Age group and sex differences in performance on a computerized neurocognitive battery in children age 8–21

Ruben C. Gur[a], **Jan Richard**[a], **Monica E. Calkins**[a], **Rosetta Chiavacci**[b], **John A. Hansen**[a], **Warren B. Bilker**[c], **James Loughead**[a], **John J. Connolly**[b], **Haijun Qiu**[b], **Frank D. Mentch**[b], **Patrick M. Abou-Sleiman**[b,d], **Hakon Hakonarson**[b,d], and **Raquel E. Gur**[a]

[a]Brain Behavior Laboratory, Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, USA

[b]Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, USA

[c]Department of Biostatistics and Epidemiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, USA

[d]Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, USA

Abstract

Objective—Examine age group effects and sex differences by applying a comprehensive computerized battery of identical behavioral measures linked to brain systems in youths that were already genotyped. Such information is needed to incorporate behavioral data as neuropsychological "biomarkers" in large-scale genomic studies.

Method—We developed and applied a brief computerized neurocognitive battery that provides measures of performance accuracy and response time for executive-control, episodic memory, complex cognition, social cognition and sensorimotor speed domains. We tested a population-based sample of 3500 genotyped youths ages 8–21 years.

Results—Substantial improvement with age occurred for both accuracy and speed, but the rates varied by domain. The most pronounced improvement was noted in executive control functions, specifically attention, and in motor speed, with some effect sizes exceeding 1.8 standard deviation units. The least pronounced age group effect was in memory, where only face memory showed a large effect size on improved accuracy. Sex differences had much smaller effect sizes but were evident, with females outperforming males on attention, word and face memory, reasoning speed and all social cognition tests and males outperforming females in spatial processing and sensorimotor and motor speed. These sex differences in most domains were seen already at the youngest age groups, and age group × sex interactions indicated divergence at the oldest groups with females becoming faster but less accurate than males.

Conclusions—The results indicate that cognitive performance improves substantially in this age span, with large effect sizes that differ by domain. The more pronounced improvement for executive and reasoning domains than for memory suggests that memory capacities have reached

Corresponding authors: Ruben C. Gur, gur@upenn.edu, Address: University of Pennsylvania, Department of Psychiatry, Brain Behavior Laboratory, 10 floor Gates Building, 3400 Spruce, Philadelphia, PA 19104. Phone: 011.215.615-3604. Fax: 011.215. 662-7903.

their apex before age 8. Performance was sexually modulated and most sex differences were apparent by early adolescence.

Keywords

cognitive development; neurocognitive endophenotypes; adolescence; social cognition; computerized testing

Studies have examined the developmental course of specific behavioral domains ranging from auditory processing (Keith, 2000) to executive-control functions, such as attention and working memory (e.g., Conklin, Luciana, Hooper, & Yarger, 2007; Goldberg, Maurer, & Lewis, 2001; Pickering, 2001), language and reasoning (e.g., Friederici & Wartenburger, 2010; Kuhl, 2010) and, more recently, social cognition (e.g., Burnett, Sebastian, Cohen-Kadosh, & Blakemore, 2010; Shaw et al., 2011). General intelligence measures have been reported as stable across development (Salthouse, 2004), but improved performance from childhood to young adulthood has been especially pronounced for executive domains of attention and working memory (Ang 2008, 2010; Gao et al., 2011). Neural substrates for these age-related changes have been examined with structural and functional neuroimaging and several consistent findings have emerged. These findings highlight childhood and adolescence, periods during which important changes occur in neural substrates associated with functional age-related changes (Casey, Duhoux & Malter, 2010; Giedd, et al., 1999; Matsuzawa et al., 2001; Shaw et al., 2008, 2011). Integrating neuroimaging with behavioral findings, Jung and Haier (2007) identified a central role for frontal and parietal regions in the neurodevelopment of cognition, and this hypothesis has received support in large-scale studies (Deary, Penke & Johnson, 2010).

Sex differences have been extensively documented in behavioral measures (e.g., Halpern, Benbow, Geary, Gur, Hyde, & Gernsbacher, 2007; Hines, 2010). Males perform better than females on spatial (Voyer, Voyer & Bryden, 1995; Linn & Petersen, 1985) and motor tasks (e.g., Moreno-Briseño, Díaz, Campos-Romo, & Fernandez-Ruiz, 2010; Thomas & French, 1985), while females perform better than males on some verbal and memory tasks (e.g., Hedges & Nowell, 1995; Hyde & Linn, 1988; Saykin et al., 1995) and measures of social cognition (Erwin et al., 1992, Williams et al., 2008). Some sex differences have been related to structural neuroimaging (e.g., De Bellis et al., 2001; Goldstein et al., 2001; Gur et al., 1999; Lenroot, et al., 2007) and functional imaging measures (e.g., Gur et al., 1982; 1995; 2000; Lenroot & Giedd, 2010), including volumetric differences in executive and memory-related areas, supporting neural substrates for sex differences in cognition. However, the developmental course of sex differences in brain-behavior relationship is still relatively unexamined.

Several shortcomings of most cognitive measures used in studies to date limit their applicability in establishing further links between brain function and behavioral domains. They tend to be broadly defined and load heavily on the "g factor" (Salthouse, 2004) and they do not separate accuracy from speed, which would preclude rigorous testing of hypotheses on the effects of brain connectivity on performance. Furthermore, the paper-and-pencil administration format precludes their use in studies that are scaled to the size required for neuroimaging genomics. More narrowly defined behavioral tasks, used in functional neuroimaging studies, have been adapted for use as computerized tests to obtain rapid and efficient quantification of individual differences (Gur, Erwin & Gur, 1992, Gur et al., 2001, 2010). No available studies have applied an identical neurocognitive test battery across a population ranging from childhood to adulthood in a prospective study aiming to integrate behavior with neuroimaging and genomics.

The present study is a collaboration between the Center for Applied Genomics (CAG) at Children's Hospital of Philadelphia (CHOP) and the Brain Behavior Laboratory at the University of Pennsylvania (Penn). The target sample consists of 10,000 youths ages 8–21 who consulted the CHOP network and volunteered to participate in genomic studies of complex pediatric disorders. All research participants undergo clinical assessment, including a neuropsychiatric structured interview, and review of electronic medical records. They are also administered a neuroscience based computerized neurocognitive battery (CNB) (Gur et al., 2001, 2010) and a subsample (1000) undergoes neuroimaging. The CNB, developed for large-scale studies, yields measures of accuracy and speed for domains of executive-control functions (abstraction, attention, working memory), episodic memory (verbal, facial, spatial), complex cognitive processing (language reasoning, nonverbal reasoning, spatial processing), social cognition (emotion identification, emotion intensity differentiation, age differentiation) and sensorimotor and motor speed. Here we present CNB testing results for the first 3,500 participants to evaluate its sensitivity and specificity for developmental age effects and sex differences. Importantly, establishing both age effects and sex differences in this age range required some adaptation of instructions and stimuli compared to earlier versions of the CNB.

In the standardization sample for the computerized battery (Gur et al., 2010), where the age range was 18–84, we found negative correlations between age and performance reflecting the sensitivity of the tests to senescence. Based on the available literature and the age range of 8 to 21 in this new sample, we hypothesized that performance will improve with increased age group, and that age effects are more pronounced for executive domains (e.g., Gao et al., 2011). Given the evidence of increased white matter volume and connectivity with age (e.g., Matsuzawa et al., 2001; Shaw et al., 2011; Deary, Penke & Johnson, 2010), we hypothesized that the age group effect will be pronounced for speed of sensory-motor response. We also hypothesized sex differences, with better performance in females on memory and social cognition tests and better performance for males on spatial and motor tests (Halpern et al., 2007). We did not have a basis to hypothesize age group × sex interaction effects on specific cognitive development, but since females mature earlier physically we expected to see earlier apex performance in females.

Method

Participants

The 3,500 participants, aged 8–21 years, were sampled from ~25,000 CHOP genotyped volunteers. The participants were from the greater Philadelphia area and selected at random after stratification by sex, age and ethnicity. Participants in the present study have been enrolled in the genetic study at CAG and they and their parents have provided informed consent (assent) to be re-contacted for additional information or participation in other studies. Participants were first mailed a letter that described the study, followed by a telephone call. The purpose of the phone call, which followed a prescribed script, was to establish that the potential participant is still interested in participation and is able to participate by meeting the following minimal *inclusion criteria*: 1. Able to provide signed informed consent. For participants under age 18 assent and parental consent were required. 2. English proficiency. 3. Physically and cognitively able to participate in an interview and computerized neurocognitive testing. The inclusion bar was set at a minimal level, just to make sure that the child can provide useful data, but children at this stage were not otherwise screened out for any specific medical or psychiatric disorder. Thus, the sample consists of children who came for pediatric care, gave blood for genomic studies, and consented to be contacted for future studies. Most came for primary care in one of the many CHOP-affiliated clinics throughout the Delaware Valley, but the sample could be somewhat enriched by children with more complicated illnesses including rare genetic disorders who

still met inclusion criteria. Thus, the sample was not screened for neurological or other deficits except for such that would result in damage severe enough to cause failure to meet the inclusion criteria (e.g., pervasive developmental disorder, mental retardation, or intracranial lesions that impact the sensory, motor or mental ability to be tested.) However, the sample is not enriched by people with behavioral disorders or those who seek out participation in research by responding to advertisements. In addition to the computerized cognitive battery described below, participants also received a structured clinical interview adapted from the Kiddie-SADS (Kaufman et al., 1997). A team of clinical investigators is scrutinizing the data and future analyses could be targeted toward specific diagnostic groups. Data from these interviews and electronic medical records will be used to evaluate the presence of subgroups with diagnosable medical or psychiatric disorders. The sample demographics are presented in Table 1.

Computerized Neurocognitive Battery

A one-hour computerized neurocognitive battery (CNB) was administered to participants using a system developed at Penn (Gur et al., 2010). The CNB consisted of 14 tests assessing 5 neurobehavioral functions: Executive-Control, Episodic Memory, Complex Cognition, Social Cognition, Sensorimotor Speed (Table 2). Except for the tests designed exclusively for measuring speed, each test provides measures of both accuracy and speed. Instructions and vocabulary for verbal stimuli were simplified from the adult CNB (Gur et al., 2010). Tests were also abbreviated based on psychometric analysis of data from other large-scale genomic studies, and the details of these analyses will be provided in separate manuscripts. The time saving allowed us to add two social cognition tests: Emotion Differentiation and Age Differentiation. A brief standardized reading test from the Wide Range Achievement Test (WRAT4, Wilkinson & Robertson, 2006) was administered first to determine participants' ability to complete the battery and to provide an estimate of IQ.

The following functional domains were assessed with the CNB using child appropriate versions of the language related tests.

Executive-Control

Abstraction and mental flexibility: Penn Conditional Exclusion Test is a measure of abstraction and concept formation. Participants decide which of 4 objects does not belong with the other 3, based on one of three sorting principles (e.g., size, shape, line thickness). The participant is guided by feedback and, after 10 successful trials demonstrating that the principle was solved, the principle is changed without informing the participant (Gur et al., 2010; Kurtz et al. 2004).

Attention: The Penn Continuous Performance Test presents 7-segment displays at a rate of 1/sec. The participant's task is to press the space bar whenever the display forms a digit (for the first half of the test) or a letter (for the second half of the test). The original Penn Continuous Performance Test (Kurtz et al., 2001, Gur et al., 2001, 2010) has been abbreviated from 6 minutes (3 minutes for digits, 3 for letters) to 3 minutes (1.5 minutes for each).

Working Memory: The Letter N-back test displays sequences of uppercase letters with stimulus duration of 500 ms (ISI 2,500 ms.) In the 0-back condition, participants respond to a single target (i.e., X). In the 1-back condition they respond if the letter is identical to that preceding it. In the 2-back condition, they respond if the letter is identical to that presented two trials back (Ragland et al., 2002, Gur et al., 2001, 2010).

Episodic Memory

Word Memory: The Penn Word Memory Test presents 20 target words that are then mixed with 20 foils equated for frequency, length, concreteness and imageability (Gur et al., 1997, 2001, 2010). The participants are asked to memorize the target words as they are presented (1/sec) and after the presentation of the target words they are asked to indicate whether a word presented was included in the target list on a 1 to 4 scale (definitely yes, probably yes, probably not, definitely not).

Face Memory: The Penn Face Memory Test presents 20 faces that are then mixed with 20 foils equated for age, sex and ethnicity (Gur et al., 1997, 2001, 2010). The presentation paradigm is otherwise identical to the verbal memory test.

Spatial Memory: The Visual Object Learning Test uses Euclidean shapes as stimuli with the same paradigm as the word and face memory (Glahn et al., 1997, Gur et al., 2001, 2010). The presentation paradigm is otherwise identical to the verbal and face memory tests.

Complex Cognition

Language and Analogical Reasoning: The Children's version of the Penn Verbal Reasoning Test consists of verbal analogy problems with simplified instructions and vocabulary (Gur et al., 1982, 2001, 2010).

Nonverbal Reasoning: The Penn Matrix Reasoning Test consists of matrices requiring reasoning by geometric analogy and contrast principles (Gur et al., 2010).

Spatial Processing: The Penn Line Orientation Test presents two lines at an angle, and participants click on a button that makes one line rotate until it has the same angle as the other. The relative location of the lines and their sizes differ across trials (Gur et al., 2010).

Social Cognition

Emotion Identification: The Penn Emotion Identification Test displays faces expressing one of 4 emotions (Happy, Sad, Anger, Fear) and Neutral faces, 8 each. The faces are presented one at a time, and the participant is asked to identify the emotion displayed from the set of four listed. The facial stimuli are balanced for sex, age, and ethnicity (Carter et al., 2008; Gur et al., 2002; RE Gur et al., 2006; Mathersul et al., 2008).

Emotion Differentiation: The Penn Emotion Differentiation Test presents pairs of emotional expressions, each pair obtained from the same individual expressing the same emotion, one more intense than the other or of equal intensity. Gradations of intensity were obtained by morphing a neutral to an emotionally intense expression and the difference between pairs of stimuli ranged between 10–60% of mixture. The task is to click on the face that displays the more intense expression or indicate that they have equal intensity. The same emotions are used as for the Emotion Identification test but the faces are different.

Age Differentiation: The Penn Age Differentiation Test requires the participant to select which of two presented faces appears older, or if they are the same age. The stimuli were generated by morphing a young person's face with that of an older person who has similar facial features. The stimuli vary by percent of difference in age (calculated based on the percentage contributed by the older face) and are balanced for sex and ethnicity.

Sensorimotor

Sensorimotor Processing Speed: The Motor Praxis task requires moving the mouse and clicking on a green square that disappears after the click. The square gets smaller and appears in unpredictable locations (Gur et al., 2001, 2010).

Motor: Motor speed is assessed with the Finger Tapping Test, which requires the participant to tap on the spacebar as quickly as possible for 10 sec with the index finger, alternating between dominant and nondominant hand for 5 trials (Gur et al., 2001, 2010).

The CNB was administered by assessors trained in a standard protocol. The four-hour didactic and hands-on training sessions included exercises in proctoring instruction, assignment of validity codes to indicate data quality, noting protocol deviation as well as technical and security issues. A training checklist including several mock administrations, an online test, and observation of CNB administrations was completed for each assessor. In addition, each assessor's first CNB was observed and rated on specific items and general rating categories including professionalism, ethics, standard administration, rapport and organization. Feedback was provided and a score of 88% was required for certification. Periodic rater reliability exercises were completed.

The web based platform for the CNB was developed using Perl CGI, HTML, a mySQL database and the Apache web server; tests were developed using Adobe Flash®. Scoring is fully automated. The web-based system was installed on Macbook Pro laptops and administered off-line in participants' homes or at the laboratory. Test data were later uploaded to Penn's web server. Data quality assurance comprised a three-stage validation process. Test administrators applied a data validity code to each test and noted anything that could affect validity of the data, including distractions, lack of participant effort, misunderstanding of the instructions and participant disabilities. Upon upload, data underwent an automated validation check for unusual response patterns or times, extreme scores, administrator notes or other indications of possible problems. About 85% of batteries were flagged as having one or more tests requiring further review. A trained validator, following specified guidelines and under supervision of a psychologist, reviewed the data and notes. About 1.6% of all tests administered were marked invalid after review, leading to the exclusion of 55 test scores due to unacceptable conditions or clear lack of cooperation, effort or understanding of the tasks.

Procedures

During an initial phone call, following up on a letter of invitation to participate in the study, participants were evaluated on whether they were interested in participation and met the inclusion criteria stipulated above. At that time logistical questions were addressed and the assessment session was scheduled at home (68.8% of participants) or in the laboratory (31.2%), according to the family's choice.

Participants met with a certified Research Coordinator who explained the research and its goals. The process included the child and the legal guardian (usually parents). After full explanation of research procedures and reading the consent form, written informed consent was obtained from the participant. For participants age <18, assent was obtained from the child/adolescent in addition to parental consent. These procedures were approved by the Institutional Review Boards of Penn and CHOP.

Standardized testing procedures were ensured with the tester sitting at the table next to the participant. Potential interference was minimized (e.g., no cell phones, television), standard instructions were read in addition to appearing on the screen and a professional testing environment was maintained. Tests were administered in a fixed order that was based on

previous experience and designed to maintain participants' engagement in the tasks and prevent fatigue. Breaks were offered approximately every 15 minutes. Tests requiring greater cognitive effort (e.g., Matrix Reasoning) were preceded or followed by a test that involved motor speed (e.g., Finger tapping) or by a break. Tests with colorful stimuli (e.g., Emotion Identification) were interspersed with monochromatic verbal stimuli. Tests requiring sustained attention (e.g., Continuous Performance Test) were placed so that they are followed by a break. Tasks with feedback regarding performance, which can be frustrating (e.g., Conditional Exclusion Test) were placed in the middle of the battery, when participants were "warmed up" but not yet fatigued, and were preceded by a break. The careful design of the battery resulted in high completion rates, even for the youngest age group: 99.5% of all participants completed the entire battery. Test order was: reading test (WRAT4), motor praxis, emotion identification, continuous performance; face memory, word memory, working memory; conditional exclusion, emotion differentiation, finger tapping, matrix reasoning; spatial memory, verbal reasoning, age differentiation, line orientation.

Data Analysis

Most tests yielded measures of accuracy and speed (response time). For the Penn Conditional Exclusion Test, the number of categories solved was multiplied by the proportion of correct choices, with 1 added to the number of categories to avoid a floor effect of 0 for individuals who did not solve any category. For the Penn Continuous Performance Test, number of true positive responses was used to calculate sensitivity. For the working memory test (Letter N-back), number of correct responses across conditions was a dependent measure of accuracy. For the episodic memory tests (Penn Word Memory Test, Penn Face Memory Test, Visual Object Learning Test), as well as the reasoning (Penn Verbal Reasoning Test, Penn Matrix Reasoning Test) and social cognition tests (Penn Emotion Identification Test, Penn Emotion Differentiation Test, Penn Age Differentiation Test), the number of correct responses was the measure of accuracy. Measures of speed included response time (in milliseconds) for correct responses on each test as well as response time for the motor praxis and number of taps on the motor test. Raw scores were standardized (z-transformed) based on the means and SDs of the entire sample. For ease of presentation, higher z-scores always reflect better performance; z-scores where higher numbers reflected poorer performance (i.e. response time) were multiplied by -1.

Accuracy and speed measures were evaluated separately. Participants were divided into 7 age groups formed using two-year intervals: 8–9 (163 males, 133 females), 10–11 (277 males, 244 females), 12–13 (227 males, 254 females), 14–15 (262 males, 269 females), 16–17 (254 males, 314 females), 18–19 (313 males, 506 females), 20–21 (101 males, 131 females). Age and sex factors were evaluated using a sex × age group ANOVA with test as a repeated measures factor. As the two largest ethnic groups included African Americans and Caucasians and because genomic analyses are conducted within ethnic groups to reduce heterogeneity, we also examined age differences in each ethnic group independently over both males and females. Here we report the data for the entire sample to optimize study power. Ethnic group data are reported separately in the Supplement.

Effect sizes were calculated for significant main effects using Cohen's d (Cohen, 1977), with the difference between the means in the numerator and the pooled standard deviation in the denominator. For age group effects, where more than two groups are included in the factor, d was calculated using the difference between the highest and lowest performing age group as numerator.

Results

Across sample performance and correlations with age

Two participants requested to withdraw from the study and their data were deleted leaving a final sample of 3448. Performance across the sample on each test and correlations with age and parental education (average of mother and father education) are provided in Table 3. The sample's estimated IQ based on the WRAT is average, with an age-adjusted mean scale score of ~101 and SD of ~15, indicating that our population-based sample has similar characteristics to that on which the WRAT was standardized. The expected correlations with age were substantial for the WRAT raw score (0.54) and, expectedly, low (−0.14) for the age-adjusted scale scores. The WRAT score also correlated with parental education (0.31 for raw and 0.42 for standardized). Of the computerized tests, correlations with age were generally positive for accuracy and negative for response time, reflecting improved accuracy and speed with age. However, the magnitude of the correlations had a wide range. The highest correlations with age were for speed of attention (0.52), motor speed (.49) and language-mediated reasoning accuracy (0.45). The lowest correlations with age were for spatial memory (0.01) and verbal memory (0.10). Notably, test accuracy scores had lower correlation with parental education than with participant age. In addition, all speed scores except the language-mediated reasoning had no association with parent education. Overall, the highest correlations with parent education were with the WRAT scale scores (0.42) and with accuracy for language-mediated reasoning (0.30) and nonverbal reasoning (0.23).

Analysis by age groups, sex and domains

The ANOVA on accuracy scores yielded main effects for age group, $F (6, 2906) = 133.28$, $p < .0001$ and domain, $F (11, 31966) = 3.62$, $p < .0001$, as well as interactions of domain × age group, $F (66, 31966) = 11.02$, $p < .0001$ and domain × sex, $F (11, 31966) = 19.78$, $p < .0001$. The sex × age group × domain interaction was also significant, $F (66, 31966) = 1.46$, $p = .0089$. Figure 1 illustrates the domain × sex interaction. As can be seen (upper panel), females across all age groups performed more accurately on several domains, notably attention and working memory, verbal and facial memory, and all social cognition measures, while males were more accurate on the abstraction and spatial memory tasks and substantially more accurate on the spatial processing task.

The ANOVA on speed scores yielded main effects for age group $F (6,2822) = 88.20$, $p < .0001$, and domain, $F (14, 39508) = 3.24$, $p < .0001$, as well as interactions of age group × sex, $F (6, 2822) = 2.52$, $p = .0197$, domain × age group, $F (84, 39508) = 23.90$, $p < .0001$, domain × sex, $F (14, 39508) = 21.03$, $p < .0001$, and domain × age group × sex, $F (84, 39508) = 1.90$, $p < .0001$. Figure 1 (bottom panel) illustrates the domain × sex interaction. Females were faster for verbal memory, language and emotion processing domains, while males were faster for working memory and motor domains.

Analysis of sub-domains—The accuracy and speed results for the univariate ANOVAS by sub-domains are presented in Table 4, which also presents effect-sizes. Figures 2–6 present data by age group and sex for the Executive, Episodic Memory, Complex Cognition, Social Cognition, and Sensorimotor speed, respectively. The hypothesized age group effect was significant for nearly all domains, with increased performance for both accuracy and speed. The exception was spatial memory, which did not show improved accuracy. Also as hypothesized, the greatest age group associated improvement was seen for Executive functions including attention and working memory where large effect sizes were evident both in accuracy and speed. However, large effect sizes were also notable for Complex Cognition and Social Cognition. Sensorimotor and Motor Speed had moderate and large effect sizes respectively. Effect sizes are on average in the large range for both accuracy

(mean ES 0.98 SDs, range .27 to 1.45 SDs) and speed (mean ES 0.93 SDs, range .29 to 1.84 SDs, excluding the reverse effect for nonverbal reasoning speed). Notably, spatial processing showed little age-related improvement in speed, and nonverbal reasoning showed decline in speed starting in adolescence. The speed decline is most likely a result of an inherent characteristic of the matrix-reasoning test, whereby the items range widely in difficulty and more difficult problems require greater processing time.

The hypothesized sex differences were evident for Episodic Memory, where females were more accurate and faster on verbal memory and more accurate in face memory. The effect sizes were small. The expected female advantage for Social Cognition was also supported with females showing better accuracy and speed with effect sizes ranging from 0.11 SD for speed of emotion differentiation to 0.33 SD for speed of emotion identification. Females also performed more accurately but more slowly on attention and were less accurate but faster in language and non-verbal reasoning. The expected better performance of males on the spatial test and on sensorimotor and motor speed was also supported by small effect sizes. Age group × sex interactions were noted only for spatial memory accuracy and speed, non-verbal reasoning accuracy and speed and all social cognition tests for speed. Motor speed also showed a significant interaction. As can be seen in Figures 3, 5 and 6, these interactions indicated that sex differences became more pronounced in the age groups following mid-adolescence. The interaction for spatial memory and nonverbal reasoning indicated that females became less accurate and faster in the older age groups. The interaction for speed on the Social Cognition tests indicated that while females continued to improve in speed into the last age group, males actually began to show decline in speed at the later age groups. The age group × sex interaction on the sensorimotor and motor speed measures indicated that males and females performed similarly until mid-adolescence and then males became faster at later age groups.

Discussion

The study examined age group effects and sex differences in performance in a large population-based sample of youths using a brief yet comprehensive set of tests that have been linked to circumscribed brain systems. This approach allowed the examination of hypotheses on age related effects on performance and their modulation by sex. Several findings stand out, some confirming hypotheses based on previous observations and some novel. With regard to age group effects, performance improved across domains, as expected, but there was substantial variability among domains in the extent and rate of improvement. As hypothesized, based on evidence for greatest maturational changes in frontal systems, Executive domains showed the largest differences between the youngest and oldest groups, with very large effect sizes (approaching or exceeding 1 standard deviation) for both accuracy and speed of attention and working memory. Both domains have been linked to prefrontal lobe function, a link demonstrated specifically for the tests used in this study (Gur et al., 2007; Ragland et al., 2002), where maturation is delayed relative to other brain regions (e.g., Giedd et al., 1999; Gogtay et al., 2004; Matsuzawa et al., 2001; Perrin et al., 2009; Pfefferbaum et al., 1994). On the other hand, abstraction and mental flexibility, also a frontal lobe domain, showed only a moderate effect size for improvement with age in accuracy and smaller effect size for speed. This domain could relate to aspects of frontal lobe functioning that mature earlier or can be supported by early maturing other brain systems.

Age related improvement in memory was considerably less pronounced than that for executive functions and was seen mostly in speed of response time rather than accuracy. The exception was face memory, which showed large effect sizes for both accuracy and speed. For verbal memory accuracy the effect size was small and spatial memory accuracy showed

no age group related improvement. Perhaps the developmental epoch sampled (based on mean age=14.8) has already missed the years of steeper developmental gains for memory, which is consistent with evidence for relatively early maturation of temporal lobe structures (Matsuzawa et al., 2001). Alternatively, ceiling effects could be a concern, since we lowered the level of word readability to accommodate the youngest age group. It is unlikely, however, to explain entirely the low correlation with age because we did find significant sex differences, and in all other domains age effects vastly overshadowed sex differences. Furthermore, the spatial memory test was as difficult as the face memory test, and showed no correlations with age. Thus, a conclusion that age effects are much less pronounced for accuracy of episodic memory than for the other domains seems justified.

Complex Cognition and Social Cognition domains showed comparable improvement that was more pronounced than for memory but not quite as pronounced as for Executive-control domains. For Complex Cognition, all domains showed improvement with age for accuracy, with very large effect sizes (all exceeding 1 standard deviation unit). For speed, the effect size was large only for language-mediated reasoning. It was small for spatial cognition, and for nonverbal (matrix) reasoning it was opposite in direction - response time for correct answers actually increased with age. Because items on the matrix-reasoning test vary greatly and appear in order of difficulty, this effect is likely due to success with more difficult items that take longer to solve. For Social Cognition, all age group related effect sizes were large for accuracy and moderate to large for speed. These results accord with evidence that heteromodal cortical association areas mature later than temporal cortex, but not quite as late as frontal lobe cortex (Gogtay et al., 2004).

As hypothesized, speed itself also improved with age even in the absence of a cognitive task. Indeed, improved speed was more pronounced for a simple motor task than for the task requiring sensorimotor integration. While there is evidence that the sensorimotor strip and cerebellum are among the earliest to mature (Gogtay et al., 2004; Yakovlev & Lecours, 1967), the present results suggest that performance continues to improve into early adulthood. Perhaps the continued physical growth interacts with brain maturation in affecting motor performance.

Sex differences were apparent both in overall performance and in age group related variation. However, these effect sizes were small compared to age group effects. The hypothesized differences favoring females for memory and social cognition tests and males on spatial and motor tests were strongly supported. Females performed more accurately and faster for the verbal memory test and more accurately for face memory, although they were less accurate on spatial memory, as has been reported in earlier studies (Saykin et al., 1995). Females were both more accurate and faster on all social cognition tests. These effects are in line with earlier reports of better performance in females for emotion processing tasks (Gur et al., 2010; Williams et al., 2008), but extend them to other social cognition measures and across a wide developmental epoch. On the other hand, males were more accurate on the spatial test and were faster on both sensorimotor and motor speed (Coleman et al., 1997; Gur et al., 1999, 2001; Halpern et al., 2007; Moreno-Briseño, Díaz, Campos-Romo, & Fernandez-Ruiz, 2010; Thomas & French, 1985). Some sex differences were unexpected. Thus, males were more accurate in abstraction and mental flexibility (a very small effect size) and females were more accurate and slower for attention and slower for working memory. The small size of the effects may explain why they have not been reported in smaller samples. Poorer accuracy in males for attention is consistent with the higher incidence of attention deficit disorder in males (Ramtekkar, Reiersen, Todorov, & Todd, 2010). Whether this difference can be explained by more males with attention deficit disorder in the current sample will be clarified when results of the clinical assessments are incorporated.

There were few age group × sex interactions. These interactions were noted only for spatial memory accuracy and speed, nonverbal reasoning accuracy and speed and all social cognition tests on speed. Motor speed also showed a significant interaction. All these interactions indicated that sex differences became more pronounced in the age groups following mid-adolescence. Across all domains, except for memory, females reached plateau before males. This finding accords with physical (Hills & Byrne, 2010), behavioral (Keulers, Evers, Stiers, & Jollies, 2010; Greenstein, Blachstein, & Vakil, 2010; Review in Yurgelun-Todd, D., 2007) and neuroimaging (Bramen et al., 2010; Tieneier, et al., 2010) data indicating earlier maturation in females. The exception in our study is for memory, where males peaked by age 18–19 whereas females continued to improve in word and face memory into the 20–21 age group. Age group × sex interactions for complex and social cognition were seen in speed, where females continued to improve while males reached a plateau in mid adolescence and then showed decline. While we are unaware of earlier studies where both social cognition and a broad range of other neurobehavioral domains have been examined across this age range for both accuracy and speed, our findings generally comport with studies examining developmental sex differences in comparable domains (e.g., Reynolds, Keith, Ridley, & Patel, 2008).

Beyond the specific findings, the present results indicate the feasibility of administering a brief yet comprehensive computerized neuropsychological battery of identical tests to a large population-based sample of children, adolescents and young adults and obtaining informative data. The testing yielded a large proportion of validated data of high quality with information pertinent to major behavioral domains that can be linked to brain function and genomics. Making this link requires adequate information on developmental effects because, as our results demonstrate, these effects are substantial and modulated by sex. Thus, while most genomic variation is fixed, the present results underscore that neurobehavioral phenotypes require demographic information to be interpretable.

Limitations and Future Directions

The study has several potential limitations. While the sample is large and demographically diverse, we have not excluded individuals with some medical conditions that may affect performance and skew some of the data. An issue can be raised concerning the extent to which the sample represents the general population. The sample was not intentionally enriched for any specific disorders. The incentive for participation was that the children gave blood for genomic analysis when they saw their pediatrician for a well-child visit or any other reason, and agreed to be re-contacted for participation in other studies. Thus, it is perhaps not as representative as a census based random sample, but more representative than a convenience sample of responders to advertisement for psychological experiments. In this regard, it is reassuring that the average WRAT score in our sample is identical to that obtained in the normative sample (Mean ~100, SD ~15). When the final sample is obtained and the results integrated with electronic medical records and updated with clinical evaluations we performed, we will have sufficient power to examine the effects of potential disease traits. Notably, the integration of neurobehavioral data with the neuropsychiatric assessment will also enable us to evaluate neurobehavioral profiles associated with psychiatric symptoms. The availability of only part of the final planned sample also presents unequal sample sizes for males and females across the age groups. Since the standardization was based on the entire sample, conceivably some effects in specific age groups could be imprecisely estimated. However, standardization across the sample is the most straightforward and easily interpretable approach and the large sample size should minimize systematic distortion of parameter estimates. Another limitation of the study is the modest number of tests administered in each neurobehavioral domain. This limitation was imposed by feasibility given the scope of the study. Obviously, the tests do not reflect the depth and

complexity of the domains sampled. For example, there is more to episodic memory than recognition of words, faces and shapes. We are also limited in the extent to which social cognition is measured and the battery does not include any auditory measures such as prosody, or other aspects of social cognition. Future longitudinal follow-up could expand to other measures pertinent to pursuing specific hypotheses in subsamples. Finally, the present study has examined a limited number of measures within each of the tests. There are alternative indices that can be derived for most tests and that can yield interesting information on cognitive strategies. For example, signal detection parameters can be applied to the attention and memory measures. Here we present data on a broad range of domains, and to contain Type I error we have limited our examination to one measure for each domain. However, the computerized format allows effective evaluations of multiple indices and relations with response time in studies focusing on individual domains.

Notwithstanding these limitations, the present large-scale study demonstrates the feasibility of obtaining reliable neuropsychological data that offer information on age and sex differences. These data can be integrated with phenotypic measures in specific behavioral domains associated with developmental disorders, with neuroimaging data on brain structure and function and with genomic parameters that can propel findings that bridge between molecular and behavioral processes. Such integration has the potential of unveiling specific genes and gene networks that drive developmental traits as well as genetic variants that may underlie key individual differences, including those related to sex differences and ethnic groups. While still underpowered at this stage, the results from future genomic analyses may contribute to understanding normal and abnormal development, with implications for intervention approaches.

Acknowledgments

References

Ang SY, Lee K. Central executive involvement in children's spatial memory. Memory. 2008; 16:918–933. [PubMed: 18802804]

Ang SY, Lee K. Exploring developmental differences in visual short-term memory and working memory. Developmental Psychology. 2010; 46:279–285. [PubMed: 20053024]

Bramen JE, Hranilovich JA, Dahl RE, Forbes EE, Chen J, Toga AW, Dinov ID, Worthman CM, Sowell ER. Puberty influences medial temporal lobe and cortical gray matter maturation differently in boys than girls matched for sexual maturity. Cerebral Cortex. 2011; 21:636–646. [PubMed: 20713504]

Burnett S, Sebastian C, Cohen Kadosh K, Blakemore SJ. The social brain in adolescence: Evidence from functional magnetic resonance imaging and behavioural studies. Neuroscience Biobehavioral Review. 2011 [Epub ahead of print].

Carter CS, Barch DM, Gur RC, Gur RE, Pinkham A, Ochsner K. CNTRICS Final Task Selection: Social Cognitive and Affective Neuroscience-Based Measures. Schizophrenia Bulletin. 2008; 35:153–162. [PubMed: 19011231]

Casey BJ, Duhoux S, Malter, Cohen M. Adolescence: what do transmission, transition, and translation have to do with it? Neuron. 2010; 67:749–760. [PubMed: 20826307]

Cohen, D. Statistical power analysis for the behavioral sciences. New York: Academic Press; 1977.

Coleman AR, Moberg PJ, Ragland JD, Gur RC. Comparison of the Halstead-Reitan and infrared light beam finger tappers. Assessment. 1997; 4:277–286.

Conklin HM, Luciana M, Hooper CJ, Yarger RS. Working memory performance in typically developing children and adolescents: behavioral evidence of protracted frontal lobe development. Developmental Neuropsychology. 2007; 31:103–128. [PubMed: 17305440]

De Bellis MD, Keshavan MS, Beers SR, Hall J, Frustaci K, Masalehdan A, et al. Sex differences in brain maturation during childhood and adolescence. Cerebral Cortex. 2001; 11:552–557. [PubMed: 11375916]

Deary IJ, Penke L, Johnson W. The neuroscience of human intelligence differences. Nature Reviews Neuroscience. 2010; 11:201–211.

Erwin RJ, Gur RC, Gur RE, Skolnick B, Mawhinney-Hee M, Smailis J. Facial emotion discrimination: I. Task construction and behavioral findings in normal subjects. Psychiatry Research. 1992; 42:231–240. [PubMed: 1496055]

Friederici AD, Wartenburger I. Language and Brain. Wiley Interdisciplinary Reviews: Cognitive Science. 2010; 1:150–159.

Glahn DC, Gur RC, Ragland JD, Censits DM, Gur RE. Reliability, performance characteristics, construct validity, and an initial clinical application of a visual object learning test (VOLT). Neuropsychology. 1997; 11:602–612. [PubMed: 9345704]

Gow AJ, Johnson W, Pattie A, Brett CE, Roberts B, Starr JM, Deary IJ. Stability and change in intelligence from age 11 to ages 70, 79, and 87: the Lothian Birth Cohorts of 1921 and 1936. Psychology and Aging. 2011; 26:232–240. [PubMed: 20973608]

Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, Paus T, Evans AC, Rapoport JL. Brain development during childhood and adolescence: A longitudinal MRI study. Nature Neuroscience. 1999; 2:861–863.

Gogtay N, Giedd JN, Lusk L, Hayashi KM, Greenstein D, Vaituzis AC, et al. Dynamic mapping of human cortical development during childhood through early adulthood. Proceedings of the National Academy of Sciences USA. 2004; 101:8174–8179.

Goldberg MC, Maurer D, Lewis TL. Developmental changes in attention: The effects of endogenous cueing and of distractors. Developmental Science. 2001; 4:209–219.

Goldstein JM, Seidman LJ, Horton NJ, Makris N, Kennedy DN, Caviness VS Jr, Faraone SV, Tsuang MT. Normal sexual dimorphism of the adult human brain assessed by in vivo magnetic resonance imaging. Cerebral Cortex. 2001; 1:490–497. [PubMed: 11375910]

Greenstein Y, Blachstein H, Vakil E. Interrelations between attention and verbal memory as affected by developmental age. Child Neuropsychology. 2010; 16:42–59. [PubMed: 19657785]

Gur RC, Alsop D, Glahn D, Petty R, Swanson CL, Maldjian JA, Turetsky BI, Detre JA, Gee J, Gur RE. An fMRI study of sex differences in regional activation to a verbal and a spatial task. Brain and Language. 2000; 74:157–170. [PubMed: 10950912]

Gur RC, Erwin RJ, Gur RE. Neurobehavioral probes for physiologic neuroimaging studies. Archives of General Psychiatry. 1992; 49:409–414. [PubMed: 1586277]

Gur RC, Gur RE, Obrist WD, Hungerbuhler JP, Younkin D, Rosen AD, Skolnick BE, Reivich M. Sex and handedness differences in cerebral blood flow during rest and cognitive activity. Science. 1982; 217:659–661. [PubMed: 7089587]

Gur RC, Mozley LH, Mozley PD, Resnick SM, Karp JS, Alavi A, Arnold SE, Gur RE. Sex differences in regional cerebral glucose metabolism during a resting state. Science. 1995; 267:528–531. [PubMed: 7824953]

Gur RC, Ragland JD, Moberg PJ, Turner TH, Bilker WB, Kohler C, Siegel SJ, Gur RE. Computerized Neurocognitive Scanning: I. Methodology and validation in healthy people. Neuropsychopharmacology. 2001; 25:766–776. [PubMed: 11682260]

Gur RC, Ragland JD, Mozley LH, Mozley PD, Smith R, Alavi A, Bilker WB, Gur RE. Lateralized changes in regional cerebral blood flow during performance of verbal and facial recognition tasks: correlations with performance and "effort". Brain and Cognition. 1997; 33:388–414. [PubMed: 9126402]

Gur RC, Richard J, Hughett P, Calkins ME, Macy L, Bilker WB, Brensinger C, Gur RE. A cognitive neuroscience-based computerized battery for efficient measurement of individual differences:

standardization and initial construct validation. Journal of Neuroscience Methods. 2010; 187:254–262. [PubMed: 19945485]

Gur RC, Sara R, Hagendoorn M, Marom O, Hughett P, Macy L, Turner T, Bajcsy R, Posner A, Gur RE. A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. Journal of Neuroscience Methods. 2002; 115:137–143. [PubMed: 11992665]

Gur RC, Turetsky BI, Matsui M, Yan M, Bilker WB, Hughett P, Gur RE. Sex differences in brain gray and white matter in healthy young adults. Journal of Neuroscience. 1999; 19:4065–4072. [PubMed: 10234034]

Gur RC, Turetsky BI, Loughead J, Waxman J, Snyder W, Ragland JD, Elliott MA, Bilker WB, Arnold SE, Gur RE. Hemodynamic responses in neural circuitries for detection of visual target and novelty: An event-related fMRI study. Human Brain Mapping. 2007; 28:263–274. [PubMed: 17133387]

Gur RE, Kohler CG, Ragland JD, Siegel SJ, Lesko K, Bilker WB, Gur RC. Flat affect in schizophrenia: Relation to emotion processing and neurocognitive measures. Schizophrenia Bulletin. 2006; 32:279–287. [PubMed: 16452608]

Halpern DF, Benbow CP, Geary DC, Gur RC, Hyde JS, Gernsbacher MA. The science of sex differences in science and mathematics. Psychological Science in the Public Interest. 2007; 8:1–51.

Hedges LV, Nowell A. Sex differences in mental test scores, variability, and numbers of high-scoring individuals. Science. 1995; 269:41–45. [PubMed: 7604277]

Hills AP, Byrne NM. An overview of physical growth and maturation. Medicine and Sport Science. 2010; 55:1–13. [PubMed: 20956856]

Hines M. Sex-related variation in human behavior and the brain. Trends in Cognitive Science. 2010; 14:448–456.

Hyde JS, Linn MC. Gender differences in verbal ability: A meta-analysis. Psychological Bulletin. 1988; 104:53–69.

Jung RE, Haier RJ. The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. Behav. Brain Sci. 2007; 30:135–154. [PubMed: 17655784]

Karama S, Colom R, Johnson W, Deary IJ, Haier R, Waber DP, Lepage C, Ganjavi H, Jung R, Evans AC. the Brain Development Cooperative Group. Cortical thickness correlates of specific cognitive performance accounted for by the general factor of intelligence in healthy children aged 6 to 18. Neuroimage. 2011; 55:1443–1453. [PubMed: 21241809]

Karama S, Ad-Dab'bagh Y, Haier RJ, Deary IJ, Lyttelton OC, Lepage C, Evans AC. the Brain Development Cooperative Group. Positive association between cognitive ability and cortical thickness in a representative US sample of healthy 6 to 18 year-olds. Intelligence. 2009; 37:145–155. [PubMed: 20161325]

Kaufman J, Bismaher B, Brent DA, Rao U, Flynn C, Moreci P, et al. Schedule for Affective Disorders and Schizophrenia for School-Aged Children - Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. Journal of the American Academy of Child and Adolescent Psychiatry. 1997; 36:980–988. [PubMed: 9204677]

Keith RW. Development and standardization of SCAN-C Test for Auditory Processing Disorders in Children. Journal of the American Academy of Audiology. 2000; 11:438–445. [PubMed: 11012239]

Keulers EHH, Evers EAT, Stiers P, Jolles J. Age, Sex pubertal phase influence mentalizing about emotions and actions in adolescents. Developmental Neuropsychology. 2010; 35:555–569. [PubMed: 20721775]

Kuhl PK. Brain mechanisms in early language acquisition. Neuron. 2010; 67:713–727. [PubMed: 20826304]

Kurtz MM, Ragland JD, Bilker WB, Gur RC, Gur RE. Comparison of the continuous performance test with and without working memory demands in healthy controls and patients with schizophrenia. Schizophrenia Research. 2001; 48:307–316. [PubMed: 11295383]

Kurtz MM, Ragland JD, Moberg PJ, Gur RC. The Penn Conditional Exclusion Test: a new measure of executive-function with alternate forms for repeat administration. Archives of Clinical Neuropsychology. 2004; 19:191–201. [PubMed: 15010085]

Lenroot RK, Gogtay N, Greenstein DK, Wells EM, Wallace GL, Clasen LS, et al. Sexual dimorphism of brain developmental trajectories during childhood and adolescence. Neuroimage. 2007; 36:1065–1073. [PubMed: 17513132]

Lenroot RK, Giedd JN. Sex differences in the adolescent brain. Brain and Cognition. 2010; 72:46–55. [PubMed: 19913969]

Linn MC, Petersen AC. Emergence and characterization of sex differences in spatial ability: a meta-analysis. Child Development. 1985; 56:1479–1498. [PubMed: 4075870]

Mathersul D, Palmer DM, Gur RC, Gur RE, Cooper N, Gordon E, Williams LM. Explicit identification and implicit recognition of facial emotions: II. Core domains and relationships with general cognition. Journal of Clinical and Experimental Neuropsychology. 2008; 19:1–14.

Matsuzawa J, Matsui M, Konishi T, Noguchi N, Gur RC, Bilker W, Miyawaki T. Age-related volumetric changes of brain gray and white matter in healthy infants and children. Cerebral Cortex. 2001; 11:335–342. [PubMed: 11278196]

Moreno-Briseño P, Díaz R, Campos-Romo A, Fernandez-Ruiz J. Sex-related differences in motor learning and performance. Behavioral and Brain Function. 2010; 6:74–77.

Perrin JS, Leonard G, Perron M, Pike GB, Pitiot A, Richer L, et al. Sex differences in the growth of white matter during adolescence. Neuroimage. 2009; 45:1055–1066. [PubMed: 19349224]

Pfefferbaum A, Mathalon DH, Sullivan EV, Rawles JM, Zipursky RB, Lim KO. A quantitative magnetic resonance imaging study of changes in brain morphology from infancy to late adulthood. Archives of Neurology. 1994; 51:874–887. [PubMed: 8080387]

Pickering SJ. The development of visuo-spatial working memory. Memory. 2001; 9:423–432.

Ragland JD, Turetsky BI, Gur RC, Gunning-Dixon F, Turner T, Schroeder L, Chan R, Gur RE. Working memory for complex figures: an fMRI comparison of letter and fractal n-back tasks. Neuropsychology. 2002; 16:370–379. [PubMed: 12146684]

Ramtekkar UP, Reiersen AM, Todorov AA, Todd RD. Sex and age differences in attention-deficit/hyperactivity disorder symptoms and diagnoses: implications for DSM-V and ICD-11. Journal of the American Academy of Child and Adolescent Psychiatry. 2010; 49:217–228. [PubMed: 20410711]

Reynolds MR, Keith TZ, Ridley KP, Patel PG. Sex differences in latent general and broad cognitive abilities for children and youth: Evidence from higher-order MG-MACS and MIMIC models. Intelligence. 2008; 36:236–260.

Salthouse TA. Localizing age-related individual differences in a hierarchical structure. Intelligence. 2004; 32:541–561.

Saykin AJ, Gur RC, Gur RE, Shtasel DL, Flannery KA, Mozley LH, Malamut BL, Watson B, Mozley PD. Normative neuropsychological test performance: effects of age, education, gender and ethnicity. Applied Neuropsychology. 1995; 2:79–88. [PubMed: 16318528]

Shaw DJ, Grosbras MH, Leonard G, Pike GB, Paus T. Development of the action observation network during early adolescence: a longitudinal study. Social Cognitive Affective Neuroscience. 2011 Jan 28. [Epub ahead of print].

Shaw P, Kabani NJ, Lerch JP, Eckstrand K, Lenroot R, Gogtay N, et al. Neurodevelopmental trajectories of the human cerebral cortex. Journal of Neuroscience. 2008; 28:3586–3594. [PubMed: 18385317]

Thomas JR, French KE. Gender differences across age in motor performance a meta-analysis. Psychological Bulletin, 1985. 1985; 98:260–282.

Tiemeier H, Lenroot RK, Greenstein DK, Tran L, Pierson R, Giedd JN. Cerebellum development during childhood and adolescence: a longitudinal morphometric MRI study. Neuroimage. 2010; 49:63–70. [PubMed: 19683586]

Voyer D, Voyer S, Bryden MP. Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. Psychological Bulletin. 1995; 117:250–270. [PubMed: 7724690]

Wilkinson, GS.; Robertson, GJ. Psychological Assessment Resources. Lutz, FL: 2006. Wide Range Achievement Test--Fourth Edition. WRAT4 Introductory Kit

Williams LM, Mathersul D, Palmer DM, Gur RC, Gur RE, Gordon E. Explicit identification and implicit recognition of facial emotions: I. Age effects in males and females across 10 decades. Journal of Clinical and Experimental Neuropsychology. 2008; 19:1–21.

Yakovlev, PL.; Lecours, AR. The myelogenetic cycles of regional maturation of the brain. In: Minkowski, A., editor. Regional Development of the Brain in Early Life. Oxford: Blackwell; 1967. p. 3-70.

Yurgelun-Todd D. Emotional and cognitive changes during adolescence. Current Opinion in Neurobiology. 2007; 17:251–257. [PubMed: 17383865]

Figure 1.
Means (±SE) of z-scores for accuracy (top panel) and speed (bottom panel) for females (red bars) and males (blue bars) across the sample on each behavioral domain. ABF=Abstraction and Mental Flexibility; ATT=Attention; WM=Working Memory; VME=Verbal Memory; FME=Face Memory; SME=Spatial Memory; LAN= Language Reasoning; NVR=Nonverbal Reasoning; SPA=Spatial Processing; EMI= Emotion Identification; EMD=Emotion Differentiation; AGD=Age Differentiation; SM=Sensorimotor Speed; MOT=Motor Speed. Stars indicate significant (p<.05) sex differences, red stars indicate better performance in females, blue stars better performance in males.

Figure 2.
Means (±SE) of z-scores for accuracy (top panel) and speed (bottom panel) for females (red bars) and males (blue bars) on executive-control functions. The results are shown by age groups and correlations with age are provided by sex (bold correlations are significant at p<.001). Stars indicate significant (p<.05) sex differences, red stars indicate better performance in females, blue stars better performance in males.
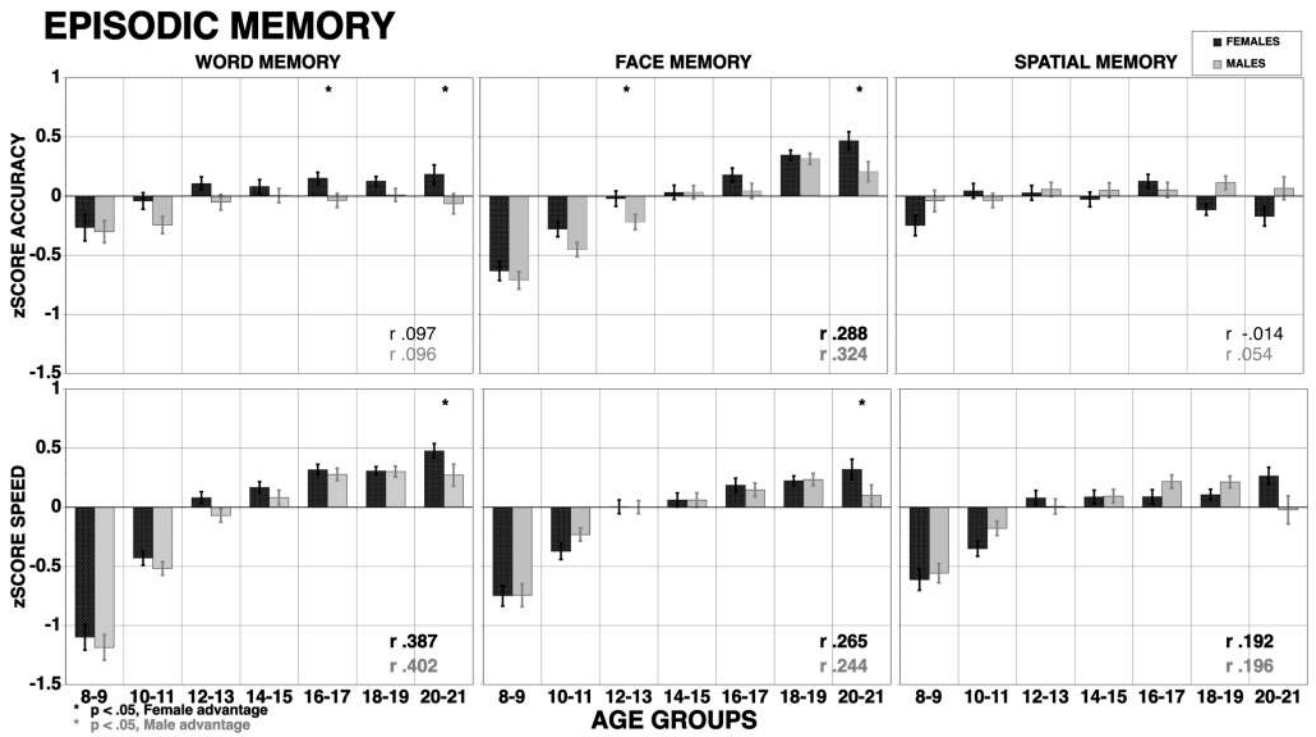
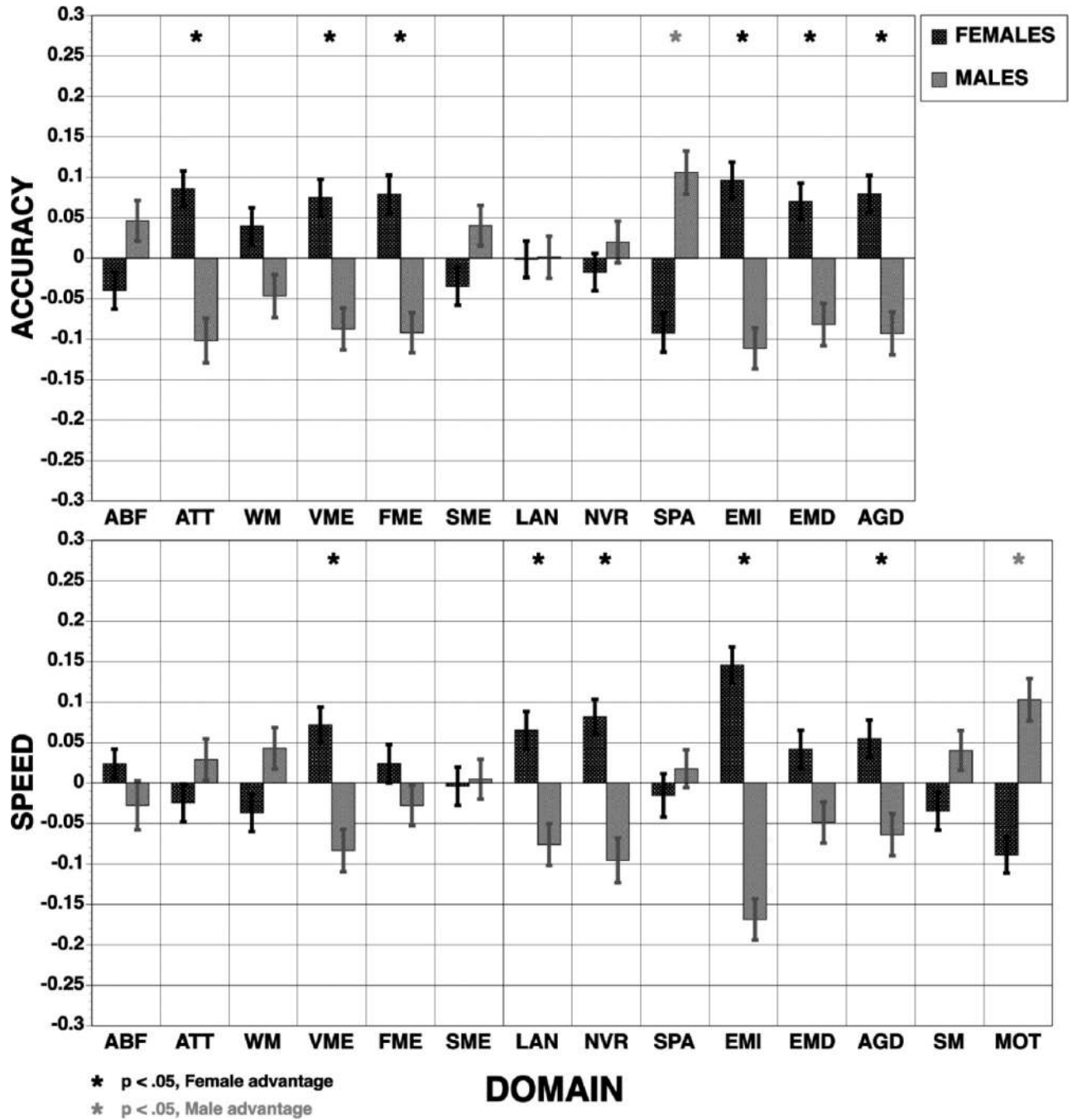**Figure 3.**
Means (±SE) of z-scores for accuracy (top panel) and speed (bottom panel) for females (red bars) and males (blue bars) on memory functions. The results are shown by age groups and correlations with age are provided by sex (bold correlations are significant at p<.001). Stars indicate significant (p<.05) sex differences, red stars indicate better performance in females, blue stars better performance in males.
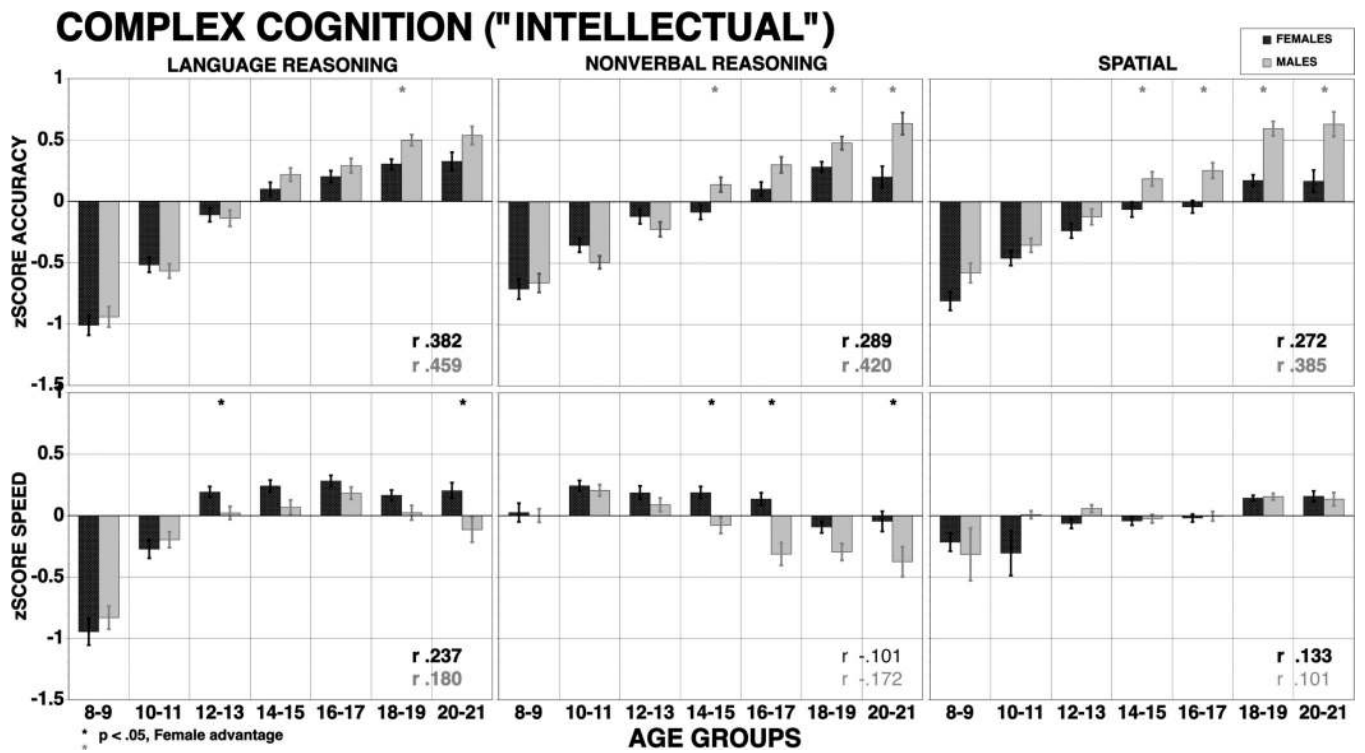
Figure 4.
Means (±SE) of z-scores for accuracy (top panel) and speed (bottom panel) for females (red bars) and males (blue bars) on complex cognitive tasks. The results are shown by age groups and correlations with age are provided by sex (bold correlations are significant at p<.001). Stars indicate significant (p<.05) sex differences, red stars indicate better performance in females, blue stars better performance in males.
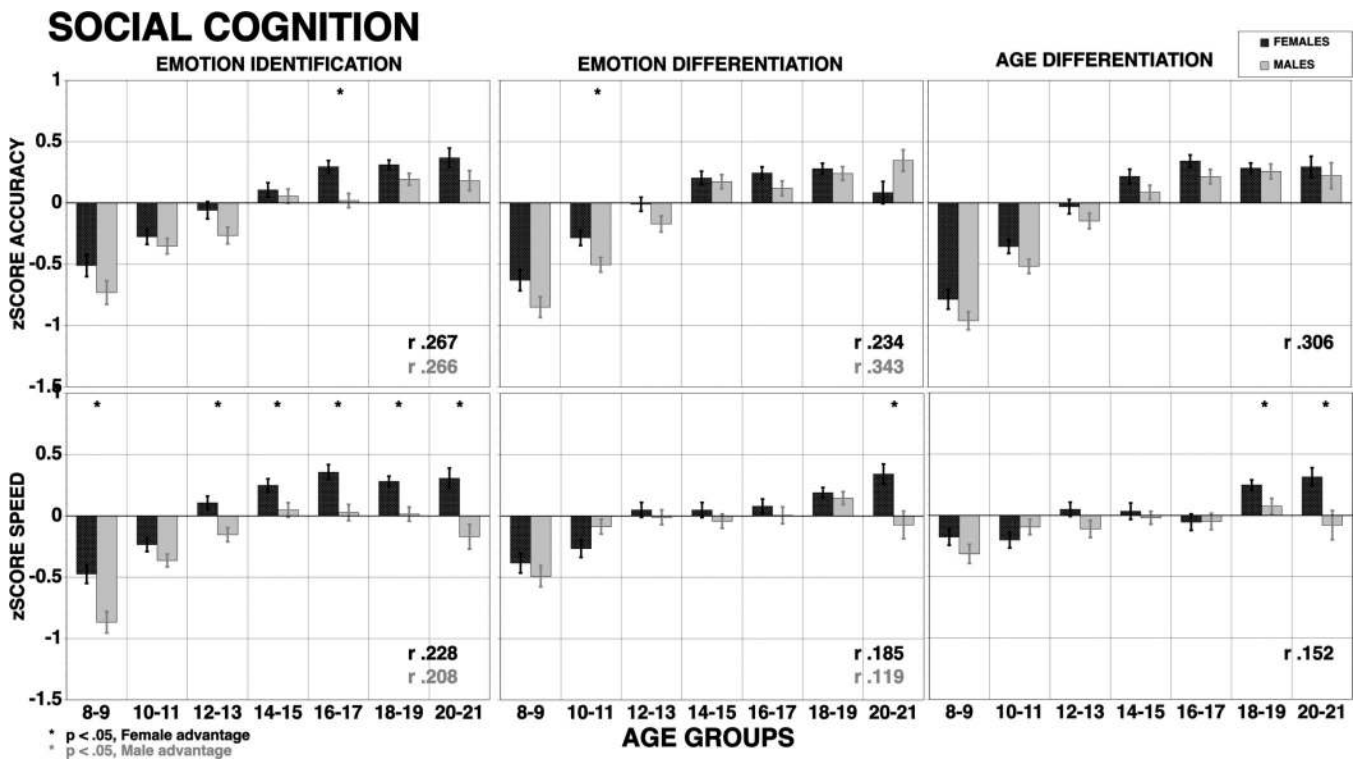
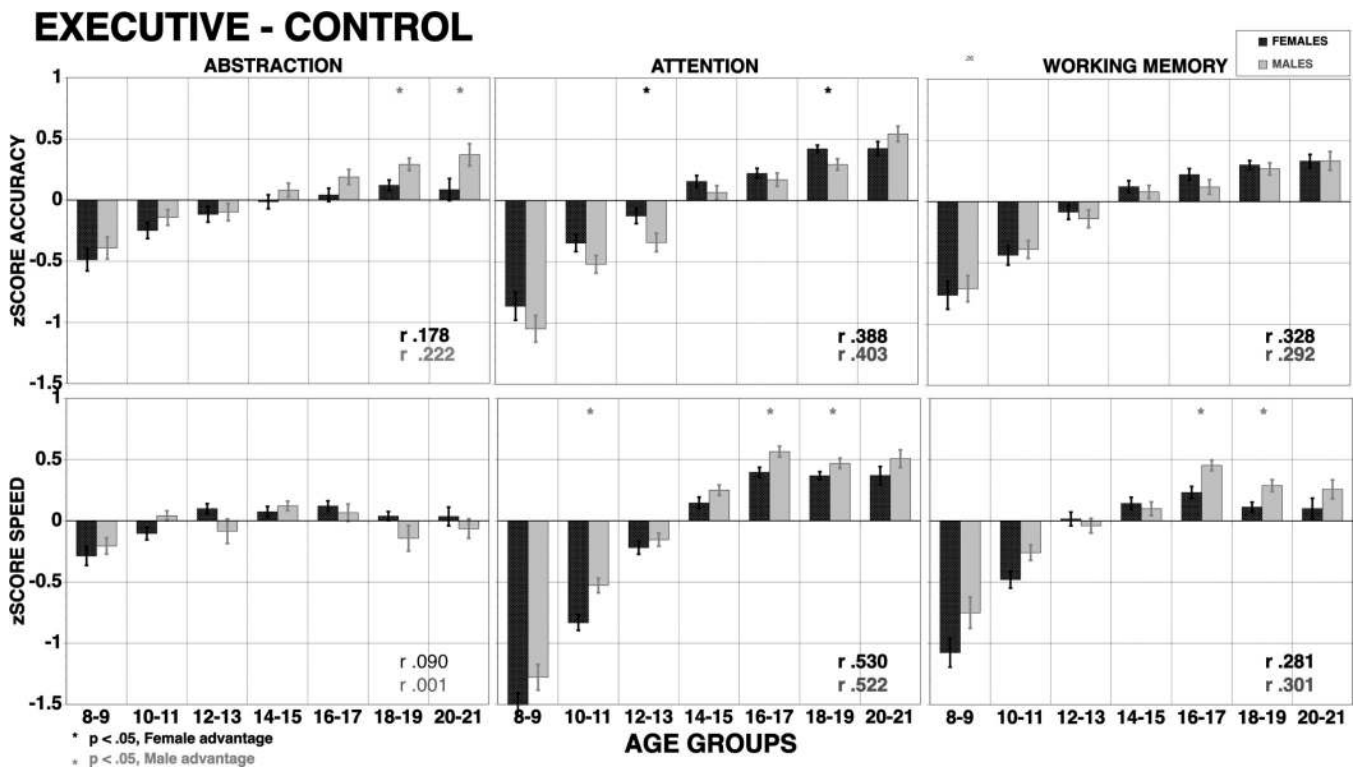**Figure 5.**
Means (±SE) of z-scores for accuracy (top panel) and speed (bottom panel) for females (red bars) and males (blue bars) on social cognition. The results are shown by age groups and correlations with age are provided by sex (bold correlations are significant at p<.001). Stars indicate significant (p<.05) sex differences, red stars indicate better performance in females, blue stars better performance in males.
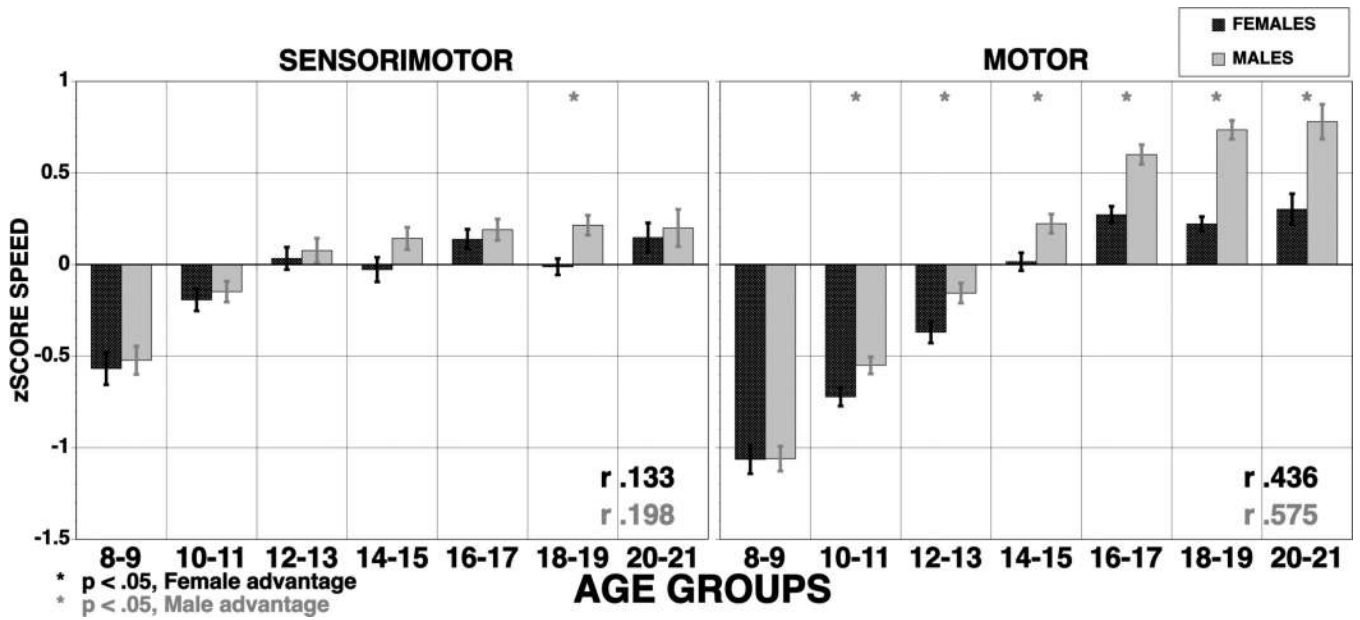
Figure 6.
Means (±SE) of z-scores for sensorimotor processing and motor speed for females (red bars) and males (blue bars). The results are shown by age groups and correlations with age are provided by sex (bold correlations are significant at p<.001). Stars indicate significant (p<.05) sex differences, red stars indicate better performance in females, blue stars better performance in males.

Table 1

Sample Demographics.

	Males		Females		Total	
Caucasian, Not Hispanic	950	27.6%	1024	29.7%	1974	57.3%
Caucasian, Hispanic	15	0.4%	18	0.5%	33	1.0%
African American, Not Hispanic	531	15.4%	707	20.5%	1238	35.9%
African American, Hispanic	6	0.2%	9	0.3%	15	0.4%
Asian, Not Hispanic	8	0.2%	16	0.5%	24	0.7%
Mixed/Other Race, Not Hispanic	64	1.9%	48	1.4%	112	3.2%
Mixed/Other Race, Hispanic	23	0.7%	29	0.8%	52	1.5%
Total	1597	46.3%	1851	53.7%	3448	100.0%
Mean Age (SD)	14.4	(3.6)	15.1	(3.5)	14.8	(3.5)

Table 2

Computerized Neurocognitive Battery Domains and Tests.

		Test Name	Label	Admin Time (minutes)	
				Mean	SD
	Reading level	WRAT4 Reading Test	WRAT4	3.2	2.2
Neurobehavioral functions	**Domain**				
Executive-Control	ABF (Abstraction/flexibility)	Penn Conditional Exclusion Test	PCET	4.1	2.6
	ATT (Attention)	Penn Continuous Performance Test	PCPT	5.3	0.5
	WM (Working memory)	Letter N-Back	LNB	9.8	1.8
Episodic Memory	VME (Verbal Memory)	Penn Word Memory	PWMT	3.3	0.4
	FME (Face Memory)	Penn Face Memory	PFMT	3.9	0.5
	SME (Spatial Memory)	Visual Object Learning Test	VOLT	2.1	0.4
Complex Cognition	LAN (Language Reasoning)	Penn Verbal Reasoning Test	PVRT	2.7	1
	NVR (Nonverbal reasoning)	Penn Matrix Reasoning Test	PMAT	4.6	3
	SPA (Spatial ability)	Penn Line Orientation Test	PLOT	5.2	1.3
Social Cognition	EMI (Emotion Identification)	Penn Emotion Identification Test	PEIT	2.3	0.5
	EMD (Emotion Differentiation)	Penn Emotion Differentiation Test	PEDT	3.1	0.8
	AGD (Age Differentiation)	Penn Age Differentiation Test	PADT	2.6	0.8
Sensorimotor	SM (Sensorimotor speed)	Motor Praxis	MP	1.5	1.2
	MOT (Motor)	Finger Tapping Test	TAP	3.7	0.3
			Total Duration	57.4	

Table 3

Means, standard deviations (SD), minimum and maximum scores on the tests administered and correlations with age and mean parental education (PAR ED).

		MEAN	SD	MIN	MAX	Corr With AGE	Corr With PAR ED
	Age	14.76	3.54	8	21		
	Parent education	14.25	2.29	8	20		
	Test						
	WRAT RAW	52.32	9.23	20	70	0.536	0.311
	SCALE	101.27	15.62	59	145	−0.138	0.417
ACCURACY	PCET	1.97	0.69	0.02	3.43	0.193	0.164
	PCPT	52.40	7.21	11	60	0.399	0.119
	LNB	17.97	2.12	4	20	0.313	0.174
	PWMT	37.18	2.83	19	40	0.103	0.087
	PFMT	31.16	4.08	16	40	0.310	0.126
	VOLT	15.57	2.32	4	20	0.014	0.159
	PVRT	76.52	17.82	13	100	0.450	0.295
	PMAT	100.96	13.94	49	132	0.347	0.230
	PLOT	42.30	20.10	0	100	0.313	0.180
	PEIT	32.17	3.32	14	40	0.271	−0.029
	PEDT	100.25	13.64	51	138	0.292	0.133
	PADT	99.40	13.40	48	135	0.335	0.062
RT	PCET	2414.1	1062.8	857	32049	−0.041	−0.098
	PCPT	483.3	59.9	307	829	−0.521	0.021
	LNB	533.9	128.5	316	1485	−0.285	−0.041
	PWMT	1456.6	291.1	928	4070	−0.399	−0.028
	PFMT	1926.3	446.2	1073	5947	−0.257	0.032
	VOLT	1761.7	398.1	838	4421	−0.192	0.029
	PVRT	6094.5	2148.1	1608	22712	−0.211	−0.180
	PMAT	6983.7	3713.3	2153	46446	0.127	0.064
	PLOT	1137.0	762.2	416	30174	−0.116	−0.034

	MEAN	SD	MIN	MAX		
PEIT	1993.8	387.6	1271	4712	−0.223	−0.039
PEDT	2984.1	769.6	1235	9441	−0.158	−0.083
PADT	2393.2	701.4	949	8098	−0.122	0.001
MP	720.71	130.1	419	1579	−0.158	−0.075
TAP	97.96	13.21	33	157	0.489	0.064

PCET= Penn Conditional Exclusion Test; PCPT=Penn Continuous Performance Test; LNB= Letter N-Back; PWMT= Penn Word Memory Test; PFMT= Penn Face Memory Test; VOLT= Visual Object Learning Test; PVRT= Penn Verbal Reasoning Test; PMAT= Penn Matrix Reasoning Test; PLOT= Penn Line Orientation Test; PEIT= Penn Emotion Identification Test; PEDT= Penn Emotion Differentiation Test; PADT= Penn Age Differentiation Test; MP=Motor Praxis; TAP= Finger Tapping Test. Nonsignificant correlations are in smaller font.

Table 4

Results of univariate ANOVAs and effect sizes for significant between-group differences by domain. Domain abbreviations as in Table 2.

		ACCURACY			SPEED		
		F^1	P	ES^2	F^1	P	ES^3
ABF	Age group	11.65	<.0001	0.50	2.38	0.0268	0.37
	Sex	9.49	0.0021	−0.09			
	A × S						
ATT	Age group	87.26	<.0001	1.45	215.61	<.0001	1.84
	Sex	14.62	<.0001	0.22	25.44	<.0001	−0.06
	A × S						
WME	Age group	44.89	<.0001	0.92	48.86	<.0001	0.93
	Sex				16.43	<.0001	−0.09
	A × S						
VME	Age group	4.05	0.0005	0.27	94.75	<.0001	1.43
	Sex	4.91	0.0268	0.13	7.00	0.0082	0.15
	A × S						
FME	Age group	45.70	<.0001	1.16	39.20	<.0001	1.04
	Sex	17.69	<.0001	0.20			
	A × S						
SME	Age group				23.43	<.0001	0.80
	Sex	4.14	0.0419	−0.07			
	A × S	2.37	0.0274		2.55	0.0185	
LAN	Age group	94.76	<.0001	1.44	35.41	<.0001	1.14
	Sex	6.46	0.0111	−0.02	13.17	0.0003	0.18
	A × S						
NVR	Age group	54.06	<.0001	1.03	10.40	<.0001	0.44 [4]
	Sex	7.60	0.0059	−0.04	19.74	<.0001	0.18
	A × S	3.39	0.0025		2.57	0.0173	
SPA	Age group	56.55	<.0001	1.06	10.77	<.0001	0.29

		ACCURACY			SPEED		
		F^1	P	ES^2	F^1	P	ES^3
	Sex	56.58	<.0001	−0.23			
	A × S						
EMI	Age group	34.68	<.0001	0.92	27.68	<.0001	0.85
	Sex	31.53	<.0001	0.25	60.26	<.0001	0.33
	A × S				2.30	0.0326	
EMD	Age group	41.21	<.0001	0.88	9.08	<.0001	0.52
	Sex	6.26	0.0124	0.15	7.67	0.0056	0.11
	A × S				3.12	0.0048	
AGD	Age group	61.90	<.0001	1.17	5.63	<.0001	0.43
	Sex	15.38	<.0001	0.18	10.62	0.0011	0.13
	A × S				3.22	0.0038	
SM	Age group				14.41	<.0001	0.70
	Sex				10.04	0.0015	−0.12
	A × S						
MOT	Age group				158.68	<.0001	1.81
	Sex				59.62	<.0001	−0.23
	A × S				5.61	<.0001	

[1] DF for Age group = 6, 2919; for Sex =1, 2919

[2] ES=Effect size; Negative ES for sex difference: males > females.

[3] While for nearly all ES estimates for accuracy the highest scoring group was also the oldest, for speed performance was fastest for 14–15 age group on ABF, 16–17 group for WME, LAN, and EMO ID, and 18–19 group for SMEM and AGE ID.

[4] For this effect size slowest performance was in the oldest age group