

Validation of an emotional stop-signal task to probe individual differences in emotional response inhibition: Relationships with positive and negative urgency

Kenneth J. D. Allen^{1,2} , Sheri L. Johnson¹, Taylor A. Burke^{3,4},
M. McLean Sammon^{2,5} , Christina Wu^{2,4,6}, Max A. Kramer^{2,7},
Jinhan Wu², Heather T. Schatten^{8,9}, Michael F. Arme^{8,9}
and Jill M. Hooley¹⁰

Brain and Neuroscience Advances

Volume 5: 1–11

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/23982128211058269

journals.sagepub.com/home/bna



Abstract

Performance on an emotional stop-signal task designed to assess emotional response inhibition has been associated with Negative Urgency and psychopathology, particularly self-injurious behaviors. Indeed, difficulty inhibiting prepotent negative responses to aversive stimuli on the emotional stop-signal task (i.e. poor *negative* emotional response inhibition) partially explains the association between Negative Urgency and non-suicidal self-injury. Here, we combine existing data sets from clinical (hospitalised psychiatric inpatients) and non-clinical (community/student participants) samples aged 18–65 years ($N=450$) to examine the psychometric properties of this behavioural task and evaluate hypotheses that emotional stop-signal task metrics relate to distinct impulsive traits among participants who also completed the UPPS-P ($n=223$). We specifically predicted associations between worse *negative* emotional response inhibition (i.e. commission errors during stop-signal trials representing negative reactions to unpleasant images) and Negative Urgency, whereas commission errors to positive stimuli – reflecting worse *positive* emotional response inhibition – would relate to Positive Urgency. Results support the emotional stop-signal task's convergent and discriminant validity: as hypothesised, poor negative emotional response inhibition was specifically associated with Negative Urgency and no other impulsive traits on the UPPS-P. However, we did not find the hypothesised association between positive emotional response inhibition and Positive Urgency. Correlations between emotional stop-signal task performance and self-report measures were the modest, similar to other behavioural tasks. Participants who completed the emotional stop-signal task twice ($n=61$) additionally provide preliminary evidence for test–retest reliability. Together, findings suggest adequate reliability and validity of the emotional stop-signal task to derive candidate behavioural markers of neurocognitive functioning associated with Negative Urgency and psychopathology.

Keywords

Affective control, behavioural research, cognitive control, emotional regulation, executive function, neuropsychological tests, reactive inhibition, self-control, stop-signal task, urgency

Received: 14 September 2021; accepted: 8 October 2021

Accumulating research has sought to explain the co-occurrence of heterogeneous psychopathologies. The ‘*p* factor’, a statistical representation of shared latent susceptibility to psychiatric disorders, is considered to represent generalised psychopathology vulnerability. Studies have consequently sought to identify risk markers that might reflect the *p* factor. Emotion dysregulation and disinhibition are candidate constructs linked to nearly all psychiatric conditions. Negative emotionality, defined by frequent and intense aversive affect, is one facet of emotion dysregulation that enhances predisposition to internalising spectrum disorders (Tackett and Lahey, 2017). Disinhibition relates similarly to externalising psychopathology (Creswell et al., 2019; Mullins-Sweatt et al., 2019).

Negative Urgency, the tendency to respond impulsively to negative affect (Whiteside and Lynam, 2001), is characterised by high levels of negative emotionality *and* disinhibition (in combination with low agreeableness; Settles et al., 2012). This trait is therefore unsurprisingly implicated across internalising, externalising, and psychotic disorders (Berg et al., 2015; Hoptman

¹Department of Psychology, University of California, Berkeley, Berkeley, CA, USA

²Department of Psychology, Oberlin College and Conservatory, Oberlin, OH, USA

³Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA

⁴Department of Psychiatry, Harvard Medical School, Boston, MA, USA

⁵Western Psychiatric Institute and Clinic, University of Pittsburgh Medical Center, Pittsburgh, PA, USA

⁶Department of Psychiatry, McLean Hospital, Belmont, MA, USA

⁷Division of the Social Sciences, The University of Chicago, Chicago, IL, USA

⁸Psychosocial Research Program, Butler Hospital, Providence, RI, USA

⁹Department of Psychiatry & Human Behavior, The Warren Alpert Medical School of Brown University, Providence, RI, USA

¹⁰Department of Psychology, Harvard University, Cambridge, MA, USA

Corresponding author:

Kenneth J. D. Allen, Department of Psychology, University of California, Berkeley, 2121 Berkeley Way, Berkeley, CA 94704, USA.
Emails: jd.allen@berkeley.edu; kenneth.jd.allen@oberlin.edu



et al., 2014; Muhtadie et al., 2013). Cyders et al. (2007) developed a Positive Urgency scale to capture corresponding reactivity to positive emotional states, which is similarly linked to transdiagnostic dysfunction (Berg et al., 2015; Cyders and Smith, 2008; Zolowski et al., 2009). Accordingly, Carver et al. (2017) articulated the idea that valence-general Urgency might be one trait manifestation of the *p* factor.

Urgency is typically measured using self-rated questionnaires, although parallel findings have been observed using ecological (Schatten et al., 2019) and informant-based (Zolowski and Smith, 2013) ratings. Complementary research seeking to identify objective markers of Urgency using neuropsychological tasks indicates fairly robust associations with response inhibition deficits, albeit with relatively small effect sizes (Carver and Johnson, 2018; Cyders and Coskunpinar, 2011, 2012; Johnson et al., 2016). Given this construct's intrinsically *affective* nature, this report focuses on *emotional* response inhibition (ERI), a proposed behavioural indicator of self-reported Urgency.

Response inhibition can be delineated into sequential stages: (1) *early* action suppression, involving 'withholding' of prepotent impulses before initiating accompanying motor responses, frequently measured using go/no-go or continuous performance tasks and (2) *late* action termination, requiring 'cancellation' of ongoing behavioural impulses *after* response initiation, often assessed via stop-signal tasks with greater inhibitory demand (Bari and Robbins, 2013; Littman and Takács, 2017; Sebastian et al., 2013). Prior work has established the transdiagnostic influence of ERI impairment on psychopathology and Negative Urgency (Gay et al., 2008). Studies have primarily investigated *early-stage* emotional reaction *suppression*, establishing early negative ERI deficits across externalising (Brugman et al., 2016; Denny and Siemer, 2012; Iria et al., 2012) and internalising spectra (Hjorft et al., 2017; Pacheco-Unguetti et al., 2012).

A few studies have specifically examined *late-stage* ERI. Growing evidence suggests that negative ERI deficits in both early and late stages may underpin emotion dysregulation, Urgency, and related characteristics – particularly in the context of childhood adversity. For example, *early* negative ERI impairment may be one mechanism through which childhood poverty exerts an indirect influence on internalising symptoms (Capistrano et al., 2016). Superior *late* negative ERI may conversely buffer the impact of early-life maltreatment on later suicidal behaviors (Allen et al., 2021). Late negative ERI dysfunction also partially explains Negative Urgency's relationship with non-suicidal self-injury (NSSI; Allen and Hooley, 2019) and prospectively predicts NSSI urges (Burke et al., 2021).

This study aims to (1) evaluate the psychometric properties of an emotional stop-signal task (ESST) designed to measure late ERI and (2) examine relationships between ERI parameters and Urgency, by analyzing aggregated data sets including this task. We hypothesise (A) adequate test–retest reliability of ESST parameters and (B) valence-specific associations of negative and positive ERI metrics with Negative and Positive Urgency scales, providing initial evidence for this task's convergent and discriminant validity.

Method

Participants

We examined a large, diverse group of participants (e.g. with and without psychiatric history) who completed the ESST in varied

contexts to provide expected ranges of task performance parameter values. However, our primary analyses focused on subsamples of participants who either completed the ESST twice (Hypothesis A) or who completed the UPPS-P (Hypothesis B). We accordingly aggregated data from multiple completed (see Allen et al., 2021; Allen and Hooley, 2015, 2019; Burke et al., 2021) and ongoing (R01 MH108610, MPI: H.T.S., Miller, and Mower-Provost; R01 MH112674, PI: M.F.A.) research studies, comprising community, student, and psychiatric samples of adults aged 18–65 years old (total $N=450$ before outlier removal described below; see Supplementary Figure S1 for additional information). Psychiatric participants were recruited from hospital inpatient units (with the permission of their treatment teams) based on histories of suicidal ideation or behaviors; individuals whose psychotic or mood symptoms were sufficiently severe to impede participation in all study procedures (e.g. ecological momentary assessment) were excluded. Hospitalised participants included psychiatric inpatients ($n=150$) with a history of (and/or current): alcohol and/or other substance use disorders ($n=95$; 63.3%), anxiety disorders ($n=25$; 16.7%), bipolar spectrum disorders ($n=23$; 15.3%), depressive disorders ($n=97$; 64.7%), feeding and eating disorders ($n=4$; 2.7%), neurodevelopmental disorders ($n=3$; 2.0%), obsessive-compulsive and related disorders ($n=2$; 1.3%), personality disorders ($n=6$; 4.0%), post-traumatic stress disorder ($n=15$; 10.0%), and psychotic disorders ($n=9$; 6.0%).

We evaluated test–retest reliability of ESST variables in a subgroup of hospitalised adult psychiatric inpatient participants ($n=61$) who completed a baseline assessment during their stay and returned for follow-up evaluation between 2 and 6 months after hospital discharge (mean (M)=111.42 days; standard deviation (SD)=51.77; median=85 days; range=59–221 days). Cross-sectional analyses included additional psychiatric inpatient participants ($n=98$) who only completed the ESST once, that is, at baseline during hospitalisation. Two other data sets consisted of community and student participants recruited from the greater Boston area, based on self-reported NSSI engagement ($n=79$) or lack of psychiatric history ($n=90$). A fourth data set comprised students enrolled in a public research university on the east coast of the United States, including those with ($n=62$) and without at least two acts of NSSI ($n=58$), plus two additional participants with complete ESST data whose NSSI history is unknown. Since the UPPS-P was not administered to inpatient participants, validity analyses were necessarily restricted to community and student participants across samples who completed *both* the ESST and UPPS-P ($n=223$ before outlier removal; see Table 1) to assess hypothesised links between ERI deficits and Urgency scales (Each research group also assessed psychopathology via clinical interviews and/or symptom checklists; data pertaining to psychiatric history and current symptoms are excluded from this report but available upon request.).

Measures

ESST. The ESST (see Figure 1) is a modified version of the original stop-signal task developed by Logan and Cowan (1984) (Allen and Hooley, 2015, 2019). This study evaluates two variants that differ only in the number of stimulus categories (four vs three). The revised version (Allen and Hooley, 2019) includes

Table 1. Sample demographic characteristics.

		Hypothesis A: longitudinal sample ($n=55$)	Hypothesis B: cross-sectional sample ($n=221$)	Total baseline ESST sample ($n=432$) ^a
		M (SD)		
	Age (years)	38.31 (13.97)	22.03 (6.11)	26.96 (11.04)
	Education (years)	13.48 (2.89)	14.43 (1.91)	14.09 (2.60)
		n (valid %)		
Gender	Female	35 (63.64)	187 (84.62)	311 (72.83)
	Male	20 (36.36)	34 (15.39)	115 (26.93)
	Non-binary	–	–	1 (0.23)
	Did not disclose/missing	–	–	5 (n/a)
Sexual orientation	Heterosexual	39 (72.22)	82 (80.39)	184 (77.64)
	LGBQ+	15 (27.78)	20 (19.61)	53 (22.36)
	Did not disclose/missing	1 (n/a)	119 (n/a)	195 (n/a)
Ethnicity/race	Asian	1 (1.82)	41 (18.64)	47 (12.95)
	Black/African American	1 (1.82)	20 (9.09)	23 (6.34)
	Multiracial/other	5 (9.09)	23 (10.46)	36 (9.95)
	Native American ^b	2 (3.64)	–	3 (0.83)
	White (Hispanic/Latinx)	–	4 (1.82)	4 (1.10)
	White (non-Hispanic)	46 (83.64)	132 (60.00)	249 (68.60)
	Did not disclose/missing	–	1 (n/a)	70 (n/a)
	Marital status	Cohabiting/engaged/married	12 (22.22)	15 (14.71)
	Dating/in a relationship	–	39 (17.65)	39 (16.46)
	Divorced/separated/widowed	14 (25.93)	1 (0.98)	36 (15.19)
	Single/not married	28 (51.85)	47 (46.08)	124 (52.32)
	Did not disclose/missing	1 (n/a)	119 (n/a)	195 (n/a)

ESST: emotional stop-signal task; SD: standard deviation; n/a: not available.

^aTotal sample includes individuals in the other two columns in combination with all other participants who completed the ESST at least once, after outlier removal (see 'Analytic procedures'), which resulted in the exclusion of 18 individuals (4.0%).

^bThis group included participants who identified as American Indian and/or native to Alaska or Hawaii (or Pacific Islander). Five participants did not disclose their age and 135 participants did not report educational history.

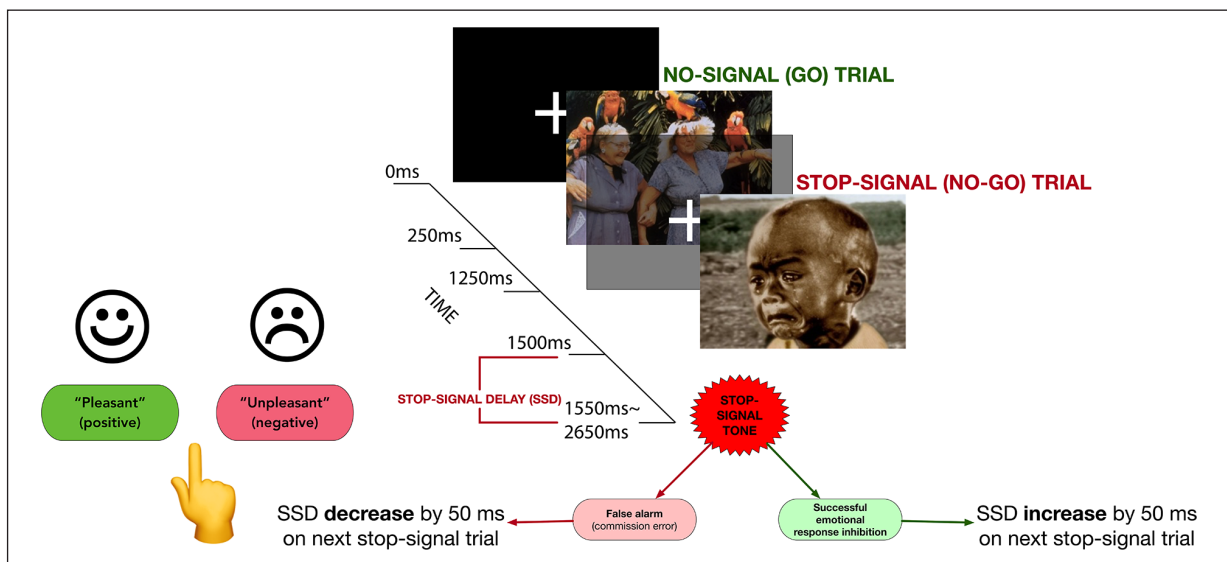


Figure 1. The emotional stop-signal task (ESST) instructs participants to rapidly indicate the valence of serially presented images by keypress, *except* on trials with an auditory stop-signal, when participants are asked to inhibit their emotional reaction and accompanying behavioural response. If participants are unable to inhibit an emotional response on a stop or 'no-go' trial, the staircase tracking algorithm decreases the stop-signal delay (SSD) on the subsequent stop trial, thereby reducing time for stimulus evaluation, response selection, and motor preparation (and vice versa).

three categories of image stimuli (Neutral, Positive, and Negative) from the International Affective Picture System (IAPS; Lang et al., 2008), whereas the original ESST, designed to study NSSI, includes a fourth category of images depicting self-harm (Allen and Hooley, 2015; Burke et al., 2021). Most participants completed the revised ESST ($n=264$; 58.7%), and given the general focus of this report, we omitted trials with self-harm stimuli from current analyses.

In this task, participants are asked to ‘quickly and accurately’ categorise (via keypress) the valence of serially presented Negative, Positive, and Neutral IAPS stimuli, randomised by valence within-block and matched for image arousal/intensity between Negative and Positive stimulus categories, as either ‘pleasant/positive’ or ‘unpleasant/negative’ on trials without a stop-signal, that is, ‘go’ or no-signal trials ($n=192$). The revised ESST comprises $N=224$ trials across four blocks (including 32 practice trials) that are roughly evenly distributed across stimulus valence, with $n=48$ trials (~16 per IAPS category) that include an auditory stop-signal, during which participants are instructed to inhibit their affective reaction and accompanying behavioural response. The temporal delay of the stop-signal is continually adjusted in 50ms increments based on individual performance via a staircase tracking algorithm. This adaptive component of the original task is necessary to estimate *stop-signal reaction time* (SSRT), which requires total commission errors (false alarms) to remain around 50% (independent of stimulus category). Maintaining an overall commission error rate around 50% (across stimulus type) also enables us to directly compare the effects of different types of IAPS stimuli on late ERI capacities, providing information regarding the relative difficulty of terminating emotional reactions to Positive, Negative, and emotionally ambiguous (i.e. Neutral) images.

The ESST provides numerous parameters relevant to affective processing, some of which are omitted here for clarity and to reduce the likelihood of statistical error in hypothesis-testing analyses. Our main variables of interest here include several measures of late ERI impairment (i.e. higher scores indicate worse ERI): (1) SSRT, calculated per traditional guidelines (Verbruggen and Logan, 2008) as the median stop-signal delay subtracted from mean reaction time during trials without a stop-signal; (2) Positive and (3a) Negative false alarm rate, that is, the percentage of commission errors (failed inhibition of negative or positive behavioural responses, matched to stimulus valence) during stop-signal trials with Positive or Negative IAPS images, respectively (see Allen and Hooley, 2015). We also derived an alternative metric of negative ERI impairment, (3b) $P(\text{Negative false alarm})$, the probability of making a behavioural response or commission error reflecting a *negative* judgment (i.e. that an image is aversive or ‘unpleasant’) on stop-signal trials *independent* of stimulus category. In other words, this variable represents the proportion of negative false alarms relative to the total number of trials with a stop-signal, but only those that *also* reflect negative reactions to IAPS stimuli that participants failed to inhibit as instructed, regardless of whether the image was classified a priori (using normative ratings; Lang et al., 2008) as Neutral, Positive, or Negative; we previously referred to this variable as *negative emotional action termination* (see Allen and Hooley, 2019). Although positive and negative ERI metrics allow us to test valence-specific predictions regarding their relationships to Positive and Negative Urgency (i.e. Hypothesis B),

SSRT may be an indicator of domain-general ERI deficits unrelated to valence; we therefore expect non-specific associations between SSRT and overall Urgency, taken as the mean of Positive and Negative Urgency scores.

We report the following ESST metrics primarily for comparison with similar tasks: (1) Accuracy, the percentage of no-signal trials with correctly-identified Positive or Negative IAPS stimuli, which is the only ESST variable where higher values reflect better performance; (2) Negativity bias, a related measure of interpretive bias and emotional identification abilities, defined as the percentage of all no-signal trials with negative judgments (i.e. behavioural responses indicating that participants identified the presented image as aversive or ‘unpleasant’), independent of actual IAPS stimulus valence; (3) No-signal reaction time, a measure of affective processing speed in the ESST, operationalised as the mean reaction time during no-signal trials, reflecting the latency of valence evaluation and emotional discrimination processes; (4) Miss rate, the percentage of total omission errors, that is, no responses on no-signal trials; (5) Total false alarm rate, the percentage of total commission errors (i.e. any responses on stop-signal trials), which should approximate 50% by design; finally, (6) Neutral false alarm rate, the percentage of commission errors (regardless of valence) during stop-signal trials with ambiguous or Neutral IAPS stimuli.

UPPS-P Impulsive Behaviour Scale. This psychometrically sound (Cyders et al., 2007; Smith et al., 2007) questionnaire consists of 59 items rated on a 4-point scale reflecting how much participants ‘agree or disagree’ with each statement (Lynam et al., 2006). The UPPS-P produces scores on five dimensions of trait-like impulsivity: (1) Positive and (2) Negative Urgency, the tendency to experience and act on strong impulses in emotional contexts, for example, ‘It is hard for me to resist acting on my feelings’; (3) (lack of) Premeditation, or the propensity to act without forethought, for example, ‘I am [not] a cautious person’; (4) (lack of) Perseverance, or the inability to remain focused on difficult or boring tasks, for example, ‘I tend to give up easily’; (5) Sensation-seeking, which refers to one’s preference for exciting or novel experiences, for example, ‘I would enjoy parachute jumping’. We calculated an additional ‘valence-general’ Urgency scale by taking the mean of Positive and Negative Urgency scores.

Analytic procedures

We first performed data cleaning and processing procedures in SPSS version 27.0 and JASP version 0.11.1 (JASP Team, 2019), which included requisite statistical assumption checks (e.g. confirming acceptable skewness and kurtosis values; see Byrne, 2010). In line with similar studies (e.g. Allen et al., 2019b; Johnson and Tottenham, 2015), we excluded participants with outlier ESST performance (i.e. more than three standard deviations outside mean values) based on the following: low valence identification Accuracy during no-signal trials with Positive or Negative IAPS stimuli; a large number of omission errors or high Miss rate; as well as negative SSRT values that indicate deliberate slowing of no-signal responses, violating a key assumption of stop-signal tasks’ estimation of inhibitory speed (e.g. Congdon et al., 2012). Exclusion criteria resulted in the removal of $n=18$

Table 2. ESST performance and UPPS-P scores.

	Hypothesis A subsample ($n=55$)		Hypothesis B subsample ($n=221$)	Total sample ($n=432$)
	Baseline (T1)	Follow-up (T2)		
ESST Accuracy (%)	90.85 (5.34)	91.51 (6.79)	93.94 (5.46)	92.94 (6.06)
ESST Negativity bias (%)	35.77 (8.30)	35.46 (7.99)	45.96 (9.66)	44.25 (11.13)
ESST Miss rate (%)	10.49 (7.01)	7.61 (8.23)	5.96 (6.01)	6.31 (6.28)
ESST No-signal reaction time (ms)	836.52 (100.29)	811.28 (124.71)	782.00 (104.85)	789.23 (104.52)
ESST Stop-signal reaction time (ms)	355.55 (194.18)	341.11 (192.99)	301.06 (90.67)	331.12 (127.93)
ESST Total false alarm rate (%)	50.07 (16.57)	52.14 (19.83)	46.96 (9.47)	49.22 (13.21)
ESST Neutral false alarm rate (%)	42.56 (20.46)	47.87 (23.13)	36.61 (16.04)	39.68 (19.10)
ESST Positive false alarm rate (%)	48.85 (18.09)	53.29 (20.95)	50.67 (15.83)	50.39 (17.56)
ESST Negative false alarm rate (%)	52.80 (17.19)	50.53 (21.47)	46.35 (16.26)	49.58 (17.35)
ESST P (Negative false alarm) (%)	21.07 (8.19)	20.29 (9.40)	23.68 (8.74)	24.03 (9.06)
UPPS-P Urgency	–	–	25.53 (8.00)	–
UPPS-P Positive Urgency	–	–	24.15 (9.22)	–
UPPS-P Negative Urgency	–	–	26.90 (7.98)	–
UPPS-P lack of Perseverance	–	–	19.98 (5.68)	–
UPPS-P lack of Premeditation	–	–	20.35 (4.79)	–
UPPS-P Sensation-seeking	–	–	31.21 (7.45)	–

ESST: emotional stop-signal task.

P (Negative false alarm): probability of negative commission error given a stop-signal, regardless of stimulus valence, an alternative metric of negative emotional response inhibition (in addition to *Negative false alarm rate*).

participants (4.0% of available sample) from descriptive analyses (Table 1), six participants (9.8% of available sample) from longitudinal reliability analyses (Hypothesis A; see Tables 1–3), and two participants (0.9% of available sample) from cross-sectional validity analyses (Hypothesis B; see, Tables 1, 2 and 4), producing final sample sizes of $n=432$ (descriptive), $n=55$ (reliability), and $n=221$ (validity), respectively.

We subsequently calculated descriptive statistics to evaluate sample characteristics (Table 1) and baseline ESST performance (Table 2) in participants who did and did not return for follow-up assessments, as well as the subgroup who also completed the UPPS-P. To examine test–retest reliability (Hypothesis A), we then conducted correlational analyses evaluating associations between baseline and follow-up ESST measures (see Figure 2 and Table 3). Finally, we evaluated task validity (Hypothesis B) using correlations of ESST and UPPS-P variables (Figure 2 and Table 4); results of *post hoc* analyses evaluating the potential effects of data collection site (see Supplementary materials) indicated comparable performance on major ERI metrics across samples included in tests of Hypothesis B.

Results

Test–retest reliability

For Hypothesis A, we examined the test–retest stability of ESST ERI metrics outlined above via linear correlations. Observed relationships between ESST performance metrics measured at two time points (at least 1 month apart) were generally consistent with predictions (see Table 3). Specifically, we found moderate test–retest correlations between baseline and follow-up ERI indicators: (1) SSRT, $r(55)=0.37$, $p=0.006$, 95% confidence interval (CI)=[0.11, 0.58], (2) Positive false alarm rate, $r(55)=0.46$,

$p<0.001$, 95% CI=[0.22, 0.65], (3a) Negative false alarm rate, $r(55)=0.37$, $p=0.006$, 95% CI=[0.12, 0.58], and (3b) P (Negative false alarm), $r(55)=0.30$, $p=0.025$, 95% CI=[0.04, 0.53], in addition to associations among test–retest indicators unrelated to ERI, that is, Accuracy, $r(55)=0.60$, $p<0.001$, 95% CI=[0.40, 0.75], Negativity bias, $r(55)=0.38$, $p=0.005$, 95% CI=[0.12, 0.58], and No-signal reaction time, $r(55)=0.36$, $p=0.007$, 95% CI=[0.11, 0.57].

As predicted, nearly all ERI variables were moderately inter-correlated between baseline and follow-up measurements (correlational effect sizes: 0.28–0.46), with a few exceptions (see Table 3 and Supplementary materials). Specifically, slower SSRT during the first ESST was associated with worse ERI across administrations and valence categories – relationships that were generally of greater magnitude than the correlation between baseline and follow-up SSRT, $r(55)=0.37$, $p=0.006$, 95% CI=[0.11, 0.58]. These results support the possibility that SSRT in the ESST may reflect ‘global’ or domain-general ERI capacities. ERI parameters were also highly intercorrelated when measured concurrently at each assessment, with medium-to-large effect sizes ranging from 0.30 to 0.88 (see Table 3). In general, results collectively suggest substantial shared variance among ESST ERI parameters – valence notwithstanding – perhaps implicating a common latent ERI factor that shows considerable stability over at least 2 months (and up to 7 months).

There were several unexpected but noteworthy associations between baseline and follow-up ESST performance parameters, in addition to some surprising *lack of* associations. Please refer to Supplementary materials for an overview of these findings, which should be considered preliminary since they were not formally tested as a priori relationships of interest. We consequently did not apply statistical correction for multiple testing and focused supplementary analyses on

Table 3. Pearson's correlations assessing ESST test-retest reliability ($n=55$).

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Accuracy T1	-												
2. Accuracy T2	0.60***	-											
3. Negativity bias T1	0.49***	0.33*	-										
4. Negativity bias T2	0.25	0.34*	0.38**	-									
5. NSRT (ms) T1	-0.28*	-0.09	-0.12	-0.19	-								
6. NSRT (ms) T2	-0.27*	-0.33*	-0.13	-0.48***	0.36**	-							
7. SSRT (ms) T1	-0.02	0.09	-0.04	0.16	-0.09	-0.17	-						
8. SSRT (ms) T2	-0.25	-0.11	-0.22	0.07	-0.19	-0.09	0.37**	-					
9. Positive false alarm rate T1	-0.03	-0.04	-0.15	0.06	-0.18	-0.13	0.58***	0.34*	-				
10. Positive false alarm rate T2	-0.07	0.13	-0.26	0.15	-0.22	-0.40**	0.44***	0.79***	0.46***	-			
11. Negative false alarm rate T1	0.24	0.21	0.10	0.29*	-0.39**	-0.28*	0.65***	0.26	0.41**	0.38**	-		
12. Negative false alarm rate T2	0.06	0.16	0.01	0.21	-0.14	-0.12	0.41**	0.75***	0.31*	0.65***	0.37**	-	
13. P (Negative false alarm) T1	0.23	0.20	0.27*	0.29*	-0.32*	-0.29*	0.61***	0.19	0.31*	0.28*	0.85***	0.30*	-
14. P (Negative false alarm) T2	0.07	0.22	0.18	0.29*	-0.13	-0.12	0.32*	0.68***	0.20	0.55***	0.31*	0.88***	0.30*

NSRT: no-signal reaction time; SSRT: stop-signal reaction time.

Correlations between ESST variables of interest measured at baseline (T1) and follow-up (T2) are highlighted in boldfaced font (Hypothesis A); we report uncorrected p -values associated with these test-retest correlations, as they did not differ substantively from significance values corrected for false discovery rate.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 4. Pearson's correlations assessing validity of baseline ERI metrics against the UPPS-P ($n=221$).

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. ESST SSRT (ms)	–								
2. ESST Positive false alarm rate (%)	0.26***	–							
3. ESST Negative false alarm rate (%)	0.27***	0.02	–						
4. ESST P (Negative false alarm) (%)	0.31***	–0.01	0.55***	–					
5. UPPS-P Urgency	0.08	–0.06	0.11	0.31***	–				
6. UPPS-P Positive Urgency	0.10	–0.06	0.05	0.30***	0.94***	–			
7. UPPS-P Negative Urgency	0.05	–0.04	0.17*	0.27***	0.92***	0.73***	–		
8. UPPS-P lack of Perseverance	0.02	–0.05	0.03	0.09	0.42***	0.32***	0.47***	–	
9. UPPS-P lack of Premeditation	0.01	–0.01	–0.04	–0.07	0.46***	0.40***	0.46***	0.36***	–
10. UPPS-P Sensation-seeking	–0.04	0.04	0.03	0.08	0.18**	0.26***	0.07	–0.10	0.17*

ERI: emotional response inhibition; ESST: emotional stop-signal task; SSRT: stop-signal reaction time.

Correlations between valence-specific ERI variables and Negative/Positive Urgency (all measured concurrently) are highlighted in boldfaced font (Hypothesis B).

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

observed links between ERI metrics and other indicators of task performance.

Convergent and discriminant validity

We used Pearson's correlations to evaluate Hypothesis B: that negative and positive ERI metrics derived from the ESST would each be specifically associated with Negative and Positive Urgency scores from the UPPS-P (see Table 4). Consistent with our hypothesis, Negative Urgency was associated with both negative ERI parameters: $r(221)=0.17$, $p=0.013$, 95% CI=[0.04, 0.29] for Negative false alarm rate and $r(221)=0.27$, $p < 0.001$, 95% CI=[0.14, 0.38] for P (Negative false alarm). P (Negative false alarm) was comparably related to Positive Urgency, $r(221)=0.30$, $p < 0.001$, 95% CI=[0.17, 0.42] and overall Urgency (i.e. the mean of Positive and Negative Urgency scores), $r(221)=0.31$, $p < 0.001$, 95% CI=[0.18, 0.42], whereas Negative false alarm rate was unrelated to the other UPPS-P scales. Contrary to prediction, Positive false alarm rate did not correlate with Positive Urgency nor with any other UPPS-P scales. We similarly found no significant associations between SSRT and UPPS-P scores.

Discussion

This study aimed to evaluate the reliability and validity of the ESST, a behavioural task designed to index neurocognitive processes associated with Urgency, that is, ERI, a core aspect of affective inhibitory control. Findings provide preliminary evidence supporting this task's psychometric characteristics. Specifically, we observed moderate correlations between ESST scores derived from baseline and follow-up assessments up to several months later, indicating test–retest reliability concordant with – and in some cases, superior to – other behavioural tasks (Cyders and Coskunpinar, 2011, 2012; Enkavi et al., 2019; Sharma et al., 2014). We also found a modest valence-specific association of our primary negative ERI parameter, Negative false alarm rate, with self-reported Negative (but not Positive) Urgency, suggesting convergent and divergent validity. An alternative indicator of negative ERI, P (Negative false alarm rate),

had even larger magnitude associations with Urgency variables; however, these relationships were valence-general. The ESST's divergent validity was nonetheless supported by the specificity of associations between Urgency and ERI parameters, which were unrelated to any other impulsive traits measured by the UPPS-P. Moreover, we observed worse negative ERI in hospitalised inpatient participants relative to community/student samples (see Supplementary materials), providing additional support for the construct validity of this proposed behavioural marker of neuropsychiatric vulnerability. Finally, inpatients hospitalised for suicidal behaviors showed greater stability of ERI parameters over time, suggesting possible ERI improvement among psychiatric inpatients *without* suicidal behaviors (see Figure 2c) – although our longitudinal sample was insufficiently powered to formally test possible moderating effects of psychopathology (or treatment) on test–retest reliability.

Participants were generally quite accurate in discriminating between negative and positive stimuli on the ESST (see Table 2), with Accuracy/error rates and reaction time metrics comparable to similar late-stage ERI tasks (e.g. Camfield et al., 2018). No-signal reaction time and SSRT were somewhat slower, yet also similar to those derived from the traditional stop-signal task (e.g. Soreni et al., 2009). Unsurprisingly, participants demonstrated worse ERI when evaluating stimuli with more obvious or intense valence, given higher false alarm rates to Negative and Positive IAPS categories compared with Neutral images. This finding is consistent with some prior studies (Camfield et al., 2018; Kalanthroff et al., 2013; Verbruggen and De Houwer, 2007 but see also Littman and Takács, 2017), suggesting an association between emotional valence intensity and inhibitory demand.

The subgroup of participants who completed the ESST twice showed relatively similar performance during both administrations, despite the considerable length of time between assessments (approximately 3 months) and the fact that most of these individuals were hospitalised when they completed the first ESST, but not the second. While the test–retest reliability of widely used behavioural tasks is not well-documented, studies with repeated response inhibition assessment generally involve much shorter intervals between task administrations (e.g. weeks as opposed to months) and typically focus either exclusively on healthy participants (e.g. Wöstmann et al., 2013) or specific

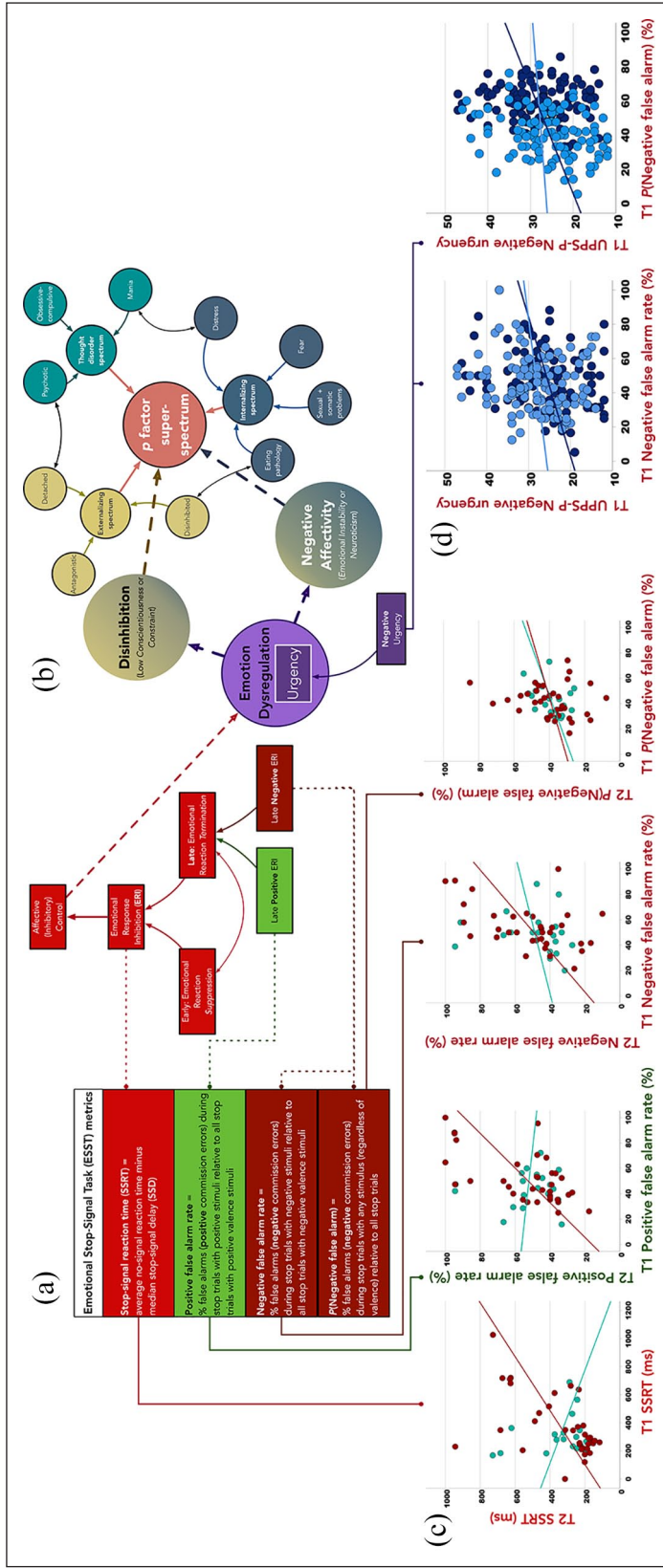


Figure 2. (a) Description of primary emotional response inhibition (ERI) metrics derived from the Emotional Stop-Signal Task (ESST). (b) Conceptual model linking neurocognitive processes underlying emotion dysregulation (i.e. affective control, which includes ERI), related personality constructs (i.e. Urgency, Disinhibition, and Negative Affectivity), and transdiagnostic latent psychopathology risk (i.e. the *p* factor). Solid lines represent the theoretical relationships between measurable indicators and latent factors, dashed lines reflect the proposed causal associations, and dotted lines tie the ESST variables to the neurocognitive constructs they reflect. (c) Visualisation of correlations among main ERI variables ($n=55$) reported in Table 3, with colour-coding of participant data corresponding to patients' history of suicidal behaviors (SBs): SB- = teal (lighter shade) and SB+ = red (darker shade). (d) Visualisation of correlations between negative ERI metrics and Negative Urgency ($n=221$) reported in Table 4, with colour-coding of participant data corresponding to ESST variant (and study site): original ESST (with NSSI stimuli) collected at 'University B' = lighter blue and revised ESST (with IAPS stimuli only) collected at 'University A' = darker blue.

psychiatric populations, oftentimes youth with attention deficit hyperactivity disorder (ADHD) (e.g. Kuntsi et al., 2005; Soreni et al., 2009). Notably, the present report is based on data aggregated from multiple studies, none of which were explicitly designed to evaluate the ESST; therefore, the modest strength of observed test–retest correlations – despite error introduced by variation in administration setting, sample characteristics, study protocols, and between-subjects factors known to modulate cognitive abilities (e.g. caffeine intake, time of day, and hormone levels), which are typically controlled in psychometric research – collectively support the ecological validity of this task.

Several questions remain regarding the interpretation of ESST metrics, particularly SSRT. Shared variance among this and other ERI parameters (across valence) suggests that SSRT in this task may tap domain-general late-stage ERI. Given this possibility, we would expect an association between SSRT and valence-general Urgency scores – which we did not find. Indeed, the lack of observed correlations between SSRT and UPPS-P variables suggests that it may capture a distinct process that does not substantively contribute to dispositional impulsivity. Additional research directly comparing the ESST with non-affective stop-signal tasks is needed to address this inconsistency. Some reviewers have additionally questioned the reliability and validity of the commission error rates we used as primary ERI metrics relative to the more widely used summary index of SSRT. We maintain that valence-specific false alarm rates in the ESST are meaningful markers of ERI, given several converging lines of evidence. Indicators of negative ERI dysfunction, including Negative false alarm rate, have demonstrated especially consistent associations with Urgency and psychopathology (here and in prior research) that do not generalise to adjacent constructs less strongly tied to emotion dysregulation, for example, lack of Premeditation and Sensation-seeking. Moreover, recent work has shown prospective influence of negative ERI deficits on NSSI (Burke et al., 2021) and suicidal behaviors (Allen et al., 2021).

We failed to find support for the hypothesised link between Positive false alarm rate and Positive Urgency. This finding is perhaps unsurprising, since the ESST was designed to measure neurocognitive processes relevant to Negative Urgency specifically in NSSI (Allen and Hooley, 2015, 2019). Potential explanations chiefly involve characteristics of the Positive IAPS stimuli, which may have been too mild, dated, or insufficiently relevant to participants to elicit strong positive affective reactions. Increased negative affect due to task demands and/or spillover effects from IAPS stimuli perceived as ‘unpleasant’ may also account for this result. Such questions motivate research focused on positive ERI, potentially examining modified ESST variants with more intense, contemporary, and/or personalised stimuli. Direct comparison of the current ESST with versions that use a valence-specific block design is also warranted to evaluate potential spillover effects between stimulus categories. Researchers also could assess participants’ mood throughout the task to help address these outstanding issues.

We acknowledge several limitations in this study. Most fundamentally, integrating multiple data sets from independent research groups resulted in a substantial amount of incomplete or missing data. For example, cross-sectional analyses evaluating relationships between ESST performance and urgency were necessarily restricted to participants drawn from community and student samples, as psychiatric inpatients did not complete the

UPPS-P; notably, *post hoc* analyses confirmed that community/student participants from different study sites performed similarly on major ERI metrics (see Supplementary materials). Relatedly, we analyzed different versions of the ESST together, although we confirmed the equivalence of our primary variables of interest between task variants, with the exception of Negative false alarm rate (see Supplementary materials). However, supplemental *post hoc* analyses suggested that this effect was fully attributable to sample type, that is, participants who completed the original version were exclusively drawn from non-clinical populations, whereas hospitalised psychiatric inpatient participants comprised the majority of those who completed the revised ESST. This explanation aligns with proposed links between negative ERI deficits and psychopathology, suggesting that elevated psychiatric symptoms among participants who completed the most recent ESST variant account for the higher rates of negative commission errors in this task compared to the original. Analyses separated by task variant indicate the superiority of the revised ESST in capturing Urgency-related neurocognitive processes (Table S1); additional research using varied (e.g. disorder-specific) stimuli in well-characterised groups of participants from healthy and clinical populations is therefore needed.

While our overall sample was diverse in several respects (e.g. psychiatric history and sexual orientation) and generally representative of demographics in the geographical areas sampled, participants were still mostly college-educated, female-identified, relatively young, and White. However, *post hoc* analyses suggested that demographic characteristics had a few effects on ESST performance, with the exception of age (see Supplementary materials). Regardless, generalisability may be further limited by high base rates of psychopathology and self-injurious behaviors in these samples. Targeted recruitment of psychiatrically ‘healthy’ participants from underrepresented groups is therefore necessary to generate more accurate estimates of normative task performance. Our longitudinal analyses were limited by the modestly sized subsample and relatedly low statistical power, and we cannot rule out the possible influence of self-selection bias on participant retention. Additional *post hoc* analyses (see Supplementary materials) suggest that follow-up attrition was unrelated to participants’ baseline characteristics, however (Table S2).

We hope to continually improve this task’s design based on this and other research to identify the most stable and relevant behavioural indices for Urgency and psychopathology. For example, we are creating ESST versions that incorporate stimuli with a larger range of valence intensity, that are compatible with functional magnetic resonance imaging (MRI) and event-related potentials. These ESST versions will provide valence-specific SSRT estimates, requiring independent stop-signal delay tracking algorithms absent from the original task.

Together, results support the use of the ESST to index negative ERI, a relatively stable and valid marker of neurocognitive mechanisms contributing to self-reported Negative Urgency and related constructs, that is, potential manifestations of the statistical *p* factor. This interpretation is consistent with recent work indicating that impaired executive functioning, which includes response inhibition deficits, is both a transdiagnostic risk factor and a consequence of psychopathology (Romer and Pizzagalli, 2021). Previous neuropsychological research has indeed linked Negative Urgency most strongly to response inhibition, a core aspect of cognitive control, which is considered a primary

mechanism for all ‘cool’ executive functions (and higher-order mental operations; see Nigg, 2017). Neuropsychiatric dysfunction particularly implicates ‘hot’ executive functions; we accordingly describe a latent factor that is conceptually tied to hot executive functioning more specifically: *affective control*, a parallel construct to cognitive control, referring to inhibitory processing in emotionally and/or motivationally salient contexts (see Allen, 2021; Allen et al., 2019a). We propose that ERI is a central feature of affective control, representing an important neurocognitive substrate for Urgency and *p*-related constructs. Affective control may also include processes such as emotional interference inhibition (e.g. Allen and Hooley, 2017; Masland et al., 2015) and emotional working memory (e.g. Schweizer et al., 2013). A few behavioural tasks purportedly measure these neurocognitive operations, and the ESST evaluated here is one of only several similar tasks designed specifically to tap late-stage ERI. In sum, current findings support the possibility that impaired inhibitory control over negative emotional reactions and accompanying motor impulses, once initiated (i.e. late-stage negative ERI), may represent an objective behavioural marker of self-reported Negative Urgency. Negative ERI deficits captured via the ESST may thus reflect underlying neuropsychiatric vulnerability, in addition to representing novel targets for prevention and intervention.

Acknowledgements

The authors thank the study participants and research assistants who contributed to this project.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by funding from the National Institute of Mental Health (K.J.D.A., H.T.S., and M.F.A., R01MH108610, MPI: H.T.S., Miller, and Mower-Provost; H.T.S. and M.F.A., R01MH112674, PI: M.F.A.), Oberlin College (PI: K.J.D.A.), and the University of California, Berkeley Research Impact Initiative (BRII; K.J.D.A. and S.L.J.).

ORCID iDs

Kenneth J. D. Allen  <https://orcid.org/0000-0001-8727-1265>

M. McLean Sammon  <https://orcid.org/0000-0002-0197-5921>

Supplemental material

Supplemental material for this article is available online.

References

Allen KJD (2021) Suicide following traumatic brain injury: Pathogenesis and neurocognitive mechanisms. In: Motta RW (ed.) *Suicide*.

London: IntechOpen. Available at: <https://www.intechopen.com/chapters/77672>

- Allen KJD and Hooley JM (2015) Inhibitory control in people who self-injure: Evidence for impairment and enhancement. *Psychiatry Research* 225(3): 631–637.
- Allen KJD and Hooley JM (2017) Negative mood and interference control in nonsuicidal self-injury. *Comprehensive Psychiatry* 73: 35–42.
- Allen KJD and Hooley JM (2019) Negative emotional action termination (NEAT): Support for a cognitive mechanism underlying negative urgency in nonsuicidal self-injury. *Behavior Therapy* 50(5): 924–937.
- Allen KJD, Bozzay ML and Edenbaum ER (2019a) Neurocognition and suicide risk in adults. *Current Behavioral Neuroscience Reports* 6(4): 151–165.
- Allen KJD, Bozzay ML, Armev MF, et al. (2021) Childhood maltreatment, emotional response inhibition, and suicide in psychiatric inpatients. *Behavior Therapy* 52(6): 1529–1542.
- Allen KJD, Fox KR, Schatten HT, et al. (2019b) Frequency of nonsuicidal self-injury is associated with impulsive decision-making during criticism. *Psychiatry Research* 271: 68–75.
- Bari A and Robbins TW (2013) Inhibition and impulsivity: Behavioral and neural basis of response control. *Progress in Neurobiology* 108: 44–79.
- Berg JM, Latzman RD, Bliwise NG, et al. (2015) Parsing the heterogeneity of impulsivity: A meta-analytic review of the behavioral implications of the UPPS for psychopathology. *Psychological Assessment* 27(4): 1129–1147.
- Brugman S, Lobbstaël J, von Borries AKL, et al. (2016) Cognitive predictors of violent incidents in forensic psychiatric inpatients. *Psychiatry Research* 237: 229–237.
- Burke TA, Allen KJD, Carpenter RW, et al. (2021) Emotional response inhibition to self-harm stimuli interacts with momentary negative affect to predict nonsuicidal self-injury urges. *Behaviour Research and Therapy* 142: 103865.
- Byrne BM (2010) *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming* (Multivariate Applications Series). New York: Taylor & Francis.
- Camfield DA, Burton TK, De Blasio FM, et al. (2018) ERP components associated with an indirect emotional stop signal task in healthy and depressed participants. *International Journal of Psychophysiology* 124: 12–25.
- Capistrano CG, Bianco H and Kim P (2016) Poverty and internalizing symptoms: The indirect effect of middle childhood poverty on internalizing symptoms via an emotional response inhibition pathway. *Frontiers in Psychology* 7: 1242.
- Carver CS and Johnson SL (2018) Impulsive reactivity to emotion and vulnerability to psychopathology. *American Psychologist* 73(9): 1067–1078.
- Carver CS, Johnson SL and Timpano KR (2017) Toward a functional view of the *p* factor in psychopathology. *Clinical Psychological Science* 5(5): 880–889.
- Congdon E, Mumford JA, Cohen JR, et al. (2012) Measurement and reliability of response inhibition. *Frontiers in Psychology* 3(37): 1–10.
- Creswell K, Wright A, Flory J, et al. (2019) Multidimensional assessment of impulsivity-related measures in relation to externalizing behaviors. *Psychological Medicine* 49(10): 1678–1690.
- Cyders MA and 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(6): 965–982.
- Cyders MA and Coskunpinar A (2012) The relationship between self-report and lab task conceptualizations of impulsivity. *Journal of Research in Personality* 46(1): 121–124.
- Cyders MA and Smith GT (2008) Emotion-based dispositions to rash action: Positive and negative urgency. *Psychological Bulletin* 134(6): 807.

- Cyders MA, Smith GT, Spillane NS, et al. (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.
- Denny KG and Siemer M (2012) Trait aggression is related to anger-modulated deficits in response inhibition. *Journal of Research in Personality* 46(4): 450–454.
- Enkavi AZ, Eisenberg IW, Bissett PG, et al. (2019) Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences* 116(12): 5472–5477.
- Gay P, Rochat L, Billieux J, et al. (2008) Heterogeneous inhibition processes involved in different facets of self-reported impulsivity: Evidence from a community sample. *Acta Psychologica* 129(3): 332–339.
- Hjordt LV, Stenbæk DS, Madsen KS, et al. (2017) State-dependent alterations in inhibitory control and emotional face identification in seasonal affective disorder. *Journal of Abnormal Psychology* 126(3): 291–300.
- Hoptman MJ, Antonius D, Mauro CJ, et al. (2014) Cortical thinning, functional connectivity, and mood-related impulsivity in schizophrenia: Relationship to aggressive attitudes and behavior. *American Journal of Psychiatry* 171(9): 939–948.
- Iria C, Barbosa F and Paixão R (2012) The identification of negative emotions through a Go/No-Go task. *European Psychologist* 17(4): 291–299.
- JASP Team (2019) *JASP* (Version 0.11.1, Computer software). Amsterdam: University of Amsterdam.
- Johnson AJ and Tottenham N (2015) Regulatory skill as a resilience factor for adults with a history of foster care: A pilot study. *Developmental Psychobiology* 57(1): 1–16.
- Johnson SL, Tharp JA, Peckham AD, et al. (2016) Positive urgency is related to difficulty inhibiting prepotent responses. *Emotion* 16(5): 750–759.
- Kalanthroff E, Cohen N and Henik A (2013) Stop feeling: Inhibition of emotional interference following stop-signal trials. *Frontiers in Human Neuroscience* 7: 78.
- Kuntsi J, Rijdsdijk F, Ronald A, et al. (2005) Genetic influences on the stability of attention-deficit/hyperactivity disorder symptoms from early to middle childhood. *Biological Psychiatry* 57(6): 647–654.
- Lang PJ, Bradley MM and Cuthbert BN (2008) *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual* (Technical Report A-8). Gainesville, FL: University of Florida.
- Littman R and Takács Á (2017) Do all inhibitions act alike? A study of go/no-go and stop-signal paradigms. *PLoS ONE* 12(10): e0186774.
- Logan GD and Cowan WB (1984) On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review* 91(3): 295–327.
- Lynam DR, Smith GT, Whiteside SP, et al. (2006) *The UPPS-P: Assessing Five Personality Pathways to Impulsive Behavior* (Technical Report). West Lafayette, IN: Purdue University.
- Masland SR, Hooley JM, Tully LM, et al. (2015) Cognitive-processing biases in individuals high on perceived criticism. *Clinical Psychological Science* 3(1): 3–14.
- Muhtadie L, Zhou Q, Eisenberg N, et al. (2013) Predicting internalizing problems in Chinese children: The unique and interactive effects of parenting and child temperament. *Development and Psychopathology* 25(3): 653–667.
- Mullins-Sweatt SN, DeShong HL, Lengel GJ, et al. (2019) Disinhibition as a unifying construct in understanding how personality dispositions undergird psychopathology. *Journal of Research in Personality* 80(12): 55–61.
- Nigg JT (2017) Annual research review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry* 58(4): 361–383.
- Pacheco-Unguetti A, Acosta A, Lupiáñez J, et al. (2012) Response inhibition and attentional control in anxiety. *Quarterly Journal of Experimental Psychology* 65(4): 646–660.
- Romer AL and Pizzagalli DA (2021) Is executive dysfunction a risk marker or consequence of psychopathology? A test of executive function as a prospective predictor and outcome of general psychopathology in the adolescent brain cognitive development study. *Developmental Cognitive Neuroscience* 51: 100994.
- Schatten HT, Allen KJD and Armey MF (2019) Assessment of emotion dysregulation using ecological momentary assessment. In: Crowell SE and Beauchaine TP (eds) *The Oxford Handbook of Emotion Dysregulation*. Oxford: Oxford University Press, pp. 411–426.
- Schweizer S, Grahn J, Hampshire A, et al. (2013) Training the emotional brain: Improving affective control through emotional working memory training. *Journal of Neuroscience* 33(12): 5301–5311.
- Sebastian A, Pohl MF, Klöppel S, et al. (2013) Disentangling common and specific neural subprocesses of response inhibition. *Neuroimage* 64: 601–615.
- Settles RE, Fischer S, Cyders MA, et al. (2012) Negative urgency: A personality predictor of externalizing behavior characterized by neuroticism, low conscientiousness, and disagreeableness. *Journal of Abnormal Psychology* 121(1): 160.
- Sharma L, Markon KE and Clark LA (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.
- Smith GT, Fischer S, Cyders MA, et al. (2007) On the validity and utility of discriminating among impulsivity-like traits. *Assessment* 14(2): 155–170.
- Soreni N, Crosbie J, Ickowicz A, et al. (2009) Stop signal and conners’ continuous performance tasks: Test-retest reliability of two inhibition measures in ADHD children. *Journal of Attention Disorders* 13(2): 137–143.
- Tackett JL and Lahey BB (2017) Neuroticism. In: Widiger TA (ed.) *The Oxford Handbook of the Five Factor Model*. Oxford: Oxford University Press, pp. 39–56.
- Verbruggen F and De Houwer J (2007) Do emotional stimuli interfere with response inhibition? Evidence from the stop signal paradigm. *Cognition and Emotion* 21(2): 391–403.
- Verbruggen F and Logan GD (2008) Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences* 12(11): 418–424.
- Whiteside SP and Lynam DR (2001) The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences* 30(4): 669–689.
- Wöstmann NM, Aichert DS, Costa A, et al. (2013) Reliability and plasticity of response inhibition and interference control. *Brain and Cognition* 81(1): 82–94.
- Zapolski TC and Smith GT (2013) Comparison of parent versus child-report of child impulsivity traits and prediction of outcome variables. *Journal of Psychopathology and Behavioral Assessment* 35(3): 301–313.
- Zapolski TC, Cyders MA and Smith GT (2009) Positive urgency predicts illegal drug use and risky sexual behavior. *Psychology of Addictive Behaviors* 23(2): 348.