HHS Public Access

Author manuscript

Dev Cogn Neurosci. Author manuscript; available in PMC 2018 August 01.

Published in final edited form as:

Dev Cogn Neurosci. 2018 August ; 32: 67–79. doi:10.1016/j.dcn.2018.02.006.

Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery

M. Luciana[a], **J.M. Bjork**[b], **B.J. Nagel**[c], **D.M. Barch**[d], **R. Gonzalez**[e], **S.J. Nixon**[f], **M.T. Banich**[g]

M. Luciana: lucia003@umn.edu; J.M. Bjork: james.bjork@vcuhealth.org; B.J. Nagel: nagelb@ohsu.edu; D.M. Barch: dbarch@wustl.edu; R. Gonzalez: raul.gonzalezjr@flu.edu; S.J. Nixon: sjnixon@ufl.edu; M.T. Banich: marie.banich@colorado.edu

[a]University of Minnesota, Minneapolis, MN, United States

[b]Virginia Commonwealth University, United States

[c]Oregon Health Sciences University, United States

[d]Washington University, St. Louis, United States

[e]Florida International University, United States

[f]University of Florida, United States

[g]University of Colorado, Boulder, United States

Abstract

Adolescence is characterized by numerous social, hormonal and physical changes, as well as a marked increase in risk-taking behaviors. Dual systems models attribute adolescent risk-taking to tensions between developing capacities for cognitive control and motivational strivings, which may peak at this time. A comprehensive understanding of neurocognitive development during the adolescent period is necessary to permit the distinction between premorbid vulnerabilities and consequences of behaviors such as substance use. Thus, the prospective assessment of cognitive development is fundamental to the aims of the newly launched Adolescent Brain and Cognitive Development (ABCD) Consortium. This paper details the rationale for ABC'lected measures of neurocognition, presents preliminary descriptive data on an initial sample of 2299 participants, and provides a context for how this large-scale project can inform our understanding of adolescent neurodevelopment.

Keywords

Adolescence; Neurocognition; NIH; Toolbox; Longitudinal; Substance use

Correspondence to: M. Luciana, lucia003@umn.edu.

1. Introduction

This paper will introduce readers to the Adolescent Brain Cognitive Development (ABCD) project's baseline neurocognitive battery, including a description of each task, the rationale for task selection in light of the goals of ABCD, and preliminary descriptive data on participants studied through May 31, 2017. As reflected throughout the papers within this issue, ABCD is a recently-initiated national consortium that includes 21 data collection sites throughout the United States. The goal is to follow a large number of children, recruited at ages 9–10, for a ten-year span through adolescence and into young adulthood in order to relate neurodevelopment to environmental exposures such as substance use as it emerges within the sample. This paper presents the rationale and design of ABCD's baseline assessment of neurocognition, its coherence with NIH's objectives for the project, and its contributions to a larger scale biopsychosocial framework for typical development. Understanding the rationale of task selection is important given that data from the ABCD project will be available to the full scientific community. We hope further that this synopsis will inform not only subsequent analyses of the ABCD study neurocognition data by data requestors, but will also foster harmonization with the designs of newly launched studies that would deploy similar measures.

The baseline cognitive battery (with our expectation that these tasks will be re-administered in future testing waves) was configured based on three principles: 1) to be sensitive to salient developmental changes in the early-to-mid-adolescent period 2) to be sensitive to cognitive processes that may be impacted by future substance misuse, and 3) to test prevailing neurodevelopmental models of adolescent risk-taking. Adolescence is a unique period of the human lifespan, characterized by strivings toward independence, increased reliance on peers as sources of social support, hormonal and physical changes, and engagement in risk-taking behaviors, such as substance use. The observed rise in risk-taking has been attributed to a number of possible factors (see reviews by Ernst et al., 2006; Geier, 2013; Luciana et al., 2012; Romer, 2010). A particularly influential model attributes such behavior to a tension between developing capacities for cognitive control and the emergence of strong incentive strivings, which interact to bias behavior toward reward pursuit (Casey et al., 2008; Steinberg, 2010; Luciana and Collins, 2012). The hypothesis is that reward-based motivations are so strong that control mechanisms cannot be consistently mobilized to regulate them, a dynamic that may be most apparent during emotional arousal (see Cohen et al., 2016). This dual systems framework, and a similar model that incorporate developmental changes in sensitivity to threat (Ernst et al., 2006), was used as a heuristic to guide our selection of measures. Our task selection thus reflects the importance of understanding the development of cognitive control, individual differences in approach, behavioral inhibition and threat sensitivities (Bjork and Pardini, 2015), and the ways in which cognition and motivation interact to influence decision-making. A particularly challenging endeavor for us in recommending a measurement strategy was to determine when, in the course of the planned ten years of assessment, each process should be measured, which tasks best capture these processes, and how the assessments can be time-efficient and non-duplicative.

Cognitive control is multi-faceted and achieved through the recruitment of fine-grained subcomponents of attention, memory and other focused processes. Examinations of

cognitive control and its constituent processes, measured when participants are in non-affective states, generally reveal linear increases through adolescent development and increasing levels of consistency in performance (Humphrey and Dumontheil, 2016; Luna et al., 2010). More basic functions, such as verbal and nonverbal skills, overall intellect, and aspects of socially-directed attention and memory, also continue to improve (Choudhury et al., 2006). The slope of developmental change, as inferred through existing studies, suggests that de-contextualized cognitive control processes (e.g., working memory, behavioral flexibility in non-emotional settings) are changing most rapidly in early adolescence before leveling off in early adulthood (Luna et al., 2010). In contrast, the extant literature suggests that positive reward strivings peak in mid-adolescence before declining into adulthood and that the impact of these strivings on decision-making and inhibitory control are likely to be most prominent in the mid-to-late adolescent period. ABCD's measurement of individual differences in approach and threat motivations is described in the paper by Barch et al. within this issue. Accordingly, our view is that a comprehensive delineation of neurocognitive development that incorporates de-contextualized measures of cognitive control early in adolescence with the subsequent addition of measures of "hot" cognition by mid-adolescence is necessary to clarify cognitive contributions to the emergence of adolescent risk-taking and its trajectory over time.

Officials at the National Institutes of Health (NIH) were cognizant of this research and the hypothesized developmental trajectories in structuring the goals of ABCD. While the primary objective of the consortium, as articulated by the National Institutes of Health (NIH) is "to establish a national, multisite, longitudinal cohort study to prospectively examine the neurodevelopmental and behavioral effects of substance use from early adolescence (approximately age 9–10) through the period of risk for substance use and substance use disorders." (https://grants.nih.gov/grants/guide/rfa-files/RFA-DA-15-015.html), the assessment of cognitive development is fundamental to the aims of the project. The NIH mandated that ABCD performance sites should incorporate a neuropsychological battery that would permit the assessment of major domains that have been associated in prior studies with substance misuse. These included attention, information processing, learning and memory, cognitive control, motivation, emotional regulation, disinhibition, and risk taking. In planning for the ABCD project, NIH convened an expert panel in May of 2014. That group identified a number of fundamental considerations in devising an appropriate assessment scheme for the project (see http://addictionresearch.nih.gov/summary-expert-panel-meeting).

Given that the broad goal of the consortium to yield information about how substance misuse impacts the brain and behavior, it was important for tasks within the neurocognitive battery to be sensitive to such effects. Recent neuroimaging studies suggest that adolescent substance use is associated with deviations in regional brain volumes (Luciana et al., 2013), as well as the structural and functional connectivity between brain regions (Feldstein Ewing et al., 2014; Jacobus and Tapert, 2013). These reports indicate that adolescents who excessively use substances, such as alcohol and marijuana, show deviant patterns of brain activation when they perform cognitive tasks. Substance use is thought to negatively impact a range of cognitive functions, including memory, attention, and visuospatial abilities (Brown et al., 2000; Gould, 2010; Solowij et al., 2002) and may compromise general

abilities over time (Meier et al., 2012). These are foundational skills that are recruited by executive systems under conditions of high demand (Baddeley, 1986) and that contribute to meaningful life outcomes. For instance, low academic achievement, often thought to reflect aspects of basic cognitive ability, is both a risk factor for (Bryant and Zimmerman, 2002), and a potential consequence of (King et al., 2006), adolescent substance misuse. While substance use might impact life outcomes through influences on basic cognitive abilities, executive functions, social systems, and family cohesion, the nature of these interactions remains largely unknown.

Substance-induced deviations in adolescent' neurocognition have been most frequently considered in relation to executive functions such as working memory, cognitive flexibility, and inhibitory control. However, this literature has significant limitations. Many studies do not comprehensively assess a range of cognitive skills so it is challenging to place group differences into a broader behavioral context. Moreover, most findings have been gleaned from cross-sectional samples of adolescents who either engage in heavy substance use, including binge drinking, or are enrolled in substance abuse treatment. These groups are often compared to non-using control samples.

Cross-sectional research, while valuable in describing case-control differences, limits the extent to which etiology can be understood. Here we use executive function as an example. Extant models suggest that executive functions, particularly control over the impulse to engage in substance use, are compromised as the process of addiction emerges (Koob and Volkow, 2016). This idea is supported by the observation that individuals with substance use disorders experience executive function deficits as compared to demographically-matched non-users (Fernández-Serrano et al., 2011; Grant and Chamberlain, 2014; Roberts et al., 2016; Smith et al., 2014; Stephan et al., 2016). Despite clear evidence of adverse neurobehavioral consequences in adults who are sustained chronic users of drugs and alcohol, few if any longitudinal studies have convincingly demonstrated that executive functions decline from their baseline levels or veer off the normative developmental trajectory as the frequency or magnitude of substance use increases in the context of adolescent neurodevelopment.

Several reports reveal relatively diminished executive functions to be predictive of use onset, use frequency, and use in heavy amounts (Heitzeg et al., 2014; Peeters et al., 2015; Pentz et al., 2015; Riggs et al., 2012) or to be representative of genetic vulnerabilities toward externalizing behaviors (Malone et al., 2014). For instance, Peeters et al. (2015) assessed working memory and inhibitory control in a high risk sample of over 500 12–15 year-olds and found that relatively weak working memory, as measured by a self-ordered pointing task, predicted both the initiation of the first alcoholic drink and the first binge drinking episode, above and beyond the effect of response inhibition, as measured by the Stroop task. Khurana et al. (2013) annually assessed a community sample over 4 consecutive years using a battery that included working memory tests, delay discounting, and measures of sensation-seeking. Baseline weaknesses in working memory predicted not only concurrent alcohol use but also an increased frequency of use over the longitudinal testing interval. However, working memory was not further compromised by increased alcohol use over time. These findings are mirrored by more direct measures of neural networks that support

executive functions. Several longitudinal studies have reported that prior to alcohol use initiation, adolescents who progress into later alcohol use, versus those who abstain, fail to activate task-relevant frontoparietal regions important for cognitive control (Jones et al., 2016; Norman et al., 2011; Wetherill et al., 2013).

Thus, whether (and to what degree) deviations in brain structure and function and relative impairments in cognition constitute vulnerabilities predating substance use, occur as a consequence of substance use, and/or occur in combination remains unclear. Finally, whether representing predispositions or consequences, ascertaining the differential vulnerability of specific domains is scientifically and clinically relevant.

2. Methodological considerations

Developmentally, the recruited baseline sample is targeted to be early pubertal (ages 9–10), allowing for frequent testing into adulthood. While a younger age of enrollment might have been preferred to capture larger numbers of pre-pubertal participants (Demerath et al., 2004; Stang et al., 2013), enrollment at ages 9–10 allows for a comprehensive pre-substance-use baseline to be obtained. Epidemiological studies suggest that a relatively small proportion of children in the United States begin to use alcohol and other substances prior to the age of 13 (Centers for Disease Control and Prevention, 2016).

Measures were selected to be neuroscientifically informed, psycho-metrically sound, and sensitive to substance use outcomes. To meet these objectives, we relied on the literature and our collective experiences as investigators to identify key tasks that could be performed reliably between ages 9 to adulthood. The battery had to be amenable to longitudinal assessment with minimal practice effects. This is important, because in the absence of a cohort-sequential design, there is the potential for age-related longitudinal performance improvements to be conflated with practice given the homogeneous age range of the sample. If tasks are chosen that have substantial practice effects, then the extent of developmental gain could be over-estimated. Other studies of large cohorts (Sullivan et al., 2017) indicate that prior experience with the testing procedures as well as the level of baseline performance substantially contribute to temporal improvements, even across a one-year retesting interval and independent of developmental effects. Accordingly, while improvements over time are expected on the basis of age-related developmental gains, the potential for practice effects is strong. Thus, tasks were selected to be sensitive to developmental effects with minimal floor and ceiling effects. We did not adopt a strict numerical cut-off in evaluating mean changes with practice for specific measures. Rather, to minimize the potential for practice effects, we aimed to select measures that did not heavily rely on rule-based learning for performance success. One exception is the NIH Toolbox® Dimensional Change Card Sort task, which was included to allow composite scores to be calculated as described below. In addition, psychometric integrity of selected measures is crucial to reliable measurement, so selection criteria emphasized that measures must show adequate reliability and validity.

Finally, the repeated-measures nature of the ABCD study precluded the repeated use of tasks such as social-exclusion paradigms (Williams and Jarvis, 2006) or location-based learning

tasks (Bechara et al., 1994) that rely on some degree of deception or where knowledge of the task at one assessment point would influence performance the next time.

There are also a number of factors to be considered when working with an epidemiological sample in the context of multi-site assessment. Members of the May 2014 expert panel stressed that methods of evaluation should be standardized across all sites and suggested they have the following properties: first, the battery should be computerized to promote standardization and ease of multi-site administration. Computerized administration also permits automated scoring, reduces data processing demands and minimizes data entry requirements that may be impractical across sites. Second, measures should be utilized that require minimal training to reduce staff burden and to further decrease the likelihood of measurement error. Because of challenges in keeping children, especially those from high risk backgrounds, motivated through multiple hours of testing, tests should be selected that have minimum durations and that are perceived as engaging. A goal was to incorporate measures that would yield multiple performance metrics (e.g., measures of trial-by-trial accuracy; reaction time) and that would be sensitive to even subtle changes in behavioral capacities over time.

3. Workgroup structure and process

A workgroup focused on neurocognition was convened (see Appendix A for a list of members). Members were invited so that the team was representative of the 21 sites within the Consortium. Weekly teleconferences were held to discuss each assessment domain, using the original Request for Applications as a guide, and each member was invited to submit proposals for task inclusion. These ideas were fully collated and potential measures were discussed in relation to the above considerations. There was a remarkable degree of consistency across members of the workgroup in suggested measures, and we were in agreement regarding the need to assess both “hot” and “cold” cognition (Zelazo and Carlson, 2012) across the lifespan of the project. Given the epidemiology of substance use initiation, our group endorsed the Consortium’s view that the age period from 9 to 12 (spanning the first and second full assessment waves for ABCD) would likely represent a substance-free baseline period for most study participants.

Our initial suggestion for the Wave 1 assessment was three hours in duration and was ultimately shortened to the tasks represented in Table 1, all of which could be administered using an iPad. The full neurocognitive battery takes approximately 70 minutes to administer. For each task, we designated a workgroup expert who would advise the Consortium on the details of task administration procedures, research assistant training, trouble-shooting, and data management. Below, we describe the rationale and structure of each task in ABCD’s baseline neurocognitive battery, and we provide initial descriptive data for each task. We refer interested readers to other papers in this issue for a full description of the study sample, recruitment strategies, and other measures.

4. ABCD baseline neurocognitive battery (Table 1)

The battery begins with an assessment of visual acuity, in which both eyes are tested together (Snellen Chart: Snellen, 1862). Legal blindness is an exclusion for participation in ABCD given the highly visual nature of the subsequent neurocognitive assessment. While it could be argued that near-vision is most relevant to the subsequent task demands, an acceptable measure that could be readily standardized across sites could not be identified. Moreover, visual acuity problems detected by the Snellen Chart could indicate deviations in neural organization that would be important to ascertain.

A measure of handedness, the brief version of the Edinburgh Handedness Inventory (Oldfield, 1971; Veale, 2014), is administered. The brief form includes four items. The child is asked to report the hand that is typically used for writing, throwing, using a spoon, and using a toothbrush. A five-point scale is implemented for each item (always right hand, usually right hand, both hands equally, usually left hand, always left hand). A laterality quotient is derived across the four responses that represents the extent to which participants are right-handed, left-handed, or ambidextrous. There are no exclusion criteria related to lateral dominance. Preliminary data from the first 2299 participants (53.8% male; 59% Caucasian) indicate that 77.7% are right-handed, 8.9% are ambidextrous, and 13.5% are left-handed. These proportions are consistent with population estimates of non-right-handedness (Lezak et al., 2012).

Other relevant measures related to histories of head injury are included as part of ABCD's screening and demographic assessment.

4.1. NIH Toolbox®–Cognition battery

The NIH Toolbox® cognition measures (herein referred to as "the Toolbox") were developed as part of the NIH Blueprint for Neuroscience Research (see http://www.nihtoolbox.org for the history of their development) and consist of seven different tasks that cover episodic memory, executive function, attention, working memory, processing speed, and language abilities and are used to generate three composite scores (Bleck et al., 2013, Gershon et al., 2013b, Hodes et al., 2013). Using state-of-the-science methods, the battery was developed to be comprehensive, amenable for use in longitudinal studies, have relatively brief administration times, and to be psychometrically sound. Normative data were generated with a nationally representative sample of close to 5000 participants. As such, the Toolbox offered many attributes desirable to ABCD and allows harmonization and comparisons of cognitive performance with numerous other studies. Indeed, as of March 2017, there are over 100 studies that include the NIH Toolbox®. The Toolbox® can be administered to children as young as age 3 (for some tasks) through age 85. The tasks were selected based on a consensus building process, and developed and validated using assessment methodologies that included item response theory (IRT) and computerized adaptive testing (CAT) where appropriate and feasible. The original Toolbox® used a desktop web-based computer interface, but subsequently moved on to iPad based administration, which is what is being used in the ABCD study. The tasks do require some tester interaction, with the results input through a standard interface into the same database. The iPad version is available through iTunes, but requires documentation of expertise before activation to ensure

that the general public does not download and practice performance on the tasks in a way that might eventually bias results. The total time for administration of the NIH Toolbox® Cognitive battery is approximately 35 minutes. Although there are both English (Casaletto et al., 2015) and Spanish (Casaletto et al., 2016, Flores et al., 2017) language versions of the Toolbox® available, the ABCD study administers only the English version to youth, as proficiency in English for the youth (but not the parent) is a requisite for study participation. Below, we briefly describe each task, the composite scores, and the performance of the ABCD sample to date.

4.1.1. Language/vocabulary comprehension—The Toolbox Picture Vocabulary Task® (TPVT) (Gershon et al., 2014, Gershon et al., 2013a) is a variant on the Peabody Picture Vocabulary Test (PPTV) and measures language as well as verbal intellect. Children hear audio files of words and are shown four pictures in a square, one of which depicts the concept, idea or object referenced by the auditorily presented words. The child is asked to touch the picture that matches the word. The task uses CAT to ensure appropriate item difficulty in an efficient format. In validation testing with children and adolescents, the TPVT showed good test-retest reliability (ICC = 0.81), expected age related effects, and strong convergent validity with the Peabody Picture Vocabulary Test (Gershon et al., 2014, Gershon et al., 2013a, Mungas et al., 2014).

4.1.2. Language/reading decoding—The Toolbox Oral Reading Recognition Task® (TORRT) is a reading test that asks individuals to pronounce single letters or words presented in the middle of the iPad screen (Gershon et al., 2014, Gershon et al., 2013a) and measures exposure to language materials as well as the cognitive skills involved in reading. This task necessitates input by the tester, as the tester must learn the correct pronunciations through training materials and must score each letter/word pronunciation as correct or not correct. As with the TPVT, the TORRT uses CAT with items calibrated using IRT to ensure appropriate item difficulty in an efficient format. In validation testing with children and adolescents, it showed excellent test-retest reliability (ICC = 0.97), expected age related effects, and strong convergent validity with the Reading Subtest of the Wide Range Achievement Test (Gershon et al., 2014, Gershon et al., 2013a, Mungas et al., 2014).

4.1.3. Processing speed—The Toolbox Pattern Comparison Processing Speed Test® (TPCPST) (Carlozzi et al., 2015, Carlozzi et al., 2014, Carlozzi et al., 2013) was modeled on the Pattern Comparison Task developed by Salthouse (Salthouse et al., 1991) and is a measure of rapid visual processing. Children are shown two pictures and asked to determine by touch input whether the pictures are the same or not. The score is based on how many items they are able to complete correctly in a specific amount of time. In validation testing with children and adolescents, the TPCPST showed good test-retest reliability (ICC = 0.84), expected age related effects, and some convergent validity with the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) or Wechsler Intelligence Score for Children (WISC) Processing Speed Composite Tasks (Carlozzi et al., 2013). However, the TPCPST showed discriminant validity correlations with the PPVT that were as strong as the convergent validity correlations with the WPPSI/WISC Processing Speed composite in a child sample (Carlozzi et al., 2013), though not in an adult sample (Carlozzi et al., 2014).

4.1.4. Working memory—The Toolbox List Sorting Working Memory Test® (TLSWMT) is a variant of the letter-number sequencing test (Gold et al., 1997) that uses pictures rather than words or letters (Tulsky et al., 2014, Tulsky et al., 2013). The basic task is to use working memory to sequence task stimuli based on category membership and perceptual characteristics. Children are presented with a series of pictures of animals or foods of different sizes. Each picture is accompanied by the animal or food name presented auditorily by the iPad. The child is asked to repeat back to the experimenter the items that were presented, but in order from smallest to largest. The TLSWMT starts by using only a single category (i.e., animals). Children are presented with a two-item list, and if they get it correct, the next trial increases to three items, and so on until a maximum of seven is reached. Children have two opportunities (different trials) to provide a correct answer at each list length, and continue on to the next length if they get at least one of the items correct. All children then progress to a next phase where the trials interleave two different categories (i.e., animals and food), regardless of how they did on the single category lists. For the interleaved trials, the child is asked to first organize and repeat back the items for one category (i.e., animals) and then the other category (i.e., food). As with the single category phase, children have two opportunities to achieve a correct response at each list length, and continue on up to a maximum length of seven if they get at least one right at the previous level. The experimenter scores the child's response as correct or incorrect. In validation testing with children and adolescents, the TLSWMT showed good test-retest reliability (ICC = 0.86), expected age related effects, and reasonable convergent validity with the WISC-IV Letter-Number sequencing task (Tulsky et al., 2013). However, the TLSWMT also showed relatively strong correlations with the Peabody Picture Vocabulary Test (PPVT), the Delis-Kaplan Executive Function System (D-KEFS) Color-Word Interference Task and the Wisconsin Card Sort –64 Perseverative Errors (r's = 0.42 to 0.45), but these were not quite as high as its convergent correlation (r = 0.57). The correlation of the TLSWMT with the PPVT was lower in an adult sample (r = 0.24) (Tulsky et al., 2014).

4.1.5. Episodic memory—The Toolbox Picture Sequence Memory Test® (TPSMT) was modeled after memory tests asking children to imitate a sequence of actions with props developed by Bauer and colleagues. In the TPSMT, participants are presented with a series of fifteen pictures depicting activities or events that could occur in a particular setting (i.e., working on a farm) (Bauer et al., 2013, Dikmen et al., 2014)). They must reproduce the sequence as it was presented. The task is not speeded.

In validation testing with children and adolescents, the TPSMT showed moderate test-retest reliability (ICC = 0.76) and expected age related effects (Bauer et al., 2013). There were moderate correlations with other memory tests such as the Rey Auditory Verbal Learning Test and the Benton Visual Retention Test (r = ~.47), but the correlation with the Peabody Picture Vocabulary Test was even higher (r = 0.58) (Bauer et al., 2013). Again in adults, correlations of the TPSMT were higher with both the Rey (r = 0.64) and the Benton (r = 0.65), and much smaller with the PPVT (r = 0.04) (Dikmen et al., 2014). Because the TPSMT does not employ delayed recall or recognition trials, the ABCD workgroup thought it critical to include a more comprehensive episodic memory measure (i.e., the RAVLT, described below).

4.1.6. Executive function/attention/inhibition—The Toolbox Flanker Task® (TFT), a variant of the Eriksen Flanker task (Eriksen and Eriksen, 1974), was adapted from the Attention Network Task (Fan et al., 2002; Rueda et al., 2004). The task assesses the degree to which participants' responses are influenced by whether surround stimuli and the target are either congruent or incongruent. There are two types of trial blocks, a somewhat easier block with fish (not used within ABCD), and then a somewhat harder block with arrows. Younger children perform the task with the fish as stimuli, and older children (starting at age 8 years) are presented with the arrows. In both blocks, the four flanking stimuli (2 on the outer left and 2 on the outer right) are all facing the same way, either left or right. The middle fish or arrow is then either facing the same way (congruent trial) or a different way (incongruent trial). Children have to push a button to indicate whether the middle stimulus is facing left or right. Scoring is based both on speed and accuracy. In validation testing with children and adolescents, based on samples of 166–188 children (Zelazo et al., 2013), the TFT showed excellent test-retest reliability (ICC = 0.92), expected improvements in performance with increasing age, and some correlation with the D-KEFs Inhibition score (r = 0.34) and a stronger correlation with WPPSI Block design (r = 0.60), though with a relatively high correlation with the PPVT (r = 0.44) (Zelazo et al., 2013). Convergent and divergent validity was better in adults, with good correlations of the TFT with D-KEFs Inhibition (r = 0.52) and not with the PPVT (r = 0.06) (Zelazo et al., 2014).

4.1.7. Executive function/cognitive flexibility—The Toolbox Dimensional Change Card Sort Task® (TDCCS) was based on the work of Zelazo and colleagues (Zelazo, 2006) and measures cognitive flexibility. In this task, children are presented with objects at the bottom of the screen (i.e., white rabbit, green boat). The participant is asked to sort a third object presented in the middle of the screen by either color or shape to match one of the two objects on the bottom of the screen (Zelazo et al., 2014, Zelazo et al., 2013). They indicate their response by touch input. After initial practice, they first do a block of trials sorting on one of the dimensions, then they do a block of trials in which they are told to switch to the other dimension. Lastly, they do a block of trials where they alternate in pseudorandom order between sorting on shape versus color. The total score is based on both accuracy and reaction time. Once again, in validation testing with children and adolescents, the TDCCS showed excellent test-retest reliability (ICC = 0.92), expected age related effects, and strong convergent validity with the D-KEFs Inhibition score (r = 0.64), with a bit lower correlation with the PPVT (r = 0.55). Convergent validity was again better in adults, with good correlations of the TFT with D-KEFs Inhibition (r = 0.55) and not with the PPVT (r = 0.06) (Zelazo et al., 2014).

4.1.8. Measures derived from the Toolbox—Each of the Toolbox® tasks produces a number of scores, some of which are adjusted based on whether information such as age, gender and race/ethnicity are entered at the start of testing. All tasks provide raw scores, uncorrected standard scores, and age-corrected standard scores based on a normative sample of 2917 children and adolescents (Casaletto et al., 2015). Standard scores have a mean of 100 and a standard deviation of 15. One can also receive Fully Corrected T-Scores that take into account certain demographic characteristics such as gender, education, and race/ethnicity, which have a mean of 50 and a standard deviation of 10. In addition to

the scores for individual tasks, the Toolbox® also provides several composites, including a Total Score Composite, a Crystalized Intelligence Composite (TPVT and TORRT) and a Fluid Intelligence Composite (TPCPST, TLSWMT, TPSMT, TFT, TDCCS) (Akshoomoff et al., 2013). These composite scores also show good test–retest reliabilities in both children and adults (Akshoomoff et al., 2013, Heaton et al., 2014)) as well as validity in children (Akshoomoff et al., 2013; Heaton et al., 2014).

4.1.9. ABCD NIH Toolbox data—To confirm that the Toolbox® was operating as expected in our sample, we examined the data collected through May 31, 2017 (n = 2259 children with complete Toolbox® data). As shown in Table 2a, the means and standard deviations for the raw (not age corrected) scores for all of the individual tests and composite scores are as expected, meaning means somewhat below 100 for domains known to show developmental improvements past age 9 (e.g., vocabulary, reading, processing speed, executive control). Of note, the age corrected scores are undergoing revision by the NIH Toolbox® and may be available for the first public data release. As shown in Table 2a, the majority of the scores have small skewness values at or below +1 or −1. The slightly more negative values for the two executive function tasks (flanker and card sorting) suggesting only slightly longer tails for lower than higher scores. All but two of the kurtosis values are positive, suggesting close to normal distributions or somewhat more peaked distributions than normal. This is particularly true for TDCCS (card sorting), which had kurtosis values above 2, suggesting a somewhat higher density of scores around the mean. Table 3 presents the inter-correlations among the individual tests and composite scores. There are stronger correlations between TPVT and TORRT (both part of the crystalized composite) than with the other measures. However, the measures in the fluid composite are not all strongly intercorrelated, with the highest correlations among processing speed, the dimensional change card sort, and flanker.

Our expectation is that the Toolbox® measures will provide critical baseline assessments of executive functions, such as working memory, which may index vulnerabilities to substance misuse. The Toolbox® measures of general verbal ability, processing speed, and inhibitory control are similarly important for the modeling of typical development and impacts of substance use.

4.2. Inter-temporal reward choice

A key goal of the ABCD study is to discover brain mechanisms of psychosocial and behavioral traits in childhood known to confer risk of substance use disorder (SUD) and other psychiatric disorders. The most replicated finding in the behavioral economics of substance abuse is the exaggerated loss (relative to controls) in the subjective value of real or hypothetical rewards the longer the subject must wait to receive them (MacKillop et al., 2011). Importantly, this tendency is also found in children at risk for drug abuse and other poor outcomes (Casey et al., 2011). Seminal fMRI studies have implicated the dorsolateral prefrontal cortex (DLPFC) in decisions to defer gratification regarding real rewards (Kable and Glimcher, 2007; McClure et al., 2004) and at least one study has found that greater proportional gray matter in dorsolateral prefrontal cortex correlates with greater preference

for larger delayed rewards (Bjork et al., 2009). For these reasons, inter-temporal reward choice behavior was viewed as an informative metric for inclusion in ABCD.

Our initial plan was to include a delay discounting measure given strong associations between temporal discounting and substance misuse (MacKillop et al., 2011). However, there were concerns about task length, and the decision was made to defer this measure to later assessments. For brevity and to capture this construct at baseline, we elected to administer the one-item Cash Choice Task (Wulfert et al., 2002). The research assistant asks the child: "Let's pretend a kind person wanted to give you some money. Would you rather have $75 in three days or $115 in 3 months?" and records the child's response on the iPad. The child indicates one of these two options or a third "can't decide" option. The two monetary alternatives were recently found to have similar subjective value in 8th graders, where choice of the smaller-sooner amount correlated with questionnaire measures of real-world externalizing behavior (Warren Bickel, personal communication).

Of 2285 participants, 41% selected the smaller-sooner reward option, 57% selected the larger-later reward option, and 1.6% could not answer (Table 2b). Use of similar tasks has revealed that while selection of the immediate choice option appears to decline with age, there is developmental stability in temporal choice behavior as well as a significant association with multiple facets of externalizing behavior, including substance use by mid-adolescence (Anokhin et al., 2011; Sparks et al., 2014). Temporal choice behavior using a one-item assessment appears to be moderately to highly heritable (Anokhin et al., 2011; Sparks et al., 2014). We expect similar associations to be evident in the ABCD sample over time.

4.3. Rey Auditory Verbal Learning Test (RAVLT)

The Rey Auditory Verbal Learning Test (RAVLT) is a widely used test of auditory learning, memory, and recognition. Although variants of the test are thought to have been around since the early 1900s, English-language adaptations were first available in the middle of the 20th century (Taylor, 1959). The most commonly used versions allow one to ascertain multiple facets of memory, including short-term memory capacity, proactive and retroactive inhibition, retention, encoding versus retrieval, and subjective organization (Strauss et al., 2006). This test requires participates to listen to and recall a list of 15 unrelated words over five learning trials. Following initial learning of the list, a distractor list of 15 words is presented, and the participant is asked to recall as many words from this second list as he/she is able. Next, recall of the initially learned list is assessed. To assess longer-term retention of the list, following a 30-min delay, during which participants engage in other tasks, recall is again assessed, as well as recognition via a forced-choice (yes/no) list of words. Care is taken to assure that the intervening tasks are not verbal in nature. Given the ease of administration and normative data across childhood and adulthood (7–89 years: Strauss et al., 2006), this test has gained acceptance as a valuable measure of learning and memory across the lifespan. Alternate forms are available to facilitate longitudinal testing.

Across childhood and adolescence, the RAVLT has been shown to be sensitive to memory deficits associated with neurodevelopmental disorders (Vakil et al., 2012) and psychopathology (Gunther et al., 2004), as well as exogenous influences, such as cannabis

use (Solowij et al., 2011), highlighting its potential efficacy for assessing auditory learning and memory in the ABCD study. Internal reliability estimates for the total score are high (coefficient alpha ~0.90), and test-retest reliability is adequate (r-values ~0.60–.70 (van den Burg and Kingma, 1999).

A challenge for the project was to incorporate a variant of the task that could be reliably measured using computerized technology. Other list-learning tasks (e.g., the California Verbal Learning Test, CVLT) are commercially available in formats amenable to automated data collection and scoring (http://www.helloq.com/overview/the-q-interactive-library/cvlt_c.html); however, the CVLT was rejected for consideration due to the categorical nature of word presentation, which may render it more susceptible to practice-related effects. A search of the literature failed to reveal an acceptable computerized variant of the RAVLT. We worked with Pearson Assessments to create a customized automated version that built upon that group's Q-interactive automated testing platform (http://www.helloq.com/home.html). Within the automated version, administered with an iPad, the investigator reads each word aloud at a pace of one word per second. The pacing is facilitated by the task program. As the participant recalls the words, the experimenter rapidly checks off each utterance using a stylus and an on-screen list. Perseverations are also marked, and intrusions can be manually logged. At the conclusion of the trial list, the program automatically calculates the number of correctly recalled words for each trial, the number of perseverations, and intrusions. Raw scores are generated, synchronized to the Pearson clinical assessments server, and then downloaded through an automated script to the ABCD central data repository.

As indicated in Table 2c, the number of items recalled steadily increases between the first and the fifth learning trial in the ABCD sample to date. Approximately 69% of items are recalled at the immediate delay and 67% are recalled at the long (30-min) delay, reflecting loss of information after consolidation. As indicated in Table 3, the RAVLT maps onto aspects of both fluid and crystallized reasoning skills.

4.4. Matrix Reasoning

The automated version of the Matrix Reasoning subtest from the Wechsler Intelligence Test for Children-V (WISC-V: Wechsler, 2014) was selected for inclusion, because it is a well-validated and reliable measure of nonverbal reasoning (Wechsler, 2014), because it can be administered using automated technology (Q-interactive: Daniel et al., 2014), and because fluid reasoning appears to be compromised by externalizing behaviors, including substance abuse (Keyes et al., 2017). On each trial of the task, the examinee views a visuospatial array consisting of a series of stimuli. The series is incomplete, and the participant must select from a list of four alternatives to complete it. The task is administered using two iPads, one of which is controlled by the examiner to present each trial while the other is viewed by the child. The iPads are synchronized via Bluetooth connection. Standardized instructions are used and 32 possible trials are available. Testing ends when the participant fails three consecutive items. The raw score (total number correct) across completed trials is tabulated and translated, using a normative database, into a standard score. The normative standard score mean is 10.0 with a standard deviation of 3.0, corresponding to an IQ score

of 100 or in the average range. The Matrix Reasoning task can be completed in under ten minutes for most participants. It measures fluid reasoning, visual intelligence, part-whole spatial reasoning, perceptual organization, attention to visual detail, and sequencing. Matrix Reasoning strongly represents the Fluid Reasoning factor of the WISC-V and correlates (r = 0.67) with general ability (*g*). Split-half reliability is strong for children in the 9–10 year-old age range (r = 0.87), and retest stability is acceptable (r = 0.78). Given recent controversies regarding the association between substance misuse and IQ decline over time (Jackson et al., 2016; Meier et al., 2012), the Consortium recognized the importance of including measures of baseline fluid (Matrix Reasoning) and crystallized (NIH Toolbox Picture Vocabulary™) reasoning with the goal of re-assessing these skills over time.

As indicated in Table 2c, the Matrix Reasoning data collected to date reinforce the notion that the ABCD sample is representative of the general population in terms of general intellectual ability. The observed scaled score mean is 10.86 with a standard deviation of 3.1, and the scores are normally distributed. To date, the Matrix Reasoning scaled score is moderately positively correlated with the NIH Tool-box®measures of fluid and crystallized reasoning and with the RAVLT (Table 3).

4.5. Little Man Task (LMT)

As indicated above, in addition to providing a comprehensive assessment of development, another objective during battery development was to ensure that cognitive domains with particular vulnerability to alcohol/drug –related compromise were interrogated. The Little Man Task (LMT) was selected on the basis of both of these objectives. Developed by Acker and Acker (1982), the task engages visual-spatial processing, specifically mental rotation with varying degrees of difficulty. Importantly, it is not a memory test. The task involves the presentation of a rudimentary male figure holding a briefcase in one hand in the middle of the screen. The figure may appear in one of four positions; right side up vs. upside down and either facing the respondent or with his back to the respondent. The briefcase may be in either the right or left hand. Using one of two buttons, respondents indicate which hand is holding the briefcase. The association between the button and response is held constant across trials and participants, i.e., there is no interference between hand and button (left hand always associated with a button to the left of the participant and which is labeled “left”). Tests of face validity conducted during task development indicated its sensitivity to visual spatial compromise (e.g., mental rotation). Furthermore, although performance on the LMT has been correlated in prior studies with performance on the Block Design subtest of the WAIS, the tasks accounted for little shared variance (Acker and Acker, 1982). Thus, the task provides information regarding an important aspect of visual spatial function not assessed elsewhere in the battery. The LMT was favored over Block Design as a measure of visuospatial reasoning, because it is automated, reducing the burden on research staff, because it is child-friendly, and because it can be administered to all participants in a highly standardized fashion within a relatively brief period of time (e.g., 6–8 minutes).

Consistent with several studies with adults with substance use disorders, current instructions urge participants to respond as quickly, yet accurately as possible, i.e., emphasizing both accuracy and reaction time. After instruction, children complete practice trials with the aid

of the research assistant. Following the practice trials, children complete 32 trials, across which the position of the figure has been counterbalanced. As noted by Kaplan (1988) and others, measures of endpoint performance, such as accuracy and reaction time, may be insensitive to subtle compromise. One of the strengths of the current performance variables is that a process variable, the efficiency ratio, derived from a ratio of the percent accurate/average reaction time for accurate responses can also be derived. Thus, in the current administration, we can review longitudinal changes in behavioral efficiency as well as track traditional measures of improvement such as increased accuracy and reduced reaction time.

Historically, the task has been used in investigations of the neurobehavioral effects of substance use in adults (Acker and Acker, 1982; Glenn and Parsons, 1991; Lawton-Craddock et al., 2003; Nixon et al., 2014). Thus, although the task had been a part of initial project pilot work, it was essential that the practicality of its use in children be determined early in the project. Two key questions were posed. First, could children grasp the task objective? Second, did performance patterns suggest sufficient variability that longitudinal change might be observed? To address these questions, we focused on data collected through May 31, 2017. As we had hoped, there was substantial performance variability, with the average percentage of correct trials being 67% (std. dev. = 0.18). The mean reaction time for correct trials was 2670 (+470) ms. When the sample was divided by overall performance (+50%), high performers had an average of 78% (+13) correct, and low performers had an average 49% (+9). For the high performers, the average RT for correct trials was 2760 (+368) ms; the numbers for low performers are 2520 (+572) ms. A summary of the key performance variable is presented in Table 2d. Accuracy is reported as the proportion of correct responses/correct + error (range 0–1). The efficiency variable is the percentage correct (of the total possible, 32)/average RT to correct responses. These data, reflecting performance of 2084 participants, suggest that the task will provide sufficient opportunity to observe age-related development, even among those who perform relatively well at initial testing. Table 3 indicates that the LMT task variables map onto aspects of both fluid and crystallized reasoning.

5. Plans for longitudinal assessment

Comprehensive measures of neurocognition are to be collected from participants every two years. The study design also includes phone calls at six-month intervals to assess aspects of substance use as well as a one-year follow-up assessment focused on mental health, substance use and other behavioral changes. In the one-year follow-up assessment, a goal is to include additional tasks that would provide insights into processes relevant to adolescent development, but which were not included in the original battery due to time constraints. The one-year neurocognitive task protocol will include only two measures: a) an emotion word –emotional face Stroop task and an adjusting delay discounting task. The rationale and background regarding inclusion for each task are described below.

5.1. Emotion–word/emotion–face stroop task

Typically, in the classic Stroop task (Stroop, 1935), individuals must attend to a less salient stimulus attribute while ignoring a more salient or automatically processed one. Hence, the

task requires executive control to enable the ability to maintain the primary task goal without interference from distracting but compelling information. Such a task is often referred to as a Color-Word Stroop task, because color is the task-relevant dimension and the word is the irrelevant dimension. Incongruent trials, requiring higher levels of executive control, are those in which the ink color and word are in conflict (e.g. the word “red” in blue ink). Congruent trials are those in which they are not (e.g., “blue” in blue ink). Incongruent versus congruent trials are associated with longer reaction times and higher error rates, a pattern often referred to as *interference*. Meta-analyses indicate that the Stroop task primarily engages brain regions within the fronto-parietal executive control network (Cieslik et al., 2015).

In our own experience with research on a large sample of over 5000 individuals (Duell, Icenogle, Silva et al., in press), there are developmental changes during adolescence on Stroop performance, suggesting maturational changes in executive control. Furthermore, in adolescents, the degree of brain activation in lateral prefrontal cognitive control regions during Color-Word Stroop performance predicts self-reported control in everyday life (Andrews-Hanna et al., 2011). Moreover, poorer performance on the Stroop (e.g., Peeters et al., 2013) is related to the subsequent use of substances in adolescents.

Because adolescence is a time of important changes both in executive control and emotional processing (e.g., Shulman et al., 2016; Casey, 2015), we wished to use a task that would require executive control in the context of highly distracting emotional information. In the color-emotion word Stroop variant, an individual identifies a word’s color, while ignoring its meaning. Unlike the classic Stroop task, there is no inherent conflict induced by the word. Rather the task involves comparing performance on identifying the color of the word when the word’s meaning is emotional in nature (e.g., “murder”) as compared to when it is not (“sum”). Because emotional words capture attention and/or are more salient, reaction times are elongated and errors are increased for emotional versus non-emotional words. However, this task often works best with clinical populations who are attuned to particular emotions (e.g., words related to threat for individuals with anxiety) (Williams et al., 1996), and thus did not seem optimal for this study.

In contrast to this approach, we used a variation of an emotional word – emotional face Stroop task (e.g., Başgöze et al., 2015) that was optimized for adolescents as the distracting task-irrelevant faces were of age-appropriate peers (Banich et al., submitted for publication). In this task, which has been programmed through the Inquisit platform (www.millisecond.com) and is publicly available, individuals decide whether a word describes a “good” feeling (happy, joyful, glad,) or a “bad” one (angry, mad, upset). As such, the task is about determining the emotional valence of a word. These decisions must be made in the context of a face positioned behind the word, which on some trials has an incongruent facial expression, and in other cases a congruent one. The task-irrelevant emotional face is likely to be a highly potent distractor, since during adolescence, as compared to either childhood or adulthood, faces become more engaging (Hare et al., 2008), perhaps due to changes in social and emotional processing in this age group. Prior research indicates that incongruent versus congruent trials lead to significant interference, and that

the task engages brain regions involved in both executive control and emotional processing (e.g., the amygdala) (Banich et al., submitted for publication).

The task was designed so that the proportion of incongruent to congruent trials varied across blocks, one in which the proportion of incongruent to congruent trials was 50/50 and another in which it was 25/75. This manipulation was included because when there is a higher proportion of incongruent trials in a task block, these incongruent trials may serve as external reminders of the task goals. Hence, they provide a stimulus-driven cue of the need for increased executive control (see Kane and Engle, 2003 for discussion). Stimulus driven engagement of control is referred to as reactive control, as compared to proactive control, which is internally guided by the individual (see Braver, 2012). With development, there is an apparent switch from more reactive to more proactive control (Munakata et al., 2012). Inclusion of this manipulation allowed us to explore this issue.

In summary, we have included this task to increase sensitivity in the battery to the relative development of cognitive control abilities in the context of information that is highly-relevant and emotionally charged to teenagers, namely faces of same-aged peers.

5.2. Delay discounting task

The second task within the one-year follow-up, the adjusting delay discounting task (ADDT) (Koffarnus and Bickel, 2014) will present the child with 42 choices between a small hypothetical reward given immediately versus a larger hypothetical $100.00 reward given at various points in the future. ADDT choices are presented in seven randomly-ordered blocks, with six choices in each. The blocks are defined by the time interval in the future when the larger $100.00 hypothetical reward would be given: 6 hours, 1 day, 1 week, 1 month, 3 months, 1 year, and 5 years. Within each block, the task automatically adjusts the amount of the small-immediate reward based on the subject[1]s previous choices. The immediate reward increases if the future reward was just chosen or decreases if the immediate reward was just chosen. At each delay, this titration rapidly converges on an indifference point of equal subjective value between a small-immediate reward, and the $100 reward at the future time interval of that block. The “indifference point” defines where a small- immediate reward is considered equal in value to a delayed but larger reward (e.g., there is no distinct preference between a $100 immediate reward and a $110 reward in a week)

These indifference points typically form a hyperbolic curve (Odum, 2011) whose steepness is defined by a discounting constant k (Bickel et al., 2012). K values can be calculated by different curve fitting methods, such as an assumption-free area under the curve metric (Myerson et al., 2001). K values derived from DD tasks typically show good test-retest reliability across short time intervals (Matusiewicz et al., 2013). Higher k values indicate more severe discounting, where k values tend to decrease from early to late adolescence (Steinberg et al., 2009) as youth become more future-oriented (see below). To enable investigators to quickly fit indifference points at the seven delay intervals as desired, indifference points at each delay (block) are calculated and output to the ABCD Data Analysis and Informatics Core (DAIC) for inclusion in the ABCD database.

Both the cash choice task and the ADDT use hypothetical rewards. Importantly, the k values calculated from choices for hypothetical and for real monetary rewards correlate well (r's = 0.80 or higher) within-subjects (Johnson and Bickel, 2002: Matusiewicz et al., 2013). More importantly, use of hypothetical rewards avoids confounding delay-based discounting with probability-discounting, where the child may choose immediate rewards for fear that the delayed rewards will not be delivered. In an early experiment (Walls and Smith, 1970), DD behavior lessened once disadvantaged children learned that other peers did in fact receive delayed rewards from the experimenter. Finally, hypothetical rewards can be larger, better discriminating substance users (Mellis et al., 2017).

The indifference point and/or the nature of the discounting function changes during the teen years. For example, individuals younger than 13 exhibit a lower indifference point than those 16 and older, with 14–15 year olds in between (Steinberg et al., 2009). Area-under-the-curve analyses also show age-related changes across adolescence (Olson et al., 2007). Meta-analyses indicate that individuals with substance use disorders exhibit steeper discounting functions than those without (MacKillop et al., 2011) and that the steeper the discounting function, the more severe the substance use and/or abuse (Amlung et al., 2017). Steeper delay discounting functions are associated with poorer outcomes in adolescents in substance abuse treatment (Stanger et al., 2012). Thus, the task is sensitive to individual differences relevant to substance use and abuse. While it has been suggested that a steeper delay function is linked to increased impulsivity (e.g., de Wit, 2009), recent evidence suggests that it may be more closely linked with a poorer ability to contemplate and plan for the future, at least in adolescents (Steinberg et al., 2009). In addition, training individuals with substance abuse to focus on the future reduced the steepness of the discounting curve as well as smoking behavior (Stein et al., 2016).

With regards to its neural bases, delay discounting, not surprisingly, appears to engage a large number of different brain regions (Frost and McNaughton, 2017; Peters and Buchel, 2011) and is associated with structural connectivity of the bilateral frontal and temporal lobes and with the integrity of major white matter pathways that interconnect the frontal lobe with other regions (Olson et al., 2009). Prevailing models propose selective activation of striatal motivational neurocircuitry by immediate rewards, and activation by DLPFC by delayed rewards, where mesial orbitofrontal cortex may integrate the competing values (McClure et al., 2004; Kable and Glimcher, 2007). One model (Peters and Buchel, 2011) emphasizes the role of three distinct regions: lateral and medial prefrontal regions involved in cognitive control; ventromedial and reward-related regions (ventral striatum, substantia nigra) involved in valuation; and medial temporal regions involved in prospection). As described above, these first two sets of regions are undergoing rapid changes during adolescence, and behavioral evidence suggests that future prospection is developing as well (Steinberg et al., 2009). Hence, the delay discounting task is sensitivity to both developmental trends and to individual differences related to substance use.

Both the delay discounting and emotional Stroop tasks will be administered using the iPad and have been programmed through Millisecond Software (www.millisecond.com).

6. Conclusions

In conclusion, the Neurocognition Workgroup of the ABCD Consortium has configured a cognitive testing regimen for the baseline assessment and “off-year” assessments that features tasks that 1) adhere to NIH preferences as articulated in the Request for Applications, 2) have linkages to adolescent development as well as substance abuse, 3) show acceptable reliability and feasibility for repeated administration, and 4) are widely available for use by other groups to enable federation of ABCD data with other projects. Moreover, the included measures have identified neural correlates within the broad literature that can be used to guide broader inquiry into brain-behavior associations as they emerge over time. Indeed, the open-access nature of ABCD affords researchers around the globe with an unprecedented opportunity to model behavior across adolescence against a variety of emerging approaches, such as applying to normative and deviant development the trial-wise modeling from the nascent field of computational psychiatry (Paulus et al., 2016).

Based on the ABCD study data collected to date, each task appears to show appropriate sensitivity to individual variations in performance that may later prove to index risk-taking vulnerabilities. Because ABCD includes a wealth of additional information on participant demographics, mental health, substance use, pubertal status and genetic predispositions (as presented elsewhere in this issue), a wide range of hypotheses regarding neurocognitive development can be tested.

There are some limitations to be considered in the design of ABCD’s baseline neurocognitive battery. In general, the measures have not been normed for sex differences that may be evident, particularly as individuals progress through puberty. While such differences have not been examined within the data collected to date, the ABCD study will provide an unprecedented opportunity to generate sex-based norms. It is heavily weighted toward the assessment of so-called “cold” cognition, in part because we recognize that these skills are still in flux during the transition from middle childhood into early adolescence. In thinking about the universe of tasks and item content that we could employ going forward across the lifespan of this project, we recognize the need to capture nuances of reward processing, decision-making under varying contingencies, and emotion-cognition interactions (such as those measured by the emotion word/face Stroop task and the Delay Discounting task).

Overall, the ABCD study provides an unprecedented opportunity to interrogate these functions in an epidemiologically-informed sample and to integrate measures of neurocognition with a rich array of other measures, as highlighted elsewhere within this issue, permitting various models of adolescent neurodevelopment to be tested.

References

Acker, W, Acker, W. Bexley Maudsley Automated Processing Screening and Bexley Maudsley Category Sorting Test Manual. NFER-Nelson Publishing; Great Britain: 1982.

Akshoomoff N, Beaumont JL, Bauer PJ, Dikmen SS, Gershon RC, Mungas D, Heaton RK. 2013; NIH toolbox cognition battery (CB): composite scores of crystallized, fluid, and overall cognition. chapter VIII. Monogr Soc Res Child Dev. 78 (4) :119–132. [PubMed: 23952206]

Amlung M, Vedelago L, Acker J, Balodis I, MacKillop J. 2017; Steep delay discounting and addictive behavior: a meta-analysis of continuous associations. Addiction. 112 (1) :51–62.

Andrews-Hanna JR, Mackiewicz Seghete KL, Claus ED, Burgess GC, Ruzic L, Banich MT. 2011; Cognitive control in adolescence: neural underpinnings and relation to self-report behaviors. PLoS One. 6 (6) :e21598. doi: 10.1371/journal.pone.0021598 [PubMed: 21738725]

Anokhin AP, Golosheykin S, Grant JD, Heath AC. 2011; Heritability of delay discounting in adolescence: a longitudinal twin study. Behav Genet. 41 (2) :175–183. [PubMed: 20700643]

Başgöze Z, Gönül AS, Baskak B, Gökçay D. 2015; Valence-based word-face Stroop task reveals differential emotional interference in patients with major depression. Psychiatry Res. 229 (3) :960–967. [PubMed: 26272019]

Baddeley, AD. Working Memory. Oxford University Press; Oxford: 1986.

Banich MT, Smolker HS, Snyder HR, Lewis-Peacock J, Godinez D, Wager TD, Hankin BL. Turning down the heat moment by moment: Multiple brain systems and individual differences predict emotional distraction during mid-adolescence.

Bauer PJ, Dikmen SS, Heaton RK, Mungas D, Slotkin J, Beaumont JL. 2013; NIH toolbox cognition battery (CB): measuring episodic memory. Chapter III. Monogr Soc Res Child Dev. 78 (4) :34–48. [PubMed: 23952201]

Bechara A, Damasio AR, Damasio H, Anderson SW. 1994; Insensitivity to future consequences following damage to human prefrontal cortex. Cognition. 50 (1–3) :7–15. [PubMed: 8039375]

Bickel WK, Jarmolowicz DP, Mueller ET, Koffarnus MN, Gatchalian MK. 2012; Excessive discounting of delayed reinforcers as a trans-disease process contributing to addiction and other disease-related vulnerabilities: emerging evidence. Pharmacol Ther. 134 :287–297. [PubMed: 22387232]

Bjork JM, Pardini DA. 2015; Who are those risk-taking adolescents? Individual differences in developmental neuroimaging research. Dev Cognit Neurosci. 11 (1) :56–64. [PubMed: 25176616]

Bjork JM, Momenan R, Hommer DW. 2009; Delay discounting correlates with proportional lateral frontal cortex volumes. Biol Psychiatry. 65 :710–713. [PubMed: 19121516]

Bleck TP, Nowinski CJ, Gershon R, Koroshetz WJ. 2013; What is the NIH Toolbox, and what will it mean to neurology? Neurology. 80 (10) :874–875. [PubMed: 23460616]

Braver TS. 2012; The variable nature of cognitive control: a dual mechanisms framework. Trends Cognit Sci. 16 :106–113. [PubMed: 22245618]

Brown S, Tapert S, Granholm E, Delis D. 2000; Neurocognitive functioning of adolescents: effects of protracted alcohol use. Alcohol: Clin Exp Res. 24 (2) :164–171. [PubMed: 10698367]

Bryant AL, Zimmerman MA. 2002; Examining the effects of academic beliefs and behaviors on changes in substance use among urban adolescents. J Educ Psychol. 94 :621–637.

Carlozzi NE, Tulsky DS, et al. 2013; NIH toolbox cognition battery (CB): measuring processing speed. Monogr Soc Res Child Dev. 78 (4) :88–102. [PubMed: 23952204]

Carlozzi NE, Tulsky DS, et al. 2014; NIH toolbox cognitive battery (NIHTB-CB): the NIHTB pattern comparison processing speed test. J Int Neuropsychol Soc. 20 (6) :630–641. [PubMed: 24960594]

Carlozzi NE, Beaumont JL, Tulsky DS, Gershon RC. 2015; The NIH toolbox pattern comparison processing speed test: normative data. Arch Clin Neuropsychol. 30 (5) :359–368. [PubMed: 26025230]

Casaletto KB, Umlauf A, Beaumont J, Gershon R, Slotkin J, Akshoomoff N, Heaton RK. 2015; Demographically corrected normative standards for the English version of the NIH Toolbox cognition battery. J Int Neuropsychol Soc. 21 (5) :378–391. [PubMed: 26030001]

Casaletto KB, Umlauf A, Marquine M, Beaumont JL, Mungas D, Gershon R, Heaton RK. 2016; Demographically corrected normative standards for the Spanish language version of the NIH Toolbox cognition battery. J Int Neuropsychol Soc. 22 (3) :364–374. [PubMed: 26817924]

Casey BJ, Jones RM, Hare TA. 2008; The adolescent brain. Ann N Y Acad Sci. 1124 :111–126. [PubMed: 18400927]

Casey BJ, Somerville LH, Gotlib IH, Ayduk O, Franklin NT, Askren MK, et al. 2011; Behavioral and neural correlates of delay of gratification 40 years later. Proc Natl Acad Sci U S A. 108 :14998–15003. [PubMed: 21876169]

Casey BJ. 2015; Beyond simple models of self-control to circuit-based accounts of adolescent behavior. Annu Rev Psychol. 66 :295–319. [PubMed: 25089362]

Centers for Disease Control and Prevention. 2016; Youth risk behavior surveillance-United States, 2015. Morbidity and Mortality Weekly Report. 65 (6)

Choudhury S, Blakemore SJ, Charman T. 2006; Social cognitive development during adolescence. Social Cognit Affect Neurosci. 1 (3) :165–174.

Cieslik EC, Mueller VI, Eickhoff CR, Langner R, Eickhoff SB. 2015; Three key regions for supervisory attentional control: evidence from neuroimaging meta-analyses. Neurosci Biobehav Rev. 48 :22–34. [PubMed: 25446951]

Cohen AO, Dellarco DV, Breiner K, Helion C, Rahdar A, Pedersen G, Casey BJ. 2016; The impact of emotional states on cognitive control circuitry and function. J Cogn Neurosci. 28 :446–459. [PubMed: 26601909]

Daniel, MH; Wahlstrom, D; Zhang, O. Equivalence of Q-interactive® and Paper Administrations of Cognitive Tasks: WISC®–V: Q-Interactive Technical Report 8. 2014. http://www.helloq.com/content/dam/ped/ani/us/helloq/media/Technical-Report_WISC-V_092514.pdf

Demerath EW, Towne B, Chumlea WC, et al. 2004; Recent decline in age at menarche: the Fels Longitudinal Study. Am J Hum Biol. 16 :453–457. [PubMed: 15214063]

de Wit H. 2009; Impulsivity as a determinant and consequence of drug use: a review of underlying processes. Addict Biol. 14 (1) :22–31. [PubMed: 18855805]

Dikmen SS, Bauer PJ, Weintraub S, Mungas D, Slotkin J, Beaumont JL, Heaton RK. 2014; Measuring episodic memory across the lifespan: NIH toolbox picture sequence memory test. J Int Neuropsychol Soc. 20 (6) :611–619. [PubMed: 24960230]

Duell N, Icenogle C, Silva K, Chein J, Steinberg L, Banich MT, et al. A cross-sectional examination of response inhibition and working memory on the Stroop task. Cogn Dev.

Eriksen BA, Eriksen CW. 1974; Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept Psychophys. 16 (1) :143–149.

Ernst M, Pine DS, Hardin M. 2006; Triadic model of the neurobiology of motivated behavior in adolescence. Psychol Med. 36 :299–312. [PubMed: 16472412]

Fan J, McCandliss BD, Sommer T, Raz A, Posner MJ. 2002; Testing the efficiency and independence of attentional networks. J Cognit Neurosci. 14 (3) :340–347. [PubMed: 11970796]

Feldstein Ewing SW, Sakhardande A, Blakemore SJ. 2014; The effect of alcohol consumption on the adolescent brain: a systematic review of MRI and fMRI studies of alcohol-using youth. Neuroimage: Clinical. 5 :420–437. [PubMed: 26958467]

Fernández-Serrano MJ, Pérez-García M, Verdejo-García A. 2011; What are the specific vs generalized effects of drugs of abuse on neuropsychological performance? Neurosci Biobehav Rev. 35 (3) :377–406. [PubMed: 20451551]

Flores I, Casaletto KB, Marquine MJ, Umlauf A, Moore DJ, Mungas D, Heaton RK. 2017 Performance of hispanics and non-hispanic whites on the NIH Toolbox cognition battery: the roles of ethnicity and language backgrounds. Clin Neuropsychol. :1–15.

Frost R, McNaughton N. 2017; The neural basis of delay discounting: a review and preliminary model. Neurosci Biobehav Rev. 79 :48–65. DOI: 10.1016/j.neubiorev.2017.04.022 [PubMed: 28465167]

Geier CF. 2013; Adolescent cognitive control and reward processing: implications for risk taking and substance use. Horm Behav. 64 :333–342. [PubMed: 23998676]

Gershon RC, Slotkin J, Manly JJ, Blitz DL, Beaumont JL, Schnipke D, Weintraub S. 2013a; NIH toolbox cognition battery (Cb): measuring language (Vocabulary comprehension and reading decoding) Chapter IV. Monogr Soc Res Child Dev. 78 (4) :49–69. [PubMed: 23952202]

Gershon RC, Wagster MV, Hendrie HC, Fox N, Cook KF, Nowinski CJ. 2013b; NIH toolbox for assessment of neurological and behavioral function. Neurology. 80 (11 Suppl 3) :S2–6. [PubMed: 23479538]

Gershon RC, Cook KF, Mungas D, Manly JJ, Slotkin J, Beaumont JL, Weintraub S. 2014; Language measures of the NIH toolbox cognition battery. J Int Neuropsychol Soc. 20 (6) :642–651. [PubMed: 24960128]

Glenn SW, Parsons OA. 1991; Effects of alcoholism and instructional conditions on speed/accuracy tradeoffs. Alcohol Clin Exp Res. 15 (4) :612–619. [PubMed: 1928635]

Gold JM, Carpenter C, Randolph C, et al. 1997; Auditory working memory and Wisconsin card sorting test performance in patients with schizophrenia. Arch Gen Psychiatry. 54 (2) :159–165. [PubMed: 9040284]

Gould TJ. 2010; Addiction and cognition. Addict Sci Clin Pract. 5 :4–14. [PubMed: 22002448]

Grant JE, Chamberlain SR. 2014; Impulsive action and impulsive choice across substance and behavioral addictions: cause or consequence? Addict Behav. 39 (11) :1632–1639. [PubMed: 24864028]

Gunther T, Holtkamp K, Jolles J, Herpertz-Dahlmann B, Konrad K. 2004; Verbal memory and aspects of attentional control in children and adolescents with anxiety disorders or depressive disorders. J Affect Disord. 82 (2) :265–269. [PubMed: 15488256]

Hare TA, Tottenham N, Galvan A, Voss HU, Glover GH, Casey BJ. 2008; Biological substrates of emotional reactivity and regulation in adolescence during an emotional go-nogo task. Biol Psychiatry. 63 (10) :927–934. [PubMed: 18452757]

Heaton RK, Akshoomoff N, Tulsky D, Mungas D, Weintraub S, Dikmen S, Gershon R. 2014; Reliability and validity of composite scores from the NIH Toolbox Cognition Battery in adults. J Int Neuropsychol Soc. 20 (6) :588–598. [PubMed: 24960398]

Heitzeg MM, Nigg JT, Hardee JE, Soules M, Steinberg D, Zubieta JK, Zucker RA. 2014; Left middle frontal gyrus response to inhibitory errors in children prospectively predicts early problem substance use. Drug Alcohol Depend. 141 (1) :51–57. [PubMed: 24882366]

Hodes RJ, Insel TR, Landis SC. On behalf of the NIH Blueprint for Neuroscience Research. 2013; The NIH toolbox: setting a standard for biomedical research. Neurology. 80 (11 Suppl 3) :S1.

Humphrey G, Dumontheil I. 2016; Development of risk-taking, perspective-taking, and inhibitory control during adolescence. Dev Neuropsychol. 41 (1–2) :59–76. [PubMed: 27070826]

Jackson NJB, Isen JD, Khoddam R, Irons D, Tuvblad C, Iacono WG, McGue M, Raine A, et al. 2016; Impact of adolescent marijuana use on intelligence: results from two longitudinal twin studies. Proc NY Acad Sci. 113 (5) :E500–E508.

Jacobus J, Tapert SF. 2013; Neurotoxic effects of alcohol in adolescence. Annu Rev Clin Psychol. 9 :703–721. [PubMed: 23245341]

Johnson MW, Bickel WK. 2002; Within-subject comparison of real and hypothetical money rewards in delay discounting. J Exp Anal Behav. 77 :129–146. [PubMed: 11936247]

Jones SA, Cservenka A, Nagel BJ. 2016; Binge drinking impacts dorsal striatal response during decision making in adolescents. Neuroimage. 129 :378–388. [PubMed: 26826511]

Kable JW, Glimcher PW. 2007; The neural correlates of subjective value during intertemporal choice. Nat Neurosci. 10 :1625–1633. [PubMed: 17982449]

Kane MJ, Engle RW. 2003; Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. J Exp Psychol Gen. 132 (1) :47–70. [PubMed: 12656297]

Kaplan, E. A process approach to neuropsychological assessment. In: Boll, T, Bryant, BK, editors. Clinical Neuropsychology and Brain Function: Research, Measurement, and Practice. American Psychological Association; Washington, D.C: 1988. 125–167.

Keyes KM, Platt J, Kaufman AS, McLaughlin KA. 2017; Association of fluid intelligence and psychiatric disorders in a population-representative sample of US adolescents. JAMA Psychiatry. 74 (2) :179–188. DOI: 10.1001/jamapsychiatry.2016.3723 [PubMed: 28030746]

Khurana A, Romer D, Betancourt L, Brodsky N, Giannetta J, Hurt H. 2013; Working memory ability predicts trajectories of early alcohol use in adolescents: the mediational role of impulsivity. Addiction. 108 (3) :506–515. [PubMed: 23033972]

King KM, Meehan BT, Trim RS, Chassin L. 2006; Substance use and academic outcomes: synthesizing findings and future directions. Addiction. 101 :1688–1689. [PubMed: 17156166]

Koffarnus MN, Bickel WK. 2014; A 5-trial adjusting delay discounting task: accurate discount rates in less than one minute. Exp Clin Psychopharmacol. 22 :222–228. [PubMed: 24708144]

Koob GF, Volkow ND. 2016; Neurobiology of addiction: a neurocircuitry analysis. Lancet. 3 (8) :760–773.

Lawton-Craddock A, Nixon SJ, Tivis R. 2003; Cognitive efficiency in stimulant abusers with and without alcohol dependence. Alcohol Clin Exp Res. 27 (3) :457–464. [PubMed: 12658111]

Lezak, MD, Howieson, DB, Bigler, ED, Tranel, D. Neuropsychological Assessment. 5. Oxford University Press; New York, NY: 2012.

Luciana M, Collins PF. 2012; Incentive motivation, cognitive control, and the adolescent brain: is it time for a paradigm shift? Child Dev Perspect. 6 (4) :392–399. [PubMed: 23543860]

Luciana M, Wahlstrom D, Collins PF, Porter JN. 2012; Dopaminergic modulation of incentive motivation in adolescence: age-related changes in signaling, individual differences, and implications for the development of self-regulation. Dev Psychol. 48 (3) :844–861. [PubMed: 22390660]

Luciana M, Collins PF, Muetzel RL, Lim KO. 2013; Effects of alcohol use initiation on brain structure in typically developing adolescents. Am J Drug Alcohol Abuse. 39 (6) :345–355. [PubMed: 24200204]

Luna B, Padmanabhan A, O'Hearn K. 2010; What has fMRI told us about the development of cognitive control through adolescence? Brain Cogn. 72 (1) :101–113. [PubMed: 19765880]

MacKillop J, Amlung MT, Few LR, Ray LA, Sweet LH, Munafò MR. 2011; Delayed reward discounting and addictive behavior: a meta-analysis. Psychopharmacology (Berl). 216 (3) :305–321. [PubMed: 21373791]

Malone SM, Luciana M, Wilson S, Sparks JC, Hunt RH, Thomas KM. 2014; Adolescent drinking and motivated decision-making: a co-twin-control investigation with monozygotic twins. Behav Genet. 44 (4) :407–418. [PubMed: 24676464]

Matusiewicz AK, Carter AE, Landes RD, Yi R. 2013; Statistical equivalence and test-retest reliability of delay and probability discounting using real and hypothetical rewards. Behav Processes. 100 :116–122. [PubMed: 23954833]

McClure SM, Laibson DI, Loewenstein G, Cohen JD. 2004; Separate neural systems value immediate and delayed monetary rewards. Science. 306 :503–507. [PubMed: 15486304]

Meier MH, Caspi A, Ambler A, Harrington H, Houts R, Keefe RSE, McDonald K, Ward A, Poulton R, et al. 2012; Persistent cannabis users show neuropsychological decline from childhood to midlife. Proc Natl Acad Sci. 109 (40) :E2657–E2664. [PubMed: 22927402]

Mellis AM, Woodford AE, Stein JS, Bickel WK. 2017; A second type of magnitude effect: reinforcer magnitude differentiates delay discounting between substance users and controls. J Exp Anal Behav. 107 :151–160. [PubMed: 28101922]

Munakata Y, Snyder HR, Chatham CH. 2012; Developing cognitive control: three key transitions. Curr Direct Psychol Sci. 21 (2) :71–77.

Mungas D, Heaton R, Tulsky D, Zelazo P, Slotkin J, Blitz D, Gershon R. 2014; Factor structure convergent validity, and discriminant validity of the NIH Toolbox Cognitive Health Battery (NIHTB-CHB) in adults. J Int Neuropsychol Soc. 20 (6) :579–587. [PubMed: 24960474]

Myerson J, Green L, Warusawitharana M. 2001; Area under the curve as a measure of discounting. J Exp Anal Behav. 76 (2) :235–243. [PubMed: 11599641]

Nixon, SJ, Prather, RA, Lewis, B. Sex differences in alcohol-related neurobehavioral consequences. In: Sullivan, Edith V, Pfefferbaum, Adolf, editors. Alcohol and the Nervous System. Vol. 125. Elsevier; Oxford, United Kingdom: 2014. 253–272. Handbook of Clinical Neurology3rd series

Norman AL, Pulido C, Squeglia LM, Spadoni AD, Paulus MP, Tapert SF. 2011; Neural activation during inhibition predicts initiation of substance use in adolescence. Drug Alcohol Depend. 119 (3) :216–223. [PubMed: 21782354]

Odum AL. 2011; Delay discounting: I'm a K: you're a K. J Exp Anal Behav. 96 :427–439. [PubMed: 22084499]

Oldfield RC. 1971; The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 9 (1) :97–113. [PubMed: 5146491]

Olson EA, Hooper CJ, Collins PF, Luciana M. 2007; Delay and probability discounting behavior in healthy adolescents: associations with age, personality style, and other measures of executive function. Personal Individ Differ. 43 (7) :1886–1897.

Olson EA, Collins PF, Hooper CJ, Muetzel R, Lim KO, Luciana M. 2009; White matter integrity predicts delay discounting behavior in adolescents: a diffusion tensor imaging study. J Cogn Neurosci. 21 (7) :1406–1421. [PubMed: 18767918]

Paulus MP, Huys QJM, Maia TV. 2016; A roadmap for the development of applied computational psychiatry. Biol Psychiatry: Cognit Neurosci Neuroimag. 1 (5) :386–392.

Peeters M, Monshouwer K, van de Schoot RAGJ, Janssen T, Vollebergh WAM, Wiers RW. 2013; Automatic processes and the drinking behavior in early adolescence: a prospective study Alcoholism. Clin Exp Res. 37 (10) :1737–1744.

Peeters M, Janssen T, Monshouwer K, Boendermaker W, Pronk T. 2015; Weaknesses in executive functioning predict the initiating of adolescents' alcohol use. Dev Cognit Neurosci. 16 :139–146. [PubMed: 25936585]

Pentz MA, Shin H, Riggs N, Unger J, Collison K, Chou C. 2015; Parent, peer, and executive function relationships to early adolescent e-cigarette use: a substance use pathway? Addict Behav. 42 :73–78. [PubMed: 25462657]

Peters J, Buchel C. 2011; The neural mechanisms of inter-temporal decision-making: understanding variability. Trends Cognit Sci. 15 (5) :227–239. [PubMed: 21497544]

Riggs N, Spruijt-Metz D, Chou C, Pentz MA. 2012; Relationships between executive cognitive function and lifetime substance use and obesity-related behaviors in fourth grade youth. Child Neuropsychol. 18 (1) :1–11. [PubMed: 21480013]

Roberts CA, Jones A, Montgomery C. 2016; Meta-analysis of executive functioning in ecstasy/polydrug users? Psychol Med. 46 (8) :1581–1596. [PubMed: 26966023]

Romer D. 2010; Adolescent risk taking, impulsivity, and brain development: implications for prevention. Dev Psychobiol. 52 :263–276. [PubMed: 20175097]

Rueda MR, Fan J, McCandliss BD, Halparin JD, Gruber DB, Lercari LP, Posner MJ. 2004; Development of attentional networks in childhood. Neuropsychologia. 42 (8) :1029–1040. [PubMed: 15093142]

Salthouse TA, Babcock RL, Shaw RJ. 1991; Effects of adult age on structural and operational capacities in working memory. Psychol Aging. 6 (1) :118–127. [PubMed: 2029360]

Shulman EP, Smith AR, Silva K, Icenogle G, Duell N, Chein J, Steinberg L. 2016; The dual systems model: review, reappraisal: and reaffirmation. Dev Cognit Neurosci. 17 :103–117. [PubMed: 26774291]

Smith JL, Mattick RP, Jamadar SD, Iredale JM. 2014; Deficits in behavioural inhibition in substance abuse and addiction: a meta-analysis. Drug Alcohol Depend. 145 (1) :1–33. [PubMed: 25195081]

Snellen, H. Also published in many languages as Optotypi ad Visum Determinandum. 1862. Letterproeven Tot Bepaling Der Gezictsscherpte, Utrecht, Weyers (Dutch Edition).

Solowij N, Stephens RS, Roffman RA, Babor T, Kadden R, Miller M, et al. 2002; Cognitive functioning of long-term heavy cannabis users seeking treatment. JAMA. 287 :1123–1131. [PubMed: 11879109]

Solowij N, Jones KA, Rozman ME, Davis SM, Ciarrochi J, Heaven PC, Lubman DI, Yücel M. 2011; Verbal learning and memory in adolescent cannabis users, alcohol users and non-users. Psychopharmacology (Berl). 216 (1) :131–144. [PubMed: 21328041]

Sparks JC, Isen JD, Iacono WG. 2014; Preference on cash-choice task predicts externalizing outcomes in 17-year-olds. Behav Genet. 44 :102–112. [PubMed: 24442381]

Stang NM, Chein JM, Steinberg L. 2013; The value of the dual systems model of adolescent risk-taking. Front Hum Neurosci. 7 :233. [PubMed: 23781181]

Stanger C, Ryan SR, Fu H, Landes RD, Jones BA, Bickel WK, Budney AJ. 2012; Delay discounting predicts adolescent substance abuse treatment outcome. Exp Clin Psychopharmacol. 20 (3) :205–212. [PubMed: 22182419]

Stein JS, Wilson AG, Koffarnus MN, Daniel TO, Epstein LH, Bickel WK. 2016; Unstuck in time: episodic future thinking reduces delay discounting and cigarette smoking. Psychopharmacology (Berl). 233 (21–22) :3771–3778. [PubMed: 27553824]

Steinberg L, Graham S, O'Brien L, Woolard J, Cauffman E, Banich M. 2009; Age differences in future orientation and delay discounting. Child Dev. 80 (1) :28–44. [PubMed: 19236391]

Steinberg L. 2010; A dual systems model of adolescent risk-taking. Dev Psychobiol. 52 :216–224. [PubMed: 20213754]

Stephan RA, Alhassoon OM, Allen KE, Wollman SC, Hall M, Thomas WJ, et al. 2016; Meta-analyses of clinical neuropsychological tests of executive dysfunction and impulsivity in alcohol use disorder. Am J Drug Alcohol Abuse. 12 :1–20.

Strauss, E, Sherman, EMS, Spreen, O. A Compendium of Neuropsychological Tests. 3. Oxford University Press; New York, New York: 2006.

Stroop JR. 1935; Studies of interference in serial verbal reactions. J Exp Psychol. 18 (6) :643–662.

Sullivan EV, Brumback T, Tapert SF, Prouty D, Fama R, et al. 2017; Effects of prior testing lasting a full year in NCANDA adolescents: contributions from age, sex, socioeconomic status, ethnicity, site,family history of alcohol or drug abuse, and baseline performance. Dev Cognit Neurosci. 24 :72–83. [PubMed: 28214667]

Taylor, EM. The Appraisal of Children with Cerebral Deficits. Harvard University Press; Cambridge, MA: 1959.

Tulsky DS, Carlozzi NE, Chevalier N, Espy KA, Beaumont JL, Mungas D. 2013; NIH toolbox cognition battery (CB): measuring working memory. Monogr Soc Res Child Dev. 78 (4) :70–87. [PubMed: 23952203]

Tulsky DS, Carlozzi N, Chiaravalloti N, Beaumont JL, Conway K, Mungas D. 2014; NIH Toolbox Cognition Battery (NIHTB-CB): list sorting test to measure working memory. J Int Neuropsychol Soc. 20 (6) :599–610. [PubMed: 24959983]

Vakil E, Blachstein H, Wertman-Elad R, Greenstein Y. 2012; Verbal learning and memory as measured by the Rey-Auditory Verbal Learning Test: ADHD with and without learning disabilities. Child Neuropsychol. 18 (5) :449–466. [PubMed: 21962025]

van den Burg W, Kingma A. 1999; Performance of 225 Dutch school children on Rey's Auditory Verbal Learning Test (AVLT): parallel test-retest reliabilities with an interval of 3 months and normative data. Arch Clin Neuropsychol. 14 (6) :545–559. [PubMed: 14590582]

Veale JF. 2014; Edinburgh handedness inventory – short form: a revised version based on confirmatory factor analysis. Laterality. 19 (2) :164–177. [PubMed: 23659650]

Walls RT, Smith TS. 1970; Development of preference for delayed reward in disadvantaged children. J Educ Psychol. 61 :118–123. [PubMed: 5532534]

Wechsler, D. Wechsler Intelligence Scale for Children®. 5. Pearson; Bloomington, MN: 2014.

Wetherill RR, Squeglia LM, Yang TT, Tapert SF. 2013; A longitudinal examination of adolescent response inhibition: neural differences before and after the initiation of heavy drinking. Psychopharmacology (Berl). 230 (4) :663–671. [PubMed: 23832422]

Williams KD, Jarvis B. 2006; Cyberball: a program for use in research on interpersonal ostracism and acceptance. Behav Res Methods. 38 :174. [PubMed: 16817529]

Williams JM, Mathews A, MacLeod C. 1996; The emotional Stroop task and psychopathology. Psychol Bull. 120 (1) :3–24. [PubMed: 8711015]

Wulfert E, Block JA, Santa Ana E, Rodriguez ML, Colsman M. 2002; Delay of gratification: impulsive choices and problem behaviors in early and late adolescence. J Pers. 70 :533–552. [PubMed: 12095190]

Zelazo PD, Anderson JE, Richler J, Wallner-Allen K, Beaumont JL, Conway KP, Weintraub S. 2014; NIH Toolbox Cognition Battery (CB): validation of executive function measures in adults. J Int Neuropsychol Soc. 20 (6) :620–629. [PubMed: 24960301]

Zelazo PD. 2006; The Dimensional Change Card Sort (DCCS): a method of assessing executive function in children. Nat Protoc. 1 (1) :297–301. [PubMed: 17406248]

Zelazo PD, Anderson JE, Richler J, Wallner-Allen K, Beaumont JL, Weintraub S. 2013; NIH toolbox cognition battery (CB): measuring executive function and attention. Monogr Soc Res Child Dev. 78 (4) :16–33. [PubMed: 23952200]

Zelazo PD, Carlson SM. 2012; Hot and cool executive function in childhood and adolescence: development and plasticity. Child Dev Perspect. 6 :354–360.

Appendix A

ABCD Neurocognition Workgroup Members Monica

Luciana, University of Minnesota, Co-Chair

Jim Bjork, Virginia Commonwealth University, Co-Chair

Marie Banich, University of Colorado, Boulder

Deanna Barch, Washington University

Lisa Freund, National Institutes of Child Health and Human Development

Stacia Friedman, National Institute of Mental Health

Raul Gonzalez, Florida International University

Rita Goldstein, Mount Sinai School of Medicine

Elizabeth Hoffman, National Institute on Drug Abuse

Joanna Jacobus, University of California, San Diego

Erin McGlade, University of Utah

Bonnie Nagel, Oregon Health & Science University

Sara Jo Nixon, University of Florida

Catherine Orr, University of Vermont

Devin Prouty, SRI International

Dana Schloesser, NIH Office of Behavioral and Social Science Research

Susan Tapert, University of California, San Diego

Ken Warrren, National Institute on Alcohol Abuse and Alcoholism

Margie Mejia Hernandez, University of California San Diego

Table 1

ABCD Baseline Neurocognition Battery: Tests and Measured Cognitive Processes.

Name of Test	Cognitive Process
Snellen Chart	Visual acuity
Edinburgh Handedness Inventory–Short Form	Handedness
Rey Auditory Verbal Learning Test	Verbal encoding; learning; memory
NIH Toolbox Flanker®	Cognitive Control/Attention
NIH Toolbox List Sorting Working Memory Test®	Working Memory; Categorization; Information Processing
NIH Toolbox Dimensional Change Card Sort®	Flexible thinking; concept formation; set shifting
NIH Toolbox Oral Reading Recognition Test®	Reading Ability; Language; Academic Achievement
NIH Toolbox Pattern Comparison Processing Speed®	Processing Speed; Information Processing
NIH Toolbox Picture Sequence Memory Test®	Visuospatial sequencing & memory
NIH Toolbox Picture Vocabulary Test®	Language; Verbal intellect
Cash Choice Task	Delay of gratification; motivation; impulsivity
Little Man Task	Visuospatial attention; Perspective-taking; mental rotation
Matrix Reasoning Test	Fluid Reasoning; Visuospatial ability; Part-whole reasoning; Visual sequencing

Table 2a

Distributional Characteristics of the NIH Toolbox® Individual Tests and Composite Scores (N = 2259) for Uncorrected Scores.

	TPVT	TORRT	TPCPST	TLSWMT	TPSMT	TFT	TDCCS	Crystalized	Fluid	Total
Mean	85.5	91.2	90.1	97.8	103.8	94.8	93.9	87.1	93.3	87.7
Median	85	91	90	97	103	96	94	87	94	88
Mode	83	91	90	105	97	97	94	85	92	91
Std. Deviation	7.97	6.68	14.3	11.4	12.0	8.80	9.09	6.79	10.24	8.74
Minimum	36	59	43	47	76	53	50	55	49	51
Maximum	119	118	132	136	136	116	120	110	123	115
Skewness	0.185	−0.073	−0.181	−0.435	0.24	−0.898	−0.719	0.056	−0.244	−0.226
Kurtosis	1.05	1.75	−0.08	0.69	−0.39	1.38	2.18	0.62	0.28	0.41
10th%	76	83	71	82	88	83	83	79	80	77
25th%	80	88	80	90	95	90	89	83	87	82
75th%	90	95	101	105	112	101	100	91	100	94
90th%	96	100	107	113	120	105	104	96	106	99

TPVT: Toolbox Picture Vocabulary Test; TORRT: Toolbox Oral Reading Recognition Test; TPCPST: Toolbox Pattern Comparison Processing Speed Test; TLSWMT: Toolbox List Sorting Working Memory Test; TPSMT: Toolbox Picture Sequence Memory Test; TFT: Toolbox Flanker Test; TDCCS: Toolbox Dimensional Change Card Sort Test.

Table 2b

Intertemporal Cash Choice Task (n = 2285).

	Smaller-Sooner Option	Larger-Later Option	Can't Answer
Number of participants	955	1294	36
Proportion of sample	0.418	0.566	0.016

Table 2c

Distributional Characteristics of the RAVLT, and Matrix Reasoning Tasks (N = 2265).

	RAVLTTr1	RAVLTTr2	RAVLTTr3	RAVLTTr4	RAVLTTr5	RAVLTListB	RAVLTImm	RAVLTDel	MRScaled
Mean	5.44	8.58	10.32	11.46	12.00	5.29	10.46	10.15	10.86
Median	5.00	9.00	11.00	12.00	12.00	5.00	11.00	10.00	11.00
Mode	5	9	11	13	14	5	12[a]	10	11
Std. Deviation	1.832	2.168	2.497	2.472	2.309	1.854	3.093	3.305	3.098
Minimum	0	0	0	0	0	0	0	0	1
Maximum	13	15	15	15	15	15	15	15	19
Skewness	0.279	−0.177	−0.558	−0.793	−1.221	0.561	−0.513	−0.469	0.019
Kurtosis	0.064	0.028	0.183	0.522	2.007	2.215	−0.112	−0.172	−0.045
10th%	3.00	6.00	7.00	8.00	9.00	3.00	7.00	6.00	7.00
25th%	4.00	7.00	9.00	10.00	11.00	4.00	8.00	8.00	9.00
75th%	5.00	9.00	11.00	12.00	12.00	5.00	11.00	10.00	11.00
90th%	7.00	10.00	12.00	13.00	14.00	7.00	13.00	13.00	13.00

RAVLTTr1 = Trial 1 total correct; RAVLTTr2 = Trial 2 Total Correct; RAVLTTr3 = Trial 3 total correct; RAVLTTr4 = Trial 4 total correct; RAVLTTr5 = Trial 5 total correct; RAVLTListB = interference list total correct; RAVLTImm = RAVLT immediate recall total correct; RAVLTDel = RAVLT delayed recall total correct; MRScaled = Matrix Reasoning Scaled Score.

Table 2d

Distributional Characteristics of the Little Man Task (n = 2084).

	Mean	Std Dev.	Min.	Max.	25th%	75th%
Accuracy[a]	0.67	0.18	0	1	0.52	0.82
Correct RT[b]	2670	470	1246	4391	2381	2991
Efficiency[c]	0.23	0.07	0	0.47	0.18	0.28

[a]Accuracy calculated as the number of correct responses divided by the sum of correct and incorrect responses (range: 0–1).

[b]Reaction time measured in milliseconds for correct responses.

[c]Efficiency measured as ratio of percentage correct/correct reaction time.

Table 3

Intercorrelations among ABCD baseline neurocognition task measures (data collected through May 2017).

	TPVT 1	TORRT 2	TPCPST 3	TLSWMT 4	TPSMT 5	TFT 6	TDCCS 7	TB- Cryst 8	TB-Fluid 9	Cash Choice 10	RVLT Tr1-5 11	RVLT ListB 12	RVLT Imm 13	RVLT Del 14	MR Scaled 15	LMT Acc 16	LMT CorrRT 17	LMT EFF 18
1	–																	
2	0.50	–																
3	0.19	0.20	–															
4	0.39	0.41	0.19	–														
5	0.21	0.22	0.19	0.32	–													
6	0.25	0.28	0.37	0.28	0.19	–												
7	0.28	0.31	0.44	0.32	0.26	0.43	–											
8	0.88	0.84	0.23	0.45	0.15	0.30	0.34	–										
9	0.39	0.43	0.71	0.63	0.61	0.64	0.71	0.47	–									
10	0.06	0.01	−0.03	−0.02	−0.08	−0.03	−0.02	0.03	−0.05	–								
11	0.23	0.23	0.13	0.29	0.29	0.15	0.19	0.28	0.32	0.02	–							
12	0.18	0.18	0.13	0.22	0.17	0.16	0.19	0.21	0.26	−0.01	0.49	–						
13	0.16	0.18	0.12	0.25	0.28	0.14	0.18	0.20	0.30	0.00	0.75	0.44	–					
14	0.19	0.20	0.10	0.25	0.28	0.14	0.16	0.23	0.28	0.00	0.72	0.38	0.78	–				
15	0.27	0.22	0.08	0.24	0.18	0.15	0.18	0.29	0.24	−0.01	0.38	0.23	0.30	0.32	–			
16	0.19	0.26	0.19	0.24	0.17	0.17	0.25	0.26	0.31	−0.04	0.19	0.14	0.17	0.17	0.21	–		
17	0.02	0.07	−0.09	0.05	0.03	−0.09	−0.07	0.06	−0.07	−0.00	0.08	0.01	0.03	0.08	0.12	0.11	–	
18	0.15	0.19	0.22	0.19	0.14	0.21	0.26	0.19	0.32	−0.03	0.14	0.13	0.15	0.12	0.13	0.82	−0.46	–

TPVT: Toolbox Picture Vocabulary Test; TORRT: Toolbox Oral Reading Recognition Test; TPCPST: Toolbox Pattern Comparison Processing Speed Test; TLSWMT: Toolbox List Sorting Working Memory Test; TPSMT: Toolbox Picture Sequence Memory Test; TFT: Toolbox Flanker Test; TDCCS: Toolbox Dimensional Change Card Sort Test; TB-Cryst = Toolbox Crystallized Composite; TB-Fluid = Toolbox Fluid Composite; Cash Choice = dichotomous item reflecting choice between two primary alternatives; RVLT Tr1–5 = summed score on RAVLT learning trials; RVLT List B = total correct on RAVLT List B distraction trial; RVLT Imm = total correct on RAVLT immediate delay; RVLT Del = total correct on RAVLT delayed memory; MR scaled = Matrix Reasoning scaled score; LMT acc = Little Man Task accuracy score; LMT Corr RT = Little Man Task correct trial reaction time; LMT Eff = Little Man Task Efficiency Score; Values represent Pearson Correlations. NIH TB® tasks are uncorrected scores.