VII. NIH TOOLBOX COGNITION BATTERY (CB): FACTOR STRUCTURE FOR 3 TO 15 YEAR OLDS

Dan Mungas, **Keith Widaman**, **Philip David Zelazo**, **David Tulsky**, **Robert K. Heaton**, **Jerry Slotkin**, **David L. Blitz**, and **Richard C. Gershon**

Abstract

Confirmatory factor analysis was used the evaluate the dimensional structure underlying the NIH Toolbox Cognition Battery (CB) and the measures chosen to serve as concurrent validity criteria for the NIH Toolbox CB. These results were used to evaluate the convergent and discriminant validity of the CB in children ranging from 3 to 15 years of age. Results were evaluated separately for a 3- to 6-year-old group and a 8- to 15-year-old group because different validation measures were used in these age groups. Three distinct dimensions were found for the 3- to 6-year-old group: Vocabulary, Reading, and Fluid Abilities. Five dimensions were found for 8–15 year olds: Vocabulary, Reading, Episodic Memory, Working Memory, and Executive Function/Processing Speed. CB measures and their validation analogues consistently defined common factors in a pattern that broadly supported the convergent and discriminant validity of the CB, but results showed higher intercorrelation and less differentiation of cognitive dimensions in younger than in older children and in older children compared with adults. Age was strongly related to the cognitive dimensions underlying test performance in both groups of children and results are consistent with broader literature showing increasing differentiation of cognitive abilities associated with the rapid brain development that occurs from early childhood into adulthood.

In this chapter, we discuss convergent and discriminant validity of the NIH Toolbox Cognition Battery (CB). This is accomplished using confirmatory factor analysis (CFA) to identify the dimensions underlying the CB tests and established validation measures, and to test the hypothesis that the CB tests measure the specific domains they were designed to measure.

Cognition undergoes rapid developmental changes across the 3- to 15-year age range related to brain development and extensive environmental input, especially formal education, which is designed to develop cognitive skills and expand knowledge. An overarching goal of the CB is to be able to assess cognitive abilities across the life span, and this presupposes that the same abilities are being measured in the same way at different ages. Consequently, a critical part of the construct validation tests whether expected relations with widely used instruments are present at different ages.

Reasons to Expect Age-Related Differences in Factor Structure

Construct validity begins with a conceptual model that describes the expected relations between domains being measured and specific tests used to measure those domains. The CB was designed to assess six specific subdomains: executive function (with tests of cognitive flexibility, inhibitory control, and attention), episodic memory, language (vocabulary), reading, working memory, and processing speed. This test development model provides a

Corresponding author: Dan Mungas, Department of Neurology, 4860 Y Street, Suite 3700, Sacramento, CA 95817, dmmungas@ucdavis.edu.

conceptual foundation for the construct validation of the CB. However, developmental changes in the structure of cognition that occur across childhood have implications for specific hypotheses deriving from this conceptual model.

A great deal of brain development occurs before birth, but it is now clear that brain development is a protracted process, with major changes taking place during the preschool years and continuing into adolescence and early adulthood. Indeed, recent technological advances have allowed unprecedented opportunities to observe detailed developmental changes in the living brain, and researchers are beginning to chart the way in which developmental changes (both progressive and regressive) in specific neural systems are related to changes in different aspects of cognitive function.

Considerable research supports the suggestion that key aspects of neurocognitive development involve the experience-dependent functional specialization of neural networks. In a pioneering series of postmortem histological studies of synaptic density in human cortex, Huttenlocher (1979, 1990) noted a general developmental pattern of initial overproduction of synapses followed by reductions to adult levels. For example, synaptic density in Layer III of the middle frontal gyrus reaches a peak at about 1 year of age that is considerably higher than the adult level, remains high until at least age 7 years, and then declines by about 40% until about age 16, when the adult level is finally attained.

Developmental neuroimaging research examining gray matter, which is comprised of neurons with dendritic and synaptic processes, as well as glia and vasculature, confirms that a prominent pattern seen in many cortical regions (especially dorsal regions) is that of increases in gray matter volume (or cortical thickness) in infancy and early childhood followed by gradual decreases that start in late childhood and continue into adulthood, when they plateau (e.g., Gogtay et al., 2004; Jernigan & Tallal, 1990; O'Donnell, Noseworthy, Levine, & Dennis, 2005; Pfefferbaum et al., 1994; Reiss, Abrams, Singer, Ross, & Denckla, 1996).

Reductions in gray matter during childhood have been attributed to synaptic pruning, which may occur in a Hebbian fashion, as a function of learning and experience (Casey, Giedd, & Thomas, 2000; Durston et al., 2001; Giedd et al., 1999), and which may result in the increasing differentiation of cognitive functions as neural regions become more specialized. A classic example of this process occurs in perceptual development. Initially, for example, occipital cortical areas involved in vision are activated by crossmodal input from other sensory modalities (reviewed in Collignon, Voss, Lassonde, & Lepore, 2009; Spector & Maurer, 2009). With normal visual experience, however, visual inputs to occipital cortex are reinforced whereas crossmodal inputs from other perceptual systems are eliminated or inhibited.

A similar process may occur more broadly in brain development, including in higher-order association areas that integrate information from lower-order, earlier developing areas such as visual cortex. According to one influential model, the Interactive Specialization model (e.g., Johnson & Munakata, 2005), neurocognitive development in general involves the increasing functional specialization of neural systems that are initially relatively undifferentiated but which become more specialized (or modularized) as part of a developmental process of adaption. Current research on executive function, for example, provides evidence that supports this suggestion. A seminal study of the factor structure of executive function in young adulthood used confirmatory factor analysis to extract three correlated latent variables from several commonly used executive function tasks, believed to represent cognitive flexibility, inhibitory control, and working memory (Miyake et al., 2000). Research with younger participants suggests that this differentiation of executive

function into three dissociable components emerges during childhood. Among pre-school-age children, research generally is consistent with a 1-factor solution (Wiebe, Espy, & Charak, 2008; Wiebe et al., 2011). Wiebe et al. (2008) used a battery of three tasks designed to measure working memory and seven tasks requiring inhibition, all of which loaded onto a single factor. This pattern was also found during the transition to adolescence (8–14 years) in a study by Prencipe et al. (2011). In contrast, several studies have found that the tripartite model of EF provides a good account of the data by middle childhood (Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Visu-Petra, Benga, & Miclea, 2007), although Huizinga, Dolan, and van der Molen (2006) found that only working memory and shifting measures (and not inhibition measures) loaded onto latent variables in 7, 11, 15, and 21 year olds. In general, research on the factor structure of executive function appears to be consistent with a shift from diffuse to more focal cortical brain activity with age (Durston et al., 2006).

Evaluating Construct Validity in the NIH Toolbox CB

Convergent and discriminant validity are important elements of construct validity and relate to the dimensions accounting for covariance among groups of tests selected to measure specific domains. Construct validity is supported when (a) the empirically observed dimensions correspond to the a priori conceptual model for the domains being measured, and (b) individual tests are strongly related to the dimensions hypothesized from the conceptual model and are not related (or are more weakly related) to other dimensions. This process is somewhat more complicated in children due to the progressive differentiation of cognitive abilities that occurs as brain systems develop. In adults, a six-dimensional model of cognitive abilities should be appropriate to explain the relations among CB and validation measures, and it would be expected that CB measures of specific domains and corresponding validation measures would define dimensions that directly correspond to the six CB domains (see Weintraub et al., Chapter 1, this volume). In children, one would expect: (a) fewer dimensions underlying intercorrelations among tests, and (b) stronger associations among differentiable dimensions of the CB and the validation tests. In particular, executive function and working memory tasks would be expected to be less differentiated from other cognitive abilities, with lesser differentiation in younger children than in older children, because of the substantial development of frontal lobe structure and function that occurs throughout childhood and adolescence, continuing into early adulthood.

Data analyses related to this chapter were designed to test systematically how well alternative, a priori defined, dimensional models account for associations among NIHTB-CB and validation tests. Alternate models ranged from a simple 1-factor model representing a single global-cognition model to a 6-factor model corresponding to the six CB subdomains. It was hypothesized that CB and corresponding validation measures would define the same factors, but that fewer factors might be needed in children than in adults, and that intercorrelations among factors would be relatively high in children compared with adults. It was further hypothesized that age would be strongly related to all factors in children.

METHOD

Participants

In addition to the child and adolescent participants in the validation study (see Weintraub et al., Chapter 1, this volume; Table 3), we examined data from 267 adults (age 20–85 years, M = 52.3, SD = 21.0) in order to enhance model estimation for the 8- to 15-year-old age group.

Measures

CB and validation tests are listed in Table 14. Development of CB tests is described in detail in the chapters for each subdomain, and validation tests also are described in more detail in individual subdomain chapters.

Data Analysis

Latent variable modeling methods were used to test convergent and discriminant validity of CB and validation measures. This process was performed separately in children aged 3–6 years and in the 8–15 year olds because different validation measures were administered to these two age groups due to a lack of established measures that are suitable across the entire age range. The basic process for both age groups was to perform a series of confirmatory factor analyses to test alternate models for the dimensions hypothesized to underlie the CB and validation tests. However, methodological limitations inherent in the design of the CB validation study led to differences in how analyses were performed.

The sample of children in the 3- to 6-year age range ($n = 119$) was sufficient to support the proposed analyses of their data, but this group received a smaller battery of tests. Consequently, not all domains had more than one observed indicator so fewer dimensions could be tested. The available sample for 8–15 year olds ($n = 88$) was relatively small for CFA purposes, but data for the same measures were available from the adults in the validation study, and these data were used to facilitate the analyses for 8–15 year olds. Specifically, the 8- to 15-year-old age group and adults ($n = 267$) were included in a multiple group CFA. In multiple group CFA modeling, a common model for both groups is specified on an a priori basis, and then group differences in individual parameters can be systematically tested. The advantage of this approach for the analysis of data from the older children is that many model parameters should be invariant across groups, and the combined sample size is used to estimate those parameters; this improves stability of estimates for the overall model. In effect, the results for the 8- to 15- year-old sample "borrow strength" from the adult sample through the use of invariance constraints on common parameter estimates, yielding a more stable pattern of results than would have occurred if the 8- to 15-year sample were analyzed separately. The focus for this study was on children and adolescents, however, so incorporation of adults when analyzing the older children data was primarily methodologically motivated. A subsequent report will address CB dimensions in adults.

The alternative models that were tested are shown in Table 15. Specific measures for each age group are presented in Table 14 along with their associated conceptual domains/ dimensions in the various models. For the 3- to 6-year-old age group, the five models shown in Table 15 were separately estimated and model fit indices were compared to identify the best fitting model. The best fitting model at this stage had a simple structure with each indicator loading on just one factor. Modification indices were then examined to identify cross loadings of CB measures on other factors that would significantly improve model fit if freely estimated. Convergent validity for a CB measure was evidenced by a strong loading on the dimension corresponding to the primary conceptual domain. Discriminant validity was shown if no loading, or a smaller loading, was required for a CB measure on a secondary dimension/domain.

Dimensional structure for the 8- to 15-year-old age group was evaluated using a multiple group CFA that included adults as the second group. The alternative dimensional models presented in Table 15 were estimated separately, the best fitting model was determined, and then cross loadings were tested. The basic process was similar to that for the 3- to 6-year age group, but the process for estimating each alternate model was different. First, a model was fitted with loadings and intercepts that were constrained to be equal in the two groups, but

common factor means, variances, and covariances and unique factor variances for individual indicators were allowed to differ across groups. Then, modification indices were used to identify noninvariant loadings and then intercepts that subsequently were freely estimated in each group. This was an iterative process. The constrained loading with the largest modification index was freely estimated first, and then the constrained loading with the largest modification index from that analysis was freely estimated. This iterative process was continued until no additional significant modification indices for loadings were identified. The same process was then followed for intercepts. This process was continued to identify any additional loadings and then any additional intercepts. Fit indices from the different alternative models at this stage of development were compared in order to identify the best fitting model. After a best fitting model was chosen, further modification to that model was achieved by including residual correlations that were conceptually justified and improved model fit in both groups. Finally, modification indices were used to identify significant cross-loadings of Toolbox measures on secondary factors.

Variables were recoded prior to analysis using the Blom rank order normalization algorithm in SAS Proc Rank. This resulted in variables with relatively normal distributions and also established a common scale of measurement of all variables. The normalization was applied separately to the 3- to 6-year-old group and the combined 8- to 15-year-old and adult groups. Scores for DKEFS Stroop Interference and Wisconsin Card Sort Total Errors were inverted so that higher scores indicated better performance on all measures. Normalized scores were multiplied by 3.0 and added to 10.0 to place them on a common scale with mean of 10.0 and standard deviation of 3.0.

Model estimation was performed with Mplus version 6.0 (Mutheén & Mutheén, 1998–2010) using a maximum likelihood estimator for continuous variables applied to a mean and covariance data structure. Latent variable modeling traditionally uses an overall chi square test of model fit, often supplemented by a number of fit indices to better characterize model fit. Commonly used fit indices include the comparative fit index (CFI; Bentler, 1990), the Tucker-Lewis index (TLI; Tucker & Lewis, 1973), the root mean square error of approximation (RMSEA; Browne & Cudek, 1993), and the standardized root mean squared residual (SRMR; Bentler, 1995). The chi-square difference test (Steiger, Shapiro, & Browne, 1985) was used to determine if fit significantly improved as a result of freeing one or more parameters in a model. Modification indices correspond to the improvement in model fit as measured by the amount the overall chi square value would decrease if a constrained parameter were freely estimated. A threshold of 6.63 was used as a standard for significant improvement in fit, which corresponds to $p = .01$ for a chi square variate with 1 degree of freedom.

RESULTS

Children 3–6 Years of Age

A 3-factor model (Vocabulary, Reading, Fluid abilities) was the best fitting of the alternate models and showed relatively good absolute fit on all indices except RMSEA (see Table 16). Fit for this 3-factor model was substantially better than either the 2-factor model or the 1-factor model. There were estimation problems for the 3-factor and 4-factor models that had separate dimensions for memory and fluid abilities because the correlation of the memory and fluid abilities latent variables were indistinguishable from 1.0.

Standardized loadings for the best fitting model are presented in Table 17. Factor loadings were strong for all factors and indicators. None of the CB measures had significant loadings on secondary factors. The correlation of the Reading and Vocabulary factors was .68 ($SE = .06$, $p < .001$), and the correlations of Fluid Abilities with Reading and Vocabulary were

both .83 (*SEs* = .04, *ps* < .001). These results indicate that reading and vocabulary are clearly differentiated from other cognitive abilities and from each other in this age group, but other cognitive abilities are not well differentiated. Results support the convergent and discriminant validity of the Toolbox reading and vocabulary measures and indicate that the other CB tests measure fluid ability that is not well differentiated in children in this age range. Within this range, however, all three factors were highly correlated with age: Reading, *r* = .75 (*SE* = .04, *p* < .001), Vocabulary, *r* = .67 (*SE* = .05, *p* < .001), and Fluid Abilities, *r* = .86 (*SE* = .03, *p* < .001).

Children 8–15 Years of Age

A 5-factor model (Vocabulary, Reading, Episodic Memory, Working Memory, Executive/ Speed) was identified as the best fitting model for the sample of 8–15 year olds (see Table 16). Estimation problems arose for the 6-factor model in the 8- to 15-year age group due to the correlation between the Executive and Speed factors being close to 1.0. (The 6-factor model provided the best fit for adults, not shown.) Model fit for the best fitting 5-factor model was good after accounting for noninvariant parameters across the two groups and including modifications to estimate covariances among unique factors for measures that overlap in methods (Wechsler Digit Symbol and Symbol Search; Toolbox Flanker Inhibitory Control and Attention Test, Toolbox DCCS, and Toolbox Pattern Comparison Processing Speed Test). The Toolbox List Sorting Working Memory Test had a significant cross loading on the Episodic Memory factor in the 8- to 15-year age group, but no other significant cross-loadings of CB variables on secondary factors were found.

Six variables had noninvariant loadings. DKEFS Stroop Interference was a stronger indicator of Executive/Speed in the 8–15 year olds than in adults (standardized loadings of . 90 vs. .80), and Wechsler Letter Number Sorting and PASAT were stronger indicators of Working Memory in 8–15 year olds (.82 vs. .66 and .91 vs. .75). The Toolbox Picture Sequence Memory Test was less strongly related to Episodic Memory in the 8- to 15-year-old group (.68 vs. .81). The Toolbox Picture Vocabulary Test was less strongly related to the Vocabulary factor, and Digit Symbol was more strongly related to Executive/ Speed in the 8–15 year olds, but the standardized loadings were minimally different (.85 vs. .91 and .82 vs. .77). Six variables had noninvariant intercepts; Wechsler Digit Symbol and PASAT were relatively easier in adults, and the Toolbox Picture Sequence Memory Test, Toolbox DCCS, WCST Errors, Toolbox Flanker and the Toolbox Pattern Comparison Processing Speed Test were relatively easier in the 8–15 year olds. That is, the expected performance for the latter five variables was better in the children than in adults after equating for the latent ability measured by the relevant factors.

Standardized loadings for the best fitting model are presented in Table 18. Loadings for the Toolbox Oral Reading Recognition Test and the Toolbox Picture Vocabulary Test were quite strong, ranging from .85 to .98. The Toolbox Picture Sequence Memory Test had a standardized loading of .68 on the Episodic Memory factor and the Toolbox DCCS had a loading of .71 on the Executive/Speed factor. The Toolbox Flanker and the Toolbox Pattern Comparison Processing Speed Test had loadings on the Executive/Speed factor in the .55–. 60 range and the Toolbox List Sorting Working Memory Test had a loading of .54 on the Working Memory factor. Toolbox List Sorting had a secondary loading of .29 on the Episodic Memory factor. Overall, these findings show evidence of excellent convergent validity. The presence of only one, relatively weak cross loading supports discriminant validity of the CB. The weakest convergent validity estimates were for the CB measures of Executive/Speed, and this is not surprising because of the relative heterogeneity of the indicators for this factor and the absence of direct analogues of the Toolbox measures as

were available for the Toolbox Oral Reading Recognition Test and the Toolbox Picture Vocabulary Test.

The intercorrelations of the five factors for the 8–15 year olds were very high, ranging from .72 to .94 (ps < .001, see Table 19). Whereas the abilities being measured by these factors were differentiable, they nevertheless were highly correlated, which is likely due to broad differences in overall development within this age group that contribute substantial, nonspecific influences on cognitive function. For comparison purposes, factor correlations for the Adult group are presented in Table 20. Correlations of the Toolbox Oral Reading Recognition Test and the Toolbox Picture Vocabulary Test with Episodic Memory, Working Memory, and Executive/Speed factors were substantially smaller. Correlations among the latter three factors were still quite high, but were smaller than were observed in the 8–15 year olds. All five factors were highly correlated with age: Reading, $r = .70$ ($SE = .05$, $p < .001$); Vocabulary, $r = .76$ ($SE = .06$, $p < .001$); Episodic Memory, $r = .53$ ($SE = .10$, $p < .001$); Working Memory, $r = .64$ ($SE = .08$, $p < .001$); and Executive/Speed, $r = .86$ ($SE = .04$, $p < .001$).

DISCUSSION

There were four primary findings from this study. First, CB measures and their corresponding validation measures consistently defined common factors in a pattern that broadly supported the convergent and discriminant validity of the CB. Second, we found fewer empirically distinct dimensions in children in the 8- to 15-year age range than in adults, and still fewer distinct dimensions in the 3–6 year olds. In the 8–15 age range, executive function and processing speed were less differentiated than in adults, and in the 3–6 year olds, measures of episodic memory, working memory, executive function, and speed all defined a common fluid abilities dimension. Third, correlations among identified dimensions were stronger in children than in adults, and this was especially evident in much stronger correlations in children of crystallized abilities (vocabulary and reading) with other abilities. Fourth, age was strongly related to the cognitive dimensions underlying test performance.

Three distinct dimensions were identified for the 3–6 years age group: vocabulary, reading, and fluid abilities. CB measures were strong indicators of these dimensions and no significant cross-loadings on secondary dimensions were found. These results support the convergent and discriminant validity of the CB measures in this age range, but suggest that the full 6-subdomain model that guided test development is less applicable in this age range because fluid abilities are not well differentiated.

Five separable dimensions were found in the 8–15 year olds. These dimensions corresponded to the subdomains in the test development model for the CB measures with the exception that Executive Function and Processing Speed were not clearly separable. The five dimensions that were observed in this age range were highly correlated, likely reflecting the broad impact of age, experience, and associated brain development in this group.

A striking and somewhat unexpected outcome was the finding that reading and vocabulary were clearly separable in both the 3–6 and 8–15 age groups. Model fit was consistently higher in both groups when vocabulary and reading measures defined separate factors as opposed to a common language or crystallized abilities factor. Although reading and vocabulary were highly correlated in both groups, they nevertheless defined distinct dimensions and were less correlated with one another than were dimensions underlying fluid abilities.

All latent factors identified in both the 3- to 6- and 8- to 15-year-old age groups were substantially correlated with age, with correlations ranging from .53 to .86. The Fluid Abilities factor in the 3–6 year olds and the Executive/Speed factor in the 8–15 year olds were very highly correlated with age. These results likely show the profound influences of brain development coupled with life experiences on cognitive abilities. The sensitivity to age suggests that the Toolbox CB will be useful for tracking cognitive development in longitudinal studies in children.

This study had a number of limitations. The sample size was relatively small for confirmatory factor analysis, and fewer and different tests were administered to the 3- to 6-year-old group. Consequently, we could not incorporate both the 3- to 6- and the 8- to 15-year-old age groups into a combined analysis. The issue of measurement invariance at different ages is especially important for the intended use of the NIH Toolbox. Being able to measure cognition on a common metric across the entire age span from 3 years to late adulthood is an important goal for the CB, and formal testing of factorial invariance in different age groups is required to show a common metric. The multiple group analysis of 8–15 year olds and adults constituted a preliminarily examination of measurement invariance. Groups sizes in the 3- to 6-year and 8- to-15-year age groups were relatively small, which likely affects stability of results, and consequently, any conclusions about measurement invariance must be considered tentative. The norming study (projected $N = 4,000$) for the Toolbox will offer a unique opportunity to formally test measurement invariance with much larger samples across the full age range from age 3 years to the end of life. The norming sample will also include a sizeable group of individuals tested in Spanish ($N = 500$) and this will provide an opportunity for evaluating measurement variance across the English and Spanish versions of the battery.

In spite of these limitations, these results show favorable evidence for the construct validity of the NIH Toolbox CB across early and mid childhood and adolesence and demonstrate how this battery can be useful for understanding the evolving structure of cognition over the course of development. Having standardized methods available for assessing cognition across the lifespan along with the other domains measured by the NIH Toolbox including emotion, motor functioning, and sensory functioning will provide an important resource for research to further our understanding of brain and cognitive development.

REFERENCES

Bentler PM. Comparative fit indices in structural models. Psychological Bulletin. 1990; 107:238–246. [PubMed: 2320703]

Bentler, PM. EQS structural equations program manual. Multivariate Software; Encino, CA: 1995.

Browne, M.; Cudek, R. Alternate ways of assessing model fit. In: Bollen, K.; Long, J., editors. Testing structural equation models. SAGE; Thousand Oaks, CA: 1993. p. 136-162.

Casey BJ, Giedd JN, Thomas KM. Structural and functional brain development and its relation to cognitive development. Biological Psychology. 2000; 54(1–3):241–257. [PubMed: 11035225]

Collignon O, Voss P, Lassonde M, Lepore F. Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. Experimental Brain Research. 2009; 192(3):343–358. [PubMed: 18762928]

Durston S, Davidson MC, Tottenham N, Galvan A, Spicer J, Fossella JA, et al. A shift from diffuse to focal cortical activity with development. Developmental Science. 2006; 9(1):1–8. [PubMed: 16445387]

Durston S, Hulshoff Pol HE, Casey BJ, Giedd JN, Buitelaar JK, van Engeland H. Anatomical MRI of the developing human brain: What have we learned? Journal of the American Academy of Child and Adolescent Psychiatry. 2001; 40(9):1012–1020. [PubMed: 11556624]

Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, et al. Brain development during childhood and adolescence: A longitudinal MRI study. Nature Neuroscience. 1999; 2(10): 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 United States of America. 2004; 101(21):8174–8179.

Huizinga MT, Dolan CV, van der Molen MW. Age-related change in executive function: Developmental trends and a latent variable analysis. Neuropsychologia (Special Issue: Advances in Developmental Cognitive Neuroscience). 2006; 44(11):2017–2036.

Huttenlocher PR. Synaptic density in human frontal cortex—Developmental changes and effects of aging. Brain Research. 1979; 163(2):195–205. [PubMed: 427544]

Huttenlocher PR. Morphometric study of human cerebral cortex development. Neuropsychologia. 1990; 28(6):517–527. [PubMed: 2203993]

Jernigan TL, Tallal P. Late childhood changes in brain morphology observable with MRI. Developmental Medicine and Child Neurology. 1990; 32(5):379–385. [PubMed: 2354751]

Johnson MH, Munakata Y. Processes of change in brain and cognitive development. Trends in Cognitive Sciences. 2005; 9(3):152–158. [PubMed: 15737824]

Lehto JE, Juujaärvi P, Kooistra L, Pulkkinen L. Dimensions of executive functioning: Evidence from children. British Journal of Developmental Psychology. 2003; 21(1):59–80.

Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. Cognitive Psychology. 2000; 41(1):49–100. [PubMed: 10945922]

Mutheén, LK.; Mutheén, BO. Mplus User's Guide. 6th. Mutheén & Mutheén; Los Angeles, CA: 1998–2010.

O'Donnell S, Noseworthy MD, Levine B, Dennis M. Cortical thickness of the frontopolar area in typically developing children and adolescents. NeuroImage. 2005; 24(4):948–954. [PubMed: 15670671]

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(9):874–887. [PubMed: 8080387]

Prencipe A, Kesek A, Cohen J, Lamm C, Lewis MD, Zelazo PD. Development of hot and cool executive function during the transition to adolescence. Journal of Experimental Child Psychology. 2011; 108(3):621–637. [PubMed: 21044790]

Reiss AL, Abrams MT, Singer HS, Ross JL, Denckla MB. Brain development, gender and IQ in children: A volumetric imaging study. Brain. 1996; 119:1763–1774. Pt 5. [PubMed: 8931596]

Spector F, Maurer D. Synesthesia: A new approach to understanding the development of perception. Developmental Psychology. 2009; 45(1):175–189. [PubMed: 19210000]

Steiger JH, Shapiro A, Browne MW. On the multivariate asymptotic distribution of sequential chi-square statistics. Psychometrika. 1985; 50:253–264.

Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. Psychometrika. 1973; 38:1–10.

Visu-Petra L, Benga O, Miclea M. Dimensions of attention and executive functioning in 5- to 12-year-old children: Neuropsychological assessment with the NEPSY battery. Cognition, Brain, Behavior (Special Issue: Developmental Cognitive Neuropsychology). 2007; 11(3):585–608.

Wiebe SA, Espy KA, Charak D. Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. Developmental Psychology. 2008; 44(2):575–587. [PubMed: 18331145]

Wiebe SA, Sheffield T, Nelson JM, Clark CA, Chevalier N, Espy KA. The structure of executive function in 3-year-olds. Journal of Experimental Child Psychology. 2011; 108(3):436–452. [PubMed: 20884004]

Table 14

MEASURES AND ASSOCIATED DOMAINS BY AGE GROUP

Age Group	Measure	Associated Domains
Both	**TPVT**	Vocabulary, Language, Crystallized/Global
	PPVT-IV	Vocabulary, Language, Crystallized/Global
	TORRT	Reading, Language, Crystallized, Global
	WRAT-R	Reading, Language, Crystallized, Global
	TPSMT	Episodic Memory, Fluid, Global
	TLSWMT	Working Memory, Fluid, Global
	TFIC + AT	Fluid, Global
	TDCCST	Fluid, Global
	TPCPST	Fluid, Global
3–6	WPPSI-III Block Design	Fluid, Global
	NEPSY-II Sentence Repetition	Episodic/Working Memory, Fluid, Global
8–15	PASAT	Working Memory, Fluid, Global
	Wechsler Letter Number Sorting	Working Memory, Fluid, Global
	Wechsler Digit Symbol	Speed, Executive/Speed, Fluid, Global
	Wechsler Symbol Search	Speed, Executive/Speed, Fluid, Global
	Wisconsin Card Sort Total Errors	Executive, Executive/Speed, Fluid, Global
	DKEFS Stroop Interference	Executive, Executive/Speed, Fluid, Global

Note. Domains are listed in order from most specific to most general. CB measures are bolded. TPVT, Toolbox Picture Vocabulary Test; PPVT-IV, Peabody Picture Vocabulary Test-4th Edition; TORRT, Toolbox Oral Reading Recognition Test; WRAT-R, Wide Range Reading Test-Revised; TPSMT, Toolbox Picture Sequence Memory Test; TLSWMT, Toolbox List Sorting Working Memory; TFIC + AT, Toolbox Flanker Inhibitory Control and Attention Test; TDCCST, Toolbox Dimensional Change Card Sort Test; TPCPST, Toolbox Pattern Comparison Processing Speed Test; WPPSI-III, Wechsler Preschool and Primary Intelligence Test; 3rd Edition; NEPSY-II, Developmental Neuropsychological Assessment, 2nd Edition; PASAT, Paced Auditory Serial Attention Test; DKEFS, Delis–Kaplan Executive Function Scales. © 2006–2013 National Institutes of Health and Northwestern University.

Table 15

Alternate Dimensional Models Underlying NIH Toolbox CB and Validation Measures

3–6 year olds	8–15 year olds
1f: Global Cognition	1f: Global Cognition
2f: Crystallized, Fluid	2f: Crystallized, Fluid
2f: Episodic/Working Memory, Nonmemory	2f: Memory, Nonmemory
	3f: Crystallized, Fluid, Memory
3f: Crystallized, Fluid, Episodic/Working Memory	3f: Language, Memory/Working Memory, Executive/Speed
3f: Vocabulary, Reading, Fluid	3f: Language, Memory, Working Memory/
4f: Vocabulary, Reading, Fluid,	Executive/Speed
Episodic/Working Memory	4f: Language, Memory, Working Memory, Executive/Speed
	4f: Vocabulary, Reading, Memory, Executive
	5f: Language, Memory, Working Memory, Executive, Speed
	5f: Vocabulary, Reading, Memory, Working Memory, Executive/Speed
	6f: Vocabulary, Reading, Memory, Working Memory, Executive, Speed

Table 16

F<small>IT</small> I<small>NDICES FOR</small> A<small>LTERNATE</small> M<small>ODELS OF</small> C<small>OGNITIVE</small> D<small>IMENSIONS IN</small> 3–6 Y<small>EAR</small> O<small>LDS AND</small> 8–15 Y<small>EAR</small> O<small>LDS</small>

Model	Overall χ^2 [df]	χ^2: 8–15	CFI	TLI	RMSEA (90% CI)	SRMR
3–6 Year Age Group						
1f: Global	214.1 [44]		.824	.780	.180 (.156–.205)	.057
2f: Crystallized, Fluid	168.3 [43]		.870	.834	.156 (.132–.182)	.090
2f: Episodic/Working Memory, Nonmemory	214.1 [43]		.823	.773	.183 (.159–.208)	.057
3f: Vocabulary, Reading, Fluid	76.8 [41]		.963	.950	.086 (.055–.115)	.039
8–15 Year Age Group						
1f: Global	1278.0 [241]	248.0	.725	.689	.156 (.147–.164)	.081
2f: Crystallized, Fluid	607.6 [245]	215.8	.904	.893	.091 (.082–.100)	.119
2f: Episodic/Working Memory, Nonmemory						
3f: Language, Episodic/Working Memory, Executive						
3f: Language Memory, Working Memory/Executive	734.9 [250]	251.2	.871	.860	.105 (.053–.113)	.106
3f: Vocabulary, Reading, Fluid	477.8 [241]	172.7	.937	.929	.074 (.053–.074)	.079
4f: Vocabulary, Reading, Memory, Working Memory/ Executive	412.8 [239]	170.6	.954	.947	.064 (.065–.084)	.109
4f: Vocabulary, Reading, Episodic/Working Memory, Executive	456.7 [239]	176.0	.942	.934	.072 (.062–.082)	.103
5f: Vocabulary, Reading, Episodic Memory, Working Memory, Executive	370.5 [230]	162.3	.963	.956	.059 (.047–.069)	.109

Note. χ^2: 8–15 shows the specific contribution of the 8- to 15-year-old group to the overall χ^2-value; CFI = comparative fit index; TLI = Tucker-Lewis index; CI = confidence interval; RMSEA = root mean square error of approximation.

Table 17

Standardized Factor Loadings (Standard Errors in Parentheses) for Best Fitting 3-Factor Model for 3- to 6-Year-Old Age Group

Latent Factor	Observed Indicator	Loading
Reading	**TORRT**	.97 (.02)
	WRAT-R	.96 (.02)
Vocabulary	**TPVT**	.75 (.05)
	PPVT-IV	.99 (.03)
Fluid Abilities	**TPSMT**	.79 (.04)
	TLSWMT	.70 (.05)
	TFIC + AT	.83 (.04)
	TDCCST	.89 (.03)
	TPCPST	.70 (.06)
	NEPSY-II Sentence Repetition	.78 (.04)
	WPPSI-III Block Design	.77 (.04)

Note. CB measures are bolded. Correlation of Reading with Vocabulary = .69 (*SE* = .06, *p* < .001), Reading with Fluid Abilities = .83 (*SE* = .04, *p* < .001), Vocabuary with Fluid Abilities = .83 (*SE* = .04, *p* < .001). TORRT, Toolbox Oral Reading Recognition Test; WRAT-R, Wide Range Reading Test: Revised; TPVT, Toolbox Picture Vocabulary Test; PPVT-IV, Peabody Picture Vocabulary Test: 4th Edition; TPSMT, Toolbox Picture Sequence Memory Test; TLSWMT, Toolbox List Sorting Working Memory; TFIC + AT, Toolbox Flanker Inhibitory Control and Attention Test; TDCCST, Toolbox Dimensional Change Card Sort Test; TPCPST, Toolbox Pattern Comparison Processing Speed Test; NEPSY-II, Developmental Neuropsychological Assessment, 2nd Edition; WPPSI-III, Wechsler Preschool and Primary Intelligence Test, 3rd Edition.

Table 18

Standardized Factor Loadings (Standard Errors in Parentheses) for Best Fitting 5-Factor Model for 8–15 Year Olds

Latent Factor	Observed Indicator	Loading
Reading	**TORRT**	.98 (.01)
	WRAT-R	.96 (.01)
Vocabulary	**TPVT**	.86 (.03)
	PPVT-IV	.97 (.02)
Episodic Memory	**TPSMT**	.68 (.08)
	RAVLT	.70 (.04)
	BVMT	.78 (.04)
	TLSWMT [a]	.29 (.07)
Working Memory	**TLSWMT**	.54 (.04)
	PASAT	.82 (.04)
	Wechsler Letter Number Sorting	.91 (.03)
Executive/Speed	**TFIC + AT**	.59 (.04)
	TDCCST	.71 (.04)
	TPCPST	.58 (.04)
	Wechsler Digit Symbol	.82 (.04)
	Wechsler Symbol Search	.74 (.03)
	Wisconsin Card Sort Total Errors	.69 (.04)
	D-KEFS Stroop Interference	.90 (.03)

Note. CB Measures are bolded. TORRT, Toolbox Oral Reading Recognition Test; WRAT-R, Wide Range Reading Test-Revised; TPVT, Toolbox Picture Vocabulary Test; PPVT-IV, Peabody Picture Vocabulary Test-4th Edition; TPSMT, Toolbox Picture Sequence Memory Test; BVMT-R: Brief Visuospatial Memory Test-Revised; RAVLT: Rey Auditory Verbal Learning Test; TLSWMT, Toolbox List Sorting Working Memory; PASAT, Paced Auditory Serial Attention Test; TFIC + AT, Toolbox Flanker Inhibitory Control and Attention Test; TDCCST, Toolbox Dimensional Change Card Sort Test; TPCPST, Toolbox Pattern Comparison Processing Speed Test; D-KEFS, Delis-Kaplan Executive Function Scales.

[a] Significant loading on secondary factor.

Table 19

I<small>NTER</small>-C<small>ORRELATION OF</small> F<small>ACTORS</small> F<small>ROM</small> B<small>EST</small>-F<small>ITTING</small> 5-F<small>ACTOR</small> M<small>ODEL FOR</small> 8–15 Y<small>EAR</small> O<small>LDS</small>

	Reading	Vocabulary	Episodic Memory	Working Memory
Vocabulary	.89 (.03)			
Episodic Memory	.72 (.07)	.85 (.06)		
Working Memory	.87 (.04)	.83 (.04)	.90 (.06)	
Executive/Speed	.91 (.03)	.90 (.03)	.89 (.06)	.94 (.03)

Note. Standard errors in parentheses ($p < .001$ for all correlations).

Table 20

Inter-Correlation of Factors From Best-Fitting 5-Factor Model for Adults

	Reading	Vocabulary	Episodic Memory	Working Memory
Vocabulary	.82 (.03)			
Episodic Memory	.30 (.06)	.12 (.07)		
Working Memory	.54 (.05)	.43 (.06)	.84 (.04)	
Executive/Speed	.41 (.06)	.23 (.07)	.80 (.04)	.90 (.03)

Note. Standard errors in parentheses ($p < .001$ for all correlations except Vocabulary with Episodic Memory, where $p = .07$). © 2006–2013 National Institutes of Health and Northwestern University.