4-2019

The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: Findings from the ABCD study's baseline neurocognitive battery

Wesley K. Thompson

Deanna M. Barch

James M. Bjork

Raul Gonzalez

Bonnie J. Nagel

See next page for additional authors

Authors

Wesley K. Thompson, Deanna M. Barch, James M. Bjork, Raul Gonzalez, Bonnie J. Nagel, Sara Jo Nixon, and Monica Luciana
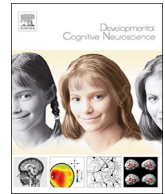
The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: Findings from the ABCD study's baseline neurocognitive battery

Wesley K. Thompson[a], Deanna M. Barch[b], James M. Bjork[c], Raul Gonzalez[d], Bonnie J. Nagel[e], Sara Jo Nixon[f], Monica Luciana[g],[*]

[a] Division of Biostatistics, Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA 92093, United States
[b] Departments of Psychological & Brain Sciences, Psychiatry and Radiology, Washington University, St. Louis, MO 63130, United States
[c] Institute for Drug and Alcohol Studies, Department of Psychiatry, Virginia Commonwealth University, Richmond, VA 23219, United States
[d] Center for Children and Families, Department of Psychology, Florida International University, Miami, FL 33199, United States
[e] Departments of Psychiatry & Behavioral Neuroscience, Oregon Health & Science University, Portland, OR 97239, United States
[f] Department of Psychiatry, University of Florida, Gainesville, FL 32611, United States
[g] Department of Psychology, University of Minnesota, Minneapolis, MN 55455, United States

ABSTRACT

The Adolescent Brain Cognitive Development (ABCD) study is poised to be the largest single-cohort long-term longitudinal study of neurodevelopment and child health in the United States. Baseline data on $N = 4521$ children aged 9–10 were released for public access on November 2, 2018. In this paper we performed principal component analyses of the neurocognitive assessments administered to the baseline sample. The neurocognitive battery included seven measures from the NIH Toolbox as well as five other tasks. We implemented a Bayesian Probabilistic Principal Components Analysis (BPPCA) model that incorporated nesting of subjects within families and within data collection sites. We extracted varimax-rotated component scores from a three-component model and associated these scores with parent-rated Child Behavior Checklist (CBCL) internalizing, externalizing, and stress reactivity. We found evidence for three broad components that encompass general cognitive ability, executive function, and learning/memory. These were significantly associated with CBCL scores in a differential manner but with small effect sizes. These findings set the stage for longitudinal analysis of neurocognitive and psychopathological data from the ABCD cohort as they age into the period of maximal adolescent risk-taking.

1. Introduction

Adolescence is a period of pronounced developmental change, including physical maturation due to puberty, changes in cortical volume and white matter microstructure, a redirection of socioemotional strivings toward peer groups, and improvements in executive function, attention, and processing speed. While these transitions are generally viewed as positive, adolescence is also a period of vulnerability given that major mental illnesses can have their onset during this time. Understanding the links between cognitive development and these vulnerabilities in an epidemiologically-informed sample of adolescents is important for structuring the timing of interventions and prevention efforts.

This paper utilized data from a large, ongoing nationally-representative cohort, the Adolescent Brain Cognitive Development

(ABCD) Study, conceived and funded by the United States' National Institutes of Health, to determine how cognitive processes in this age group are structured and how they relate to trait-level vulnerabilities, such as internalizing and externalizing tendencies, that are often associated with later risk-taking behaviors and emotional distress.

Individual differences in cognitive abilities can take many forms, including variation in intellectual capacity (e.g., IQ), attention, cognition-emotion integration, decision-making, memory, and executive function. When ABCD was conceptualized, the consortium was tasked with measuring neurocognitive abilities across this diverse array of functions in a maximally efficient manner with minimal subject burden. For the baseline assessment, this goal was achieved through the use of an automated task battery that includes the NIH Toolbox measures of cognition (seven subtasks), the Matrix Reasoning task from the Wechsler Intelligence Scale for Children, the Rey Auditory Verbal

---

Learning test (RAVLT), and a measure of spatial reasoning, the Little Man Task (see Luciana et al., 2018 for detailed task descriptions). While these measures were selected because they seemingly reflect distinct cognitive abilities, it may be that relatively few latent dimensions underlie performance across tasks, and these may dynamically change with development.

Accordingly, the structure of cognition, as assessed through factor analytic approaches (Miyake and Friedman, 2012; Friedman and Miyake, 2017; Brydges et al., 2014; Mungas et al., 2013) is thought to change markedly between early childhood and adolescence. Efforts to understand the nature of this change have primarily focused on measures of executive functions (EF), their development, and differentiation over time. One influential model (Miyake and Friedman, 2012) based on laboratory-based measures of non-affective EF suggests that three latent constructs (working memory updating, inhibitory control, cognitive flexibility) underpin EF, and these are related, given that the three components correlate moderately with one another, and distinct, given that a single component is not adequate in explaining the overall variance in adult EF. This does not appear to be the case in young children. In children ages 3–6, a single factor emerges as the best-fitting model in confirmatory factor analyses of cognitive task data (Mungas et al., 2013; Visu-Petra et al., 2007; Wiebe et al., 2011). In adolescents aged 13 and older (Xu et al., 2013) and in young adults (Mungas et al., 2014), the three factor structure of EF is more clearly evident, consistent with the notion that neural substrates of these functions consist of a series of overlapping but partially distinct networks (McKenna et al., 2017).

The dimensional structure of cognition in middle childhood is complex and not easily discernible, because most studies have tended to combine pre-adolescents with older or younger children, leading to inconsistent findings. For instance, (Lehto et al., 2003) studied 8–13 year-olds using tasks from the CANTAB battery and found, via exploratory and confirmatory factor analysis, evidence for three distinct EF factors. In contrast, (Xu et al., 2013) examined a sample of 457 7–15 year-old Chinese children from low SES backgrounds and found that a single factor explained performance in 7–9 and 10–12 year-olds but a three-factor structure emerged by ages 13-15. Some studies, while reporting a multi-dimensional structure of EF, indicate that measures of inhibitory control do not cohere during childhood and adolescence despite the emergence of working memory and shifting/flexibility factors (Huizinga et al., 2006). In a study of 102 8–15 year-olds who completed the Iowa Gambling Task, a Color Word Stroop task, a Delay Discounting task, and a Digit Span task, an exploratory factor analysis indicated that performance could be explained by a single factor (Prencipe et al., 2011).

Nevertheless, these findings have led to the differentiation hypothesis, which states that the structure of cognition becomes more differentiated as development advances (Mungas et al., 2013; Shing et al., 2010) and that the observed correlations among multiple factors (if they emerge) diminishes with age. A difficulty in evaluating and replicating this literature is that there is little consensus regarding which tasks are optimal within and across age groups to address the differentiation hypothesis. Studies that have utilized the NIH Toolbox together with other validation measures (Mungas et al., 2013) to assess the structure of cognition have found fewer factors, even above and beyond the construct of EF, in young children, ages 3–6 relative to 8–15 year-olds, though the same measures were not utilized across groups. In young children who were administered eleven total measures, three factors emerged that appeared to reflect vocabulary knowledge, reading, and fluid reasoning skills. Conventional EF tasks loaded together onto the fluid reasoning factor. In contrast, five factors were evident in 8–15 year-olds, though this group completed fifteen total measures. Fluid reasoning skills differentiated into episodic memory, working memory, and executive function.

Beyond the realm of EF, the differentiation hypothesis has also been studied in relation to aspects of general intellect, including crystallized and fluid abilities. Li et al. (2004) assessed 291 individuals who were represented by six age groups ranging from 6 to 89 years of age and who completed a battery of intellectual assessments. Associations between crystallized and fluid reasoning measures in adolescent (ages 12–17), young adult, and middle adult groups were smaller in magnitude than those observed in young children (ages 6–11) and older adults. The lifespan perspective of this report is unique in suggesting differentiation of ability into middle adulthood with de-differentiation in old age.

The ABCD study provides an opportunity to further assess the differentiation hypothesis in the span between middle childhood and early adulthood given its longitudinal design and the use of measures of cognition that encompass aspects of working memory, inhibitory control, and cognitive flexibility as well as episodic memory, spatial reasoning, oral reading, and verbal intellect. In addition, the current examination of cognitive function in the ABCD study may also inform risk prediction models for mental disorders and problem behaviors more broadly. For example, meta-analytic findings suggest that different mental illnesses share common cognitive abnormalities as well as common neural substrates (Goodkind et al., 2015; McTeague et al., 2017), including regions involved in emotion regulation (Peters et al., 2016), response inhibition (Aron et al., 2014), and conflict monitoring (Ridderinkhof et al., 2004). Moreover, many mental illnesses tend to show impaired behavioral inhibition in laboratory performance tasks (McTeague et al., 2016), and it has been argued that deficits in executive function may be a general risk factor for psychopathology (McTeague et al., 2017, 2016). Thus, cross-sectional studies suggest that different facets of neurocognitive performance could underpin, or be markers for, broader emotional and behavioral dysregulation, that could in turn confer risk for substance use disorder and other mental illnesses (Belcher et al., 2014). For instance, the early initiation of substance use is associated with high levels of externalizing tendencies (Dodge et al., 2009; King et al., 2004; McGue et al., 2001), which are also associated with an increased risk for substance use disorders (Marmorstein and Iacono, 2001; Riggs et al., 1995). Moreover, high levels of externalizing behavior are robustly associated, even from a preschool age, with executive dysfunction (Schoemaker et al., 2013; Woltering et al., 2016; Young et al., 2009). One of these studies, a meta-analysis of twenty-two studies including an overall sample of 4025 preschoolers, found a modest association between externalizing and overall EF (effect size r = 0.22) (Schoemaker et al., 2013). Another smaller study (Woltering et al., 2016) reported an association primarily with "hot" EF measures. A large-scale meta-analysis of 14,786 antisocial individuals (Ogilvie et al., 2011) found significant variation in effect sizes across studies; the largest effects were observed between criminality and EF (d = 0.54). A recent special issue focused on associations between externalizing and EF (Sulik, 2017) emphasizes through a number of longitudinal studies that strong EF ability buffers against later externalizing behaviors in vulnerable children. Similar studies that have examined cognition beyond EF are relatively sparse, though a recent population-based study of over 1100 children from the Generation R cohort (Blanken et al., 2017) reported associations between mother-reported CBCL scores and cognition as measured by the NEPSY-II, administered over one year later. Externalizing was significantly associated with poor attention/executive function as well as with poor sensorimotor function. Internalizing, which has been associated in case-control studies with executive dysfunction (Klimes-Dougan and Garber, 2016), was associated with poor attention, poor executive functioning, decrements in language skills and poor memory and learning. All associations were small in magnitude after adjusting for potential confounds (coeff = −0.05 to 0.11).

Thus, associations among cognitive function, externalizing, internalizing, and problem behaviors are important to quantify developmentally, before the onset of risk-taking behaviors such as substance misuse. Important steps are to quantify baseline levels of cognitive function in a large epidemiological early adolescent cohort in the

context of a planned prospective assessment, to determine how cognition is structured this group, and to associate major domains of cognition with externalizing and internalizing traits. Finally, associations between cognition and problem behaviors are influenced by socio-economic factors (Atherton et al., 2016; Deater-Deckard et al., 1998; Lawson et al., 2018; Whitesell et al., 2013), which are often difficult to model in the context of small sample studies and difficult to interpret in the absence of longitudinal assessment. The detailed assessment of socioeconomic factors and the large sample size of ABCD make it possible to address this concern.

Indeed, the sample size of ABCD is large enough to reliably detect and accurately estimate even small effects related to cognitive and neural development. It will therefore directly address the over-estimation of effect sizes and the replication crisis inflicting current neuroscience research (Button et al., 2013). Moreover, ABCD will collect data on a rich variety of genetic, environmental, and biomarker-based measures germane to neurocognition, substance use, and mental health, enabling the construction of realistically-complex etiological models incorporating factors from many domains simultaneously. Even if the effects of individual characteristics are small, as has been the case in other large epidemiological samples (Klimes-Dougan and Garber, 2016; Miller et al., 2016), cumulatively they may explain a sizeable proportion of the variation in neurodevelopmental trajectories, a scenario which has recently played out in genome-wide association analyses of complex traits (Boyle et al., 2017).

In the current paper, we performed principal component analyses of the baseline neurocognitive battery, including a within-sample replication, to identify latent components that in turn can be related to broader traits of internalizing and externalizing symptomatology. Notably, use of component scores mitigates potential method variance that can result from reliance on a single task score as the metric for an entire cognitive construct (Snyder et al., 2015). We utilized a Bayesian Probabilistic Principal Components Analysis (BPPCA) that incorporates nesting of subjects within families and families within data collection sites to account for these aspects of the ABCD study design. We extracted component scores from the model and associated them with Child Behavior Checklist (CBCL) externalizing and internalizing symptoms recorded for each participant. Association analyses controlled for demographic and socio-economic factors. The validation of the cognitive battery through the BPPCA, and the examination of associations with psychopathology in a large epidemiologically-informed sample represent novel elements of this work.

2. Methods

2.1. The ABCD study design and sample

Information regarding funding agencies, recruitment sites, investigators, and project organization can be obtained at http://abcdstudy.org. A baseline cohort of 11,872 children between the ages of 9–11 (and their parents/guardians) has been recruited across 21 data collection sites (see Garavan et al., 2018) and will be followed for at least ten years. The study closely matches the US population of 9–10 year-old children on several key demographic variables, including gender, race/ethnicity, household income, and parental education and marital status. Thus, ABCD will be capable of estimating US population norms of developmental neurocognitive trajectories. The recruitment catchment areas of the 21 participating sites encompass over 20% of the entire US population of nine and ten year-olds. The sociodemographic sample size targets for the ABCD baseline cohort come from a combination of two sources: 1) the American Community Survey (ACS), a large-scale survey of approximately 3.5 million households conducted annually by the U.S. Census Bureau; and 2) annual 3rd and 4th grade school enrollment data maintained by the National Center for Education Statistics (NCES). The ACS is one of the primary sources of demographic data for the nation as a whole and for smaller areas as well.

The NCES data sources provide aggregate counts of students for simple demographic classifications of children at the school district and individual school level.

At each ABCD data-collection site, participants were predominantly recruited through local elementary and charter schools (Garavan et al., 2018). ABCD employed a probability sampling strategy to identify schools within the 21 catchment areas as the primary method for contacting and recruiting eligible children and their parents. This method has been utilized within other large national studies (e.g., Monitoring the Future Bachman et al., 2011); the Add Health Study (Chantala and Tabor, 1999); the National Comorbidity Replication-Adolescent Supplement (Conway et al., 2016); the National Education Longitudinal Studies (Ingels et al., 1990)). A minority of participants were recruited through non-school-based community outreach and word-of-mouth referrals. Twins were recruited from birth registries (see Garavan et al., 2018; Iacono et al., 2017) for participant recruitment details). Across recruitment sites, inclusion criteria included being in the desired age range (9–10 years of age) and able to provide informed consent (parents) and assent (child). Exclusions were minimal and were limited to lack of English language proficiency in the children, the presence of severe sensory, intellectual, medical or neurological issues that would impact the validity of collected data or the child's ability to comply with the protocol, and contraindications to MRI scanning. Parents must be fluent in either English or Spanish. Sample demographics are presented in Table 1.

2.2. Neurocognitive measures

The neurocognitive battery was designed to be completed in 70 min (see Luciana et al., 2018). Participants first completed the Snellen vision chart (Snellen, 1862) as a measure of visual acuity. Legal blindness (with vision correction) was a study exclusion. A brief handedness inventory, consisting of four self-report questions, was also administered (Oldfield, 1971; Veale, 2014). The neurocognitive testing battery, comprised of ten measures, was then initiated (Luciana et al., 2018). All tests were administered using an iPad with one-on-one monitoring by a research assistant.

2.2.1. NIH Toolbox® cognition measures

The NIH Toolbox® cognition measures (herein referred to as "the Toolbox") were used by ABCD to foster harmonization of common data elements across federally funded studies and were developed as part of the NIH Blueprint for Neuroscience Research (http://www.nihtoolbox.org). The battery consists of seven different tasks that cover episodic memory, executive function, attention, working memory, processing speed, and language abilities (Bleck et al., 2013; Gershon et al., 2013a; Hodes et al., 2013). The Toolbox® was normed for samples between the ages of 3 and 85 years. The total administration time for the NIH Toolbox® Cognitive battery is approximately 35 min. Despite the availability of a Spanish language version (Casaletto et al., 2016; Flores et al., 2017), the ABCD study administers only the English language version (Casaletto et al., 2015) to youth given that English fluency is an inclusion criterion.

The Toolbox Picture Vocabulary Task® (Gershon et al., 2014, 2013b) measures language skills and verbal intellect. The Toolbox Oral Reading Recognition Task® is a reading test that asks individuals to pronounce single words. The Toolbox Pattern Comparison Processing Speed Test® (Carlozzi et al., 2013, 2014; Carlozzi et al., 2015) is a measure of rapid visual processing. The Toolbox List Sorting Working Memory Test® requires participants to use working memory to sequence task stimuli based on category membership and perceptual characteristics. The Toolbox Picture Sequence Memory Test® was modeled after memory tests asking children to imitate a sequence of actions using props (Bauer et al., 2013; Dikmen et al., 2014). The Toolbox Flanker Task®, a variant of the Eriksen Flanker task (Eriksen and Eriksen, 1974), is a response inhibition/conflict monitoring task that measures the ability to

Table 1
Demographics for Complete-Data and Incomplete-Data Subjects.

	Complete mean (sd)	Incomplete mean (sd)	p-value
n	4093 (90.6%)	428 (9.4%)	
Age	10.00 (0.61)	9.98 (0.61)	0.409
Female = yes (%)	1935 (47.3)	211 (49.3)	0.455
Race/Ethnicity (%)			0.256
Hispanic	802 (19.6)	85 (20.0)	
White	2417 (59.2)	234 (54.9)	
Black	394 (9.6)	50 (11.7)	
Asian	95 (2.3)	8 (1.9)	
Other	378 (9.3)	49 (11.5)	
Highest Parental Education (%)			0.025
< HS Diploma	156 (3.8)	22 (5.2)	
HS Diploma/GED	281 (6.9)	44 (10.3)	
Some College	1010 (24.7)	113 (26.5)	
Bachelor	1117 (27.3)	105 (24.6)	
Post Graduate Degree	1525 (37.3)	143 (33.5)	
Household Married (%)	2918 (71.6)	286 (67.0)	0.053
Household Income (%)			0.037
[< 50 K]	923 (24.5)	118 (29.6)	
[> =50 K & < 100 K]	1139 (30.2)	124 (31.1)	
[> =100 K]	1704 (45.2)	157 (39.3)	
Relationship (%)			0.221
single	2993 (73.1)	326 (76.2)	
sibling	287 (7.0)	33 (7.7)	
twin	801 (19.6)	69 (16.1)	
triplet	12 (0.3)	0 (0.0)	
Site (%)			< 0.001
site01	113 (2.8)	17 (4.0)	
site02	209 (5.1)	24 (5.6)	
site03	252 (6.2)	27 (6.3)	
site04	275 (6.7)	56 (13.1)	
site05	98 (2.4)	10 (2.3)	
site06	200 (4.9)	11 (2.6)	
site07	28 (0.7)	7 (1.6)	
site08	125 (3.1)	6 (1.4)	
site09	131 (3.2)	19 (4.4)	
site10	267 (6.5)	20 (4.7)	
site11	114 (2.8)	3 (0.7)	
site12	110 (2.7)	2 (0.5)	
site13	280 (6.8)	26 (6.1)	
site14	290 (7.1)	8 (1.9)	
site15	142 (3.5)	28 (6.5)	
site16	383 (9.4)	47 (11.0)	
site17	246 (6.0)	20 (4.7)	
site18	123 (3.0)	3 (0.7)	
site19	242 (5.9)	25 (5.8)	
site20	235 (5.7)	26 (6.1)	
site21	230 (5.6)	43 (10.0)	

Group differences are tested using two-sample t-test with equal variance assumption for continuous variables and χ^2 tests for discrete variables.

modulate responding under congruent versus incongruent stimulus contexts. The Toolbox Dimensional Change Card Sort Task® measures cognitive flexibility (Zelazo et al., 2013, 2014). Each of the Toolbox® tasks produces a number of scores, some of which are adjusted based on participant demographics. All tasks provide raw scores, uncorrected standard scores, and age-corrected standard scores (Casaletto et al., 2015). Uncorrected task scores were used in our analyses.

2.2.2. Rey auditory verbal learning test

The Rey Auditory Verbal Learning Test (RAVLT) measures auditory learning, memory, and recognition. A customized automated version, created through the Q-interactive platform of Pearson assessments (Daniel et al., 2014) was used. 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. Recall following a 30-min delay (during which participants engage in other non-verbal tasks), permits longer term retention to be assessed.

2.2.3. Little man task

This task engages visual-spatial processing, specifically mental rotation, with varying degrees of difficulty (Acker and Acker, 1982). 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. Respondents indicate by button press which hand is holding the briefcase.

2.2.4. Other measures

Two additional cognitive measures were administered to participants but were not included in our analysis. One measure (the Cash Choice Task; see Luciana et al., 2018) is a single-item delay of gratification measure with dichotomous scoring. Preliminary analyses suggested that it did not load onto any of the observed factors. The second measure, the Matrix Reasoning Task, from the Wechsler Intelligence Test for Children-V (WISC-V (Wechsler, 2014)) was administered using automated technology (Q-interactive (Daniel et al., 2014)). Standard score distributions (see Table 2) affirm that the sample is normally distributed with respect to fluid cognitive abilities. Preliminary analyses indicated that the Matrix Reasoning task scores were distributed across all components and that the general PCA solution was equivalent with and without inclusion of the task. In the interest of parsimony, we excluded it from our final models.

2.3. Child behavior checklist

Externalizing and internalizing behaviors were reported by the parent using an automated version of the Child Behavior Checklist (Achenbach, 2009) (CBCL). The CBCL is comprised of 113 items that measure aspects of the child's behavior across the past six months. This assessment did not include the instrument's open-ended questions but relied on those that could be rated using a three-point rating scale (not true, somewhat or sometimes true, very often or always true). Internalizing and externalizing scores are derived from the following syndrome scores: anxious/depressed, withdrawn/depressed, somatic complains, social problems, thought problems, rule-breaking behavior, and aggressive behavior. Competencies across several social domains are also measured.

2.4. Statistical approach

We implemented a principal component analysis (PCA) algorithm on the ABCD neurocognitive battery. We instantiated the PCA algorithm using Bayesian Probabilistic PCA (BPPCA (Tipping and Bishop,

Table 2
Neurocognitive Assessments.

	Complete mean (sd)	Incomplete mean (sd)	p-value
n	4093	428	
Pic Vocab	85.68 (7.96)	84.65 (8.00)	0.015
Flanker	94.95 (8.74)	93.79 (9.44)	0.014
List	98.17 (11.19)	97.67 (12.08)	0.410
Card Sort	93.59 (9.04)	93.32 (9.33)	0.577
Pattern	89.09 (14.33)	89.09 (14.98)	1.000
Picture	103.59 (12.00)	104.59 (12.09)	0.120
Reading	91.57 (6.57)	90.48 (6.83)	0.002
RAVLT	45.23 (9.68)	43.29 (10.27)	0.001
WISC-V	10.13 (2.92)	9.79 (2.89)	0.055
LMT	0.60 (0.17)	0.59 (0.17)	0.393
Externalizing	4.21 (5.50)	4.72 (5.88)	0.071
Internalizing	5.05 (5.47)	4.87 (5.07)	0.515
Stress	2.81 (3.27)	2.84 (3.13)	0.868

Group differences are tested using two-sample t-test with equal variance assumption.

1999; Bishop, 1999)) with random effects for site and for family to account for correlation among subjects in factor scores and in residuals caused by the nested structure of data collection in ABCD. Prior distributions were mildly regularizing for the component loading matrix and otherwise minimally informative. Primary advantages of this algorithm include posterior credible intervals for component loadings and other parameters that properly account for the nested data collection design of ABCD, and model selection metrics for the number of components such as the Leave-One-Out Information Criterion (Vehtari et al., 2017) (LOOIC). The model was implemented using the Bayesian inference engine stan (Carpenter et al., 2017) and using the R package rstan (Team, 2016) to interface with R Version 3.4.2 (R Core Team, 2017). Details of the BPPCA model and the model-fitting algorithm are given in the Supplementary materials.

The nine neurocognitive measures for the complete-data subjects were first standardized to have zero mean and unit variance before being placed into the BPPCA algorithm along with data collection site and family membership. A Monte Carlo Markov Chain (MCMC) was run for 1000 iterations on each of three chains with random starting values. The first 500 iterations were discarded. Model selection for number of components retained in the model was performed using the LOOIC as implemented in the R package loo (Vehtari et al., 2017). After selecting the number of principal components, the BPPCA solution is rotationally invariant. To address this issue, we performed *post hoc* orientation of the loading matrix and component scores using the method described in (Lockwood et al., 2015). Factor loadings and scores were then further rotated using the varimax criterion. We report the varimax-rotated solution in the main text and the unrotated and promax-rotated solutions in the Supplementary materials. We assessed the stability of the chosen model by randomly splitting the data in half, running the model on each half separately, and comparing the component loadings across models. We also examined the effect of missing data by imputing missing neurocognitive measures and re-running the BPPCA algorithm on the completed data.

To examine the association of principal components with CBCL measures, varimax rotated component scores were then extracted for each subject and correlated (Spearman's rho) with CBCL Internalizing, Externalizing, and Stress Reactivity scores. Since these initial correlations did not include consideration of demographic factors that might impact observed associations, we then input the component scores as independent variables in Generalized Linear Mixed-effects Models (GLMMs) with CBCL measures as dependent variables. GLLMs include age, sex at birth, household income, highest household education, race/ethnicity, and household marital status as fixed effects, and with data collection site and family as random effects. Missing component scores and demographic variables were imputed to produce five completed datasets using the R using the package mice (van Buuren and Groothuis-Oudshoorn, 2011). GLLMs were implemented in R using the package gamm4 (Wood and Scheipl, 2014). Results from running the GLLMS across the five imputations were combined using Rubin's formula (Rubin, 2004).

3. Results

3.1. Source of the ABCD data

The ABCD Study's Curated Annual Release 1.1 was made publicly available on November 2, 2018, and can be accessed through the NIMH Data Archive (NDA, https://data-archive.nimh.nih.gov/abcd/query/abcd-annual-releases.html). This release contains baseline data from 4521 subjects. After obtaining permissions as described there, data files can be downloaded in csv format; R scripts for merging these files and including some initial processing (e.g., computing the demographic categories used in this paper) can be found at https://github.com/ABCD-STUDY/analysis-nda17. These scripts produce an. Rds file which can then be used with the R, stan, and R Markdown scripts available

online at https://github.com/ABCD-STUDY/ to reproduce the results (and the entire manuscript) presented here precisely.

3.2. Descriptive data

Means and standard deviations for neurocognitive task data and CBCL scores can be found in Table 2. Histograms for neurocognitive assessments are presented in Supplementary Fig. 1. Histograms of CBCL outcomes are presented in Supplementary Fig. 2.

3.3. Bayesian probabilistic principal components analysis

The BPPCA algorithm was first implemented on subjects with complete data on all nine neurocognitive measures. This reduced the sample to 4093 subjects. The demographic breakdown of subjects included in the analysis (Complete) and subjects excluded from analyses because of one or more missing neurocognitive measures (Incomplete) are given in Table 1. Complete data subjects do not differ meaningfully from subjects with one or more missing neurocognitive measures in terms of demographics. Table 2 presents summaries of the nine neurocognitive measures included in analyses, again by Complete and Incomplete status. Levels of neurocognitive measures are similar but slightly lower for some measures in the Incomplete group. Histograms of the standardized neurocognition measures are displayed in Supplementary Fig. 1. A sensitivity analysis displaying the BPCCA factor loadings on the full dataset of 4521 subjects after one missing data imputation is given in Supplementary Table 7.

The model was run for each of $D = 1$, 2, 3, and 4, where denotes the number of retained components. There was a substantial improvement from (LOOIC = 98473, sd = 341) to (LOOIC = 97773, sd = 341) and from to (LOOIC = 97419, sd = 346). However, while the LOOIC for the four models was smallest for (LOOIC = 97231, sd = 352), the LOOIC for this model was well within one standard devation of the LOOIC for the model with. We thus proceeded with the model on the principle of parsimony. In future work will investigate the replicability and predictive power of BPPCA models with more than three components. Component loadings for the three-component model after a varimax rotation are shown in Table 3, along with their posterior credible intervals. The posterior median variance explained by these three components was 59.5% (posterior credible interval: [56.2%, 63.1%]). The unrotated solution is presented in Supplementary Table 1b. Communalities and uniquenesses for the three-component varimax-rotated model are given in Supplementary Table 3.

As shown in Table 3, the BPPCA findings indicated (a) a *General Ability* component (variance explained = 21.1%, [19.9%, 22.5%]) with strongest loadings for the Toolbox Picture Vocabulary and Oral Reading tests, and more moderate loadings for the List Sort Working Memory and Little Man tasks, (b) an *Executive Function* component (20.4% [19.4%, 21.5%]) with strongest loadings from the Toolbox Flanker task, the Toolbox Dimensional Change Card Sort task, and the Toolbox Pattern Comparison Processing Speed task, and (c) a *Learning/Memory* component (18.0% [16.9%, 19.2%]) with strongest loadings from the Toolbox Picture Sequence Memory task and the RAVLT total number correct. The Toolbox List Sort Working Memory task was represented on both the *General Ability* component and the *Learning/Memory* component but not the *Executive Function* component.

The posterior median of the data collection site random effect variance was = 0.055 [0.035, 0.087] and the posterior median of the family random effect variance = 0.524 [0.460, 0.592]). Thus, component scores of subjects within the same site who are not siblings had a significant but low median correlation 0.055, and scores of subjects within families in the same site had a moderately high correlation of 0.580 (= 0.055 + 0.5243). The random effects variances for the residuals of data collection site (median = 0.006 [0.004, 0.009]) and family (median = 0.094 [0.075, 0.114]) similarly demonstrated an over 10-fold higher covariance due to family than due to data collection site.

Table 3
Varimax Rotated Loadings for Three-Factor Model.

	PC1			PC2			PC3		
	.025	0.50	.975	.025	0.50	.975	.025	0.50	.975
Pic Vocab	0.706	**0.754**	0.799	0.029	0.065	0.102	0.133	0.19	0.252
Flanker	0.161	0.213	0.26	0.668	**0.712**	0.754	0.013	0.067	0.119
List	0.4	**0.471**	0.538	0.105	0.148	0.195	0.416	**0.493**	0.563
Card Sort	0.163	0.205	0.252	0.668	**0.71**	0.751	0.184	0.232	0.287
Pattern	−0.029	0.015	0.055	0.771	**0.813**	0.85	0.039	0.085	0.135
Picture	−0.023	0.012	0.049	0.102	0.135	0.171	0.816	**0.863**	0.904
Reading	0.782	**0.82**	0.86	0.084	0.12	0.16	0.067	0.122	0.173
RAVLT	0.253	0.306	0.364	0.085	0.125	0.163	0.663	**0.712**	0.76
LMT	0.424	**0.5**	0.57	0.246	0.299	0.36	0.002	0.068	0.144

Pic Vocab = Toolbox Picture Vocabulary; Flanker = Toolbox Flanker Test; List Sort = Toolbox List Sort Working Memory Task; Card Sort = Dimensional Change Card Sort Task; Pattern = Toolbox Pattern Comparison Processing Speed Task; Picture = Toolbox Picture Sequence Memory Task; Reading = Toolbox Oral Reading Test; RAVLT = Rey Auditory Verbal Learning Task, total correct; LMT = Little Man Task percent correct. For Toolbox measures, uncorrected scores were entered into the analysis. Loadings above 0.40 are highlighted; this is an arbitrary cutoff intended solely to assist with simple description of the factors, and does not enter into follow-up analyses in any fashion. Quantiles are from the posterior draws of the MCMC algorithm for each factor loading after varimax rotation and give the middle 95% of the distribution of the loadings (i.e., 95% posterior credible intervals).

See the Supplementary materials for a detailed description of these parameters.

Convergence for this model was acceptable for all parameters of interest (Gelman et al., 2003; also see Supplementary Fig. 3 for variance components trace plots). To examine the stability of the three-component model, we randomly split the data in two halves and ran the BPPCA model separately on each half. Resulting loadings, displayed in Supplementary Tables 5 and 6, were quite similar to the component loadings produced from the full data. Results of the completed data after imputation, given in Supplementary Table 7, are likewise quite similar to the results on the listwise complete sample of 4093 subjects.

3.4. Basic associations between PCA scores and CBCL measures

PCA scores of the varimax-rotated components were extracted and correlated (Spearman's rho) with three CBCL measures: Externalizing, Internalizing, and Stress Reactivity. Data summaries of these variables are given in Table 2, and histograms displayed in Supplementary Fig. 2. Results of the correlational analysis are displayed in Fig. 1. As shown in the figure, CBCL measures were strongly intercorrelated. *General Ability* (PC1) was modestly negatively correlated with Externalizing and Stress Reactivity. *Executive Function* (PC2) was negatively associated with Stress Reactivity most strongly with more minimal associations with Internalizing and Externalizing. *Learning/Memory* (PC3) exhibited the largest correlations (though these were still small in magnitude) and was negatively associated with Externalizing and Stress Reactivity.

3.5. Association with CBCL problem behaviors after controlling for demographic variables

Next, the PCA scores were placed as independent variables in GLMMs with the three CBCL measures, including fixed effects of socioeconomic and demographic variables and random effects of data collection site and family. Complete cases had slightly lower values on average for all three CBCL measures, most pronounced with Externalizing. Thus, we performed multiple imputation for missing component scores and demographic variables and fit the GLMMs of five completed datasets, combining estimates using Rubin's formula for standard errors. Because the CBCL measures were highly skewed, we modeled them using the Gamma distribution with a log link function. As indicated in Table 4, lower *General Ability* (PC1) predicted higher levels of Externalizing, as well as Stress Reactivity after controlling for relevant demographic factors. Lower *Executive Function* (PC2) was predictive of higher Internalizing as well as Stress Reactivity but was, unexpectedly, not associated with Externalizing tendencies. Lower

levels of *Learning/Memory* (PC3) were predictive of higher externalizing tendencies. Demographic variables were also associated with CBCL outcomes. For instance, even within this narrow age band, older age was associated with higher Internalizing and Stress Reactivity symptoms. Males demonstrated higher levels of Stress Reactivity and Externalizing symptoms. Economic indicators of socioeconomic difficulty (e.g., lower incomes; single parent households) were generally associated with higher levels of problem behaviors. The change in adjusted R-squared from the baseline model (demographics and random effects of site and family) to the full model (baseline model plus all three component scores) were as follows: Externalizing: = 0.64%; Internalizing: = 0.01%; and Stess Reactivity = 1.03%. Results of the GLMMs were thus consonant with correlations observed in Fig. 1.

We also performed the same GLMM analyses using 10-fold cross-validation to obtain out-of-sample prediction accuracy. The change in out-of-sample squared correlation from the baseline model to the full model were as follows: Externalizing: = 0.4%; Internalizing: = 0.24%; and Stress Reactivity = 0.28%.

3.5.1. BPPCA model for NIH toolbox measures

Because some researchers may be interested in using the Toolbox measures in isolation and to inform future studies, we repeated the analyses described above but limiting the analysis to only the seven NIH Toolbox measures. There were 4456 complete-data subjects in this analysis. Using the same LOOIC criterion for BPPCA model selection, we chose the model with 3 components. Factor loadings for the three-component model after a varimax rotation are shown in Table 5. The unrotated solution is given in the Supplementary Table 2. Communalities and uniquenesses for the three-component model are given in Supplementary Table 4.

The posterior median variance explained by the three-component model was 76.5% [72.2%, 81.2%]). The observed component structure was highly similar to what is described for the full set of measures: a *General Ability* component (19.2% [18.2%, 20.3%]) with strongest loadings on Oral Reading, Picture Vocabulary, and List Sort Working Memory tasks, an *Executive Function* component (20.2% [19.3%, 21.2%]) with strongest loadings on the Flanker, Dimensional Change Card Sort, and Pattern Comparison Processing Speed tasks, and a *Memory* component (13.1% [12.3%, 14.0%]) with strongest loadings on the Picture Sequence Memory and List Sort Working Memory tasks. The correlations between these components and those obtained from the more inclusive set of measures were as follows: *General Ability*; *Executive Function*; *Learning/Memory*.

Spearman correlations of the 7-item BPPCA with CBCL outcomes and with the 9-item BPPCA solution are displayed in Fig. 1, and
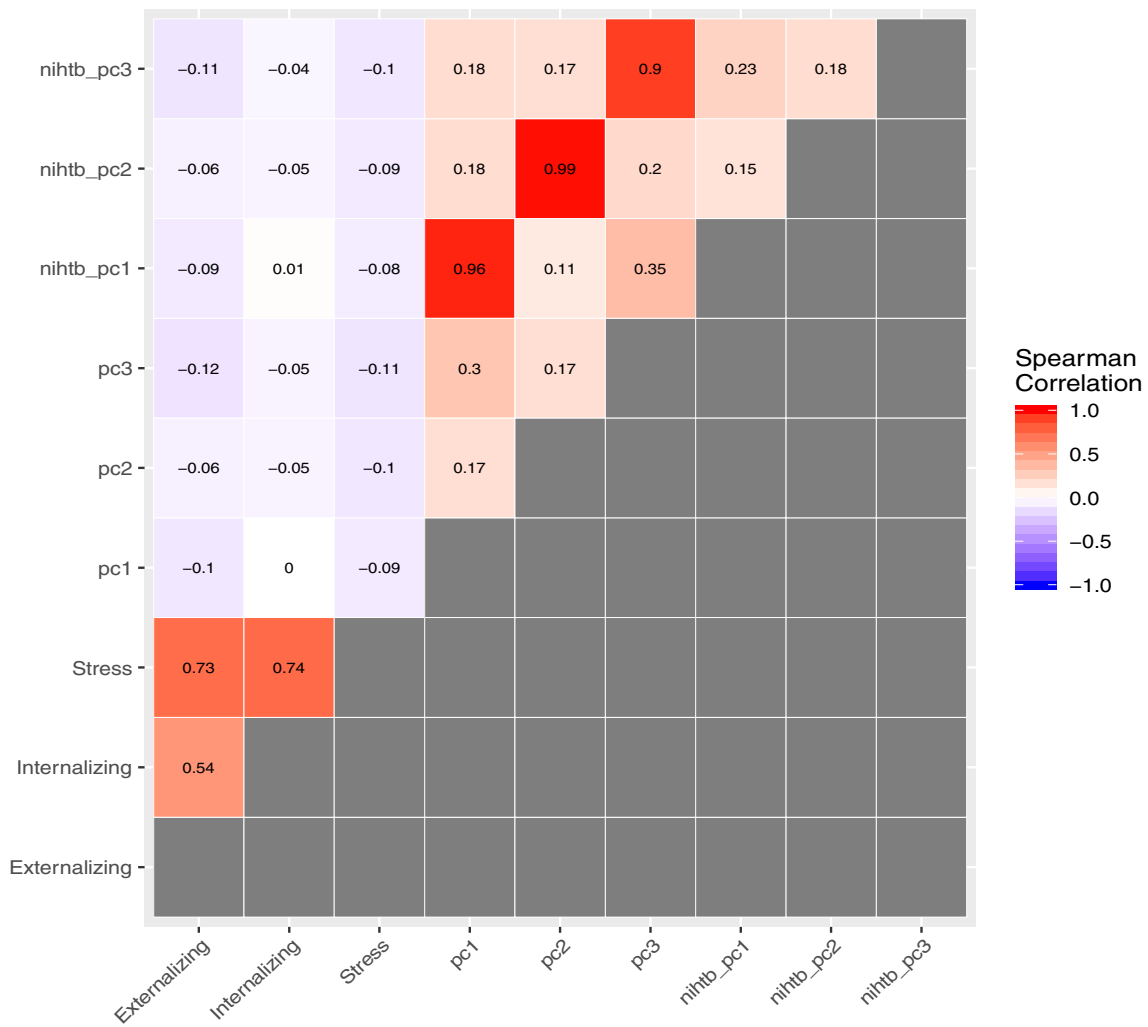
Fig. 1. Spearman Correlation of Scores from BPPCA and CBCL Outcomes.
PC1 = General Ability Factor; PC2 = Executive Function Factor; PC3 = Learning/Memory Factor; NIHTB_PC1 = Toolbox-derived General Ability factor; NIHTB_PC2 = Toolbox-derived Executive Function factor; NIHTB_PC3 = Toolbox-derived Learning/Memory factor; Total = CBCL Total Problem score; Externalizing = CBCL externalizing score; Internalizing = CBCL Internalizing; Stress = CBCL Stress Reactivity score. Heat maps represent the magnitude of Spearman's rho correlation coefficients. P-values are not presented.

Table 4
Regression of CBCL Measures on Varimax-Rotated Factors.

Variable	Externalizing			Internalizing			Stress Reactivity		
	coef	se	p-value	coef	se	p-value	coef	se	p-value
pc1	−0.193	0.04	4.89e-07	0.021	0.03	0.522	−0.222	0.03	1.85e-10
pc2	−0.015	0.03	0.633	−0.046	0.03	0.111	−0.158	0.03	1.41e-07
pc3	−0.133	0.04	0.0006	−0.001	0.03	0.962	−0.088	0.04	0.012
Age	0.017	0.04	0.652	0.079	0.03	0.019	0.163	0.04	3.15e-06
Female	−0.389	0.05	2.22e-15	0.088	0.04	0.041	−0.296	0.04	3.73e-11
White	0.301	0.08	0.0003	0.077	0.07	0.287	0.236	0.08	0.001
Black	−0.025	0.11	0.825	−0.499	0.1	4.94e-07	−0.402	0.1	9.18e-05
Asian	0.135	0.19	0.484	−0.164	0.17	0.332	0.075	0.18	0.666
Other	0.227	0.11	0.046	0.043	0.1	0.667	0.145	0.1	0.159
HS Diploma/GED	0.066	0.17	0.694	−0.057	0.15	0.702	0.048	0.15	0.752
Some College	0.125	0.15	0.412	0.020	0.13	0.876	0.138	0.14	0.319
Bachelor	0.104	0.16	0.522	0.002	0.14	0.987	0.155	0.15	0.293
Post-Graduate	0.085	0.17	0.611	0.018	0.15	0.899	0.128	0.15	0.398
Married	−0.281	0.07	7.65e-05	−0.183	0.06	0.003	−0.197	0.06	0.002
> = 50 K & < 100 K	−0.002	0.09	0.976	−0.023	0.07	0.748	−0.023	0.08	0.760
> = 100 K	−0.183	0.1	0.067	−0.224	0.09	0.009	−0.223	0.09	0.012

Table 5

NIH Toolbox Varimax Rotated Loadings for Three-Factor Model.

	PC1			PC2			PC3		
	.025	0.50	.975	.025	0.50	.975	.025	0.50	.975
Pic Vocab	0.77	0.805	0.84	0.084	0.116	0.147	0.027	0.074	0.121
Flanker	0.178	0.219	0.264	0.685	0.727	0.768	−0.027	0.033	0.092
List	0.485	0.548	0.605	0.121	0.159	0.198	0.398	0.48	0.565
Card Sort	0.167	0.206	0.243	0.697	0.737	0.776	0.137	0.193	0.249
Pattern	−0.035	−0.003	0.032	0.782	0.817	0.852	0.026	0.076	0.128
Picture	0.043	0.082	0.125	0.107	0.13	0.153	0.912	0.943	0.972
Reading	0.79	0.825	0.858	0.129	0.162	0.193	0.02	0.062	0.112

Pic Vocab = Toolbox Picture Vocabulary; Flanker = Toolbox Flanker Test; List Sort = Toolbox List Sort Working Memory Task; Card Sort = Dimensional Change Card Sort Task; Pattern = Toolbox Pattern Comparison Processing Speed Task; Picture = Toolbox Picture Sequence Memory Task; Reading = Toolbox Oral Reading Test. Loadings above 0.40 are highlighted; this is intended solely to assist with simple description of the factors, and does not enter into follow-up analyses in any fashion. Quantiles are from the posterior draws of the MCMC algorithm for each factor loading after varimax rotation and give the middle 95% of the distribution of the loadings (i.e., 95% posterior credible intervals).

associations of the PCA scores with CBCL outcomes from GLMMs are given in Table 6. The pattern of basic intercorrelations between the Toolbox-based components and CBCL variables were highly similar to what was observed for the full battery as are the GLMMs that control for demographic factors. For these GLMMS, the change in adjusted R-squared from the baseline model to the full model were as follows: Externalizing: = 0.49%; Internalizing: = 0.09%; and Stress Reactivity = 0.73%.

As with the 9-item BPPCA, we performed the same GLMM analyses using 10-fold cross-validation to obtain out-of-sample prediction accuracy. The change in out-of-sample squared correlation from the baseline model to the full model were as follows: Externalizing: = 0.49%; Internalizing: = 0.15%; and Stress Reactivity = 0.26%.

4. Discussion

The goals of the current analyses were to establish the latent structure of the cognitive tests being administered as part of the ABCD study in middle childhood and examine associations between individual differences in these components and individual differences in domains of problem behavior at the study's baseline, prior to the onset of more worrisome forms of adolescent risk-taking behavior.

Using a PCA model, we found evidence for three broad components that appear to represent General Ability, Executive Function, and Learning/Memory. Our approach was exploratory in nature but with a robust solution that was replicated in a split-half analysis of the full sample. With respect to executive function, recent research using a latent variable approach suggests both unity and diversity of function given that separable but correlated factors have emerged in other studies that represent inhibition, behavioral flexibility, and working memory updating, respectively (see Miyake and Friedman, 2012 for review).

Consistent with the notion that there is differentiation of cognitive ability from infancy to early adolescence, we found that executive function abilities in 9–10 year-olds are distinct from general cognitive abilities and from learning/memory processes (Brydges et al., 2014). Thus, a single unitary factor does not underpin all abilities. However, this differentiation does not yet approach what is observed in adults given that measures of inhibition (Flanker Task), set-shifting (Dimensional Change Card Sort), and processing speed (Pattern Comparison Processing Speed task) load onto a common component indicating unity of executive function in this age group as indexed by the NIH Toolbox measures of these processes. Other studies have similarly reported more "unity" versus "diversity" of executive function in young children (Wiebe et al., 2011; Weintraub et al., 2013). On the other hand, our finding that working memory, as measured by the List Sort Working Memory task, did not cohere with other measures of executive function was unexpected and suggests that this aspect of EF may begin to differentiate at an earlier age relative to other processes such as inhibition and cognitive flexibility.

Prior confirmatory factor analyses of the NIH Toolbox® in a sample of 267 8–15 year-olds (Mungas et al., 2013) and adults (Mungas et al., 2014), including numerous other validation measures such as the RAVLT, measures of visuospatial memory, achievement tests, and the Peabody Picture Vocabulary test (see Weintraub et al., 2013), found

Table 6

Regression of CBCL Measures on NIH Toolbox Varimax-Rotated Factors.

	Externalizing			Internalizing			Stress Reactivity		
Variable	coef	se	p-value	coef	se	p-value	coef	se	p-value
pc1	−0.155	0.04	2.23e-05	0.046	0.03	0.153	−0.195	0.03	6.67e-09
pc2	−0.036	0.03	0.258	−0.053	0.03	0.060	−0.174	0.03	3.86e-09
pc3	−0.112	0.04	0.005	0.002	0.04	0.943	−0.055	0.04	0.131
Age	0.002	0.04	0.938	0.069	0.03	0.040	0.148	0.04	2.23e-05
Female	−0.391	0.05	1.55e-15	0.087	0.04	0.043	−0.297	0.04	2.81e-11
White	0.295	0.08	0.0004	0.081	0.07	0.267	0.218	0.08	0.004
Black	−0.007	0.11	0.946	−0.492	0.1	7.20e-07	−0.388	0.1	0.0001
Asian	0.128	0.19	0.506	−0.171	0.17	0.311	0.039	0.18	0.822
Other	0.229	0.11	0.045	0.056	0.1	0.573	0.141	0.1	0.174
HS Diploma	0.044	0.17	0.794	−0.059	0.15	0.690	0.020	0.15	0.893
Some College	0.097	0.15	0.526	0.010	0.13	0.937	0.107	0.14	0.441
Bachelor	0.062	0.16	0.702	−0.022	0.14	0.874	0.102	0.15	0.491
Post-Graduate	0.033	0.17	0.840	−0.019	0.15	0.892	0.055	0.15	0.713
Married	−0.302	0.07	1.94e-05	−0.195	0.06	0.001	−0.222	0.06	0.0005
> =50 K & < 100 K	0.019	0.09	0.822	−0.031	0.07	0.667	0.002	0.08	0.979
> =100 K	−0.177	0.1	0.072	−0.193	0.09	0.025	−0.151	0.09	0.0862

evidence for five broad factors that bear some similarities to those that emerged from our analyses (Mungas et al., 2013). In contrast to our findings, the Toolbox® Picture Vocabulary and Oral Reading tests loaded on separate factors (Mungas et al., 2013). However, an Episodic Memory factor emerged that was similar to the Learning/Memory factor reported here with loadings from the Toolbox Picture Sequence Memory Test, the RAVLT, the Toolbox List Sort Working Memory Test, and other measures of visuospatial memory. An Executive Function factor also emerged that included the same three Toolbox® measures as found in our analyses. Mungas et al. (Mungas et al., 2013) found that the Toolbox List Sort Working Memory task loaded onto a separate Working Memory factor, together with other measures in this domain such as the Wechsler Letter-Number Sequencing Task and the Paced Serial Addition Task, neither of which is incorporated into the ABCD baseline battery. Thus, the Executive Function factor as well as the Episodic Memory factor observed in the Toolbox validation study of adolescents was replicated here. A similar confirmatory analysis that was focused only on adults (N = 268) found evidence for the invariance of the originally-reported five factor structure between the ages of 20 and 85 years (Mungas et al., 2014). In addition to the five-factor structure observed in adolescents and adults, the Toolbox® measures have been described by a secondary composite score framework (Akshoomoff et al., 2013), which proposes a segregation into crystallized (Oral Reading, Picture Vocabulary) versus more fluid (all five other measures) components. This broad differentiation may not adequately characterize individuals in middle childhood and early adolescence given that fluid abilities segregated in our analysis into two separate components. Overall, our findings from a considerably larger sample (albeit within a narrow age band) provide further validation of the factor structure of the NIH Toolbox® cognition battery. The three-component structure that we observed was maintained when the Toolbox® measures were analyzed in isolation, which is informative for those who might choose to use the battery in the absence of other measures. ABCD is capturing cognition at a precise developmental stage which may influence the cognitive structure and relevant associations that emerge from this sample given that working memory and other executive functions are rapidly developing during early adolescence (Luciana et al., 2005; Luna et al., 2004). We anticipate that the longitudinal design of ABCD will enable granular characterization of dynamic change in the factor structure of cognitive ability across adolescent development in future assessment waves.

A second major set of findings concern the basic associative structure between the latent factor scores and CBCL-based measures of problem behaviors. Prior to accounting for any sociodemographic factors, we found evidence for small-in-magnitude associations between cognitive abilities and domains of problem behaviors. When these patterns were examined in more detail through the use of GLMMs that modeled the impacts of demographic factors, these basic associations were generally maintained. The pattern that we observed for the prediction of externalizing behavior was somewhat unexpected given that externalizing was predicted by lower levels of General Ability as well as Learning/Memory but not by Executive Function. The general ability component was most strongly represented by crystallized functions such as oral reading and picture vocabulary that may depend on educational attainment. Both components are represented by measures of working memory (the List Sort Working Memory Task), conceptualized more broadly within the literature as an executive function. Poor working memory has been associated with externalizing behavior (Schoemaker et al., 2013; Ogilvie et al., 2011). In contrast, internalizing tendencies were associated with low levels of executive function (but not with poor learning/memory) in keeping with recent findings from both a meta-analysis of adult findings (McTeague et al., 2016) and clinical samples of adolescents suggesting that depression and anxiety are associated with executive dysfunction, particularly in relation to cognitive flexibility or lack thereof (Klimes-Dougan and Garber, 2016; Snyder et al., 2015). A novel contribution of the current study is the association of multiple domains of cognition with these realms of problem behavior.

However, effect sizes for all of these respective associations are small in magnitude. One possibility for the modest effect sizes may lie in a restriction of range, in that despite intentional targeting for recruitment of children from low-SES schools or other high-risk contexts, due to the substantial commitment required by enrollment in the ABCD study, the ABCD sample may nevertheless over-represent higher-functioning children and families. Notably, the mean scores of CBCL factors (especially externalizing) observed here are substantially lower than in conduct disorder or other clinical samples. This does not diminish their potential significance given public health and educational implications at the population level, where even small effects suggest avenues for prevention and intervention. Indeed, the extent to which the patterning observed here is at odds with the literature is difficult to ascertain. Using one recent meta-analysis (Schoemaker et al., 2013) of 4021 preschoolers across 22 studies as an exemplar, ten of the included studies involved case-control categorical comparisons of EF function in children diagnosed with various externalizing disorders, such as Attention-Deficit Hyperactivity Disorder. This type of study is representative of the literature as a whole. Twelve of the included studies of children with behavior problems examined externalizing in a more dimensional fashion through methods such as symptom counts for discrete domains of psychopathology. The range of EF tasks varied from study-to-study from 1 to 6 measures. Importantly, it was observed that studies focused on referred clinical samples observed larger effect sizes (mean) than those based on community samples (mean) as did studies that included predominantly male samples. A stronger association between EF and externalizing was observed as behavior problems became more severe, suggesting non-linear associations between the two constructs. We hypothesize that this same patterning would be observed in the ABCD sample over time. Our findings suggest that in segments of the population where externalizing problems have not necessarily reached clinical levels of magnitude, the linear associations between externalizing and aspects of EF that include conflict monitoring, processing speed, and flexibility are non-significant.

Within the broad literature, potential associations between internalizing and cognitive function have not been as well characterized, though there is increasing interest in case-control comparisons of EF in clinical samples with mood and anxiety disorders (Klimes-Dougan and Garber, 2016). Those studies have focused on aspects of EF that include cognitive flexibility but also affective regulation, which is not among the constructs assessed to date in the ABCD study sample. It may be that effect sizes will be stronger between domains of problem behavior and more affectively salient EF measures, as observed in some studies (Woltering et al., 2016).

These findings also provide evidence for very modest premorbid associations between cognitive function and mood/behavior abnormalities found in cross-sectional comparisons in adults (McTeague et al., 2016), where the relationships we describe here are devoid of the confound of neurotoxicity from substance use that is endemic in mental illness. A major question in the field has concerned whether associations between cognitive functions (e.g., executive function) and problem behaviors such as externalizing and internalizing exist prior to the onset of risk-taking behaviors such as substance use that frequently emerge in later adolescence and if so, how strong the associations are in the context of the pubertal transition. Our findings indicate that these associations are small in magnitude. Prior to the onset of substance use, cognition does not appear to be severely compromised by individual differences in behaviors captured by the CBCL in a largely healthy sample. ABCD's comprehensive neuroimaging findings may enable discovery of variations in neurocircuitry that contribute to the interpretation of these relationships.

Our findings reinforce the merits of controlling, via population-based sampling and sophisticated statistical modeling, for the influences of sociodemographic factors. While not the primary focus of this

paper, we note that indices, such as participant sex, parental marital status, family income, and parental education levels, were independently predictive of children's problem behaviors, accounting in many cases for larger proportions of the variance in these outcomes than cognitive functioning. As the ABCD study continues to follow this cohort, it will be important to remain cognizant of the importance of these sociodemographic factors in the interpretation of health outcome behaviors and their correlates over the course of development.

4.1. Limitations

While the ABCD baseline assessment is comprehensive, it does not include the full range of validation measures that were included in the NIH Toolbox® validation studies. Thus, the factor structure that we observed may be more nuanced when evaluated in the context of a more representative set of measures, particularly those that more strongly capture working memory performance as well as inhibitory control. That said, the overlap between these findings and those reported in the validation studies is encouraging and lends confidence in the abbreviated, automated, and economically-feasible assessment that ABCD has undertaken. Another limitation concerns the limited age range of the sample. While the reported age band is narrow, there is a wide range of individual differences that can be assessed for their influence on developmental outcomes as the project advances. Parent-based CBCL ratings were used as outcome variables in this analysis, and these ratings may be biased in some respects (e.g., in their associations with socioeconomic variables) or less sensitive in others (e.g., with respect to the reporting of internalizing versus externalizing behaviors). At subsequent assessment points, children will provide self-report ratings in addition to parent reports, allowing further validation of parental reports. Finally, a potential confound within our analyses is that participant race/ethnicity is conflated with socioeconomic status in that racial minorities who are ABCD participants tend to be in relatively lower income and education groups while volunteers in the racial/ethnic majority tend to be of higher education and income levels. This is a sampling bias and not reflective of the full U.S. population. The data presented here are derived from ABCD's first wave of assessment, representing less than half of the full baseline sample of 11,872 individuals. A disentangling of these influences was a goal in recruiting the second half of the baseline sample. Finally, this analysis focuses on main effects without consideration of interactions that may moderate the associations between cognition and problem behaviors.

4.2. Conclusions

This analysis sets the stage for future analyses of cognitive ability in an epidemiologically-informed cohort of young adolescents, supporting the existence of separable measures of general ability, executive function, and learning/memory. Associations with problem behaviors are, at present, small in magnitude, which is an important finding given that such associations may increase in size as the sample ages into the period of higher risk for behaviors such as substance misuse.

Conflict of interest

None.

Acknowledgments

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.dcn.2018.12.004.

References

Achenbach, T.M., 2009. Achenbach System of Empirically Based Assessment (Aseba): Development, Findings, Theory, and Applications. University of Vermont, Research Center of Children, Youth & Families.

Acker, W., Acker, C., 1982. Bexley Maudsley Automated Processing Screening and Bexley Maudsley Category Sorting Test Manual. NFER-Nelson, Windsor, England.

Akshoomoff, N., et al., 2013. VIII. NIH toolbox cognition battery (CB): composite scores of crystallized, fluid, and overall cognition. Monogr. Soc. Res. Child Dev. 78, 119–132.

Aron, A.R., Robbins, T.W., Poldrack, R.A., 2014. Inhibition and the right inferior frontal cortex: one decade on. Trends Cognit. Sci. 18, 177–185.

Atherton, O.E., Conger, R.D., Ferrer, E., Robins, R.W., 2016. Risk and protective factors for early substance use initiation: a longitudinal study of mexican-origin youth. J. Res. Adolesc. 26, 864–879.

Bachman, J.G., Johnston, L.D., O'Malley, P.M., Schulenberg, J.E., 2011. The Monitoring the Future Project after Thirty-Seven Years: Design and Procedures.

Bauer, P.J., et al., 2013. III. NIH toolbox cognition battery (CB): measuring episodic memory. Monogr. Soc. Res. Child Dev. 78, 34–48.

Belcher, A.M., Volkow, N.D., Moeller, F.G., Ferré, S., 2014. Personality traits and vulnerability or resilience to substance use disorders. Trends Cognit. Sci. 18, 211–217.

Bishop, C.M., 1999. Bayesian PCA. Advances in Neural Information Processing Systems. pp. 382–388.

Blanken, L.M., et al., 2017. Cognitive functioning in children with internalising, externalising and dysregulation problems: a population-based study. Eur. Child Adolesc. Psychiatry 26, 445–456.

Bleck, T.P., Nowinski, C.J., Gershon, R., Koroshetz, W.J., 2013. What is the NIH toolbox, and what will it mean to neurology? Neurology 80 (10), 874–875.

Boyle, E.A., Li, Y.I., Pritchard, J.K., 2017. An expanded view of complex traits: from polygenic to omnigenic. Cell 169, 1177–1186.

Brydges, C.R., Fox, A.M., Reid, C.L., Anderson, M., 2014. The differentiation of executive functions in middle and late childhood: a longitudinal latent-variable analysis. Intelligence 47, 34–43.

Button, K.S., et al., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14, 365.

Carlozzi, N.E., Tulsky, D.S., Kail, R.V., Beaumont, J.L., 2013. VI. NIH toolbox cognition battery (CB): measuring processing speed. Monogr. Soc. Res. Child Dev. 78, 88–102.

Carlozzi, N.E., et al., 2014. NIH toolbox cognitive battery (NIHTB-CB): the NIHTB pattern comparison processing speed test. J. Int. Neuropsychol. Soc. 20, 630–641.

Carlozzi, N.E., Beaumont, J.L., Tulsky, D.S., Gershon, R.C., 2015. The nih toolbox pattern comparison processing speed test: normative data. Arch. Clin. Neuropsychol. 30, 359–368.

Carpenter, B., et al., 2017. Stan: a probabilistic programming language. J. Stat. Softw. 76.

Casaletto, K.B., et al., 2015. Demographically corrected normative standards for the english version of the NIH toolbox cognition battery. J. Int. Neuropsychol. Soc. 21, 378–391.

Casaletto, K.B., et al., 2016. Demographically corrected normative standards for the spanish language version of the nih toolbox cognition battery. J. Int. Neuropsychol. Soc. 22, 364–374.

Chantala, K., Tabor, J., 1999. National Longitudinal Study of Adolescent Health. Strategies to Perform a Design-Based Analysis Using the Add Health Data.

Conway, K.P., Swendsen, J., Husky, M.M., He, J.-P., Merikangas, K.R., 2016. Association of lifetime mental disorders and subsequent alcohol and illicit drug use: results from the national comorbidity survey–adolescent supplement. J. Am. Acad. Child Adolesc. Psychiatry 55, 280–288.

Daniel, M.H., Wahlstrom, D., Zhang, O., 2014. Equivalence of Q-Interactive™ and Paper Administrations of Cognitive Tasks: WISC–V. Q-Interactive Technical Report 8.

Deater-Deckard, K., Dodge, K.A., Bates, J.E., Pettit, G.S., 1998. Multiple risk factors in the development of externalizing behavior problems: group and individual differences. Dev. Psychopathol. 10, 469–493.

Dikmen, S.S., et al., 2014. Measuring episodic memory across the lifespan: NIH toolbox picture sequence memory test. J. Int. Neuropsychol. Soc. 20, 611–619.

Dodge, K.A., et al., 2009. A dynamic cascade model of the development of substance-use onset. Monogr. Soc. Res. Child Dev. 74 vii–119.

Eriksen, B.A., Eriksen, C.W., 1974. Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept. Psychophys. 16, 143–149.

Flores, I., et al., 2017. Performance of hispanics and non-hispanic whites on the nih toolbox cognition battery: the roles of ethnicity and language backgrounds. Clin. Neuropsychol. 31, 783–797.

Friedman, N.P., Miyake, A., 2017. Unity and diversity of executive functions: individual differences as a window on cognitive structure. Cortex 86, 186–204.

Garavan, H., et al., 2018. Recruiting the abcd sample: DABCD sample: design considerations and procedures. Dev. Cognit. Neurosci. 32, 16–22.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2003. Bayesian Data Analysis,

(Chapman & Hall/CRC Texts in Statistical Science).

Gershon, R.C., et al., 2013a. NIH toolbox for assessment of neurological and behavioral function. Neurology 80, S2–S6.

Gershon, R.C., et al., 2013b. IV. NIH toolbox cognition battery (CB): measuring language (vocabulary comprehension and reading decoding). Monogr. Soc. Res. Child Dev. 78, 49–69.

Gershon, R.C., et al., 2014. Language measures of the NIH toolbox cognition battery. J. Int. Neuropsychol. Soc. 20, 642–651.

Goodkind, M., et al., 2015. Identification of a common neurobiological substrate for mental illness. JAMA Psychiatry 72, 305–315.

Hodes, R.J., Insel, T.R., Landis, S.C., et al., 2013. The nih toolbox setting a standard for biomedical research. Neurology 80, S1.

Huizinga, M., Dolan, C.V., van der Molen, M.W., 2006. Age-related change in executive function: developmental trends and a latent variable analysis. Neuropsychologia 44, 2017–2036.

Iacono, W.G., et al., 2017. The utility of twins in developmental clinical neuroscience research: Hhow twins strengthen the abcdABCD research design. Dev. Cognit. Neurosci. 32, 30–42. https://doi.org/10.1016/j.dcn.2017.09.001. Epub 2017 Sep 12.

Ingels, S., Abraham, S., Karr, R., Spenser, B., Frankel, M., 1990. National Education Longitudinal Survey of 1988. Technical Report. National Opinion Research Center, University of Chicago.

King, S.M., Iacono, W.G., McGue, M., 2004. Childhood externalizing and internalizing psychopathology in the prediction of early substance use. Addiction 99, 1548–1559.

Klimes-Dougan, B., Garber, J., 2016. Regulatory Control and Depression in Adolescents: Findings from Neuroimaging and Neuropsychological Research. J. Clin. Chil Adolesc. Psychol. 45 (1), 1–5. https://doi.org/10.1080/15374416.2015.1123637.

Lawson, G.M., Hook, C.J., Farah, M.J., 2018. A meta-analysis of the relationship between socioeconomic status and executive function performance among children. Dev. Sci. 21, e12529.

Lehto, J.E., Juujärvi, P., Kooistra, L., Pulkkinen, L., 2003. Dimensions of executive functioning: evidence from children. Br. J. Dev. Psychol. 21, 59–80.

Li, S., et al., 2004. Lifespan transformations in the couplings of mental abilities and underlying cognitive processes. Psychol. Sci. 15, 155–163.

Lockwood, J., Savitsky, T.D., McCaffrey, D.F., et al., 2015. Inferring constructs of effective teaching from classroom observations: an application of Bayesian exploratory factor analysis without restrictions. Ann. Appl. Stat. 9, 1484–1509.

Luciana, M., Conklin, H.M., Hooper, C.J., Yarger, R.S., 2005. The development of nonverbal working memory and executive control processes in adolescents. Child Dev. 76, 697–712.

Luciana, M., et al., 2018. Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (abcdABCD) baseline neurocognition battery. Dev. Cognit. Neurosci. 32, 67–79.

Luna, B., Garver, K.E., Urban, T.A., Lazar, N.A., Sweeney, J.A., 2004. Maturation of cognitive processes from late childhood to adulthood. Child Dev. 75, 1357–1372.

Marmorstein, N.R., Iacono, W.G., 2001. An investigation of female adolescent twins with both major depression and conduct disorder. J. Am. Acad. Child Adolesc. Psychiatry 40, 299–306.

McGue, M., Iacono, W.G., Legrand, L.N., Malone, S., Elkins, I., 2001. Origins and consequences of age at first drink. I. Associations with substance-use disorders, disinhibitory behavior and psychopathology, and P3 amplitude. Alcohol.: Clin. Exp. Res. 25, 1156–1165.

McKenna, R., Rushe, T., Woodcock, K.A., 2017. Informing the structure of executive function in children: a meta-analysis of functional neuroimaging data. Front. Hum. Neurosci. 11, 154.

McTeague, L.M., Goodkind, M.S., Etkin, A., 2016. Transdiagnostic impairment of cognitive control in mental illness. J. Psychiatr. Res. 83, 37–46.

McTeague, L.M., et al., 2017. Identification of common neural circuit disruptions in cognitive control across psychiatric disorders. Am. J. Psychiatry 174, 676–685.

Miller, K.L., et al., 2016. Multimodal population brain imaging in the uk biobank prospective epidemiological study. Nat. Neurosci. 19, 1523.

Miyake, A., Friedman, N.P., 2012. The nature and organization of individual differences in executive functions: Four general conclusions. Curr. Direct. Psychol.l Sci. 21, 8–14.

Mungas, D., et al., 2013. VII. nih toolbox cognition battery (cb): factor structure for 3 to 15 year olds. Monogr. Soc. Res. Child Dev. 78, 103–118.

Mungas, D., et al., 2014. Factor structure, convergent validity, and discriminant validity of the nih toolbox cognitive health battery (nihtb-chb) in adults. J. Int. Neuropsychol. Soc. 20, 579–587.

Ogilvie, J.M., Stewart, A.L., Chan, R.C., Shum, D.H., 2011. Neuropsychological measures of executive function and antisocial behavior: a meta-analysis. Criminology 49, 1063–1107.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the edinburgh inventory. Neuropsychologia 9, 97–113.

Peters, S.K., Dunlop, K., Downar, J., 2016. Cortico-striatal-thalamic loop circuits of the salience network: a central pathway in psychiatric disease and treatment. Front. Syst. Neurosci. 10, 104.

Prencipe, A., et al., 2011. Development of hot and cool executive function during the transition to adolescence. J. Exp. Child Psychol. 108, 621–637.

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Ridderinkhof, K.R., Ullsperger, M., Crone, E.A., Nieuwenhuis, S., 2004. The role of the medial frontal cortex in cognitive control. Science 306, 443–447.

Riggs, P.D., Baker, S., Mikulich, S.K., Young, S.E., Crowley, T.J., 1995. Depression in substance-dependent delinquents. J. Am. Acad. Child Adolesc. Psychiatry 34, 764–771.

Rubin, D.B., 2004. Multiple Imputation for Nonresponse in Surveys, vol. 81 John Wiley & Sons.

Schoemaker, K., Mulder, H., Deković, M., Matthys, W., 2013. Executive functions in preschool children with externalizing behavior problems: a meta-analysis. J. Abnorm. Child Psychol. 41, 457–471.

Shing, Y.L., Lindenberger, U., Diamond, A., Li, S.-C., Davidson, M.C., 2010. Memory maintenance and inhibitory control differentiate from early childhood to adolescence. Dev. Neuropsychol. 35, 679–697.

Snellen, H., 1862. In: In: Greven, J. (Ed.), Letterproeven, Tot Bepaling Der Gezigtsscherpte, vol.1.

Snyder, H.R., Miyake, A., Hankin, B.L., 2015. Advancing understanding of executive function impairments and psychopathology: bridging the gap between clinical and cognitive approaches. Front. Psychol. 6, 328.

Sulik, M.J., 2017. Introduction to the special section on executive functions and externalizing symptoms. J. Abnorm. Child Psychol. 45, 1473–1475.

Team, S.D., 2016. RStan: The R Interface to Stan. R Package Version 2.14.1.RStan: The R Interface to Stan. R Package Version 2.14.1.

Tipping, M.E., Bishop, C.M., 1999. Probabilistic principal component analysis. J. R. Stat. Soc.Ser. B: Stat. Methodol. 61, 611–622.

van Buuren, S., Groothuis-Oudshoorn, K., 2011. Mice: multivariate imputation by chained equations in R. J. Stat. Softw. 45, 1–67.

Veale, J.F., 2014. Edinburgh handedness inventory–short form: a revised version based on confirmatory factor analysis. Laterality: Asym. Body Brain Cogn. 19, 164–177.

Vehtari, A., Gelman, A., Gabry, J., 2017. Practical bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat. Comput. 27, 1413–1432.

Visu-Petra, L., Benga, O., Miclea, M., 2007. Dimensions of attention and executive functioning in 5-to 12-years-old children: neuropsychological assessment with the NEPSY battery. Cogn. Brain Behav. 11.

Wechsler, D., 2014. Wechslerintelligence Scale for Children–Fifth Edition (WISC–V): Technical and Interpretive Manual. Pearson Clinical Assessment, Bloomington, MN.

Weintraub, S., et al., 2013. I. NIH toolbox cognition battery (CB): introduction and pediatric data. Monogr. Soc. Res. Child Dev. 78, 1–15.

Whitesell, M., Bachand, A., Peel, J., Brown, M., 2013. Familial, social, and individual factors contributing to risk for adolescent substance use. J. Addict. 2013.

Wiebe, S.A., et al., 2011. The structure of executive function in 3-year-olds. J. Exp. Child Psychol. 108, 436–452.

Woltering, S., Lishak, V., Hodgson, N., Granic, I., Zelazo, P.D., 2016. Executive function in children with externalizing and comorbid internalizing behavior problems. J. Child Psychol. Psychiatry 57, 30–38.

Wood, S., Scheipl, F., 2014. Gamm4: Generalized Additive Mixed Models Using mgcv and lme4. R Package Version 0.2-3.

Xu, F., et al., 2013. Developmental differences in the structure of executive function in middle childhood and adolescence. PLoS One 8, e77770.

Young, S.E., et al., 2009. Behavioral disinhibition: liability for externalizing spectrum disorders and its genetic and environmental relation to response inhibition across adolescence. J. Abnorm. Psychol. 118, 117.

Zelazo, P.D., et al., 2013. II. NIH toolbox cognition battery (CB): measuring executive function and attention. Monogr. Soc. Res. Child Dev. 78, 16–33.

Zelazo, P.D., et al., 2014. NIH toolbox cognition battery (CB): validation of executive function measures in adults. J. Int. Neuropsychol. Soc. 20, 620–629.