIS scores with a very large effect, indicating individuals with higher MIS scores (averaged across 10 days of assessment) endorsed more temporal instability of MIS scores across the assessment period. After accounting for aggregate MIS scores, partial correlations between the MSSD of MIS scores and UPPS-P impulsivity facets were not statistically significant (see Table 3).

Supporting Hypothesis 4, the within-person correlation between daily quantity of alcohol consumption and MIS scores

[4] Copyright © [2002–2014] SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

[5] To our knowledge, the first-order autocorrelation (iACORR(1)) and the variance of a time series (iVAR) have not been examined with MIS scores. Across samples, the first-order autocorrelations of MIS scores were comparable (Sample 1 iACORR(1) = -0.01; Sample 2 iACORR(1) = -0.14). However, the variances of MIS scores in a time series were not (Sample 1 = iVAR = 4.30; Sample 2 iVAR = 5.38). By extension, the MSSD of MIS scores are not directly comparable between samples because of these differing variances.

Table 1
ML-CFA standardized factor loadings and 95% confidence intervals across samples.

MIS Item	Between-Person λ	95% CI	Within-Person λ	95% CI
Sample 1				
Said without thinking	0.94	[0.90, 0.98]	0.67	[0.62, 0.72]
Spent more money	0.80	[0.74, 0.86]	0.60	[0.55, 0.65]
Impatient	0.80	[0.74, 0.86]	0.38	[0.34, 0.42]
Spur of the moment	0.93	[0.90, 0.96]	0.76	[0.72, 0.80]
Sample 2				
Said without thinking	0.86	[0.81, 0.91]	0.44	[0.39, 0.49]
Spent more money	0.79	[0.74, 0.84]	0.52	[0.47, 0.57]
Impatient	0.75	[0.69, 0.81]	0.21	[0.16, 0.26]
Spur of the moment	0.86	[0.81, 0.91]	0.53	[0.48, 0.58]

Note. ML-CFA = multilevel-confirmatory factor analysis. All factor loadings (λ) are standardized. MIS = Momentary Impulsivity Scale; CI = confidence interval. Sample 1 = daily diary sample; Sample 2 = EMA sample.

Table 2
Descriptive Statistics of Study Variables across Samples.

	Sample 1			Sample 2			
Variable	*M*	*SD*	Range	Variable	*M*	*SD*	Range
MSSD	21.46	22.56	0–136	MSSD	10.27	8.63	0.17–43.31
MIS	7.48	2.54	4.00–14.50	MIS[a]	4.03	2.21	0.25–11.50
NU	2.36	0.50	1.00–3.33	NU	2.33	0.59	1.17–3.83
PU	2.00	0.59	1.00–3.43	PU	1.84	0.59	1.00–4.00
SS	3.15	0.53	1.67–4.00	SS	3.08	0.57	1.42–4.00
LPer	1.83	0.44	1.00–3.10	Per	3.11	0.54	1.70–4.00
LPlan	1.93	0.47	1.09–3.18	Plan	3.17	0.45	1.27–4.00
bGS	31.04	18.50	0.00–80.00				
fGS	24.55	18.58	0.00–70.00				
bIMT	0.35	0.15	0.06-0.68				
fIMT	0.33	0.16	0.03-0.74				
bTCIP	16.73	12.80	0.00–50.00				
fTCIP	21.40	18.73	0.00–50.00				

Note. *M* = mean; *SD* = standard deviation; MSSD = mean squared successive difference of MIS scores; MIS = aggregate MIS scores (averaged across 10 days of assessment); NU = UPPS-P negative urgency; PU = UPPS-P positive urgency; SS = UPPS-P sensation seeking; LPer = UPPS-P lack of perseverance; Per = UPPS-P perseverance; LPlan = UPPS-P lack of planning; Plan = UPPS-P planning; bGS = baseline GoStop percentage of response inhibition failures; fGS = follow-up GoStop percentage of response inhibition failures; bIMT = baseline IMT ratio; fIMT = follow-up IMT ratio; bTCIP = baseline number of immediate choices; fTCIP = follow-up number of immediate choices; Sample 1 baseline n = 75–77; Sample 1 follow-up n = 63–66. Sample 2 baseline n = 126–147.

[a] MIS scores from Sample 2 (i.e., using a 0–100 sliding scale) were rescaled by dividing by 20 to facilitate direct comparisons between samples.

Table 3
Between-Person bivariate correlations between UPPS-P impulsivity facets and MIS scores.

	MSSD[a]	MIS	NU	PU	SS	Per	Plan
MSSD[a]	–	0.46**	−0.02	0.11	0.15	−0.01	−0.04
		[0.32, 0.58]	[−0.33, 0.29]	[−0.21, 0.40]	[−0.17, 0.44]	[−0.32, 0.30]	[−0.34, 0.27]
MIS	0.68**	–	0.42*	0.43**	0.00	−0.26**	−0.24**
	[0.53, 0.78]		[0.28, 0.55]	[0.29, 0.56]	[−0.16, 0.17]	[−0.41, −0.10]	[−0.39, −0.08]
NU	0.02	0.58**	–	0.65**	−0.04	−0.39**	−0.25**
	[−0.21, 0.25]	[0.41, 0.71]		[0.55, 0.74]	[−0.20, 0.12]	[−0.52, −0.24]	[−0.40, −0.09]
PU	0.18	0.57**	0.70**	–	0.20*	−0.22**	−0.30**
	[−0.05, 0.40]	[0.39, 0.70]	[0.56, 0.80]		[0.04, 0.35]	[−0.37, −0.06]	[−0.44, −0.14]
SS	−0.10	0.25*	0.38**	0.39**	–	0.17	−0.14
	[−0.33, 0.13]	[0.02, 0.45]	[0.16, 0.55]	[0.17, 0.56]		[−0.00, 0.32]	[−0.30, 0.02]
LPer	−0.06	0.11	0.33**	0.13	0.05	–	0.42**
	[−0.28, 0.18]	[−0.12, 0.33]	[0.11, 0.52]	[−0.10, 0.34]	[−0.18, 0.27]		[0.28, 0.55]
LPlan	0.03	0.28*	0.46**	0.44**	0.26*	0.39**	–
	[−0.20, 0.26]	[0.05, 0.48]	[0.26, 0.62]	[0.24, 0.60]	[0.03, 0.45]	[0.18, 0.56]	

Note. ** $p < .01$, * $p < .05$; MSSD = mean squared successive difference of MIS scores; MIS = aggregate MIS scores (averaged across 10 days of assessment); NU = UPPS-P negative urgency; PU = UPPS-P positive urgency; SS = UPPS-P sensation seeking; LPer = UPPS-P lack of perseverance; LPlan = UPPS-P lack of planning. Correlations from Sample 1 (daily diary sample) are included below the diagonal, whereas correlations from Sample 2 (EMA sample) are included above the diagonal. UPPS-P perseverance (per) and UPPS-P planning (plan) were scored such that higher scores reflect higher perseverance and planning in Sample 2, as opposed to (lack of) perseverance and planning. Ninety-five percent confidence intervals are contained in brackets.

[a] Correlations with MSSD of MIS scores reflect partial correlations after accounting for aggregate MIS scores consistent with Dejonckheere et al. (2019). Sample 1 Level-2 n = 74–76; Sample 2 Level-2 n = 126–147.

was positive with a large-to-very large effect (r = 0.36; 95% CI = 0.29–0.43; $p < .01$). Similarly, the within-person correlation between negative affect and MIS scores was also positive and large (r = 0.33; 95% CI = 0.25-0.40; $p < .01$). By contrast and inconsistent with hypotheses, positive affect and MIS scores were unrelated at the within-person level (r = −0.00; 95% CI = −0.08 to 0.08; p = .99).

Supporting Hypothesis 6 (Hypothesis 5 is only applicable to Sample 2), bivariate correlations among aggregate MIS scores, MSSD of MIS scores, and three laboratory-based task outcomes were inconsistent and not statistically significant. Specifically, aggregate MIS scores exhibited small, positive correlations with response initiation errors, as assessed by the IMT, at both baseline and follow-up. MIS scores showed a positive and very small-to-small correlation, across individuals, with prepotent response inhibition, as assessed by GoStop, at baseline, as well as a positive, very small correlation with prepotent response inhibition assessed at follow-up. By contrast, aggregate MIS scores demonstrated a *negative* and very small-to-small correlation with delay discounting, as assessed by the TCIP, at follow-up (but not at the baseline assessment). Partial correlations between the MSSD of MIS scores and laboratory-based impulsivity tasks were statistically non-significant (see Table 4).

4. Methods (Sample 2: EMA Sample)

4.1. Participants

Participants ($N = 169$; $M = 18.48$, $SD = 0.84$) were undergraduate students with an internet-enabled smartphone at a university in the Pacific Northwest who received course credit for survey participation. Participants who reported weekly alcohol or cannabis use on a screening survey administered to the participant subject pool were recruited for the current study. Participants self-identified as follows: 53% female; 63% White, 34% Asian/Pacific Islander, and 1% Black.

4.2. Study protocol

Eligible undergraduates completed an in-laboratory baseline survey and were trained on an EMA protocol. Starting on a Friday, participants received three text messages per day (at random times during the morning, mid-day, and evening) with a link to a brief web-based survey for 10 days. Participants received surveys at least two hours apart and were provided two hours to complete each survey with one text message reminder for each survey. Only five participants (3.5%) completed fewer than 50% of EMAs; the response rate among those completing more than 50% of EMAs was 88.2%. To increase the reliability of the EMA data, we only included data from participants who completed at least three full days of data collection (i.e. $\geq$9 total EMAs) in the final analysis (EMA $n = 147$). Eight excluded participants completed the baseline but declined the EMA portion of the study or had missing data on negative urgency, while four participants completed only one to three EMAs. There were no significant differences between included and excluded participants on the baseline data.

4.3. Measures

4.3.1. Demographics

Eligible participants completed demographic questions at the baseline visit, including self-reported age, sex assigned at birth, gender, race, and ethnicity.

4.3.2. Self-reported impulsivity-like facets

Consistent with Sample 1 (see Section 2.3.2.), we assessed five impulsivity-like facets via the 59-item UPPS-P Impulsive Behavior Scale (Lynam et al. 2007). Alphas ranged from good-to-excellent across scales ($\alpha = 0.82–0.92$).

4.3.3. Baseline alcohol consumption

Past-30-day alcohol consumption was assessed via the Daily Drinking Questionnaire-Revised (DDQ-R; Collins, Parks, & Marlatt, 1985) at the baseline assessment. The DDQ-R asks individuals to self-report the number of drinks on each of the 7-calendar days in the past 30 days' typical week and the past 30 days' heaviest drinking week. These measures were then summarized by taking the mean number of drinks per drinking day, resulting in a measure of daily alcohol consumption during a typical week and a measure of alcohol consumption during the heaviest drinking week.

4.3.4. Baseline alcohol problems

The 18-item Rutgers Alcohol Problem Index (RAPI; White & Labouvie, 2000) was used to assess past-year alcohol-related consequences (e.g., "got into fights with other people") at baseline using a 4-point Likert-type scale ranging from 0 (*none*) to 3 (*more than 5 times*). A mean score was computed for the 18 items ($\alpha = 0.85$).

4.3.5. Momentary impulsivity

The 4-item MIS (Tomko et al., 2014) was assessed at each assessment. Participants read each item and selected which number on a 0–100 sliding scale[6] best described their behavior. Anchors on the sliding scale were as follows: "strongly agree," "somewhat agree," "somewhat disagree," and "strongly disagree." At each time point, a randomly selected subset of MIS items was presented to lessen the respondent burden and reduce participant response fatigue resulting from seeing an identical survey many times over. This missing completely at random (MCAR) design allows for assumptions of missing data techniques to be met (Little & Rubin, 1989; Rubin, 1976). In correlational analyses, MIS scores were aggregated to the daily level. See below for between- and within-person reliabilities.

4.3.6. Daily alcohol consumption

During the morning assessment, participants were asked to report the number of standard alcoholic drinks consumed since yesterday's evening report (i.e., "Since the last assessment, how many alcoholic drinks have you had? Remember: one standard drink is 12 oz of beer, 8–9 oz. of malt liquor, 5 oz. of wine, or 1.5 oz. hard liquor.").

4.3.7. Daily perceived alcohol intoxication

At the morning assessment, participants reported on their perceived intoxication (i.e., "If you drank alcohol since the last assessment, how drunk/intoxicated did you get?"). Response options included "not at all/I didn't drink," "somewhat drunk," and "very drunk" using a 0–100 sliding scale.

4.3.8. Daily affect

At each assessment, participants were asked to report on their positive (5 items) and negative affect (5 items) using a 10-item version of the PANAS (Watson et al., 1998). The following positive affect items were included: "cheerful," "happy," "engaged," "calm," "friendly." Negative affect items included, "Irritable," "unhappy," "bored," "anxious," and "angry." Response options ranged from "not at all" to "very much" using a 0–100 sliding scale. Daily affect measures were combined to create mean scores representing positive affect (5 items) and negative affect (5 items). Positive and negative affect scores were aggregated to the daily level in correlational analyses.

[6] Sample 2 MIS scores were rescaled by dividing by 20 to make more direct comparisons between samples on indices of temporal instability. The rescaled MIS was only used when calculating the MSSD, ACORR, and WPV. The original, sliding scale variable was used in the ML-CFA and all correlational analyses.

Table 4
Between-Person bivariate correlations between laboratory-based tasks and MIS scores.

	MSSD[a]	MIS	bTCIP	bGS	bIMT	fTCIP	fGS
MSSD	–						
MIS	0.68** [0.53, 0.78]	–					
bTCIP	−0.08 [−0.30, 0.16]	0.04 [−0.19, 0.26]	–				
bGS	0.08 [−0.15, 0.30]	0.07 [−0.16, 0.29]	0.00 [−0.22, 0.23]	–			
bIMT	−0.04 [−0.27, 0.19]	0.11 [−0.12, 0.33]	−0.03 [−0.26, 0.19]	0.38** [0.16, 0.55]	–		
fTCIP	0.12 [−0.12, 0.36]	−0.09 [−0.33, 0.16]	0.46** [0.24, 0.63]	−0.04 [−0.28, 0.20]	0.13 [−0.12, 0.36]	–	
fGS	−0.11 [−0.34, 0.14]	0.05 [−0.20, 0.29]	0.06 [−0.19, 0.29]	0.47** [0.26, 0.64]	0.38** [0.15, 0.57]	0.12 [−0.13, 0.35]	–
fIMT	−0.10 [−0.34, 0.16]	0.11 [−0.09, 0.40]	−0.02 [−0.28, 0.23]	0.35** [0.11, 0.55]	0.75** [0.61, 0.84]	0.04 [−0.21, 0.29]	0.40** [0.17, 0.59]

Note. $^{**}p < .01$, $^{*}p < .05$; MSSD = mean squared successive difference of MIS scores; MIS = aggregate MIS scores (averaged across 10 days of assessment); bGS = baseline GoStop percentage of response inhibition failures; fGS = follow-up GoStop percentage of response inhibition failures; bIMT = baseline IMT ratio; fIMT = follow-up IMT ratio; bTCIP = baseline number of immediate choices; fTCIP = follow-up number of immediate choices. Ninety-five percent confidence intervals are contained in brackets.

[a] Correlations with MSSD of MIS scores reflect partial correlations after accounting for aggregate MIS scores consistent with Dejonckheere et al. (2019). Study 1 baseline n = 74–77; Study 1 follow-up n = 62–66.

4.4. Data analytic plan

Unless indicated otherwise, identical analytic procedures were used for Sample 2 (see Section 2.4). For Sample 2, ML-CFA was conducted in Mplus using maximum likelihood (ML) estimation, given the relatively continuous nature of item responses in this sample. Given that alcohol outcomes were only assessed at the first EMA survey of each day, variables examined in within-person correlational analyses (i.e., alcohol consumption, perceived level of intoxication, positive affect, negative affect, and MIS scores) were aggregated to the daily level.

4.4.1. Factor structure, instability, and reliability

See Section 2.4.1 for the analytic procedure, including mathematical equations used to estimate between- and within-person reliabilities of MIS items and temporal instability of MIS scores.

4.4.2. Content and criterion validity

Content validity analyses were computed identically to Sample 1 (see Section 2.4.2). In Sample 2, criterion validity was assessed at between- and within-person levels. First, we computed bivariate correlations between alcohol quantity (typical and heaviest week), alcohol problems, aggregate MIS scores, and the MSSD of MIS scores in SAS™ using Pearson's *r*. Next, we computed within-person correlations among MIS scores, reported alcohol quantity, perceived alcohol intoxication, positive affect, and negative affect in R.

5. Results (Sample 2: EMA)

5.1. Factor structure, instability, and reliability

Consistent with Sample 1 and Hypothesis 1, the one-factor solution of MIS items exhibited excellent fit to the data at between- and within-person levels ($\chi^2(4)$ = 8.12, p = .09; CFI = 0.99, RMSEA = 0.02, within-person SRMR = 0.02, between-person SRMR = 0.02). Variance attributed to between-person differences ranged from 24.2% to 33.5%. See Table 1 for standardized factor loadings and 95% CIs. Like Sample 1, the "impatient" item loaded poorly on the within-person factor (λ = 0.21; see Table 1). Consistent with Sample 1 and Hypothesis 2, the between- and within-person reliabilities were high using a GT approach (R_{KF} = 0.99; R_{CN} = 0.84). Using Coefficient Omega, between-person reliability remained high for Sample 2 (ω = 0.97), though the within-person reliability was low (ω = 0.49). See Table 2 for descriptive statistics of study variables from Sample 2.

5.2. Content and criterion validity

See Table 3 (above the diagonal) for between-person bivariate correlations among UPPS-P impulsivity traits, aggregate MIS scores, and the MSSD of MIS scores. Supporting Hypothesis 3, aggregate MIS scores (averaged across EMAs) evinced positive and very large correlations with negative and positive urgency. Aggregate MIS scores also exhibited medium-to-large correlations with (lack of) perseverance and (lack of) planning in the expected direction, whereas the correlation between MIS scores and sensation seeking was nonexistent in this sample. Partial correlations between the MSSD of MIS scores and UPPS-P impulsivity facets were statistically non-significant after accounting for aggregate MIS scores.

Consistent with Hypothesis 5 and providing support for between-person criterion validity, aggregate MIS scores positively correlated with alcohol-related problems with a medium-to-large effect. Aggregate MIS scores showed a minimal, negative correlation with typical alcohol consumption and a very small, negative correlation with heaviest alcohol consumption. Partial correlations between the MSSD of MIS scores and between-person alcohol indices were statistically non-significant after accounting for mean levels of MIS scores (see Table 5). For within-person criterion-validity, daily MIS scores and quantity of alcohol consumption exhibited a statistically non-significant correlation at the intraindividual level (r = 0.05; 95% CI = −0.08 to 0.18; p = .42). The within-person correlation between MIS scores and perceived intoxication was very small-to-small (r = 0.07; 95% CI = 0.01–0.13; p < .01). Consistent with Sample 1 and Hypothesis 4, negative affect evinced a medium-to-large, within-person correlation with MIS scores (r = 0.21; 95% CI = −0.13 to 0.26; p < .01). Finally, counter to Hypothesis 4, within-person positive affect exhibited a very small, inverse correlation with daily MIS scores (r = −0.06; 95% CI = −0.12 to −0.01; p = .02).

Table 5
Between-Person bivariate correlations between alcohol-related outcomes and MIS scores.

	RAPI	DDQ Typical	DDQ Heaviest	MSSD	MIS
RAPI	–				
DDQ Typical	0.27** [0.10, 0.43]	–			
DDQ Heaviest	0.27** [0.10, 0.43]	0.79** [0.71, 0.85]	–		
MSSD[a]	0.06 [−0.26, 0.36]	0.06 [−0.25, 0.37]	0.16 [−0.15, 0.45]	–	
MIS	0.27** [0.11, 0.42]	−0.04 [−0.20, 0.14]	−0.05 [−0.22, 0.12]	0.46** [0.32, 0.58]	–
M	1.44	4.89	7.10	10.27	24.96
SD	0.36	2.65	3.86	8.63	13.50

Note. ** $p < .01$, * $p < .05$; RAPI = past-year alcohol-related problems, as assessed by the Rutgers Alcohol Problem Index; DDQ Typical = past-30-day alcohol consumption over a typical week, as assessed by the DDQ-R; DDQ Heaviest = past-30-day alcohol consumption over a heavy week, as assessed by the DDQ-R; MSSD = mean squared successive difference of MIS scores; MIS = aggregate MIS scores (averaged across 10 days of assessment). Ninety-five percent confidence intervals are contained in brackets.

[a] Correlations with MSSD of MIS scores reflect partial correlations after accounting for aggregate MIS scores consistent with Dejonckheere et al. (2019). Sample 2 Level-2 n = 136–147.

6. Discussion

Despite its widespread adoption in the literature, we are the first study to attempt to further validate the psychometric properties of the MIS outside of the original study sample by using two intensive longitudinal samples of young adults. We also are the first to examine criterion validity for MIS scores at between- and within-person levels as well as the first to examine associations between MIS scores and laboratory-based impulsivity assessments thought to capture state-level impulsivity. Findings from the present study are briefly summarized as follows: (1) support for the MIS's one factor between- and within-person structure, with the "impatient" item loading poorly within-person; (2) high between-person reliabilities using GT and Coefficient Omega approaches; (3) high within-person reliabilities using GT and low within-person reliabilities using Coefficient Omega; (4) higher aggregate MIS scores linked to more temporal instability across the assessment window; (5) statistically non-significant partial correlations between the MSSD of MIS scores and UPPS-P facets, laboratory-based task outcomes, and between-person alcohol indices after accounting for average MIS scores; (6) large correlations between aggregate MIS scores and both positive and negative urgency facets; (7) aggregate relation between MIS scores and alcohol problems, but not consumption, at the between-person level in Sample 2; (8) inconsistent within-person correlations between MIS scores and alcohol indices across samples; (9) within-person correlation between MIS scores and negative affect across samples, with inconsistent within-person correlations between MIS scores and positive affect; and (10) minimal-to-small and inconsistent relations between aggregate MIS scores and laboratory-based assessments in Sample 1. Below, we fully integrate findings from our two young adult samples. We then discuss these findings in the context of the findings from Tomko et al. (2014) and other extant research in this area.

6.1. Factor structure, reliability, and temporal instability of the MIS

Across samples and at between- and within-person levels, the MIS's one-factor structure exhibited excellent fit to the data, with comparable between-person standardized factor loadings. Notably, the "impatient" item loaded poorly within-person in both samples (Sample 1: λ = 0.38, 95% CI = 0.34–0.42; Sample 2: λ = 0.21, 95% CI = 0.16–0.26), suggesting this item is potentially problematic, at least in young adult samples. This low factor loading may reflect the potential multidimensionality of state-level impulsivity that is not currently captured in its one-factor structure. Unlike the other three items, which appear to reflect impulsive *behaviors*, "I felt impatient" does not refer to an observable impulsive behavior, but rather a self-reflective perception. Speculatively, the "impatient" item may reflect affective impulsivity more directly, as opposed to a specific impulsive behavior, which may explain why this item loads poorly on the purported one-factor within-person MIS structure. In fact, intraindividual MIS item-level correlations with negative and positive affect provide preliminary support for this speculative notion. The "impatient" item showed a positive and large-to-very large correlation, within individuals, with negative affect in Sample 1 (r = 0.37), as well as a positive and medium-to-large within-person correlation in Sample 2 (r = 0.27). In Sample 2, the "impatient" item also evinced a small and negative within-person correlation with positive affect (r = −0.11; see Supplementary Table 2 for item-level correlations with positive and negative affect).

Differences in within-person standardized factor loadings were found between samples for the remaining three items: "said things without thinking" (Sample 1: λ = 0.67, 95% CI = 0.62–0.72; Sample 2: λ = 0.44, 95% CI = 0.39–0.49), "spent more money" (Sample 1: λ = 0.60, 95% CI = 0.55–0.65; Sample 2: λ = 0.52, 95% CI = 0.47–0.57), and "spur of the moment" (Sample 1: λ = 0.76, 95% CI = 0.72–0.80); Sample 2: λ = 0.53, 95% CI = 0.48–0.58). Speculatively, these findings may be due to different methodologies (i.e., daily diary vs. EMA) and assessment approaches used across samples (e.g., a Likert-type scale vs. a sliding scale), which may have impacted intraindividual response consistency.

In both samples, between-person reliabilities of MIS scores were very high for both samples when using a GT approach, consistent with Tomko et al. (2014), and when using the more contemporary Coefficient Omega, as recommended by Geldhof et al. (2014). Within-person reliabilities of MIS scores were also high when examined using a GT approach; however, within-person reliabilities were low in both samples when examined with Coefficient Omega estimates. It is possible that the within-person reliability estimate using GT may overestimate internal consistency. However, more research is needed to determine the internal consistency of MIS scores across samples, though, as it stands, there is limited evidence supporting the within-person internal consistency of MIS scores in young adults.

6.2. Validity of the MIS

6.2.1. Content validity

In both samples, aggregate MIS scores evinced very large correlations with both positive and negative urgency, suggesting the

MIS is linked to affective impulsivity, at least in young adult samples. Additional UPPS-P impulsivity facet correlations with aggregate MIS scores were relatively inconsistent across samples. For example, sensation seeking demonstrated a medium-to-large correlation with aggregate MIS scores in Sample 1, whereas this same correlation was nonexistent in Sample 2. Although there was an inconsistent pattern across samples for some impulsivity facets, MIS scores appear to be consistently linked to affective impulsivity, which was also found in Tomko et al. (2014) using the original urgency facet from the UPPS.

6.2.2. Criterion validity

Relations between MIS scores and alcohol outcomes were inconsistent across levels of analysis and samples. A medium-to-large between-person correlation between aggregate MIS scores and alcohol problems was found in Sample 2, whereas MIS scores showed small relations with other alcohol outcomes at within- and between-person levels in this sample. Conversely, a positive and large-to-very large within-person correlation between daily MIS scores and daily alcohol consumption was observed in Sample 1. Further research is necessary to determine sources contributing to differential relations between MIS scores and alcohol outcomes.

Across samples, MIS scores correlated with negative urgency with at least medium effect sizes, as well as with within-person fluctuations in negative affect. Conversely, within-person relations with positive affect were small and inconsistent across samples, suggesting the link between MIS scores and measures capturing negative mood states is relatively more robust in these data. Future work should examine the generalizability of these conclusions across various methodological approaches and types of samples.

Examinations of convergent validity in the present study were hindered by a dearth of state-level impulsivity assessments. Nevertheless, MIS scores exhibited inconsistent and minimal-to-small correlations with laboratory-based measures of impulsivity that were not statistically significant. These null results are also consistent with much work in this area demonstrating small correlations ($r < 0.15$) between self-report and laboratory-based assessment methods (Sharma et al., 2014). Given the psychometric limitations of laboratory-based tasks discussed elsewhere (e.g., Cyders & Coskunpinar, 2011; Fillmore & Weafer, 1999; Weafer, Baggott, & de Wit, 2013; see King et al., 2019), we refrain from concluding that the convergent validity of the MIS was not supported, though future research is needed to more accurately determine the MIS's convergent validity using well-validated measures in naturalistic studies.

6.2.3. MSSD of MIS scores

In both samples, the MSSD of MIS scores was not associated with self-report impulsivity facets, laboratory-based impulsivity task outcomes, or alcohol indices over and above mean-levels of MIS scores. This is not surprising given strong correlations between the MSSD of MIS scores and mean-level MIS scores in both samples (Sample 1 $r = 0.68$, $p < .01$; Sample 2 $r = 0.46$, $p < .01$). This is also consistent with recent evidence by Dejonckheere et al. (2019) that found, for positive and negative affect, the construct's mean and standard deviation explained a large majority of the variance in the construct's MSSD and that the MSSD did not contribute unique information in predicting psychological well-being over and above mean levels of positive and negative affect. These findings call into question the predictive utility of the MSSD beyond conventional descriptive statistics, such as the mean and standard deviation. More research in this area is needed to determine if this finding is consistent across samples, across assessment windows, and across frequency of assessments within an EMA framework.

6.3. Young adult vs. clinical adult MIS

Based on model fit indices, our young adult samples corroborated the MIS's one-factor between- and within-person structure that was first demonstrated in Tomko et al. (2014) adult clinical sample using exploratory factor analysis (EFA). However, there appear to be differences in standardized factor loadings across levels for young adults compared to Tomko et al. (2014). Between-person standardized factor loadings for "spent more money" and "impatient" items were larger for young adults than for clinical adults (Tomko et al.'s "spent more money" $\lambda = 0.59$, 95% CI = 0.56–0.62; Tomko et al.'s "impatient" $\lambda = 0.57$, 95% CI = 0.50–0.64). Differences in standardized factor loadings were also found within individuals, such that "said without thinking," "impatient," and "spur of the moment" items all exhibited lower within-person factor loadings in young adults compared to clinical adults (Tomko et al.'s "said without thinking" $\lambda = 0.84$, 95% CI = 0.82–0.86; Tomko et al.'s "impatient" $\lambda = 0.57$, 95% CI = 0.54–0.60; Tomko et al.'s "spur of the moment" $\lambda = 0.89$, 95% CI = 0.87–0.91). These inconsistent findings may reflect different analytic procedures being used across studies. Specifically, Tomko et al. (2014) used EFA to develop the original 4-item scale, whereas we are the first, to our knowledge, to employ ML-CFA using the MIS. Additionally, methodologies differed considerably across the three samples (e.g., daily diary vs. three EMAs daily vs. six EMAs daily), which could, in part, explain these equivocal findings.

The MIS appears to exhibit high between-person reliability across young adult samples and Tomko et al. (2014) adult clinical sample across approaches for internal consistency (i.e., internal consistency derived from GT vs. internal consistency determined by Coefficient Omega). This consistency is not surprising and reflects a reliable stability of impulsivity between individuals that is akin to trait impulsivity reflecting a person's mean-level of impulsivity across situations. Within-person reliability estimates differed from Tomko et al. (2014). When using the GT approach detailed by Tomko et al. (2014), within-person reliability estimates were higher in our two young adult samples. When using Coefficient Omega, which is an increasingly preferred reliability index (Dunn, Baguley, & Brunsden, 2014; Geldhof et al., 2014; Sijtsma, 2009), the within-person reliability estimates decreased, particularly for Sample 2. Thus, our findings provide additional evidence for the between-person reliability of MIS scores; however, evidence for the within-person reliability of MIS scores is inconclusive.

6.4. Research and clinical implications

Given state-level impulsivity assessment is in its nascent stages, these findings have important research and clinical implications. A central focus of this study was to reexamine the psychometric properties of the MIS in young adults using various methodological approaches. We are the first, to our knowledge, to do so using an MTMM approach with the inclusion of self-report and laboratory-based measures of impulsivity. Although findings indicate inconsistent and minimal-to-small, statistically nonsignificant relations between aggregate MIS scores and laboratory-based tasks of impulsivity, results support the extant literature in this area demonstrating little shared variance between the two assessment methods (i.e., self-report versus laboratory-based tasks; see Cyders & Coskunpinar, 2011; Sharma et al., 2014). As researchers continue to question if laboratory-based tasks truly capture "impulsivity" (see Sharma et al., 2014), particularly "state impulsivity," our results contribute to a growing body of literature suggesting laboratory-based assessments of impulsivity do not share considerable overlap with self-report assessments.

Our findings also have significant clinical implications. One of the most desirable qualities of assessing state-level impulsivity is to capture fluctuations in impulsivity and concurrent risky behavior. To date, clinicians have few options for assessing impulsivity in a clinical setting, despite impulsivity being the most transdiagnostic symptom criterion in the *DSM*-5 (American Psychiatric Association, 2013). Considering the MIS items are readily available and the scale itself is brief, this measure could be incorporated into clinical settings to track fluctuations in state impulsivity and impulsive behavior over time. This 4-item measure likely would not add considerable time to clinicians' and patients' time commitments and could also be implemented in time-limited therapeutic settings, such as integrated primary care or partial hospital settings. Likewise, just-in-time adaptive interventions (JITAIs; Spruijt-Metz & Nilsen, 2014) and ecological momentary interventions (EMI; Heron & Smyth, 2010) could include the MIS as an assessment of state-level impulsivity to provide real-time, evidence-based coping skills to effectively manage elevations and fluctuations in impulsivity and associated impulsive behaviors. For example, distress tolerance skills could be provided via mobile-phone interventions in response to elevated or increasingly variable MIS scores, given the apparent link between MIS scores and affective impulsivity. If deemed effective, this may reduce state-level impulsivity and associated impulsive behaviors (e.g., binge drinking; see Linehan, 1993).

6.5. Limitations and future directions

Study findings should be interpreted considering their limitations. Both samples predominantly consisted of individuals who self-identified as White and female; thus, generalization to other races and ethnicities may be limited. The Sample 1 sample size was also modest (though comparable to Tomko et al., 2014]); future replication studies should prioritize using large samples, as well as a longer assessment window to ensure adequate opportunity for fluctuations in impulsivity to occur. Samples 1 and 2 were also limited with respect to criterion outcomes, and future research should consider additional seemingly impulsive behaviors (e.g., unprotected sexual behavior, illicit substance use, pathological gambling) to further examine its criterion validity. Samples 1 and 2 also did not use identical items to assess positive and negative affect, which may influence the reliability of our comparisons between samples for positive and negative affect. Analytic approaches used in both samples were primarily correlational, and we cannot infer temporal ordering from within-person correlations or disaggregate between- and within-person effects (Curran & Bauer, 2011). Therefore, research should prioritize using time-varying effect modeling to understand the nuances of these dynamic associations (Tan, Shiyko, Li, Li, & Dierker, 2012). Finally, Monte Carlo confidence intervals for Coefficient Omega estimates could not be computed, as recommended by Geldhof et al. (2014), given the low within-person Omega found for Sample 2. We recommend that future studies prioritize using Coefficient Omega when examining multilevel internal consistency as this reliability estimate is increasingly preferred, particularly compared to Cronbach's alpha that assumes tau-equivalence (Dunn et al., 2014; Geldhof et al., 2014; Shrout & Lane, 2012; Sijtsma, 2009).

More broadly, self-report assessments continue to be used in ambulatory methodology given their convenience and psychometric qualities. Future EMA studies could be more intentional with assessment timing (e.g., event-based vs. scheduled assessments) to fully capture and disentangle the complex impulsivity-risk behavior relation. Studies utilizing cognitive interviewing could also inform the conceptualization of within- and between-person state-level impulsivity, given individuals and researchers may interpret questions and answers differently (which can, of course, impact study findings; see Schwarz, 1999).

6.5.1. Limitations of the MIS

Though we supported the one-factor structure of the 4-item MIS at between- and within-person levels in both samples, it is important to recognize that this is counter to the extant impulsivity literature showing that trait impulsivity is multidimensional with at least four facets (see Cyders, 2015; see also Smith et al., 2007). Indeed, it may be that four items are insufficient to fully capture the potential multidimensionality of momentary impulsivity, but little is known about the latent structure of momentary impulsivity, which may or may not mirror the latent structure of trait impulsivity. Existing evidence suggests that MIS scores strongly correlate, within- and between-persons, with UPPS urgency scores (Halvorson et al., under review), which is consistent with our between-person correlations from two young adult samples. At present, the MIS appears to reliably reflect UPPS urgency (Halvorson et al., under review; Tomko et al., 2014), though more work is needed to understand the latent structure of momentary impulsivity as it compares to trait impulsivity.

Three of the four MIS items also reflect specific impulsive behaviors (e.g., "I said things without thinking"). Behavioral items require opportunity to perform a given behavior within the assessment window and assessing observable momentary impulsive behaviors as a proxy for latent momentary impulsivity may underestimate true score differences related to this construct. Thus, it is recommended that the current 4-item MIS be used with caution, recognizing its important limitations. We also recommend several changes for future iterations of the MIS. First, it is recommended that additional items be considered in future revisions to reflect more breadth of impulsivity-like facets at the momentary level, consistent with our theoretical understanding of trait impulsivity (i.e., Halvorson et al., under review). Second, research should measure both discrete impulsive actions as well as individuals' perception of their impulsive states.

As stressed above, more research across various samples and methodological approaches is needed to: (1) reliably understand the meaning of state-level impulsivity as assessed by the MIS's four items at between- and within-person levels, (2) determine if and how the MIS's underlying, internal structure aligns with latent trait impulsivity models, and (3) clarify the dimensionality of state-level impulsivity through revisions to the MIS. In service of these goals, one potential next step is to investigate dynamic relations between MIS scores, UPPS-P impulsivity facets, and criterion outcomes at momentary and/or daily levels.

7. Conclusion

We employed intensive longitudinal designs in two young adult samples, and our findings provide important contributions to the burgeoning area of state impulsivity. Overall, we supported the MIS's one-factor between- and within-person structure and identified one potentially problematic item (i.e., "I felt impatient"), at least for young adult samples. Our findings suggest the MIS is best capturing emotion-driven impulsivity and is closely linked to negative affect within individuals, which has relevance to the etiology of problematic substance use (see Baker, Piper, McCarthy, Majeskie, & Fiore, 2004). With more replication and revision, the MIS shows promise for continued use in naturalistic studies and potential use in ambulatory clinical intervention research, especially considering its brevity. An important next step for research in this area is to consider the potential multidimensionality of state-level impulsivity that is not currently reflected.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jrp.2020.103914.

References

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Ansell, E. B., Laws, H. B., Roche, M. J., & Sinha, R. (2015). Effects of marijuana use on impulsivity and hostility in daily life. *Drug and Alcohol Dependence, 148*, 136–142.

Asparouhov, T., & Muthén, B. (2010). Weighted least squares estimation with missing data. Technical report. Retrieved from https://www.statmodel.com/download/GstrucMissingRevision.pdf.

Bagge, C. L., Littlefield, A. K., Rosellini, A. J., & Coffey, S. F. (2013). Relations among behavioral and questionnaire measures of impulsivity in a sample of suicide attempters. *Suicide and Life-Threatening Behavior, 43*, 460–467. https://doi.org/10.1111/sltb.12030.

Bakdash, J. Z., & Marusich, L. R. (2018). rmcorr: repeated measures correlation. R package version 0.3.0. https://CRAN.R-project.org/package=rmcorr.

Baker, T. B., Piper, M. E., McCarthy, D. E., Majeskie, M. R., & Fiore, M. C. (2004). Addiction motivation reformulated: An affective processing model of negative reinforcement. *Psychological Review, 111*(1), 33–51.

Barker, V., Romaniuk, L., Cardinal, R. N., Pope, M., Nicol, K., & Hall, J. (2015). Impulsivity in borderline personality disorder. *Psychological Medicine, 45*(9), 1955–1964.

Barlow, D. H., Sauer-Zavala, S., Carl, J. R., Bullis, J. R., & Ellard, K. K. (2014). The nature, diagnosis, and treatment of neuroticism: Back to the future. *Clinical Psychological Science, 2*(3), 344–365. https://doi.org/10.1177/2167702613505532.

Blanchard, B. E., Stevens, A. K., Littlefield, A. K., Talley, A. E., & Brown, J. L. (2017). Examining the link between nonmedical use of sedatives, tranquilizers, and pain relievers with dispositions toward impulsivity among college students. *Addictive Behaviors, 69*, 8–13. https://doi.org/10.1016/j.addbeh.2017.01.003.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105. https://doi.org/10.1037/h0046016.

Carver, C. S., Johnson, S. L., & Timpano, K. R. (2017). Toward a functional view of the p factor in psychopathology. *Clinical Psychological Science, 5*(5), 880–889. https://doi.org/10.1177/2167702617710037.

Collins, R. L., Parks, G. A., & Marlatt, G. A. (1985). Social determinants of alcohol consumption: The effects of social interaction and model status on the self-administration of alcohol. *Journal of Consulting and Clinical Psychology, 53*(2), 189–200.

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62*, 583–619. https://doi.org/10.1146/annurev.psych.093008.100356.

Cyders, M. A. (2015). The misnomer of impulsivity: Commentary on "choice impulsivity" and "rapid-response impulsivity" articles by Hamilton and colleagues. *Personality Disorders: Theory, Research, and Treatment, 6*, 204–205. https://doi.org/10.1037/per0000123.

Cyders, M. A., & Coskunpinar, A. (2011). Measurement of constructs using self-report and behavioral lab tasks: Is there overlap in nomothetic span and construct representation for impulsivity? *Clinical Psychology Review, 31*, 965–982. https://doi.org/10.1016/j.cpr.2011.06.001.

Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019). Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature Human Behaviour, 3*(5), 478–491.

DeYoung, C. G. (2015). Cybernetic big five theory. *Journal of Research in Personality, 56*(33), 58. https://doi.org/10.1016/j.jrp.2014.07.004.

Dougherty, D. M., Marsh, D. M., & Mathias, C. W. (2002). Immediate and delayed memory tasks: A computerized behavioral measure of memory, attention, and impulsivity. *Behavior Research Methods, Instruments, & Computers, 34*, 391–398.

Dougherty, D. M., Marsh-Richard, D. M., Hatzis, E. S., Nouvion, S. O., & Mathias, C. W. (2008). A test of alcohol dose effects on multiple behavioral measures of impulsivity. *Drug and Alcohol Dependence, 96*, 111–120. https://doi.org/10.1016/j.drugalcdep.2008.02.002.

Dougherty, D. M., Mathias, C. W., Marsh, D. M., & Jagar, A. A. (2005). Laboratory behavioral measures of impulsivity. *Behavior Research Methods, 37*, 82–90.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399–412.

Evenden, J. (1999). Impulsivity: A discussion of clinical and experimental findings. *Journal of Psychopharmacology, 13*, 180–192. https://doi.org/10.1177/026988119901300211.

Fillmore, M. T., & Weafer, J. (1999). Behavioral inhibition and addiction. In J. MacKillop & H. de Wit (Eds.), *The Wiley-Blackwell handbook of addiction psychopharmacology* (pp. 135–167). West Sussex, UK: John Wiley and Sons Limited.

Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science, 13*(2), 83–87.

Fleeson, W. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. *Journal of Personality, 75*, 825–862. https://doi.org/10.1111/j.1467-6494.2007.00458.x.

Fleeson, W. (2017). The production mechanisms of traits: Reflections on two amazing decades. *Journal of Research in Personality, 69*, 4–12. https://doi.org/10.1016/j.jrp.2017.07.003.

Fleeson, W., & Jayawickreme, E. (2015). Whole trait theory. *Journal of Research in Personality, 56*, 82–92. https://doi.org/10.1016/j.jrp.2014.10.009.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156–168. https://doi.org/10.1177/2515245919847202.

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*(1), 72–91.

Halvorson, M. A., Pedersen, S. L., Field, M. C., Lengua, L. J., Molina, B. S. G., & King, K. M. (in preparation). Impulsive states and impulsive traits: A study of the multilevel structure and validity of a multifaceted measure of impulsivity states.

Heron, K. E., & Smyth, J. M. (2010). Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psychology, 15*(1), 1–39. https://doi.org/10.1348/135910709X466063.

Jahng, S. (2008). *A mixed model for variance of successive difference of stationary time series: Modeling temporal instability in intensive longitudinal data*. University of Missouri-Columbia.

Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods, 13*, 354–375. https://doi.org/10.1037/a0014173.

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist, 43*(1), 23–34.

King, K. M., Feil, M. C., Seldin, K., Smith, M. R., Halvorson, M. A., & Kuehn, K. S. (2019). Assessment of risky and impulsive behaviors. In E. Youngstrom & M. Prinstein (Eds.), *Assessment of childhood disorders* (5th ed.. Guilford Press.

Linehan, M. M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. New York, NY, US: Guilford Press.

Little, R. J., & Rubin, D. B. (1989). The analysis of social science data with missing values. *Sociological Methods & Research, 18*(2–3), 292–326. https://doi.org/10.1177/0049124189018002004.

Littlefield, A. K., & Sher, K. J. (2016). Personality and substance use disorders. In K. J. Sher (Ed.), *The Oxford handbook of substance use and substance use disorders: Two-volume set*. https://doi.org/10.1093/oxfordhb/9780199381678.013.006.

Littlefield, A. K., Stevens, A. K., Ellingson, J. M., King, K. M., & Jackson, K. M. (2016). Changes in negative urgency, positive urgency, and sensation seeking across adolescence. *Personality and Individual Differences, 90*, 332–337. https://doi.org/10.1016/j.paid.2015.11.024.

Lynam, D. R., Smith, G. T., Cyders, M. A., Fischer, S., & Whiteside, S. A. (2007). The UPPS-P: A multidimensional measure of risk for impulsive behavior. Unpublished technical report.

Muthén, L. K., & Muthén, B. O. (1998–2018). Mplus user's guide (8th ed.). Los Angeles, CA: Muthén & Muthén.

National Institute on Alcohol Abuse and Alcoholism (NIAAA) (2005). Helping patients who drink too much: A clinician's guide. NIH Pub. No. 07–3769. Bethesda, MD: NIAAA.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), 943. https://doi.org/10.1126/science.aac4716.

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology, 51*, 768–774.

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Founding for Statistical Computing http://www.R-project.org/.

Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality, 43*(2), 137–145.

Rubin, D. B. (1976). Multivariate matching methods that are equal percent bias reducing, II: Maximums on bias reduction for fixed sample sizes. *Biometrics*, 121–132. https://doi.org/10.2307/2529343.

SAS Institute Inc (2010). *The SAS system for Windows. Release 9.4*. Cary, NC: SAS Inst.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*(2), 93–105. https://doi.org/10.1037/0003-066X.54.2.93.

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. *Psychology Bulletin, 140*, 274–408. https://doi.org/10.1037/a0034418.

Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 620–635). New York, NY: Guilford Press.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107.

Simons, J. S., Dvorak, R. D., Batien, B. D., & Wray, T. B. (2010). Event-level associations between affect, alcohol intoxication, and acute dependence symptoms: Effects of urgency, self-control, and drinking experience. *Addictive Behaviors, 35*, 1045–1053. https://doi.org/10.1016/j.addbeh.2010.07.001.

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.

Spruijt-Metz, D., & Nilsen, W. (2014). Dynamic models of behavior for just-in-time adaptive interventions. *IEEE Pervasive Computing, 13*(3), 13–17. https://doi.org/10.1109/MPRV.2014.46.

Stevens, A. K., Blanchard, B. E., & Littlefield, A. K. (2018). Impulsive dispositions and alcohol: What we know, how we know it, and where to go from here. *Borderline Personality Disorder and Emotion Dysregulation, 5*(1), 4. https://doi.org/10.1186/s40479-018-0081-0.

Stevens, A. K., Littlefield, A. K., Talley, A. E., & Brown, J. L. (2017). Do individuals higher in impulsivity drink more impulsively? A pilot study within a high risk sample of young adults. *Addictive Behaviors, 65*, 147–153. https://doi.org/10.1016/j.addbeh.2016.10.026.

Tan, X., Shiyko, M. P., Li, R., Li, Y., & Dierker, L. (2012). A time-varying effect model for intensive longitudinal data. *Psychological Methods, 17*(1), 61–77. https://doi.org/10.1037/a0025814.

Tomko, R. L., Solhan, M. B., Carpenter, R. W., Brown, W. C., Jahng, S., Wood, P. K., & Trull, T. J. (2014). Measuring impulsivity in daily life: The momentary impulsivity scale. *Psychological Assessment, 26*, 339–349. https://doi.org/10.1037/a0035083.

Trull, T. J., Wycoff, A. M., Lane, S. P., Carpenter, R. W., & Brown, W. C. (2016). Cannabis and alcohol use, affect and impulsivity in psychiatric out-patients' daily lives. *Addiction, 111*(11), 2052–2059. https://doi.org/10.1111/add.13471|.

Vazire, S., & Sherman, R. A. (2017). Introduction to the special issue on within-person variability in personality. *Journal of Research in Personality*.

Watson, D., Clark, L. A., & Tellegen, A. (1998). Development and validation of brief measures of positive and negative affect. *Journal of Personality and Social Psychology, 54*, 1063–1070.

Weafer, J., Baggott, M. J., & de Wit, H. (2013). Test–retest reliability of behavioral measures of impulsive choice, impulsive action, and inattention. *Experimental and Clinical Psychopharmacology, 21*, 475–481.

White, H. R., & Labouvie, E. W. (2000). Longitudinal trends in problem drinking as measured by the Rutgers Alcohol Problem Index. *Alcoholism: Clinical and Experimental Research, 24*(5 Suppl 1).

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*, 669–689. https://doi.org/10.1016/S0191-8869(00)00064-7.

Wilson, R. E., Thompson, R. J., & Vazire, S. (2017). Are fluctuations in personality states more than fluctuations in affect?. *Journal of Research in Personality, 69*, 110–123.