American Psychological Association
Psychological Assessment

ISSN: 1040-3590
2020, Vol. 32, No. 4, 336–347
http://dx.doi.org/10.1037/pas0000791

Factor Structure, Measurement and Structural Invariance, and External Validity of an Abbreviated Youth Version of the UPPS-P Impulsive Behavior Scale

Ashley L. Watts
University of Missouri

Gregory T. Smith
University of Kentucky

Deanna M. Barch
Washington University

Kenneth J. Sher
University of Missouri

The current study examines the measurement properties and validity of a novel, abbreviated youth version of the UPPS-P Impulsive Behavior Scale that was developed to maintain measurement consistency with the existing adult short form. Specifically, we examined this scale's (a) factor structure; (b) measurement and structural invariance across four demographic characteristics: gender, ethnicity, household income, and parental education; and (c) correlates using a subset of 4,521 preadolescent (9- and 10-year old) children (53% male) from the baseline wave of the Adolescent Brain and Cognitive Development (ABCD) study, a large, community-based sample. Our findings supported a correlated 5-factor model, as well as a hierarchical model that recaptured the covariation among these 5 lower-order factors in three higher-order factors. Both of these models are consistent with the commonly observed structure of the UPPS-P among adults. We established measurement invariance across all demographic characteristics. Finally, our UPPS-P scales evidenced good convergent and discriminant validity with a broad swath of theoretically relevant external criteria, including self- and parent-reported personality and psychopathology, as well as lab-based neurocognitive tasks. Our findings indicate that we can assess multidimensional impulsivity in children reliably and validly by means of self-report, allowing assessment of this critical domain at early stages of development. We hope that this measure will facilitate the study of impulsivity in large-scale samples to begin to understand the evolution and long-term consequences of impulsivity.

Public Significance Statement
Multidimensional impulsivity can be assessed in children by means of self-report. This measure can be used to study impulsivity in large-scale samples and trace its evolution across the life span.

Keywords: impulsivity, self-report, children and adolescents, validity

Supplemental materials: http://dx.doi.org/10.1037/pas0000791.supp

This article was published Online First December 16, 2019.
Ashley L. Watts, Department of Psychology, University of Missouri; Gregory T. Smith, Department of Psychology, University of Kentucky; Deanna M. Barch, Departments of Psychological and Brain Sciences, Psychiatry, and Radiology, Washington University; Kenneth J. Sher, Department of Psychology, University of Missouri.
Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9–10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health and additional federal partners under Awards U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, U24DA041147, U01DA041093, and U01DA041025. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/wp-content/uploads/2019/04/Consortium_Members.pdf. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in analysis or writing of this report. This article reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from http://dx.doi.org/10.15154/1412097.
Correspondence concerning this article should be addressed to Ashley L. Watts, Department of Psychology, University of Missouri, 200 South 7th Street, Columbia, MO 65201. E-mail: ashleylwatts@gmail.com

Impulsivity, or the tendency toward rash action, is a broad construct thought to manifest as a diverse array of dysfunctional behaviors, including various forms of psychopathology. Contemporary models of impulsivity construe it as multidimensional, comprising five separable but correlated factors: (a) lack of perseverance, the inability to sustain attention or motivation to complete a task; (b) lack of premeditation, the tendency to not plan ahead and behave without thinking; (c) sensation seeking, the inclination toward seeking out novel, thrilling experiences; (d) negative urgency, the tendency to act hastily when in an extreme negative mood state; and (e) positive urgency, the tendency to act hastily when in an extreme positive mood state (Cyders et al., 2007; Whiteside & Lynam, 2001). Notably, positive urgency was not included in original conceptions of this model but was added later (Cyders et al., 2007). These five factors can be organized hierarchically to include three higher-order factors: Sensation Seeking; Deficits in Conscientiousness, which comprises lack of premeditation and lack of perseverance; and Urgency, which comprises negative and positive urgency (e.g., Cyders & Smith, 2007). Considerable research has demonstrated that distinguishing among these factors has merit, inasmuch as they relate differentially with theoretically relevant external criteria, including risky behavior (e.g., substance use) and psychological dysfunction (Smith et al., 2007).

The UPPS-P Impulsive Behavior Scale (UPPS-P; Lynam, Smith, Cyders, Fischer, & Whiteside, 2007) was developed to assess these five factors. It was subsequently validated for use with child, adolescent, and adult samples, as well as across samples ranging in psychiatric severity. One important limitation of the UPPS-P is its length (59 items), which can preclude its inclusion in large-scale investigations that require efficient use of resources. The original authors of the UPPS created a short UPPS-P scale for adults (Lynam, 2013) but no such scale exists for youth. Others have created UPPS scales for children and adolescents (Zapolski, Stairs, Settles, Combs, & Smith, 2010) but this measure omits Positive Urgency (Cyders et al., 2007).

Study Aims

We examined the measurement properties and external validity of a novel, abbreviated youth version of the UPPS-P scale developed by the 2nd author (Smith) for use in the Adolescent Brain and Cognitive Development Study (ABCD; see Barch et al., 2018). To do so, we culled items from the long-form UPPS-P (Lynam et al., 2007), with the goal of harmonizing our scale with the existing abbreviated adult scale (Lynam, 2013) to the extent possible, and with the loftier goal of facilitating the assessment of impulsivity in large-scale, longitudinal data collection efforts across development. The present study's aims were as follows. First, we examined the factor structure of this new scale using a mix of confirmatory and exploratory factor analysis. Second, we used the candidate factor model to test for measurement invariance across gender, ethnicity, and two proxies for socioeconomic status, household income, and parental education. Third, we examined the candidate factor model's relations with an array of theoretically relevant external criteria assessed by means of self- and parent-reported questionnaire and interview-based assessments, as well as laboratory tasks. External criteria included behavioral inhibition and activation; dimensional indicators of internalizing, externalizing, mania, and prodromal schizophrenia; diagnostic indicators of mood and anxiety disorders; parental history of alcoholism; and lab tasks probing neurocognitive functioning.

Hypotheses

Regarding factor structure, we hypothesized that we would identify a correlated five-factor structure of the UPPS-P commensurate with those identified older samples and with more thorough item sets (e.g., Lynam et al., 2007). That is, we expected that this model would fit well in children and with a more limited number of items per UPPS-P factor that would minimize assessment burden. We were agnostic with regards to the extent of measurement invariance of the UPPS-P factors across demographic characteristics.

Regarding external validity, all hypotheses were informed by the existing child and adult literatures to the extent possible. First, we expected relatively high convergence ($r > .5$) between UPPS-P Negative Urgency and BIS Inhibition, given that both are thought to reflect, at least in part, tendencies towards avoidance and negative emotionality. We also expected high convergence among UPPS-P Sensation Seeking and BAS Drive, Fun Seeking, and Reward Responsiveness ($r > .5$), given that each are imbued with novelty seeking and related approach-oriented traits (Segarra, Poy, López, & Moltó, 2014; Whiteside & Lynam, 2001). Second, we expected all UPPS-P dimensions to exhibit moderate convergence ($r = .2$; Funder & Ozer, 2019) with externalizing, but also expected that the UPPS-P Lack of Perseverance and Lack of Premeditation scales would exhibit the most pronounced relations with externalizing (Segarra et al., 2014; Whiteside & Lynam, 2001). Third, we expected that UPPS-P scales, with the exception of Negative Urgency, would exhibit null to small relations with internalizing ($r = .00$ to $.10$). In contrast, given that Negative Urgency assesses negative emotionality and poor emotional control, we predict that it would manifest medium-sized relations with internalizing psychopathology ($r = .20$ to $.30$; e.g., Berg, Latzman, Bliwise, & Lilienfeld, 2015). Lastly, we expected small positive associations between UPPS-P scales and lab tasks ($r = .10$ to $.20$; Cyders & Coskunpinar, 2011) given meager convergence between self-reported and laboratory-based tasks more broadly (Block, 1977).

Method

Item Selection

As we noted earlier, the items included in our abbreviated UPPS-P-Youth Version were culled from the UPPS-P and chosen to maintain consistency with Lynam's (2013) existing adult short form to the extent possible. Alterations to this form were informed by analysis of G.T. Smith's large-scale, longitudinal sample of children ($N = 1906$; Gunn & Smith, 2010) at the approximate age of those in the ABCD baseline sample (5th grade). All items from Lynam's (2013) Negative Urgency, Lack of Perseverance, and Sensation Seeking scales were retained. The items comprising Lynam's Lack of Premeditation scale did not exhibit adequate internal consistency among children. Smith replaced two poorly functioning items from the Lynam adult short-form with two items from the full-length child form that had the highest corrected item-total correlations in his sample. Similar internal consistency issues arose for the Positive Urgency scale, so one item was replaced with another that had the highest corrected item-total correlation. All told, this strategy permitted reasonable harmonization across abbreviated youth and adult forms of the UPPS-P but included items that appear more developmentally appropriate for youth (see online supplemental materials for a more thorough description of the item selection process). The final measure comprised 20 items (see Figure 1 for all items), 4 for each UPPS-P dimension, rated on a 1 (agree strongly) to 4 (disagree strongly) scale. Cronbach's alphas were as follows: .74 (Lack of Premeditation), .69 (Lack of Perseverance), .63 (Negative Urgency), .50 (Sensation Seeking), and .78 (Positive Urgency). Mean corrected item-total correlations (MCITCs) were as follows: .53 (Lack of Premeditation), .48 (Lack of Perseverance), .41 (Negative Urgency), .30 (Sensation Seeking), and .58 (Positive Urgency).

Participants and Procedure

Participants were 4,521 9- and 10-year-olds from the United States ($M_{\text{age}} = 120$ months, $SD = 7$; 53% male) from the baseline sample of the ABCD Study (Data Release 1.1); data on the UPPS were missing for 3 of the 4,524 participants from the baseline sample. Fifty-nine percent identified as White, 20% Hispanic, 10% Black, 2% Asian, and 10% other. Combined household income was collapsed into 3 groups: less than 50,000 (23%), 50- to 100,000 (28%), and over 100,000 (41%) dollars per year (8% did not report). Parental education was collapsed into 5 groups: less than a high school diploma (4%), high school diploma or GED (7%), some college (25%), bachelor's degree (27%), and post graduate degree (37%). All parents provided written informed consent, and all children provided assent to a research protocol approved by the institutional review board at each data collection site (https://abcdstudy.org/study-sites/).

Measures

The necessary brevity of this article precludes a detailed description of this battery but we direct readers to the following papers, where the ABCD study battery is summarized more thoroughly (personality and mental health: Barch et al., 2018; neurocognitive: Luciana et al., 2018).

Child-reported external criteria. Youth completed several well-validated instruments assessing personality and psychopathology, each of which have been validated for use in youth samples, including abbreviated Behavioral Inhibition and Activation scales (BIS/BAS; Pagliaccio et al., 2016), the Kiddie Schedule

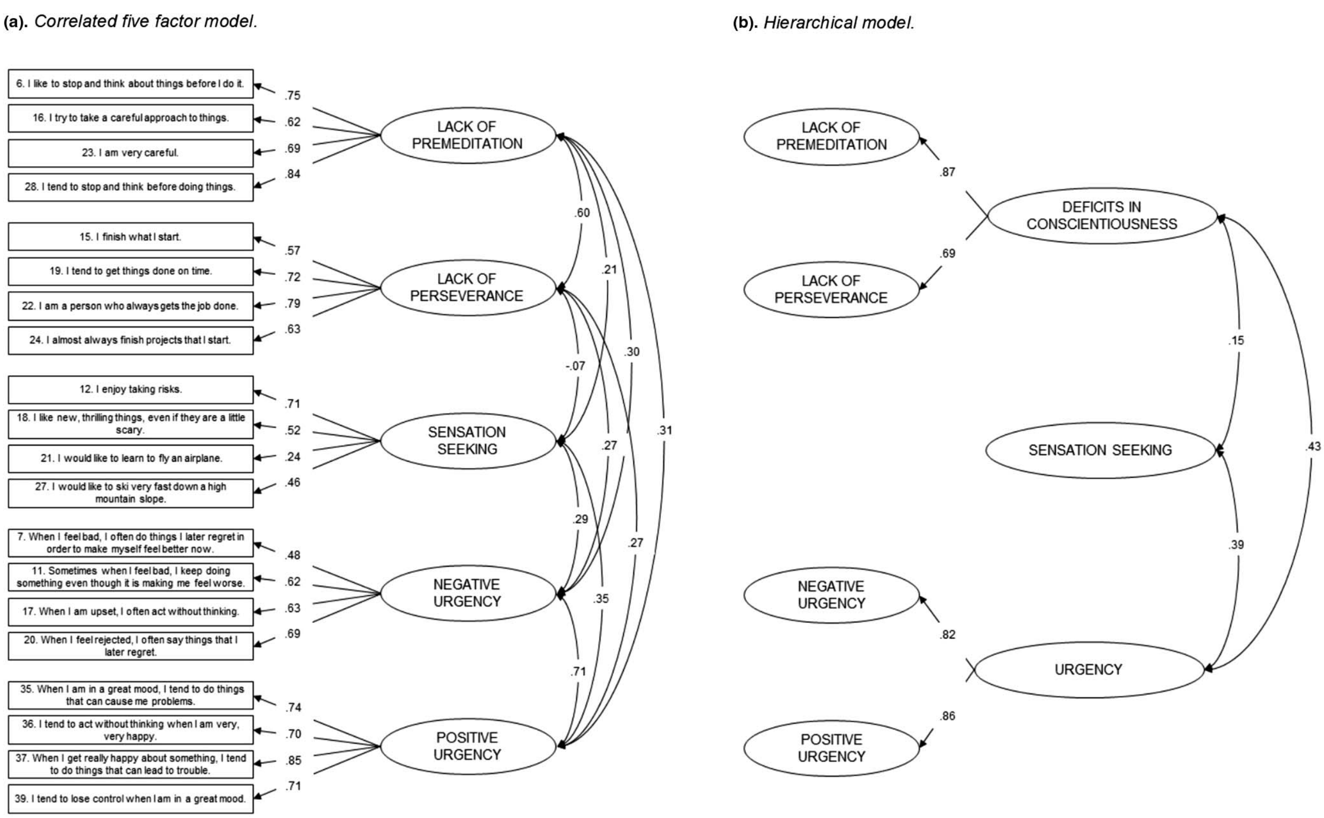

Figure 1. Lower-order and higher-order confirmatory models of the abbreviated UPPS-P-Youth Version scales. The hierarchical model included all UPPS-P items loading onto all lower-order UPPS-P factors but are excluded from the figure for simplicity. All items' loadings onto the lower-order factors were statistically equivalent to those presented in (a). Factor loadings and intercorrelations are standardized and significant at $p <$ .001.

for Affective Disorders and Schizophrenia for *DSM–5* (KSADS-5; Townsend et al., 2019), and the Prodromal Questionnaire-Brief Version (PQ-B; Loewy, Pearson, Vinogradov, Bearden, & Cannon, 2011).

BIS/BAS. The BIS/BAS comprises 20 items assessing two broad motivational systems, the behavioral inhibition (BIS) and behavioral activation (BAS) systems (Gray, 1982). BIS system is sensitive to signals of punishment and nonreward, novel stimuli, and innate fear stimuli, resulting in avoidance and negative emotionality, whereas BAS is sensitive to positive reinforcement and the absence of punishment, resulting in approach and positive emotionality. The BIS/BAS includes one subscale for BIS and three for BAS: Drive, Fun Seeking, and Reward Responsiveness (coefficient alphas ranged from .63 [Inhibition] to .77 [Reward Responsiveness]; MCITCs ranged from .33 [Inhibition] to .59 [Drive]). Items were rated on a 0 ("Not True") to 3 ("Very True") scale.

KSADS-5. The KSADS-5 (Kobak, Kratochvil, Stanger, & Kaufman, 2013) was administered to children with the help of a research assistant using a computer-based structured interview designed to assess current psychopathology in children and adolescents (Townsend et al., 2019). Youth reported on *DSM–5* diagnoses for current and past mood and anxiety disorders. Because base rates for individual diagnoses were generally extremely low (ranged from 0.2% [Child-reported Current Bipolar II] to 3.5% [Child-reported Past Bipolar I]), we collapsed current and past disorder diagnoses into lifetime disorder diagnoses, and further collapsed these diagnoses into dichotomous (present/absent) lifetime mood and lifetime anxiety disorder indicators.

PB-Q. The PB-Q comprises 21 items designed to assess symptoms associated with subclinical manifestations of psychosis. We report prodromal symptom counts as opposed to a severity score that takes into account symptoms and impairment (Karcher et al., 2018). Items are answered in a dichotomous response format (True/False). Total scores reflect number of endorsed symptoms (α = .86; MCITC = .45).

Parent-reported external criteria. Parents rated their child on several additional, well-validated instruments assessing psychopathology, including the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), an abbreviated scale assessing dimensional mania symptoms (Youngstrom, Frazier, Demeter, Calabrese, & Findling, 2008) adapted from the Parent General Behavior Inventory (P-GBI; Youngstrom, Findling, Danielson, & Calabrese, 2001), the KSADS-5, and the Family History Assessment Module Screener (FHAM-S; Rice et al., 1995).

CBCL. The CBCL includes 118 items that coalesce into two broad scores for Externalizing (includes subscales for Rulebreaking Behavior & Aggressive Behavior; .92, MCITC .40) and Internalizing (includes subscales for Social Withdrawal, Somatic Complains, and Anxiety/Depression) that comprise 2 and 3 subscales, respectively. Items are rated on a 0 (not true) to 2 (very true or often true) scale.

Dimensional mania. The dimensional mania scale comprises 10 mania items taken from the P-GBI, a longer inventory that comprises 73 items pertaining to mood (e.g., mania, depression, mixed depression); only these 10 mania items were used in the ABCD study. Items were summed into a single composite (α = .85; MCITC = .58). Items are rated on a 0 ("Never or Hardly Ever") to 3 ("Very Often or Almost Constantly") scale.

KSADS-5. Parents reported on their child's current and past mood, affect, externalizing (i.e., conduct disorder, oppositional defiant disorder, and attention-deficit/hyperactivity disorder), and eating disorders (i.e., anorexia nervosa, bulimia, binge eating disorder), as well as associated psychotic symptoms (i.e., delusions, hallucinations). Again due to low base rates, we collapsed current and past disorder diagnoses into lifetime disorder diagnoses, and then collapsed these diagnoses into lifetime mood, anxiety, externalizing, and eating disorder categories (present/absent), as well as an associated psychotic symptoms category (present/absent).

FHAM-S. We extracted a dichotomous indicator of parental history of alcoholism (no history/at least one parent with history) from the FHAM-S, in which parents reported on the presence/absence of symptoms associated with alcohol use disorder in both of the child's biological parents.

Laboratory-based neurocognitive functioning tasks. Youth completed selected tasks from the NIH Toolbox Cognitive battery (see http://www.nihtoolbox.org), which assess various aspects of neurocognitive functioning ranging from attention to executive functioning. Included in this battery were card sorting, list sorting, and pattern comparison tasks, as well as a flanker task. The card sorting task assesses cognitive flexibility, the list sorting task assesses capacity to sequence stimuli based on category and perceptual characteristics, the pattern comparison task assesses rapid visual processing, and the flanker task assesses the ability to inhibit attention to peripheral stimuli. Additionally, participants were administered a one-item cash choice task to probe delay of gratification, in which they chose between receiving $75 in 3 days or $115 in 3 months (see Luciana et al., 2018, for a more thorough description of each task).

Data Analysis

We conducted most all analyses except where specified using Mplus Version 7.2 (Muthén & Muthén, 2012). Design effects for neither site nor family necessitated use of the cluster option (i.e., they were less than 2); the average design effect across UPPS-P items was .72 for site (observations per site cluster: M = 215, SD = 101, range = 1–428) and 1.15 for family (observations per family cluster: M = 1.15, SD = .37, range = 1–5). Nevertheless, we clustered by site to ensure comparability between our analyses and those conducted elsewhere using the ABCD sample.

Factor analysis. Confirmatory and exploratory factor analyses (CFA, ESEM, EFA) were estimated using WLSMV.[1] WLSMV is a robust estimator that does not assume normally distributed variables and provides the best option for modeling categorical or ordered data (Brown, 2006). Factor models were specified by fixing latent factor means to 0 and variances to 1 as opposed to fixing the first factor loading to 1. We supplemented EFAs conducted in Mplus with several packages from R Version 3.5.1, namely the psych (Revelle, 2017), GPArotation (Bernaards & Jennrich, 2012), and nFactors (Raiche & Magis, 2010) packages

[1] Rates of missingness were low: .2% for UPPS-P items, and less than 2% for all external criteria, with the exception of parental history of alcoholism (3.6%) and the neurocognitive tasks (9.6%). Neither UPPS-P items and factors nor external criteria were univariate normal on the basis of the Shapiro-Wilk test. UPPS-P items were not multivariate normal on the basis of the Mardia test.

for Horn's parallel analysis (Horn, 1965) and the test of Very Simple Structure (Revelle & Rocklin, 1979). We examined UPPS-P factors' relations with external criteria while simultaneously estimating the UPPS-P structural model, obviating the need to save factor scores.

We considered the following goodness-of-fit statistics to evaluate model fit: the χ^2 test statistic and its associated degrees-of-freedom (*df*), the Comparative Fit Index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). Adequacy of model fit was based on the following guidelines suggested in the literature: CFI and TLI greater than 0.95 for reasonably good fit (Hu & Bentler, 1999); and RMSEA less than or equal to .08 for adequate fit and less than or equal to .05 for close fit (Browne & Cudeck, 1992; MacCallum, Browne, & Sugawara, 1996). Although the χ^2 values, their *p* values, and their degrees of freedom are presented for each model tested, we did not rely on them to determine adequacy of model fit because χ^2 significance tests (which are highly sensitive to N) would be virtually certain to be rejected given our large sample size (Brown, 2006). Instead, the fit of a single model was evaluated using the combination of CFI, TLI and RMSEA, as each individual fit index has its strengths and limitations, and no consensus exists regarding the use of a single fit index to evaluate the adequacy of model fit (e.g., Loehlin, 2004).

To adjudicate EFA and ESEM solutions, we considered the interpretability of the solution and its incremental utility over the CFA, in addition to model fit. We relied on these additional criteria in light of evidence that ESEM solutions often yield preferential fit over CFA solutions not due to increased validity but to increased parameterization and flexibility in the model (Herrmann & Pfister, 2013).

Measurement invariance. Conceptually, measurement invariance is important to establish to rule out the possibility that observed differences in a construct across groups are due to measurement. Put another way, tests of measurement invariance establish the extent to which a measure assesses the same latent trait across groups. We relied on the general procedure outlined by Widaman and Reise (1997) to test for measurement invariance (see also Byrne, Shavelson, & Muthén, 1989; Cheung & Rensvold, 2002). The baseline model of comparison was one that specified the same factor structure (i.e., number of factors and the pattern of factor-indicator relationships) across groups (equal form, or configural invariance). All models freely estimated UPPS-P factor means, variances, and covariances across groups as opposed to constraining them to equality to facilitate later tests of structural invariance; latent factor means and variances were constrained to 0 and 1, respectively, in a reference group.

Increasing equality constraints on multigroup confirmatory factor models that tested for levels of measurement invariance included (in increasing order): item (a) loadings and thresholds (strong or scalar invariance), which informs the extent to which each UPPS-P factor is assessed the same way across groups; and (b) residual variances (strict invariance), which informs the extent to which residual (error) variances of observed UPPS-P items are equivalent across groups. Due to the number of thresholds per item (i.e., three) and the fact that we estimated, as opposed to constrained, latent means and variances across groups, tests of measurement invariance must constrain item loadings and thresholds simultaneously rather than in sequence because models that equate one and not the other are not identified. Equating loadings and thresholds in tandem is debated in the literature, with most researchers arguing that they should be constrained in tandem given that they jointly contribute to item functioning (Lubke & Muthén, 2004; Sass, 2011).[2]

Of levels of measurement invariance, establishing strong invariance (e.g., equating item loadings and thresholds across groups) is arguably the only necessary step (e.g., Byrne et al., 1989). For instance, establishing equivalent item loadings and thresholds across groups (strong invariance) ensures that the UPPS-P factors are assessed the same across groups, thereby allowing for group comparisons. Further establishing that item residual variances can be equated across groups (strict invariance) indicates that differences in factor variances are not due to item variances.

We tested for scalar and strict invariance to provide as much information as possible, and to facilitate further structural invariance tests, which consider whether groups differ in terms of UPPS-P factor variances and/or means. In the presence of established measurement invariance, observed differences in UPPS-P factor variances, covariances, and means across groups are thought to reflect *substantive* differences in UPPS-P factors, as opposed to *artifactual* differences that arise from measurement. We were interested in the question of structural invariance given well-demonstrated gender differences in impulsivity (Cyders, 2013) and the general lack of attention to group mean differences in impulsivity across other demographic characteristics. Equality constraints tested sequentially included equating latent factor (4) variances, (5) covariances, and (6) means.

To determine whether a model was invariant across groups, we relied on change in two absolute model fit indices, CFI and RMSEA. Specifically, we primarily relied upon the ΔCFI criterion of −.002 (Meade, Johnson, & Braddy, 2008) but also report ΔRMSEA criterion of +.015 (Chen, 2007); preference was given to the Meade criterion, which is generally regarded as more stringent and is more widely accepted in the measurement invariance literature. We did not rely on $\Delta\chi^2$ on nested models because it is sensitive to factors unrelated to changes in invariance targeted constraints (e.g., sample size), namely, it is overly sensitive to small, unimportant deviations in fit in large samples (Chen, 2007; Cheung & Rensvold, 2002).

Results

Factor Structure

We used a mix of confirmatory and exploratory factor analysis to determine the structure of the UPPS-P. We began with confirmatory analyses that tested an a priori, five correlated factors structure, as well as a higher-order structure that reorganizes the same five factors into three higher-order ones, given that the factor structure of the UPPS-P has received considerable empirical val-

[2] Differences in the item loadings and thresholds across groups were relatively small in the configural model where both were freely estimated. Average differences in item loadings across groups were .04 for gender (average SE .02), .06 for ethnicity (.04), .05 for household income (.02), and .06 for parental education (.03). Average differences in item thresholds across groups were .08 for gender (.04), .17 for ethnicity (.06), .14 for household income (.04), and .14 for parental education (.05).

idation. To do so, we used both CFA and ESEM with target rotation, the latter of which is a quasi-confirmatory approach whereby (1) the number of factors and (2) hypothesized (target) loadings of items onto factors are specified, but (3) all possible cross-loadings of items onto factors are allowed (Marsh, Morin, Parker, & Kaur, 2014). This approach therefore tests whether the same general a priori UPPS-P structure can be retained but reveals potentially important deviations from simple structure not identified in a CFA solution (Marsh et al., 2014). Using these two approaches facilitates testing the same general UPPS-P structure across CFA and ESEM, but the former assumes simple structure whereas the latter does not.

We then followed-up with purely exploratory analyses using EFA. Although this sequencing (i.e., exploratory analyses *after* confirmatory ones) may appear counterintuitive, we did so to put forth an even riskier test (Popper, 1959) of our a priori confirmatory model by examining the extent to which it could be recovered with no specifications (or manipulations of the data) whatsoever. McDonald (1999) advocated for this approach for the following reasons: "By construction we may have a clear confirmatory restrictive hypothesis to fit to the data. If the fit is poor, a follow-up exploratory analysis might serve to diagnose the failure of the design. If the fit is good, a companion exploratory analysis will still check if we have missed anything" (p. 188). Indeed, we view this as a useful final step given that a number of nonequivalent although equally as well-fitting models can be observed in the same data (Tomarken & Waller, 2003). That is, in our view, this (confirmatory then exploratory) sequencing subjects our hypothesized model to a strong test by means of fully exploratory inquiry.

Confirmatory models. A correlated five-factor model corresponding to those observed in the existing youth and adult literatures (see Figure 1a and Tables S1 and S2) generally fit the data well (χ^2 = 941, *df* = 160, $p < .001$; RMSEA = .03; CFI = .95; TLI = .94). Each of the factors were significantly intercorrelated, with high correlations between (1) Lack of Premeditation and Lack of Perseverance, $r = .60$, $p < .001$) and (2 Negative Urgency and Positive Urgency, $r = .71$, $p < .001$. A model collapsing the Negative Urgency and Positive Urgency factors into one factor fit the data significantly more poorly ($\Delta\chi^2$ = 720, *df* = 1, $p < .001$), as did a model collapsing Lack of Perseverance and Lack of Premeditation factors ($\Delta\chi^2$ = 1053, *df* = 1, $p < .001$). Each factor was well-defined by its items, with loadings exceeding .40, with one exception ("I would like to learn to fly an airplane" = .24 on Sensation Seeking). It is possible that this latter item functions poorly as an indicator of sensation seeking. It is also possible that loading on Sensation Seeking increases with age as children become more experienced with air travel. We encourage subsequent investigations of this issue in future waves of ABCD data collection.

The ESEM with target rotation that specified five correlated (geomin rotated; Asparouhov & Muthén, 2009) factors with hypothesized (target) loadings in accord with the CFA model also fit the data well (χ^2 = 433, *df* = 100, $p < .001$; RMSEA = .03; CFI = .98; TLI = .96), and significantly better than the CFA ($\Delta\chi^2$ = 650, *df* = 60, $p < .001$). There was little evidence of appreciable cross-loadings (all cross-loadings < .22), each factor was well-defined by its items (loadings > .40, with the exception of "I would like to fly an airplane" = .37 on Sensation Seeking), and patterns of covariation among UPPS-P factors were comparable across the two models such that Negative and Positive Urgency ($r = .62$) and Lack of Premeditation and Perseverance ($r = .58$) were equally as highly correlated across the models (see Tables S2, S14). Taken together, the ESEM with target rotation, although associated with improved fit, generated a largely similar model to the CFA. This finding suggests that this preferential fit likely occurred due to increased parameterization and flexibility in the model (Herrmann & Pfister, 2013).

We next tested a hierarchical CFA model with three higher-order factors: Deficits in Conscientiousness, which comprised Lack of Perseverance and Premeditation subscales; Urgency, which comprised Negative and Positive Urgency subscales; and Sensation Seeking (see Figure 1). This model fit the data well (χ^2 = 1019, *df* = 163, $p < .001$; RMSEA = .03; CFI = .95; TLI = .94) but slightly worse than the five correlated factors CFA model ($\Delta\chi^2$ = 143, *df* = 3, $p < .001$). The loadings of the UPPS-P items onto the second-order factors were nearly identical to those in the correlated factors model. The second-order Deficits in Conscientiousness factor was strongly represented by Lack of Premeditation ($\lambda = .87$) and Lack of Perseverance ($\lambda = .69$), with slightly stronger representation by the former. The second-order Urgency factor was equally and strongly represented by Negative Urgency ($\lambda = .82$) and Positive Urgency ($\lambda = .86$). All higher-order factors (Deficits in Conscientiousness, Urgency, and Sensation Seeking) were significantly correlated, with the magnitudes of these effects ranging from small (.15: Deficits in Conscientiousness-Sensation Seeking) to moderate (.43: Deficits in Conscientiousness-Urgency).

Exploratory models. We next tested a series of EFAs allowing for one to 10 factors, all with goemin rotation and factor loadings freely estimated to allow maximum flexibility in the exploration of factor structure (Asparouhov & Muthén, 2009). Methods to determine the optimal number of factors yielded from these solutions (i.e., Horn's Parallel Analysis, Test of Very Simple Structure) conflicted at least somewhat (all results reported in online supplemental materials). For instance, Horn's parallel analysis (Horn, 1965) suggested 7 factors and 5 components, and the test of Very Simple Structure (VSS; Revelle & Rocklin, 1979) suggested 3 factors (factor loadings and intercorrelations for the 5-, 3-, 4-, and 7-factor solutions are presented in the online supplemental materials).

We believe the EFA supports a 5-factor solution for the following reasons. First, the 3-factor solution was consistent with the hierarchical (higher-order) structure of the UPPS-P, yielding Deficits in Conscientiousness, Urgency, and Sensation Seeking factors. Previous examinations have revealed that the constituent UPPS-P scales within Deficits in Conscientiousness and Urgency factors yield discriminating profiles of relations with theoretically relevant external criteria, however, which supports their being disentangled at a lower level of the hierarchy (Cyders & Smith, 2007, 2008; and see the "Differential external validity among UPPS-P factors" section, where we cannot consistently equate their relations with external criteria). Second, the 4-factor solution, which collapsed Negative and Positive Urgency and retained all other UPPS-P factors, was not raised as the optimal one by parallel analysis or the VSS test. Moreover, the Negative and Positive Urgency factors in the 5-factor EFA solution were highly correlated ($r = .61$) but not redundant with one another (see also Cyders

& Smith, 2008, and see the "Differential external validity" section, where we cannot consistently equate their relations with external criteria). Third, the 7-factor solution largely supported the 5-factor solution, inasmuch as it yielded five substantive factors, two of which were each decomposed into two method factors.[3]

Although we settled on the 5-factor EFA structure, it is worthwhile to note that each of the potential solutions indicated by the various criteria by which we adjudicate EFA solutions were consistent with various rungs of the UPPS-P hierarchy; indeed, the sequential extraction of factors highlighted the hierarchical unfolding of the UPPS-P model that is evidenced in confirmatory tests and is well-demonstrated in the adult literature (Cyders & Smith, 2007). The 5-factor EFA model fit the data well and slightly better than the 5-factor CFA solution ($\chi^2 = 432.85$, $df = 100$, $p < .001$; RMSEA = .03; CFI = .98; TLI = .96; SRMR = .02). As was the case in ESEM with targeted rotation model, there were few significant cross-loadings of items onto multiple factors (cross-loadings never exceeded .25). Given these findings, we elected to move forward with the five correlated factors CFA model for all invariance and external validity analyses.

Measurement Invariance

We next examined the extent to which various properties of the correlated five-factor CFA model could be constrained to equality across levels of four demographic characteristics (gender, ethnicity, household income, and parental education). We achieved strict invariance for gender, indicating that the abbreviated UPPS-P-Youth Version had equal form for males and females, and that we could further equate the item thresholds, loadings, and residual variances. We achieved strong invariance for all other demographic variables (ethnicity, household income, and parental education), indicating that we could equate item thresholds and loadings, but not residual variances (see Table 1); again, note that equating residual variances is not necessary for establishing measurement invariance.

Structural Invariance

There were significant differences in latent means and variances on UPPS-P factors on all demographic characteristics, indicating *substantive* as opposed to *artifactual* differences in latent means and variances on these demographic characteristics in UPPS-P factors in the presence of established measurement invariance (Table S7). Relative to females, males reported significantly higher levels of all UPPS-P factors (males were on average .28 *SD*s higher on UPPS-P dimensions compared with females), which is observed consistently in the literature (Cyders, 2013). Relative to White Americans, Black and Hispanic Americans reported significantly lower levels of UPPS-P Lack of Premeditation and Lack of Perseverance (.28 *SD*s on average). Moreover, again relative to White Americans, Hispanic Americans reported significantly lower levels of Sensation Seeking and Black Americans reported significantly higher levels of Positive Urgency (.22 and .31 *SD*s, respectively). Levels of UPPS-P Lack of Premeditation tended to increase as a function of parental education and combined household income (.56 and .31 *SD*s from lowest to highest groups), UPPS-P Sensation Seeking tended to increase as a function of parental education (.55 *SD*s from lowest to highest groups), and UPPS-P Positive Urgency tended to decrease as a function of parental education and household income (.30 and .36 *SD*s, respectively, from lowest to highest groups).

External Validity

Convergent and discriminant validity. As described above, age and gender were sometimes significantly associated with UPPS-P factors (*r*s ranged from −.06 to .09 for age and from −.03 to .21 for gender; see Table 2). As such, we conducted all external validity analyses covarying age and gender (but see Table S8 for all external validity analyses not adjusting for age and gender). Due to the number of tests conducted and our sample size, we focus on relations that were statistically significant at $p < .001$ to balance Type I and Type II error rates. We present zero-order correlations between UPPS-P factors and external criteria, as well as partial correlations derived from regressions in which single external criteria were regressed onto all UPPS-P factors, yielding unique relations for each UPPS-P factor. We focus our exposition of the results on zero-order correlations given the sometimes substantial correlations between UPPS-P factors, and the subsequent interpretational difficulties associated with partial correlations in the presence of multicollinearity (e.g., Lynam, Hoyle, & Newman, 2006). We refer to regression coefficients of .10, .20, .30, and .40 as small, medium, large, and very large, respectively (Funder & Ozer, 2019).

BIS/BAS. As expected, Lack of Premeditation manifested medium positive associations with BAS Fun Seeking ($r = .18$) and Lack of Perseverance manifested small negative associations with BAS Reward Responsiveness ($r = -.12$). Again as expected, Negative Urgency and Positive Urgency manifested comparable relations with BIS/BAS scales (see the "Differential external validity" section, for direct tests), with effects ranging from small to medium for BAS Reward Responsiveness and BAS Drive (*r*s ranged from .11 to .24), and from medium to very large for BIS Inhibition and BAS Fun Seeking (*r*s ranged from .24 to .40). That (1) Negative Urgency manifested relatively more pronounced relations with BIS Inhibition and that (2) Positive Urgency manifested relatively more pronounced relations with BAS Fun Seeking is consistent with these scales being imbued with neuroticism/negative emotionality and extraversion/positive emotionality, respectively (e.g., Segarra et al., 2014; Whiteside & Lynam, 2001). Finally, as expected, Sensation Seeking manifested only medium associations with BIS Inhibition ($r = .19$), and large to very large associations with BAS Reward Responsiveness, Drive, and Fun Seeking (*r*s ranged from .26 to .57). The especially pronounced associations between Sensation Seeking and BAS Fun Seeking were expected given that (1) BAS Fun Seeking appears to reflect an amalgam of reward responsivity, impulsivity, and positive emotionality (Smillie, Jackson, & Dalgleish, 2006) and (2) Sensa-

[3] The 7-factor solution yielded three substantive factors reflecting Negative Urgency, Lack of Perseverance, and Positive Urgency. Two factors comprised two Sensation Seeking items each. Two additional factors comprised two items each from Lack of Premeditation, one for positively worded items and one for negatively worded items. These four factors appeared to reflect method factors that coalesce into two substantive factors reflecting Sensation Seeking and Lack of Premeditation. Note also that parallel analysis tends to suggest overextracting factors in samples of this size (i.e., $N > 500$; Revelle & Rocklin, 1979).

Table 1
Measurement and Structural Invariance by Demographic Characteristics

Demographic characteristics	$\chi^2(df)$	CFI	TLI	RMSEA	WRMR	ΔCFI	ΔRMSEA
Gender							
Confirmatory models by group							
Females	588.94*** (160)	.958	.950	.035 (.032, .038)	2.096		
Males	578.73*** (160)	.944	.933	.033 (.030, .036)	2.151		
Measurement invariance							
Equal form	1272.26*** (360)	.949	.946	.033 (.032, .035)	3.245		
Equal item loadings and thresholds	1201.04*** (370)	.953	.952	.032 (.030, .034)	3.082	.004	−.001
Equal item residual variances	1188.29*** (390)	.955	.956	.030 (.028, .032)	3.144	.002	−.002
Structural invariance							
Equal latent factor variances	1189.29*** (395)	.955	.957	.030 (.028, .032)	3.251	.000	.000
Equal latent factor covariances	1151.66*** (405)	.958	.961	.029 (.027, .031)	3.402	.003	−.001
Equal latent factor means	**1220.64*** (410)**	**.954**	**.958**	**.030 (.028, .032)**	**3.626**	**−.004**	**.001**
Ethnicity							
Confirmatory models by group							
Black	235.69*** (160)	.933	.921	.033 (.023, .041)	1.221		
Hispanic	246.37*** (160)	.948	.938	.025 (.018, .031)	1.293		
Other	243.74*** (160)	.944	.934	.035 (.026, .043)	1.157		
White	582.96*** (160)	.959	.952	.032 (.029, .034)	2.182		
Measurement invariance							
Equal form	1516.30*** (760)	.950	.950	.030 (.027, .032)	3.618		
Equal item loadings and thresholds	1480.19*** (790)	.954	.956	.028 (.026, .030)	3.482	.004	−.002
Equal item residual variances	**1668.30*** (850)**	**.946**	**.951**	**.029 (.027, .031)**	**3.888**	**−.008**	**.001**
Structural invariance							
Equal latent factor variances	1634.14*** (865)	.949	.955	.028 (.026, .030)	3.922	.003	−.001
Equal latent factor covariances	1625.33*** (895)	.952	.959	.027 (.025, .029)	4.087	.003	−.001
Equal latent factor means	1673.21*** (910)	.949	.958	.027 (.025, .029)	4.291	−.003	.000
Household income							
Confirmatory models by group							
Low	357.00*** (160)	.932	.920	.034 (.030, .039)	1.609		
Medium	372.04*** (160)	.951	.941	.032 (.028, .037)	1.639		
High	496.32*** (160)	.962	.955	.034 (.030, .037)	1.995		
Measurement invariance							
Equal form	1382.31*** (560)	.950	.949	.033 (.030, .035)	3.392		
Equal item loadings and thresholds	1379.11*** (580)	.951	.952	.032 (.029, .034)	3.333	.001	−.001
Equal item residual variances	**1569.89*** (620)**	**.942**	**.947**	**.033 (.031, .035)**	**3.719**	**−.009**	**.001**
Structural invariance							
Equal latent factor variances	1525.26*** (630)	.945	.950	.032 (.030, .034)	3.743	.003	−.001
Equal latent factor covariances	1451.26*** (650)	.951	.957	.030 (.028, .032)	3.834	.006	−.002
Equal latent factor means	1497.84*** (660)	.949	.956	.030 (.028, .032)	4.033	−.002	.000
Parental education							
Confirmatory models by group							
Less than high school	183.63 (160)	.933	.920	.026 (.000, .041)	.910		
High school/GED	225.04*** (160)	.924	.910	.023 (.022, .041)	1.102		
Some college	347.96*** (160)	.960	.952	.030 (.026, .035)	1.602		
Bachelor's degree	432.41*** (160)	.950	.941	.035 (.031, .039)	1.750		
Graduate degree	364.51*** (160)	.962	.955	.032 (.028, .037)	1.561		
Measurement invariance							
Equal form	1813.48*** (960)	.950	.951	.031 (.029, .034)	3.864		
Equal item loadings and thresholds	1748.96*** (1000)	.956	.959	.029 (.027, .031)	3.692	.006	−.002
Equal item residual variances	**1978.40*** (1080)**	**.948**	**.954**	**.030 (.028, .032)**	**4.126**	**−.008**	**−.026**
Structural invariance							
Equal latent factor variances	1969.71*** (1100)	.949	.956	.030 (.027, .032)	4.178	.001	.027
Equal latent factor covariances	1974.28*** (1140)	.952	.960	.028 (.026, .031)	4.378	.003	−.002
Equal latent factor means	2044.68*** (1160)	.949	.958	.029 (.027, .031)	4.636	−.003	.001

Note. Best-fitting models are bolded.
*** $p < .001$.

tion Seeking and BAS Fun Seeking are thought to reflect overlapping latent constructs (Gray, 1982; see also Segarra et al., 2014).

Psychopathology. Each of the UPPS-P scales were generally equally associated with CBCL Internalizing and Externalizing, dimensional mania symptoms, and KSADS-5 lifetime mood disorder diagnoses, and parent-reported lifetime anxiety disorder diagnoses, and parental history of alcoholism. The exceptions were that Sensation Seeking was unrelated to CBCL Internalizing and

Table 2
External Validity of the Abbreviated UPPS-P-Youth Version

External criteria	Lack of premeditation		Lack of perseverance		Negative urgency		Sensation seeking		Positive urgency		Model R^2
	r (*SE*)	ß (*SE*)	*r* (*SE*)	ß (*SE*)	*r* (*SE*)	ß (*SE*)	*r* (*SE*)	ß (*SE*)	*r* (*SE*)	ß (*SE*)	
Age in months	.02 (.02)	*.09* (.03)	.00 (.02)	−.04 (.03)	−.02 (.02)	−.01 (.03)	.00 (.02)	.04 (.02)	−.03 (.02)	−.06 (.03)	.01
Gender	**.18** (.02)	*.11* (.03)	**.15** (.02)	**.05 (.04)**	**.16** (.02)	**.06 (.04)**	**.21** (.02)	**.14 (.03)**	**.15** (.02)	**−.03 (.04)**	**.05**
Child-reported BIS/BAS											
BIS Inhibition	.00 (.02)	**−.22** (.03)	.02 (.02)	.00 (.03)	**.24** (.02)	**.48** (.03)	**.19** (.02)	**−.08** (.02)	**.20** (.02)	−.07 (.03)	**.19**
BAS Reward Responsiveness	−*.07* (.02)	**−.14** (.03)	**−.12** (.02)	**−.14** (.03)	**.12** (.02)	*.09* (.03)	**.26** (.03)	**.17** (.03)	**.11** (.02)	*.09* (.03)	**.11**
BAS Drive	.07 (.02)	−.05 (.02)	.01 (.02)	**−.13** (.03)	**.24** (.02)	.07 (.03)	**.38** (.02)	**.11** (.02)	**.24** (.02)	**.20** (.03)	**.11**
BAS Fun Seeking	**.18** (.02)	−.04 (.02)	*.08* (.02)	−.08 (.03)	**.40** (.02)	.02 (.03)	**.57** (.02)	**.39** (.02)	**.36** (.02)	**.20** (.03)	**.26**
Parent-reported CBCL											
Internalizing	**.08** (.02)	−.03 (.03)	**.10** (.02)	*.07* (.03)	**.10** (.02)	**.15** (.03)	.01 (.03)	**−.10** (.02)	**.08** (.02)	−.03 (.03)	**.04**
Externalizing	**.21** (.02)	**.11** (.02)	**.24** (.02)	.03 (.03)	**.25** (.02)	**.17** (.03)	**.24** (.02)	−.03 (.02)	**.24** (.02)	.00 (.03)	**.07**
Total problems	**.19** (.02)	.04 (.03)	**.23** (.02)	**.11** (.03)	**.22** (.02)	**.14** (.03)	**.19** (.03)	−.04 (.02)	**.20** (.02)	.02 (.03)	**.06**
Child-reported PB-Q prodromal symptoms	**.15** (.02)	−.05 (.03)	**.14** (.02)	.05 (.04)	**.27** (.02)	**.19** (.03)	**.27** (.03)	.05 (.04)	**.25** (.02)	.07 (.04)	**.07**
Parent-reported dimensional mania symptoms	**.08** (.01)	−.01 (.01)	**.10** (.01)	.04 (.02)	**.14** (.02)	**.03** (.03)	**.10** (.03)	−.04 (.02)	**.13** (.02)	**.10** (.03)	**.02**
Child-reported KSADS-5											
Lifetime mood disorder	**.16** (.03)	.04 (.05)	**.14** (.03)	−.07 (.04)	**.28** (.03)	.07 (.05)	**.28** (.04)	−.06 (.04)	**.27** (.03)	**.25** (.05)	**.08**
Lifetime anxiety disorder	*.13* (.05)	−.12 (.09)	**.14** (.04)	.10 (.09)	**.24** (.04)	*.28* (.09)	.16 (.08)	−.06 (.09)	**.20** (.04)	.00 (.01)	*.08*
Parent-reported KSADS-5											
Lifetime mood disorder	*.10* (.03)	−.04 (.04)	**.12** (.03)	.08 (.04)	**.13** (.03)	.07 (.04)	.07 (.03)	−.04 (.03)	**.12** (.03)	.07 (.04)	.02
Lifetime anxiety disorder	.02 (.03)	−.09 (.06)	.06 (.03)	.15 (.06)	−.01 (.03)	.06 (.07)	**−.10** (.03)	−.08 (.05)	−.01 (.03)	−.04 (.06)	.03
Lifetime conduct disorder	*.11* (.03)	.08 (.04)	.10 (.04)	−.07 (.05)	**.18** (.04)	.05 (.07)	**.15** (.03)	.11 (.08)	**.16** (.04)	−.01 (.05)	.05
Lifetime oppositional defiant disorder	**.18** (.02)	**.16** (.03)	**.19** (.02)	−.03 (.03)	**.20** (.03)	*.12* (.04)	**.17** (.03)	.00 (.05)	**.18** (.03)	−.06 (.04)	**.06**
Lifetime attention-deficit/hyperactivity disorder	**.27** (.03)	.01 (.03)	**.33** (.04)	**.22** (.04)	**.24** (.02)	.03 (.05)	**.19** (.02)	.10 (.06)	**.25** (.02)	.02 (.03)	**.13**
Lifetime eating disorder	.09 (.07)	−.15 (.10)	.12 (.07)	.21 (.10)	.14 (.08)	.06 (.12)	.13 (.06)	.01 (.09)	.13 (.07)	.12 (.08)	.05
Lifetime psychotic symptoms	.10 (.05)	.02 (.06)	.11 (.06)	.01 (.09)	.14 (.05)	−.09 (.07)	.16 (.06)	*.20* (.06)	*.14* (.05)	.01 (.05)	.03
Parental history of alcoholism	*.09* (.03)	−.04 (.04)	*.08* (.03)	*.09* (.03)	*.08* (.03)	−.05 (.05)	*.10* (.04)	.08 (.04)	*.09* (.03)	.07 (.05)	.02
Neurocognitive functioning											
Cash choice	−.03 (.02)	.00 (.03)	−.03 (.02)	.00 (.03)	−.04 (.02)	−.03 (.04)	.00 (.03)	.06 (.03)	−.03 (.02)	−.03 (.04)	.01
Card sorting	−.02 (.02)	**.11** (.03)	−.05 (.02)	−*.08* (.02)	−*.05* (.02)	.05 (.03)	.02 (.01)	*.07* (.02)	−*.06* (.02)	**−.15** (.03)	**.03**
Flanker task	.02 (.02)	.05 (.03)	−.01 (.02)	−.01 (.03)	−.01 (.02)	.01 (.03)	*.06* (.02)	**.12** (.03)	−.02 (.02)	−*.10* (.03)	**.02**
List sorting	−.02 (.03)	**.11** (.03)	−.04 (.03)	−.03 (.02)	**−.08** (.02)	.06 (.03)	−.01 (.02)	**.12** (.03)	**−.09** (.02)	**−.23** (.04)	**.04**
Pattern comparison	−.03 (.02)	*.08* (.03)	−.05 (.02)	**−.11** (.03)	−.04 (.02)	**.11** (.03)	−.01 (.02)	.00 (.03)	−.05 (.02)	**−.14** (.03)	**.04**

Note. *SE* = standard error; BIS/BAS = Behavioral Inhibition/Activation Scales; CBCL = Child Behavior Checklist; PB-Q = Prodromal Questionnaire-Brief Version; KSADS = Kiddie Schedule for Affective Disorders and Schizophrenia for *DSM–5*. Bolded are $p < .001$ and italicized are $p < .01$. *r* values reflect correlations after covarying age and gender. ßs were entered as simultaneous predictors of external criteria. Age and gender were included in all models.

parent-reported mood disorder diagnoses (*r*s were .01 and .07, respectively). Effects were positive and small to large for internalizing-related indices (i.e., CBCL Internalizing, KSADS-5 lifetime mood disorder, dimensional mania symptoms), with effects being more pronounced for child-reported lifetime mood disorder diagnoses, and small for parental history of alcoholism (*r*s ranged from .07 to .28). Effects were positive and medium to large for externalizing-related indices (i.e., CBCL Externalizing, KSADS-5 lifetime oppositional defiant disorder diagnoses, lifetime attention-deficit/hyperactivity disorder diagnoses; *r*s ranged from .17 to .33). UPPS-P scales were essentially unrelated to parent-reported KSADS-5 lifetime eating disorder and lifetime psychotic symptoms (*r*s ranged from .09 to .16). That UPPS-P scales yielded relatively more pronounced relations with externalizing compared with internalizing is expected given that impulsivity is regarded as an especially salient transdiagnostic risk factor for externalizing psychopathology (e.g., Krueger, Markon, Patrick, & Iacono, 2005; Smith et al., 2007).

There was also evidence for discriminant relations with psychopathology indices. UPPS-P Sensation Seeking manifested small negative relations with lifetime anxiety disorder ($r = -.10$), whereas all other UPPS-P scales were either unrelated to or slightly positively associated lifetime anxiety disorder diagnoses. Negative Urgency, Positive Urgency, and Sensation Seeking manifested small to medium positive relations with KSADS-5 lifetime conduct disorder diagnoses (*r*s ranged from .15 to .18). Similarly, Negative Urgency, Positive Urgency, and Sensation Seeking manifested medium to large positive relations with PB-Q prodromal symptoms (*r*s ranged from .25 to .27), whereas the relations for Lack of Premeditation and Lack of Perseverance were small to medium in magnitude (*r*s were .15 and .14).

Lab tasks probing neurocognitive functioning. UPPS-P scales were generally unrelated to performance on neurocognitive lab tasks, with the following exceptions. Negative Urgency and Positive Urgency manifested small negative associations with list sorting (*r*s were −.08 and −.09).

Differential external validity of UPPS-P factors. Given the considerable overlap between Lack of Premeditation and Lack of Perseverance, on the one hand, and Negative Urgency and Positive Urgency, on the other, we conducted a series of subsidiary analyses in which we equated these pairs of scales' relations with external criteria. We then conducted chi-square differences tests comparing these models against those that freely estimated each factor's relations with external criteria. Differential relations with external criteria were indicated by a significant chi-squared difference test (see Tables S9 and S10).

Around half of these tests (48%) for Negative and Positive Urgency were significant (see Table S9; Cyders & Smith, 2008). Although their relations with external criteria were generally of similar magnitude, we could *not* equate Negative and Positive Urgency's relations with BIS Inhibition, BAS Drive and Fun Seeking, CBCL Internalizing and Externalizing, prodromal symptoms, child-reported mood disorder, parent-reported lifetime psychotic symptoms, and all neurocognitive tasks with the exception of the cash choice task; these findings indicate that Negative and Positive Urgency manifested significantly different relations with these indicators. Fifty-six percent of these tests for Lack of Premeditation and Lack of Perseverance were significant (see Table S10; Smith et al., 2007). We could *not* equate these scales' relations with BIS/BAS scales; CBCL Internalizing; parent-reported anxiety, conduct, and oppositional defiant disorders; and all neurocognitive tasks with the exception of the cash choice task; these findings again indicate that Lack of Perseverance and Lack of Premeditation manifested significantly different relations with these indicators. In general, there was little evidence that (1) Negative and Positive Urgency and (2) Lack of Premeditation and Lack of Perseverance related differentially with diagnostic indicators of psychopathology.

Discussion

Our abbreviated UPPS-P-Youth Version generally demonstrated excellent structural validity (Loevinger, 1957). The items' loadings on their respective factors were appreciable, and factor intercorrelations were of expected magnitudes. Through a systematic set of tests of alternative structural models using a mix of confirmatory and exploratory approaches, we determined that the factor structure of our UPPS-P measure mirrored that of its adult counterpart (Cyders & Smith, 2007) such that we detected five lower-order impulsivity factors that coalesced into three higher-order ones. A particular advantage of the hierarchical model is that accommodates the natural covariation among UPPS-P factors housed within Deficits in Conscientiousness and Urgency higher-order factors, thereby allowing researchers to examine correlates of both narrow and broad impulsivity dimensions.

Our findings indicate that our abbreviated UPPS-P scales are assessing the same five latent impulsivity factors, with the same basic measurement properties, in youth across varying levels of gender, ethnicity, and socioeconomic status. As such, it is not expected that UPPS-P factors would bear differential correlates across levels (or groups) of these demographic characteristics, meaning that UPPS-P factors' relations with external criteria should not be moderated by gender, ethnicity, or socioeconomic status. Given that we established adequate measurement invariance, group comparisons on these demographic characteristics are tenable and are thought to reflect *substantive* as opposed to *artifactual* differences in latent constructs across groups.

Our detection of significant differences in latent means and variances in UPPS-P factors across demographic groups replicates well-replicated research on mean level differences in impulsivity across gender (Cyders, 2013), and contributes to the literature by establishing relatively novel evidence for ethnicity- and socioeconomic-based mean-level differences in impulsivity. There are number of interpretations of these ethnicity- and socioeconomic-based mean level differences. Ultimately, our findings have the potential to inform investigations of the sources of impulsigenic traits, including but not limited to etiological covariation among ethnicity, socioeconomic status, and impulsivity.

The relations between our abbreviated UPPS-P-Youth Version factors and external criteria were generally similar to those observed in the adult literature (e.g., Smith et al., 2007). For instance, all UPPS-P factors were associated with increased internalizing and externalizing psychopathology, with, in general, effects being significantly more pronounced for externalizing (Krueger et al., 2005). Negative and Positive Urgency, in particular, tended to be more robustly associated with psychological maladjustment, including relations with potentially more severe forms of psychopathology (i.e., bipolar disorder, prodromal schizophrenia), as well as internalizing psychopathology and indicators imbued with trait negative emotionality (e.g., BIS; Cyders & Smith, 2008).

There was mixed evidence that Negative and Positive Urgency, on the one hand, and Lack of Premeditation and Perseverance, on the other, manifested distinct relations with external criteria (see also Berg et al., 2015). Nevertheless, we were well-powered to detect small differences in the magnitudes of UPPS-P factors' relations with external criteria, and as such were able to detect statistically significant albeit small differences. Moreover, these pairs of scales' (i.e., Negative Urgency-Positive Urgency, Lack of Perseverance-Lack of Premeditation) lack of differential validity is potentially consistent with their relations with external criteria being mediated through the broader Urgency and Deficits in Conscientiousness factors, although these tests are not necessarily probative of this possibility.

One less than ideal quality of our external validity findings was the general lack of overlap between child-reported UPPS-P factors and lab tasks probing impulsivity-related neurocognitive processes. Some authors have offered that this lack of overlap occurs because self-reports and lab tasks assess discrete aspects of impulsivity (Cyders & Coskunpinar, 2011). At the same time, low effect sizes between these two methods are often expected due to lack of method covariance (e.g., Block, 1977). Moreover, these effects, albeit very small in magnitude, are theoretically and empirically consistent with the existing literature (Cyders & Coskunpinar, 2011).

Conclusion

There is now considerable evidence that dispositional impulsivity can be assessed reliably and validly early in the life span, and that impulsivity goes onto predict dysfunction later in life. This emphasizes the need for early identification of impulsigenic traits, as well as the need to trace their trajectories and correlates across the life span. In the present study, we demonstrated that preadolescent children as young as nine years of age can self-report reliably and validly on multidimensional impulsivity, and the structure of these traits mirrored that of adults. These traits were concurrently associated with a host of theoretically relevant external criteria including other impulsigenic and reward sensitivity-related personality traits, psychopathology, and neurocognitive functioning. Ultimately, we hope that the development of this abbreviated UPPS-P scale for youth will facilitate the study of impulsigenic traits in large-scale, longitudinal data collection efforts. Such efforts could include population-based studies that track impulsivity across development. Doing so has the potential to facilitate progress in understanding the sources of impulsivity, its evolution, its antecedents and consequences, and its manifold implications for everyday functioning.

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles*. Burlington: University of Vermont.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16,* 397–438. http://dx.doi.org/10.1080/10705510903008204

Barch, D. M., Albaugh, M. D., Avenevoli, S., Chang, L., Clark, D. B., Glantz, M. D., . . . Sher, K. J. (2018). Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. *Developmental Cognitive Neuroscience, 32,* 55–66. http://dx.doi.org/10.1016/j.dcn.2017.10.010

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

Bernaards, C., & Jennrich, R. (2012). GPArotation: GPA factor rotation (R package version 3.1) [Computer software]. Los Angeles, CA: Author.

Block, J. (1977). Advancing the psychology of personality: Paradigmatic shift or improving the quality of research? In D. Magnusson & N. S. Endler (Eds.), *Personality at the cross-roads: Current issues in interactional psychology* (pp. 37–63). Hillsdale, NJ: Erlbaum.

Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21,* 230–258. http://dx.doi.org/10.1177/0049124192021002005

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105,* 456–466. http://dx.doi.org/10.1037/0033-2909.105.3.456

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14,* 464–504. http://dx.doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233–255. http://dx.doi.org/10.1207/S15328007SEM0902_5

Cyders, M. A. (2013). Impulsivity and the sexes: Measurement and structural invariance of the UPPS-P Impulsive Behavior Scale. *Assessment, 20,* 86–97. http://dx.doi.org/10.1177/1073191111428762

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. http://dx.doi.org/10.1016/j.cpr.2011.06.001

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

Cyders, M. A., & Smith, G. T. (2008). Emotion-based dispositions to rash action: Positive and negative urgency. *Psychological Bulletin, 134,* 807–828. http://dx.doi.org/10.1037/a0013341

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

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,* 156–168. http://dx.doi.org/10.1177/2515245919847202

Gray, J. A. (1982). *The neuropsychology of anxiety*. New York, NY: Oxford University Press.

Gunn, R. L., & Smith, G. T. (2010). Risk factors for elementary school drinking: Risk factors for elementary school drinking: Pubertal status, personality, and alcohol expectancies concurrently predict fifth grade alcohol consumption. *Psychology of Addictive Behaviors, 24,* 617–627. http://dx.doi.org/10.1037/a0020334

Herrmann, A., & Pfister, H. R. (2013). Simple measures and complex structures: Is it worth employing a more complex model of personality in Big Five Inventories? *Journal of Research in Personality, 47,* 599–608. http://dx.doi.org/10.1016/j.jrp.2013.05.004

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179–185. http://dx.doi.org/10.1007/BF02289447

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6,* 1–55. http://dx.doi.org/10.1080/10705519909540118

Karcher, N. R., Barch, D. M., Avenevoli, S., Savill, M., Huber, R. S., Simon, T. J., . . . Loewy, R. L. (2018). Assessment of the Prodromal Questionnaire-Brief Child Version for Measurement of Self-reported Psychoticlike Experiences in Childhood. *Journal of the American Medical Association Psychiatry, 75,* 853–861. http://dx.doi.org/10.1001/jamapsychiatry.2018.1334

Kobak, K. A., Kratochvil, C. J., Stanger, C., & Kaufman, J. (2013). *Computerized screening of comorbidity in adolescents with substance or psychiatric disorders*. La Jolaa, CA: Anxiety Disorders and Depression.

Krueger, R. F., Markon, K. E., Patrick, C. J., & Iacono, W. G. (2005). Externalizing psychopathology in adulthood: A dimensional-spectrum conceptualization and its implications for *DSM-V*. *Journal of Abnormal Psychology, 114,* 537–550. http://dx.doi.org/10.1037/0021-843X.114.4.537

Loehlin, J. C. (2004). *Latent variable models: An introduction to factor, path, and structural equation analysis*. Hillsdale, NJ: Erlbaum.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3,* 635–694. http://dx.doi.org/10.2466/pr0.1957.3.3.635

Loewy, R. L., Pearson, R., Vinogradov, S., Bearden, C. E., & Cannon, T. D. (2011). Psychosis risk screening with the Prodromal Questionnaire—Brief version (PQ-B). *Schizophrenia Research, 129,* 42–46. http://dx.doi.org/10.1016/j.schres.2011.03.029

Lubke, G. H., & Muthén, B. O. (2004). Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. *Structural Equation Modeling, 11,* 514–534. http://dx.doi.org/10.1207/s15328007sem1104_2

Luciana, M., Bjork, J. M., Nagel, B. J., Barch, D. M., Gonzalez, R., Nixon, S. J., & Banich, M. T. (2018). Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. *Developmental Cognitive Neuroscience, 32,* 67–79. http://dx.doi.org/10.1016/j.dcn.2018.02.006

Lynam, D. R. (2013). *Development of a short form of the UPPS-P Impulsive Behavior Scale*. Unpublished technical report. Retrieved from http://www.impulsivity.org/measurement/UPPS_P

Lynam, D. R., Hoyle, R. H., & Newman, J. P. (2006). The perils of partialling: Cautionary tales from aggression and psychopathy. *Assessment, 13,* 328–341. http://dx.doi.org/10.1177/1073191106290562

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. Retrieved from http://www.impulsivity.org/measurement/UPPS_P

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1,* 130–149.

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical*

Psychology, 10, 85–110. http://dx.doi.org/10.1146/annurev-clinpsy-032813-153700

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence, Erlbaum.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93,* 568–592. http://dx.doi.org/10.1037/0021-9010.93.3.568

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Pagliaccio, D., Luking, K. R., Anokhin, A. P., Gotlib, I. H., Hayden, E. P., Olino, T. M., . . . Barch, D. M. (2016). Revising the BIS/BAS Scale to study development: Measurement invariance and normative effects of age and sex from childhood through adulthood. *Psychological Assessment, 28,* 429–442. http://dx.doi.org/10.1037/pas0000186

Popper, K. (1959). *The logic of scientific discovery*. London, England: Hutchinson.

Raiche, G., & Magis, D. (2010). nFactors: An R package for parallel analysis and non graphical solutions to the Cattell scree test (R package version 2.3) [Computer software]. Montreal, Quebec, Canada: Author.

Revelle, W. (2017). psych: Procedures for personality and psychological research (R package version 1) [Computer software]. Evanston, IL: Author.

Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research, 14,* 403–414. http://dx.doi.org/10.1207/s15327906mbr1404_2

Rice, J. P., Reich, T., Bucholz, K. K., Neuman, R. J., Fishman, R., Rochberg, N., . . . Begleiter, H. (1995). Comparison of direct interview and family history diagnoses of alcohol dependence. *Alcoholism: Clinical and Experimental Research, 19,* 1018–1023. http://dx.doi.org/10.1111/j.1530-0277.1995.tb00983.x

Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment, 29,* 347–363. http://dx.doi.org/10.1177/0734282911406661

Segarra, P., Poy, R., López, R., & Moltó, J. (2014). Characterizing Carver and White's BIS/BAS subscales using the Five Factor Model of personality. *Personality and Individual Differences, 61–62,* 18–23. http://dx.doi.org/10.1016/j.paid.2013.12.027

Smillie, L. D., Jackson, C. J., & Dalgleish, L. I. (2006). Conceptual distinctions among Carver and White's (1994). BAS scales: A reward-reactivity versus trait impulsivity perspective. *Personality and Individual Differences, 40,* 1039–1050. http://dx.doi.org/10.1016/j.paid.2005.10.012

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,* 155–170. http://dx.doi.org/10.1177/1073191106295527

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology, 112,* 578–598. http://dx.doi.org/10.1037/0021-843X.112.4.578

Townsend, L., Kobak, K., Kearney, C., Milham, M., Andreotti, C., Escalera, J., . . . Kaufman, J. (2019). Development of three web-based computerized versions of the Kiddie Schedule for Affective Disorders and Schizophrenia Child Psychiatric Diagnostic Interview: Preliminary validity aata. *Journal of the American Academy of Child & Adolescent Psychiatry*. Advance online publication. http://dx.doi.org/10.1016/j.jaac.2019.05.009

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. http://dx.doi.org/10.1016/S0191-8869(00)00064-7

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, and S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Youngstrom, E. A., Findling, R. L., Danielson, C. K., & Calabrese, J. R. (2001). Discriminative validity of parent report of hypomanic and depressive symptoms on the General Behavior Inventory. *Psychological Assessment, 13,* 267–276. http://dx.doi.org/10.1037/1040-3590.13.2.267

Youngstrom, E. A., Frazier, T. W., Demeter, C., Calabrese, J. R., & Findling, R. L. (2008). Developing a 10-item mania scale from the Parent General Behavior Inventory for children and adolescents. *The Journal of Clinical Psychiatry, 69,* 831–839. http://dx.doi.org/10.4088/JCP.v69n0517

Zapolski, T. C., Stairs, A. M., Settles, R. F., Combs, J. L., & Smith, G. T. (2010). The measurement of dispositions to rash action in children. *Assessment, 17,* 116–125. http://dx.doi.org/10.1177/1073191109351372

Received February 21, 2019
Revision received October 30, 2019
Accepted November 1, 2019 ■