Journal of Experimental Child Psychology 189 (2020) 104681

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Developmental transformations in the structure of executive functions

Johanna Hartung [a,*], Laura E. Engelhardt [b], Megan L. Thibodeaux [b], K. Paige Harden [b,c], Elliot M. Tucker-Drob [b,c]

[a] *Institute of Psychology and Education, Department of Psychology, Ulm University, 89081 Ulm, Germany*
[b] *Department of Psychology, University of Texas at Austin, Austin, TX 78712, USA*
[c] *Population Research Center, University of Texas at Austin, Austin, TX 78712, USA*

A R T I C L E   I N F O

Article history:
Received 16 January 2019
Revised 4 August 2019
Available online 21 October 2019

Keywords:
Executive function
Age differentiation
Inhibition
Development
Local structural equation models
Moderated factor analysis

A B S T R A C T

Comparisons of results from studies of executive function (EF) in early childhood to those of EF in middle and late childhood suggest that individual differences in EFs may differentiate from a unitary factor in early childhood to an increasingly multidimensional structure in middle childhood and adolescence. We tested whether associations among EFs strengthened from middle childhood to adolescence using cross-sectional data from a population-based sample of 1019 children aged 7–15 years (M = 10.79 years). Participants completed a comprehensive EF battery consisting of 15 measures tapping working memory, updating, switching, and inhibition domains. Moderated factor analysis, local structural equation modeling, and network modeling were applied to assess age-related differences in the factor structure of EF. Results from all three approaches indicated that working memory and updating maintained uniformly high patterns of covariation across the age range, whereas inhibition became increasingly differentiated from the other three domains beginning around 10 years of age. However, consistent with past research, inhibition tasks were only weakly intercorrelated. Age-related differences in the organization of switching abilities were mixed.

* Corresponding author.
E-mail address: johanna.hartung@uni-ulm.de (J. Hartung).

https://doi.org/10.1016/j.jecp.2019.104681

Introduction

Executive functions (EFs) are a constellation of cognitive functions that supervise and manipulate other cognitive processes. In both childhood and adulthood, EFs have been implicated in learning, academic achievement, psychiatric health, and everyday functioning (Best, Miller, & Naglieri, 2011; Diamond, 2013). EFs include inhibition of prepotent responses, updating of information, shifting between cognitive operations, and short-term storage and processing of information in working memory (Miyake et al., 2000; Unsworth & Engle, 2007). These are commonly referred to as EF domains.

From a long history of studies, it is well known that individuals differ in their executive abilities. Individuals who have good performance in a given EF task (e.g., *N*-Back) tend to do well on other tasks tapping the same EF (e.g., updating), which shows that these tasks capture reliable individual differences. Studies of individual differences in EFs in adulthood indicate that performance on tasks designed to tap a specific EF domain (e.g., inhibition) is correlated with, but also separable from, performance on tasks tapping other EF domains (e.g., switching). One statistical method for understanding these correlations is latent factor analysis. Under this framework, correlations in task performance arise because the tasks tap the same underlying, but not directly observed, ability. This unobserved (or latent) ability can be interpreted as what is common across the tasks. This method is advantageous for studying relations among multiple measures because the latent factors do not include any error variance specific to each task. So variance in performance on any given task can be explained by common, specific, and error/unexplained variance. Confirmatory factor analysis (CFA) is a form of latent factor analysis that allows for the evaluation of different models reflecting different theories by different assignments of indicators (observed performance on a task) to latent factors (Brown & Moore, 2012). The latent factors represent variance shared across constituent indicators (common variance). Therefore, CFA helps to identify the number of underlying constructs of a task battery by testing how well different measures cluster together (by evaluating fit of a specified factor model to actual data on task performance as well as loading patterns in the model). When applied to performance on multiple EF tasks, CFA has revealed that a common EF factor accounts for high degrees of variance shared across EF domains, but some domain-specific variance independent of common EF remains (Miyake et al., 2000). In other words, EF domains are statistically separable—performance is not perfectly correlated—but they still share a significant portion of their variance.

The development of EF across childhood and adolescence has gained increasing attention in recent years. In particular, the question of the extent to which EF domains are differentiable across childhood and adolescence remains highly contested. Studies attempting to reproduce the hierarchical structure of latent EF factors in child and adolescent samples have generated mixed support for a fully differentiated EF framework (for a review, see Table 1 in Lee, Bull, & Ho, 2013). Studies of individual differences in EFs in early childhood (~2–6 years of age) have routinely reported a single EF factor, such that performance on multiple tasks does not cluster into separable components (e.g., Shing, Lindenberger, Diamond, Li, & Davidson, 2010; Wiebe, Espy, & Charak, 2008; Willoughby, Wirth, & Blair, 2012). Studies in middle childhood and adolescence have tended to report two-, three-, and four-factor solutions, with latent factors most commonly representing working memory, inhibition, and shifting (e.g., Agostino, Johnson, & Pascual-Leone, 2010; McAuley & White, 2011; Wu et al., 2011). These studies provide circumstantial evidence that performance on EF becomes increasingly differentiated in middle childhood and adolescence. In their review of the development of EF, Best and Miller (2010) supported this conjecture, stating that "the degree of unity or diversity of EF varies from age to age" (p. 1652). Nevertheless, there have been reports of fewer EF factors in samples of older children (e.g., Huizinga, Dolan, & van der Molen, 2006) as well as more EF factors in younger children (e.g., Wu et al., 2011).

The hypothesis that the factor structure of cognitive abilities differentiates with age stems from early work by Garrett (1946), who suggested that, over the course of development, factor structure changes from a strong general ability to separable correlated abilities. According to this hypothesis, a general factor of cognitive ability explains less variance in lower-level abilities with increasing age, such that the relations among specific cognitive abilities weaken. This age differentiation hypothesis has been examined extensively with respect to psychometric measures of intelligence (Deary et al., 1996; Filella, 1960; Hartmann, 2006; Hildebrandt, Lüdtke, Robitzsch, Sommer, & Wilhelm, 2016; Hülür, Wilhelm, & Robitzsch, 2011; Juan-Espinosa et al., 2002; Molenaar, Dolan, Wicherts, & van der Maas, 2010; Tucker-Drob, 2009). Few studies, however, have directly examined this hypothesis with respect to the structure of EFs, and no study to date has applied newer multivariate methods for modeling EF differentiation continuously as a function of age. Challenges in the measurement of EFs in childhood, as well as discrepancies in the content and coverage of measurement batteries, contribute to inconsistent conclusions about the differentiation of EF across age.

Few studies intent on testing the age differentiation hypothesis have used a single age-heterogeneous sample that is representative of the general population or that has completed the same battery of EF tasks. Exceptions are Lee et al. (2013) and Xu et al. (2013). However, Lee et al. (2013) used the same tasks as indicators for inhibition and switching, making interpretation of the differentiation of these factors difficult (even though they used different trials as indicators for the two factors), and Xu et al. (2013) split participants into age groups. The risk of missing a nonlinear relation between model parameter and age is substantial when age is arbitrarily categorized (MacCallum, Zhang, Preacher, & Rucker, 2002). In summary, the small number of studies investigating the age differentiation hypothesis directly—that is, using a single age-heterogeneous sample with a multivariate task battery—relied on categorization of age into discrete bins (Deary et al., 1996). To our knowledge, no previous study has examined developmental transformations in the structure of EF using methods that allow for a continuous moderation of model parameters.

The EF factor structure may transform with age as certain components become more central to general executive processing and, therefore, have a greater influence on other EF processes. It has been suggested that inhibitory control is particularly important for allowing young children to employ other EFs in pursuit of goals, and when an individual's baseline levels of inhibitory control increase with age, inhibition no longer plays a central role in regulating other EFs (Best, Miller, & Jones, 2009; Huizinga & van der Molen, 2007; Isquith, Gioia, & Espy, 2004). Others have extended this theory, suggesting that once a certain level of inhibition has been developed, working memory capacity plays a critical role as a processing resource throughout adolescence and adulthood (Best et al., 2009; Davidson, Amso, Anderson, & Diamond, 2006; Huizinga & van der Molen, 2007; Isquith et al., 2004; Vandierendonck, 2012). Language ability has also been discussed as a possible driver of EF development (Bishop, Nation, & Patterson, 2014), although a recent test of this theory by Messer et al. (2018) did not find that EF factor structure differed on the basis of verbal and nonverbal tasks.

Drawing on task performance on a multidimensional battery of EF tasks administered to more than 1000 children ranging from 7 to 15 years of age, the current study examined both quantitative and qualitative transformations in the structure of EF over the transition to adolescence. At the quantitative level of analysis, we tested whether the proportion of variance in individual EFs that is explained by a common EF factor decreased with age, as would be predicted by the age differentiation hypothesis. We examined age-related shifts in the relative contributions of domain-common and domain-general variance to individual differences in EF performance. In other words, we tested whether the core of common EF was differentially representative of EF domains at different ages.

We addressed these questions using three complementary and innovative methods for estimating continuous age moderation in multidimensional space: moderated factor analysis (MFA; Cheung, Harden, & Tucker-Drob, 2015; Molenaar et al., 2010; Tucker-Drob, 2009), local structural equation modeling (LSEM; Hildebrandt, Wilhelm, & Robitzsch, 2009), and network modeling (Epskamp, Rhemtulla, & Borsboom, 2017). MFA and LSEM are based on a measurement model of EF testing whether age moderates the strength of associations between tasks in the same EF domain as well as the strength of associations across EF domains. Both analytical approaches allow for continuous age moderation of factor model parameters. Whereas the shape of the moderation effect needs to be specified a priori in MFA (e.g., a linear function in the case of linear moderation), this is not necessary for the nonparametric LSEM analysis. As a baseline model, we relied on the factor structure previously established for EFs generally (Miyake & Friedman, 2012; Miyake et al., 2000) and in this sample specifically (Engelhardt, Briley, Mann, Harden, & Tucker-Drob, 2015; Engelhardt et al., 2016) (more details about the measurement model can be found in the Method section). Network analysis is used to investigate the interrelations of tasks without a predefined structure. In summary, the current study moved beyond previous studies that used individual CFAs for different age groups, namely by our ability to investigate factor structure across continuously measured age.

Method

Sample

Data were drawn from 1019 participants from the Texas Twin Project (Harden, Tucker-Drob, & Tackett, 2013), a population-based sample that included children spanning the range of functioning so long as they were able to understand the instructions and complete the tasks encompassed in the protocol. For this reason, children with very severe disabilities and delays were not eligible to participate. The average full scale intelligence quotient (FSIQ) of participants in the sample measured by the Wechsler Abbreviated Scale of Intelligence (Wechsler, 2011) was 103.99, with a standard deviation of 14.09. These scores are normed to have a mean of 100 and a standard deviation of 15 in the general population of the United States. Therefore, the sample was very close to average IQ (slightly above) and nearly as variable as (slightly less than) the nationwide standardization sample. The current sample consisted of children in Grades 3 to 8 between 7.8 and 15.3 years of age (M = 10.79 years, SD = 1.76). This sample, of which 50.4% was female, included 479 twin pairs, 19 triplet sets, and 1 quadruplet set. In terms of race/ethnicity, 59.1% of the sample identified as non-Hispanic Caucasian, 15.0% as Hispanic, 6.7% as African American, 4.2% as Asian, 0.6% as another race or ethnicity, and 14.3% as multiple races or ethnicities.

Measures

In total, 15 tasks were administered to assess four EF domains: inhibition, switching, updating, and working memory (see Table 1). The battery initially consisted of 12 tasks (three per domain). After approximately 2 years of data collection, three of the tasks were replaced by versions that were amenable for use in a magnetic resonance imaging (MRI) scanner as follows: Stop Signal auditory was replaced by Stop Signal visual, 2-Back was replaced by *N*-Back (consisting of 1- and 2-Back), and Plus–Minus was replaced by Cognitive Flexibility. The remaining nine tasks were collected for the whole sample. Data used in this analysis were collected following a planned missing data design (Little & Rhemtulla, 2013).[1] The tasks constituting the EF battery are described in Appendix A (Table A1).

Inhibition was measured with the following tasks: Animal Stroop (Wright, Waterman, Prescott, & Murdoch-Eaton, 2003), Mickey (Lee et al., 2013), and Stop Signal auditory and visual (Logan, Schachar, & Tannock, 1997; Verbruggen, Logan, & Stevens, 2008). Differences in reaction times for inhibit versus noninhibit trials were used as dependent variables for Animal Stroop and Mickey. The dependent variable for both Stop Signal tasks was Stop Signal reaction time, which was calculated by subtracting the average Stop Signal delay (time between presentation of the "go" stimulus and the "stop" stimulus) from mean go reaction time. Block-level Stop Signal scores (64 trials per block) were computed and omitted on the basis of consistent stop failures, misidentification of arrow direction, failure to respond to go trials, and implausibly low Stop Signal reaction times (Congdon et al., 2010). Scores from the remaining blocks were then averaged to produce a final Stop Signal reaction time.

Switching was assessed with Trail Making (Salthouse, 2011), Local–Global (Miyake et al., 2000), Plus–Minus (Miyake et al., 2000), and Cognitive Flexibility (based on Heaton, 1993). The outcome of interest for Trail Making, Local–Global, and Plus–Minus was switch cost or the difference in reaction times between switch and nonswitch conditions. To correct for positive skew, the scores of Trail Making and Local–Global were log-transformed. For Cognitive Flexibility, the reaction time for all correct trials was used as the dependent variable.

Working memory (WM) was assessed with Digit Span Backward (Wechsler, 2003), Symmetry Span (Kane et al., 2004), and Listening Recall (Daneman & Carpenter, 1980). All memory tasks were scored

[1] Results from MFA and LSEM analyses using only the old tasks did not substantially differ from the analysis including all tasks.

Table 1
Descriptive statistics for executive function tasks.

Task	*n*	Mean (ms)	*SD* (ms)	Reliability α
Inhibition				
Stroop	1016	238.77	239.91	.86[a]
Mickey	780	35.40	70.87	46[a]
Stop Signal auditory	594	322.32	90.74	.40[b]
Stop Signal visual	326	267.07	65.71	.31[b]
Switching				
Trail Making	862	1219.57	783.62	.84[a]
Local–Global	1002	1436.68	746.63	.74[a]
Plus–Minus	582	669.55	1200.83	.69[a]
Cognitive Flexibility	374	1122.84	192.89	.80[b]
Updating				
Running Memory	833	18.84	8.34	.75[c]
Keeping Track	1004	6.57	2.36	.52[c]
2-Back	599	3.77	1.07	.84[b]
N-Back	396	4.47	1.73	.90[b]
Working memory				
Symmetry Span	993	19.90	8.69	.78[c]
Listening Recall	954	23.35	8.12	.79[c]
Digit Span Backward	880	7.02	1.83	.60[c]

Note. Total $N = 1019$.
[a] Reliability estimates were calculated based on difference scores formed by subtracting reaction time on nonswitch (or noninhibit) blocks from reaction time on switch (or inhibit) blocks for each possible pair of switch (inhibit) and nonswitch (noninhibit) blocks.
[b] Reliability estimates were calculated across blocks.
[c] Reliability estimates were calculated across trials.

with the number of items correctly recalled. The square root of Listening Recall scores was taken to correct for skew.

Updating was assessed with 2-Back, *N*-Back (Jaeggi et al., 2010), Keeping Track (Miyake et al., 2000), and Running Memory for Letters (Broadway & Engle, 2010). For 2-Back and *N*-Back, we computed the number of correctly identified matches minus incorrectly identified nonmatches (i.e., hits minus false alarms). Square root transformations were applied to 2-Back and *N*-Back scores to correct for skew. The outcome of interest for the remaining updating tasks was the number of items correctly recalled.

Table 1 reports descriptive statistics for the EF indices, including the number of observations per task. As discussed in Engelhardt et al. (2015), low reliability estimates for many of the EF indices are due, in part, to the scoring of these tasks. Building a difference score between executive and nonexecutive conditions leads to the compounding of measurement error (Cronbach & Furby, 1970). This assumption is substantiated by the high reliabilities for the individual task conditions that underlie the difference scores (Engelhardt et al., 2015, 2016).

Analyses

As described earlier, age-related differences in the EF covariance structure were examined using three different analytical methods: MFA (Molenaar et al., 2010; see also Cheung et al., 2015), LSEM (Hildebrandt et al., 2009; also see Briley, Harden, & Tucker-Drob, 2015), and network modelling (Epskamp et al., 2017). A comparison of the three methods can be found in Table 2. Structural equation models were fit in Mplus Version 7.4 (Muthén & Muthén, 1998–2012). Mplus uses a full-information likelihood estimation to account for missing data. In all analyses, to correct standard errors for the nonindependence of observations from individuals within the same family, a sandwich estimator was implemented using the Complex Survey option.[2]

[2] A sensitivity analysis of the analyses was done with data from one sibling per family. Results can be found in Appendix A.

Table 2
Comparison of characteristics of moderated factor analysis, local structural equation modeling, and network analysis approaches used in the current study.

	MFA	LSEM	Network analysis
Parametric	✓	X	X
Inferential tests	✓	X	X
Detection of onset	X	✓	✓
Confirmatory	✓	✓	X
Continuous moderator	✓	✓	✓

Note. A check mark indicates that it is a characteristic of this approach, a cross notates it is not; MFA, moderated factor analysis; LSEM, local structural equation modeling.

Measurement model

Engelhardt et al. (2015) compared a series of competing models, concluding that a hierarchical factor structure with a common EF factor summarizing the common variance of four factors representing four broad EF domains (inhibition, switching, WM, and updating) represents the data best. Following Molenaar et al. (2010), we examined differentiation at multiple levels within a hierarchical factor structure. In this specific case, we used a higher-order factor model in which observed task scores are related to first-order factors (first-order loadings) and the first-order factors are regressed on a second-order factor (second-order factor loadings). The second-order factor explains the intercorrelations among the first-order constructs (see Fig. 1). Hence, the tasks are influenced by the four first-order factors representing the constructs inhibition, switching, updating, and WM. However, tasks are also influenced by task-specific factors (i.e., residual variances) that are independent of first- and second-order factors. The shared variance of the four EF domains is accounted for by a second-order factor that represents the higher-order construct EF. The age differentiation hypothesis is typically conceptualized that the correlations among EF domains decrease with age. However, it is necessary to investigate all levels of the model so as to test the age differentiation hypothesis because differentiation can arise on the task level (Molenaar et al., 2010). We examined age trends in the magnitudes of the residual variances of the individual EF tests, loadings of the individual tests onto the first-order EF domains, residual variances of the first-order domains, and loadings of the first-order domains onto the higher-order common EF factor. If EF follows the age differentiation hypothesis, we would anticipate residual variances of individual EF tests and/or first-order EF factors to increase systematically across age and/or first-/second-order loadings to decrease systematically across age

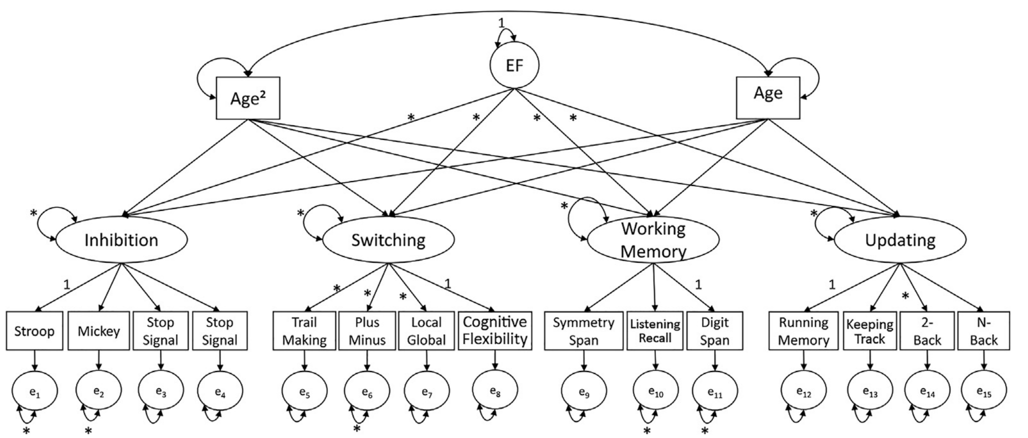

Fig. 1. Path diagram for moderated factor analysis. Asterisks indicate age-moderated parameters. Moderation of first-order loadings and test-specific residuals was specified only when age moderation was significant in initial domain-specific models. EF, executive function.

depending on the level on which differentiation occurs. Unsystematically changing higher-order factor loadings could be indicative for a shift in the meaning of the common EF factor, with differing EF domains getting more or less important. Using the same factor model across the whole age range allows us to investigate effects on parameters in an explicit model.

Moderated factor analysis

MFA allows for the examination of model parameters along age as continuous moderator according to a parametric function (e.g., a linear function in the case of linear moderation). The parameters (e.g., loadings of an indicator) are predicted by age, resulting in regression equations for each moderated parameter. The interpretation of the coefficients is similar to a regression analysis; the intercept is the expected parameter at the mean of the variable age (because age is centered) and the slope indicates a linear increase or decrease in the model parameter with increasing age. When the chosen parametric function conforms to the shape of the true moderation function, MFA is advantageous for increasing power while decreasing false discovery. The restriction to linear effects can produce unintuitive estimates for some values of the moderator. Prior to fitting MFA within a full hierarchical structure, we fit separate MFA models for subsets of variables from each of the four EF domains (inhibition, switching, updating, and WM). Age was centered by subtracting its mean. To scale the metric of each factor, its variance was fixed to 1. We included main effects of age and age squared on the latent factor and allowed for age-moderated factor loadings and loadings onto age-moderated residual variances. We went on to fit MFA within a fully hierarchical structure that was informed by the domain-specific analyses. In this model, we allowed for main effects of age and age squared on the first-order latent factors and allowed all second-order factor loadings and loadings onto residual factors to be linearly moderated by age. Moderation of the factor loadings and loadings onto residual factors of the individual EF measures was allowed when these parameters showed significant moderation by age in the domain-specific models. In the hierarchical factor model, we scaled the metric of each first-order factor by fixing a single loading to 1.0 for an indicator that had demonstrated no age moderation in the initial domain-specific models. This procedure is data driven and needed to have enough power to estimate parameters in the full model. However, it also means that not all possible moderation parameters are estimated in the full model.

The focus of MFA is the significance of slope parameters, reflecting moderation of the respective parameter. A significant coefficient indicates a linear moderation of the respective parameter by age, whereas the parameter is assumed to be not (linearly) moderated by age if the slope coefficient is not significantly different from 0.

Local structural equation modeling

LSEM is a method for modeling a parameter of a structural equation model as a continuous nonparametric function of a moderator (in this case age). LSEM is more descriptive and flexible than MFA because a priori specification of the age function of the parameter estimates (e.g., linear, quadratic, exponential) is not necessary. Therefore, it is particularly useful for detecting nonlinear trends (for an illustration, see Fig. 5 in Olaru, Schroeders, Hartung, & Wilhelm, 2019) by visualizing transformations in factor model parameters. LSEM estimates a series of models along a range of moderator values (i.e., ages), with observations weighted as a function of their proximity to a focal age point. The weighting function was based on the recommendations of Hildebrandt et al. (2009). At each defined value of the age variable, called focal points, a Gaussian kernel function is used to weight observations, with the highest weight at the age point and decreasing weights for observations further from this age point. Because of the nonparametric character of LSEM, onset of transformation can be detected. We fit LSEM to examine age trends in the parameters for the same hierarchical model of EF for which we fit the MFA models. As described by Briley et al. (2015), the *MplusAutomation* package in R (Hallquist, 2016; R Development Core Team, 2001) was used to automate the fitting of the multiple models.

We investigated trends in standardized loadings. Standardized first-order loadings (from EF domain factors to indicator tasks) indicate how first-order loadings are defined by the indicators while also allowing for statements about how reliable indicators measure the construct. How stable the common EF factor is defined was investigated with standardized second-order loadings (loadings from the EF domain factors to common EF).

Network modeling

Network modeling is the most exploratory approach to representing the factor structure of a set of indicators because it imposes no particular structure on variable interrelations. In a conventional psychometric network model, a correlation matrix of the task scores is used to build a network representing the interrelations between tasks. Here, we extended the network approach to investigate age trends in the EF task covariance network. We estimated age-specific, fully saturated correlation matrices using the LSEM weighting approach. Thus, correlation matrices for each weighted sample are formed, and a network illustrating these relations can be extracted. This is the most exploratory of the three approaches and has the strongest potential to illuminate qualitative transformations in the relative patterns of covariation among EF variables across age. Due to the data collection design described earlier, correlations between the three old and three new tasks were necessarily missing, and including all tasks could lead to inaccurate depiction of the relations. Based on the larger sample size for the old tasks and a slightly more evenly distributed age in this subsample, we excluded the three new tasks from this analysis. The network includes 12 nodes, which represent the 12 EF tasks, and 66 edges, which are the connections between nodes.

We aimed to evaluate the centrality of specific nodes, meaning the interrelation of performance on one task with that on all the other tasks. Although different indices can be used to measure centrality, they are typically computed on the basis of the edges, such that nodes sharing a lot of variance with other nodes have high centrality (Opsahl, Agneessens, & Skvoretz, 2010). One such index is the closeness of a node, which is computed as the inverse of the total length of all the shortest paths between the selected node and all other nodes in the network. Another centrality index, strength, is based on summing up the weights of the direct relations of a node to all other nodes. Higher closeness centrality indicates that a task is related to more other tasks, and strength indicates that a task is related more strongly with other tasks. Other measures of centrality were not reported based on their limitations (e.g., *degree* ignores the structure of the network) or redundancy with one of the former mentioned indices (e.g., *betweenness* is redundant with the closeness index).

Results

Descriptive statistics

Correlations among age and scores on the individual tasks can be found in Table 3. Age significantly correlated with performance on all tasks. Most variables were positively related to each other. Scatterplots depicting the mean value trajectories of the indicators across age can be found in Appendix A (Fig. A1).

Moderated factor analyses

MFA produces estimates reflecting the age moderation of model parameters that can be tested for differing significantly from 0. According to the age differentiation hypothesis, EF domains should share less variance with one another across age. This would be supported by negative moderation coefficients for associations between the common EF factor and the domain-specific factors and by positive moderation coefficients for residual loadings.

We began by fitting domain-specific moderation models in which we allowed for an interaction parameter constraint to be equal across all loadings and an interaction parameter constraint to be equal across variances (see Cheung et al., 2015). For each domain, we compared models with and without interaction terms by conducting chi-square difference tests based on log likelihood values and scaling correction factors. In the next step, we included specific interaction coefficients[3] for each

[3] All coefficients reported for the moderated factor analyses are unstandardized due to the model constraints.

Table 3
Correlation matrix of task performance and age.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Stroop															
2. Stop Signal auditory	.02														
3. Stop Signal visual	.09	NA													
4. Mickey	.10	.11	.10												
5. Trail Making	.21	.07	.08	.16											
6. Plus–Minus	.25	.06	NA	.09	.20										
7. Local–Global	.27	.04	.18	.15	.39	.22									
8. Cognitive Flexibility	.26	NA	.13	.04	.29	NA	.30								
9. Running Memory	.25	.07	.06	.17	.44	.21	.39	.27							
10. 2-Back	.22	.12	NA	.18	.39	.14	.34	NA	.54						
11. *N*-Back	.27	NA	.11	.18	.39	NA	.43	.38	.58	NA					
12. Keeping Track	.21	.07	.23	.13	.38	.15	.31	.24	.47	.42	.36				
13. Digit Span Backward	.18	.05	.03	.14	.36	.16	.28	.13	.44	.33	.29	.35			
14. Symmetry Span	.27	.02	.11	.18	.40	.21	.35	.30	.46	.38	.45	.36	.32		
15. Listening Recall	.28	.02	.15	.21	.44	.25	.40	.35	.62	.45	.50	.45	.42	.52	
16. Age	.34	.18	.23	.26	.43	.24	.43	.31	.44	.35	.40	.36	.30	.50	.53

Note. Zero-order Pearson correlation coefficients are shown. All task scores are standardized. Cells marked "NA" are due to the change in task administration (participants were administered only one of the tasks).

indicator; interaction coefficients applied to the factor loadings, the loadings onto residual factors, or both, depending on the previous model results. Table 4 summarizes the estimates from the four separate models, one for each EF domain. We demonstrate how to interpret an MFA parameter using the example of Plus–Minus. The main effect value (.22) is the expected loading of Plus–Minus on the switching factor at mean age (because age is centered). The linear age moderation effect, which is negative (−.18), signifies a decrease in the loading with a one-unit increase in age. The effects on the loading onto the residual variance can be interpreted in the same manner. Results from the domain-specific models (see Table 4) indicate that (a) for the inhibition indicators, the residual variances for Stroop and Mickey were significantly moderated by age, with both decreasing over age; (b) for switching, the loadings of Trail Making, Plus–Minus, and Local–Global onto their respective lower-order factors decreased over age, as did the residual variance for Plus–Minus; (c) the loading of 2-Back onto the updating factor increased over age; and (d) the residual variances of WM indicators Digit Span Backward and Running Memory increased slightly over age.

We went on to construct a hierarchical EF model, allowing for age moderation of EF factor loadings and domain-specific variances as well as for task-specific loadings and variances that were significant in the domain-specific models. A path diagram of the MFA model, including the common EF factor, is depicted in Fig. 1. Coefficient estimates for this model are summarized in Table 5. Loadings onto the residual variances for the inhibition and switching factors were moderated by age, as was the factor loading of switching onto the common EF factor. On the other hand, higher-order loadings and residual variances for WM and updating were stable across age. At the indicator level, the largest age trend corresponded to residual variances for Stroop and Plus–Minus. The estimates for the EF domains can be interpreted as follows. For inhibition, the loading onto EF at 10.79 years of age (which is the mean age in the sample) is .16 with no significant linear moderation, with a residual variance of .005 (the loading onto the residual variance factor needs to be equivalent to the residual variance) at mean age and .04 at about age 14. For switching, the loading onto EF at mean age is .27 with a significant decrease of −.04 per year, which results in a factor loading of .14 (notice that differences occur due to rounding errors) at about age 14; the residual variance of .01 at mean age decreases linearly with increasing age, resulting in a negative residual variance of −.02 at about age 14. Factor loadings of updating and WM are not significantly moderated.

Fig. 2 depicts the proportion of variance in each of the four EF domains that was explained by the general EF factor is depicted as a function of age. For a given EF domain and age, this was computed as $\text{loading}^2 \div (\text{loading}^2 + \text{residual variance}^2)$. The proportion of variance in the inhibition factor that could be explained by general EF decreases after about 11 years of age. Age trends for switching were

Table 4
Coefficient estimates from the domain-specific moderated factor analysis models.

		Loading on latent domain-specific factor		Loading on residual variance	
		Main effect	Linear age moderation	Main effect	Linear age moderation
Inhibition	Stroop	.27	–	.78***	–.20**
	Mickey	.20**	–	.96***	–.09***
	Stop Signal auditory	.13**	–	.98***	–.01
	Stop Signal visual	.19**	–	.96***	–.04
Switching	Trail Making	.37***	–.15*	.82***	–.01
	Plus–Minus	.22***	–.18***	.69***	–.38***
	Local–Global	.40***	–.19**	.81***	–.01
	Cognitive Flexibility	.32	–.16	.88***	.00
Updating	Running Memory	.64***	.03	.61***	–.01
	Keeping Track	.48***	.03	.81***	–.01
	2-Back	.55***	.13**	.73***	–.04
	N-Back	.54***	–.01	.72***	–.01
Working memory	Symmetry Span	.49***	–	.73***	–.01
	Listening Recall	.57***	–	.61***	.06*
	Digit Span	.35***	–	.87***	.04*

Note. Note that models were estimated separately for each domain.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 5
Coefficient estimates from the higher-order moderated factor analysis model.

	Main effect age	Main effect age^2	Factor loading		Loading on residual factor	
			Main effect	Linear age moderation	Main effect	Linear age moderation
Inhibition	.21***	–.02*	.16***	–.02	.07	.10*
Switching	.20***	–.03**	.27***	–.04**	.12**	–.10***
Updating	.28***	–.04***	.62***	.01	.17*	.02
Working memory	.21***	–.02**	.35***	–.01	.12*	–.00
Stroop			1	–	.81***	–.19**
Mickey			.63***	–	.96***	–.09***
Stop Signal auditory			.35*	–	.99***	–
Stop Signal visual			.58***	–	.97***	–
Trail Making			1.37***	.12	.78***	–
Plus–Minus			.74***	–.02	.81***	–.23***
Local–Global			1.30***	.06	.79***	–
Cognitive Flexibility			1	–	.87***	–
Running Memory			1	–	.60***	–
Keeping Track			.76***	–	.80***	–
2-Back			.84***	.08**	.74***	–
N-Back			.83***	–	.74***	–
Symmetry Span			1.29***	–	.75***	–
Listening Recall			1.52***	–	.62***	.03*
Digit Span Backward			1	–	.85***	.03*

* $p < .05$.
** $p < .01$.
*** $p < .001$.

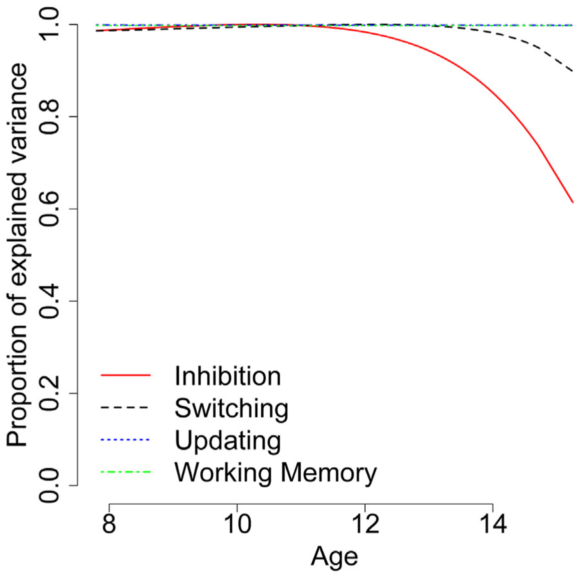

Fig. 2. Proportion of variance in the first-order factors explained by the general executive function factor across age, as measured by moderated factor analysis.

flatter, with the decrease starting at about age 13. WM and updating were not significantly moderated by age.

According to the age differentiation hypothesis, general EF is expected to account for less variance in the EF domains with age, resulting in a decrease of second-order factor loadings. We do not find a general decrease in second-order factor loadings, indicating that the variance among all four EF domains does not decrease generally. What we find is a more specific pattern where the higher-order EF factor consistently accounts for individual differences in WM and updating but not in inhibition and switching.

Local structural equation modeling

LSEM allows for visual detection of nonlinear relations between model parameters and age. According to the age differentiation hypothesis, we should observe decreases in standardized loadings of domain-specific EF onto the common EF factor with increasing age. For LSEM, focal age points ranged from 8.0 to 14.0 with 0.1 increments. These upper and lower limits were selected to reduce boundary bias based on the age distribution of the sample (Hildebrandt et al., 2009, 2016). In total, we estimated 61 weighted models. Standardized loadings are reported to facilitate interpretation. Fig. 3 shows the trajectories of the standardized higher-order factor loadings, and Fig. 4 shows the first-order factor loadings. LSEM indicated that the standardized loadings of the switching, updating, and WM factors on common EF were relatively stable over the age range. The standardized loading of inhibition onto common EF was 1.0 at earlier ages. After approximately 10 years of age, this loading dropped precipitously, reaching a lower magnitude of approximately .40. This result indicates that inhibition shared less variance with the other factors after about age 10.

The first-order loadings of specific tasks onto the updating and WM factors were stable with two exceptions. First, after 12 years of age, the loading of Symmetry Span onto WM decreased, whereas the loadings of the other two tasks slightly increased. Second, the loading of 2-Back onto the updating factor increased. Both of these trends were also observed in the MFA results. The loading of Cognitive Flexibility onto the latent switching factor showed a reverse U-shaped relation with age. Between about 10 and 12 years of age, Cognitive Flexibility and Trail Making had the highest factor loadings,

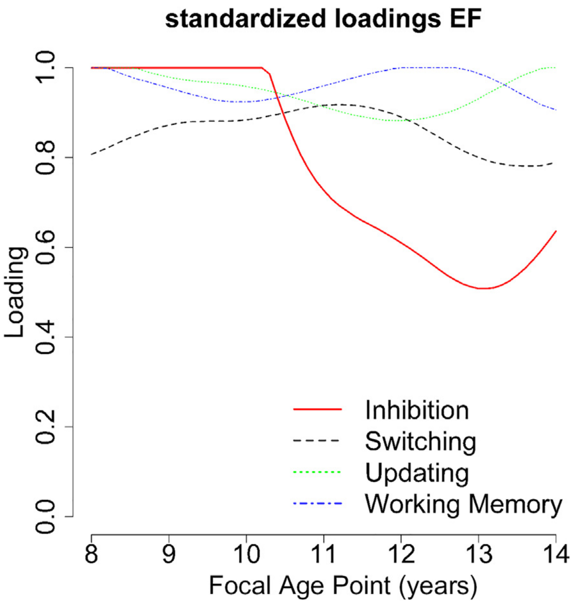

Fig. 3. Standardized second-order factor loadings across age from local structural equation modeling. EF, executive function.

whereas at younger and older ages, Trail Making and Local–Global had the highest loadings. Loadings of individual tasks onto the inhibition factor remained negative or close to 0 at younger ages, complementing MFA results concerning instability of the inhibition factor across age. Both Stop Signal tasks showed considerable increase in factor loadings in the LSEM model.

Counter to the age differentiation hypothesis, results of this analysis do not show a general decrease in second-order factor loadings, nor is there an overarching pattern at the level of first-order factor loadings.

Network models

The main goal of network analysis is to investigate relations at the task level. According to the age differentiation hypothesis, we should observe a general decrease in relations among all tasks. We used the ages and steps from the LSEM analysis to estimate the saturated correlation matrices on which we based the network models. (A visual depiction of network transformations across age can be found in the online supplementary material.) In a visual representation of a network model, the thickness of the lines corresponds to the magnitude of the interrelation. Network models indicated that the inhibition tests became less strongly related to one another from 8 to 13 years of age. This pattern had the effect of reducing variance shared by inhibition with the other first-order domains. However, it is worth remembering that loadings of the inhibition tasks onto their first-order factor were sometimes negative at younger ages. Because the network was based only on the *strength* of interrelations, node valence was not taken into account; in the supplemental visual depiction of network transformations, inhibition indicators are closer to each other even when the relation is negative. Fig. 5 plots the centrality indices by age, with strength capturing how strongly a task is directly connected to other tasks by summing up the absolute values of all correlations of a task with the other tasks and with closeness describing each task's indirect closeness to all other tasks by taking the inverse of the sum of all the shortest paths between a task and all other tasks in the network. We see that, for many of the nodes, centrality decreased over the age range of the sample. This is particularly the case for the Stroop and Plus–Minus tasks. Higher centrality before 10 years of age for the inhibition indicators might be explained by higher negative relations at younger ages and slightly positive relations after age 10.

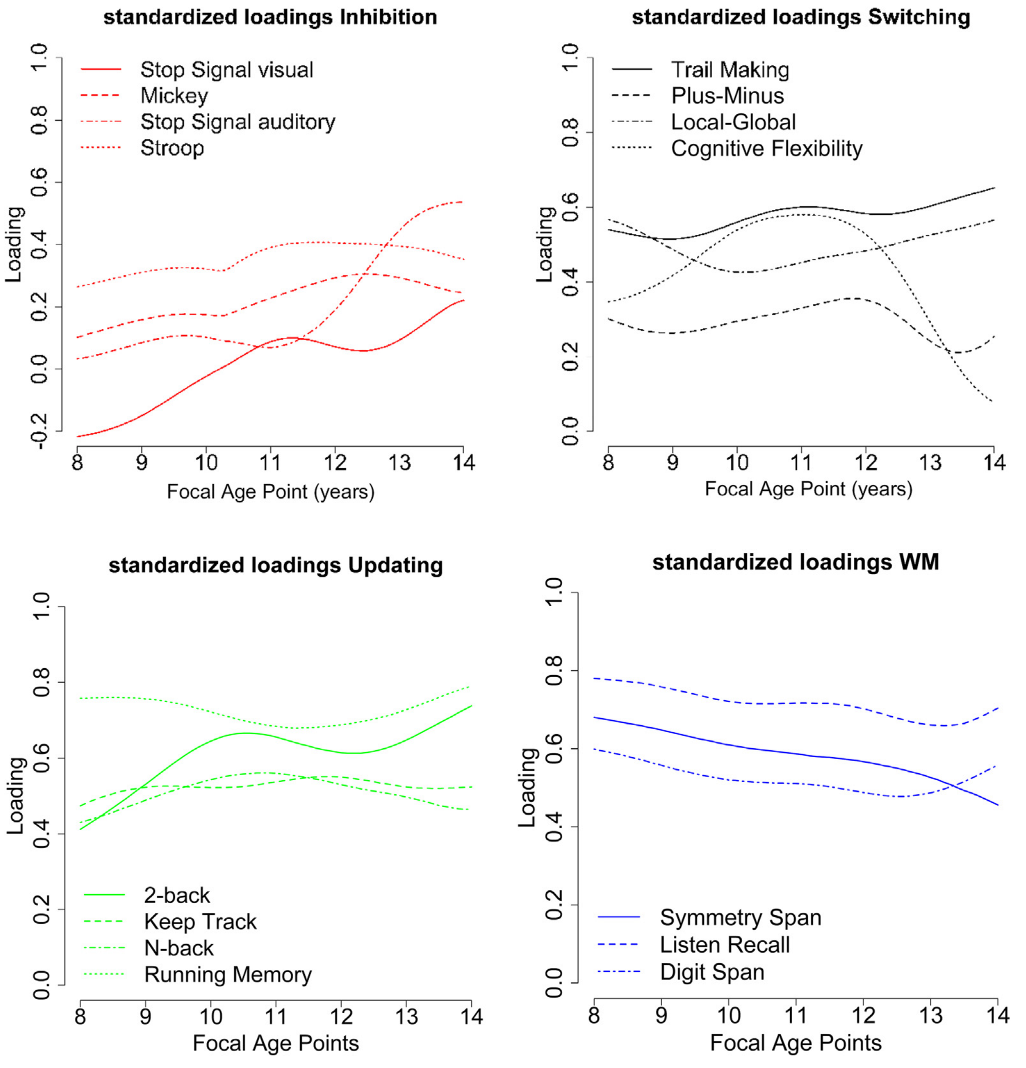

Fig. 4. Standardized first-order loadings on first-order factor loadings from local structural equation modeling. WM, working memory.

The decrease in centrality of all indicators until approximately 13 years of age could indicate support for the age differentiation hypothesis. However, taking a closer look on the network or the correlation coefficients, this decrease is largely driven by the inhibition indicators and by Plus–Minus (see online supplementary material). Steep decreases in correlations of the Stroop and Plus–Minus tasks would lead to lower centrality indices for all indicators if they were not simultaneously becoming more related to each other. Furthermore, the increase in centrality after age 13 is contrary to the age differentiation hypothesis. Therefore, the network analysis suggests the same implication as the previous two approaches: There is no general decrease in task correlations, but specific tasks have an age-moderated relationship with the remaining EF tasks.

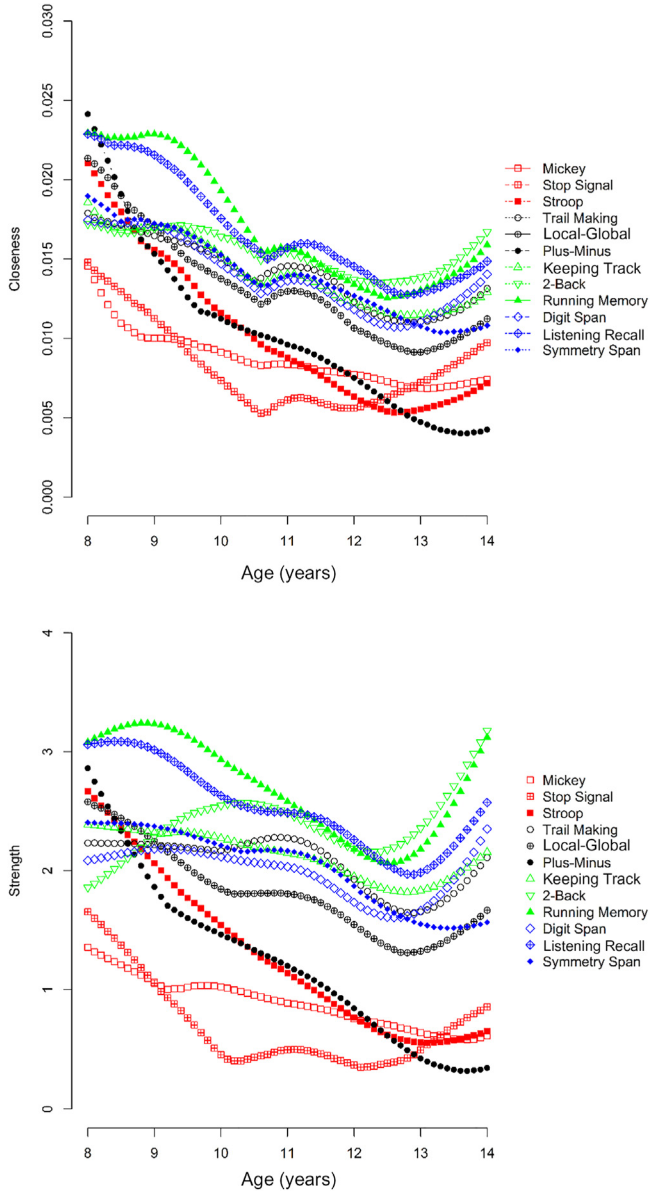

Fig. 5. Closeness and strength of the 12 nodes in the network models across age.

Sensitivity analyses using one twin per family

Although we statistically accounted for the dependency between individuals from the same family in all previously described analyses, a sensitivity analysis is useful to determine whether the pattern of results was unduly influenced by the dependency between siblings. Sensitivity analysis were conducted by repeating the analyses with a subsample consisting of one sibling per family only. In summary, the results from the sensitivity analyses indicated that the pattern of results did not change substantially if a subsample consisting of only independent observations was used. Due to the lower power in the subsample, linear moderation of inhibition was hard to discover within the subsample, but nonparametric and correlational analyses did not differ substantially from the complete sample. Refer to Appendix A for a detailed description of the results from the sensitivity analysis.

Discussion

The aim of this project was to investigate age trends in the multivariate structure of children's executive abilities. We used three complementary techniques, exploiting their respective strengths. In contrast to the age differentiation hypothesis, none of the three methods indicated a global pattern of decreasing EF interrelations with age. MFA and LSEM did not show consistent decreases in factor loadings or increases in residual variances, nor did network analysis indicate uniform decreases in relations between nodes across age. Rather, we observed a more nuanced pattern marked by consistent covariance patterns for WM and updating across the age range and differentiation of inhibition from the other EFs with age. Trends for switching were less consistent across analytic approaches.

Results of MFA indicated that parameters linking a common EF factor to the lower-order WM and updating factors were stable over the age range. In contrast, inhibition and switching factors differentially loaded onto the common EF factor across the age range. LSEM highlighted the relative stability of the switching loading across the age range but substantiated the age trends detected for the loading of the inhibition factor onto common EF. In particular, the common EF factor explained 100% of the variance in the inhibition factor (loading = 1) through approximately 10 years of age and dropped in magnitude to approximately .50 by the teenage years. This effect could be driven by the small to negative loadings of the tasks onto the inhibition factor at earlier ages, reflecting a lack of shared variance or high task specificity. In the network models, the inhibition tasks and one switching task (Plus–Minus) became more remote from the other tasks through approximately 13 years of age, indicating that variance shared between these and other tasks decreased, then subsequently increased. Despite nuanced differences in the results as a function of the three analytic approaches, all approaches indicated substantial age differences in the structure of the inhibition factor across the age range, with greater consistency across age observed for the WM, updating, and switching factors.

It is less clear whether age differences affected the *domain-specific* factor structure. The network approach indicated that at younger ages there were negative relations specific to the inhibition tasks. Very low to negative relations between inhibition tasks have been found in other studies. For instance, in Huizinga et al. (2006), a common inhibition factor could not be extracted due to low correlations of the tasks. The low task loadings onto the inhibition factor in the current study's LSEM and MFA, in addition to decreasing age-moderated residual variance in the Stroop and Mickey tasks, are in line with this observation. All three methods cast doubt on the existence of a unified inhibition factor at younger ages. Earlier analyses based on the same tasks and partly the same sample also found low correlations and low factor loadings of the inhibition indicators (Engelhardt et al., 2015, 2016). In addition, reliability of the current inhibition tasks was relatively low, which is not unusual for inhibition tasks. Because inhibition tasks are scored by subtracting performance on one condition from performance on another, measurement error is compounded, leading to low reliability. Indeed, Engelhardt et al. (2015) previously reported relatively high reliability of individual conditions within inhibition tasks but lower reliability for difference scores across conditions. Future studies could investigate the effects of different scoring procedures on the presented results.

The high loadings of WM and updating onto the common EF factor are in line with studies seeking to understand the nature of a common EF factor. These studies have shown a high overlap between WM capacity and common EF, with some referring to this common ability as executive attention (McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010). According to Miyake and Friedman (2012), "Common EF is about one's ability to actively maintain task goals and goal-related information and use this information to bias lower-level processing" (p. 11). Similarly, Engle (2002, 2018) hypothesized that a major feature of both EFs and fluid intelligence is the capacity for controlled attention. More insight about the nature of the common EF factor is provided by studies suggesting that it is highly related to, but distinct from, both general intelligence and fluid intelligence specifically (Engelhardt et al., 2016; Friedman et al., 2008).

Interestingly, in the current study, all three methods indicate at least some developmental transformations in associations within and between EF domains. Engelhardt et al. (2015) investigated age invariance of their model with a subsample drawn from the Texas Twin Project ($N = 505$). They conducted a multiple group analysis with two age groups. They reported measurement invariance on the first-order level; a model in which first-order loadings and intercepts were constrained to be invariant across age groups fit the data as well as a model in which these parameters were free to differ across groups. However, second-order invariance was not tested, and the downsides of categorizing participants into arbitrary groups were discussed in the Introduction of the current article.

The measurement model used in the current analysis is based on previous studies that were performed on data drawn from the Texas Twin Project (Engelhardt et al., 2015, 2016). In this model, updating and WM are modeled as separate factors, although others have argued that the two factors assess the same underlying construct (e.g., Chein, Moore, & Conway, 2011; St Clair-Thompson & Gathercole, 2006). In the current sample, the model in which updating and WM tasks loaded onto a single factor exhibited the same fit as the primary model that included separate factors (standardized root mean square residual [SRMR] = .038 for both models). Due to the superiority of the four-factor model in previous analysis and a balanced representation of factors, the four-factor model provides a better basis for the current analysis. The number and nature of components of EF is a subject of an ongoing debate. As defined in the current analysis, WM can be interpreted as storage or memory maintenance measured with simple span tasks, whereas updating reflects storage-plus-processing or updating. Although there are studies arguing for a common ability underlying both domains (Schmiedek et al., 2009), other studies show that they are conceptually related but differentiable in terms of cognitive processes (Ecker, Lewandowsky, & Oberauer, 2014; Ecker, Lewandowsky, Oberauer, & Chee, 2010). Nevertheless, it should be noted that the WM and updating factors are substantially correlated (.74).

Best et al. (2009) theorized that a certain level of inhibition needs to emerge first to facilitate other EFs. Our results indicated that whereas inhibition tasks become more correlated with one another with increasing age, the inhibition factor becomes less central in the overall EF framework. Therefore, the inhibition factor might increasingly convey specific variance that is unique from other EF factors, whereas the little variance shared by the inhibition indicators in young age can be completely explained by common EF. The results of the current study are in line with Lee et al. (2013) and Xu et al. (2013), who found less differentiable factors in CFAs of younger children's EFs in the sense that we find correlations between second-order factors to be highest at younger ages. A possible explanation for the low loadings of inhibition indicators for the younger participants could be a stronger effect of the task impurity problem (Snyder, Miyake, & Hankin, 2015). This means that a greater proportion of variance in inhibition tasks can be attributed to specific task contexts such as articulation speed in the Stroop task. This concern with respect to the low convergent validity of inhibition tasks is not a new one. Weak correlations between inhibition tasks have been reported in many studies (Salthouse, Atkinson, & Berish, 2003; Shilling, Chetwynd, & Rabbitt, 2002; van der Sluis, de Jong, & van der Leij, 2007). Nevertheless, studies with data from older children or young adults have been able to identify inhibition factors, albeit with weak to moderate factor loadings (Miyake et al., 2000; Wu et al., 2011). The current results, therefore, are consistent with the finding that inhibition tasks converge only weakly on a common source of individual differences. Here, we expand on this finding by showing that the shared variance between inhibition tasks increases with increasing age. More generally, the results of the current study suggest that task impurity poses a major problem in the measurement of inhibition in earlier childhood.

The current study was based on a broad EF battery containing at least three tasks per EF domain and compared inferences derived from three different analytic approaches (Engle, 2018; Miyake et al., 2000). The initial selection of EF measures was based on a broad review of the literature on individual differences in EFs and has been empirically validated using confirmatory factor analytic methods (Engelhardt et al., 2015, 2016). Because of advantages and disadvantages of different analytic techniques, it is reasonable to combine them in examining age moderation. Nevertheless, the extent to which the inferences drawn here will generalize to other EF tasks, domains, or analytic approaches is unclear.

It is also important to consider whether results are affected by a sample composed of twins and multiples. Importantly, we statistically accounted for the dependency that arises from including multiple individuals per family. Furthermore, in a sensitivity analysis using only one sibling per family, results indicated patterns similar to those obtained from the entire dataset. Regarding generalizability from twin samples to the broader population of singletons, we cannot be certain because we have not tested singletons using this battery. Nevertheless, studies by the Miyake and Friedman group used a similar battery in both twins and non-twins and reported very similar factor structures (Friedman et al., 2008; Miyake & Friedman, 2012; Miyake et al., 2000). Furthermore, the FSIQ of the current sample was quite similar to that of the general U.S. population. Of course, any one study cannot be definitive in its conclusions. Therefore, it will be important for future work—employing different protocols, methods, and sampling strategies—to examine the question of age differentiation of EFs.

In summary, results of three different modeling strategies did not produce clear evidence for a global pattern of decreasing interrelations among domain-specific EF factors. Rather, we found that WM and updating are domains of EF with consistently high centrality to the EF construct space with age and that inhibition tasks increased in their within-domain coherence while at the same time increasing in between-domain separation over age. At the level of the constructs, EFs may be relatively stable in their dimensionality with childhood age. However, challenges may remain in the consistent measurement of some domains of EF in early childhood.

Acknowledgments

This project was supported by a National Institutes of Health grant (R01HD083613). The Population Research Center at the University of Texas at Austin is supported by a National Institutes of Health grant (R24 HD042849). The first author was funded by the equal opportunity office of Ulm University (Financial Support Programs for Female Scientists; Application No. 36).

Appendix A

Sensitivity analysis with only one sibling per family

The analyses described in the article were repeated using data from only one randomly chosen sibling per family ($n = 498$). The analysis followed the same procedure as that of the complete sample. Results from the domain-specific MFA models did not differ for inhibition and updating. The domain-specific model for switching differed from the results of the overall data set; only the residual factor loading of Plus–Minus was significantly moderated; and factor loadings for Trail Making, Plus–Minus, and Local–Global were not significantly moderated. However, in the higher-order model using the overall data set, these factor loading moderations also did not reach the significance value. In the domain-specific model for working memory, only the residual factor loading of Listening Recall was significantly moderated, whereas the residual factor loading of Digit Span Backward was not significantly moderated. In the higher-order model of the complete sample, both moderation estimates were quite small.

Table A2 shows the results from the higher-order MFA using the subset. Predictably, fewer estimates reached significance because power in this subsample is lower. Because of the lower power, we are not focusing on significance but rather report on substantial differences between parameter estimates. As can be seen in Table A2, most estimates were very close to the full sample. The most notable difference is the moderation of the loading on the inhibition residual factor, which was 10 times larger in the full sample. Fig. A2 shows the proportion of explained variance in the four EF domains.

Fig. A3 shows the trajectories of the standardized higher-order factor loadings, and Fig. A4 shows the first-order factor loadings from local structural equation modeling. Results did not differ substantively from the results of the complete sample.

Table A1
Descriptions of the tasks.

Task	Source	Paradigm
Inhibition		
Stroop	Wright et al. (2003)	Verbally identify animals from line drawings. In the *congruent* condition, the face of the animal matches the body. In the *incongruent* condition, the face does not match the body and identification should be based on the body. In the *neutral* condition, the face area is blank and identification should be based on the body
Mickey	Lee et al. (2013)	Indicate on which side of a computer screen a cartoon Mickey Mouse face appears while ignoring any squares that flash on-screen before Mickey. In the *congruent* condition, a square flashes on the same side where Mickey appears. In the *incongruent* condition, a square flashes on the opposite side. In the *neutral* condition, squares flash on both sides
Stop Signal auditory	Verbruggen et al. (2008)	Indicate which way an arrow points, but do not respond if a tone (stop signal) sounds after the arrow is presented
Stop Signal visual	Verbruggen et al. (2008)	Indicate which way an arrow points, but do not respond if a red X (stop signal) appears on top of the arrow
Switching		
Trail Making	Salthouse (2011)	Connect circles containing numbers in numerical sequence and circles containing letters in alphabetical order. In the two *simple* conditions, only numbers or letters are presented. In the two alternating conditions, both numbers and letters are presented, and the circles should be connected in an alternating sequence (numbers–letters: 1-A-2-B, etc.; letters–numbers: A-1-B-2, etc.)
Local–Global	Miyake et al. (2000)	Verbally identify letters and shapes composed of smaller letters and shapes. In the two *local* conditions, participants should name the small constituent letters or shapes. In the two *global* conditions, they should name the large overall letter or shape. In the *alternating* condition, they should alternate between naming the constituent and overall letters or shapes
Plus–Minus	Miyake et al. (2000)	Complete simple addition and subtraction problems on paper. In the *adding* condition, participants should add 1 to each provided number. In the *subtracting* condition, they should subtract 1 from each number. In the *alternating* condition, they should alternate between adding 1 and subtracting 1
Cognitive Flexibility		Pay attention to which of two possible rules (shape or color) is cued on a given trial. When a target stimulus appears, select one of two response choices that matches the target stimulus on the cued rule
Updating		
Running Memory	Broadway and Engle (2010)	View a sequence of single letters and identify the last *n* digits in order of their presentation
Keeping Track	Miyake et al. (2000)	Listen to words falling under four categories and recall the most recent word from a given category
2-Back	Jaeggi et al. (2010)	View a sequence of individual shapes and indicate when the current shape matches the shape from two trials prior
N-Back	Jaeggi et al. (2010)	View a sequence of individual shapes and indicate when the current shape matches the shape from *n* trials prior. Blocks consist of 1- and 2-back targets
Working memory		
Symmetry Span	Kane et al. (2004)	View and encode a square flashing on a grid. On alternating trials, indicate whether the geometric display is symmetrical. Later, recall the locations and order of the flashing squares on the preceding trials
Listening Recall	Daneman and Carpenter (1980)	Listen to single letters and sentences presented in alternation trials. On sentence trials, determine whether the sentences make sense. Later, recall the order of the letters on the preceding trials
Digit Span Backward	Wechsler (2008)	Repeat increasingly long strings of numbers backward

Note. This table is based on Table 1 from Engelhardt et al. (2015).

Table A2
Coefficient estimates from the higher-order moderated factor analysis model using the subset with only one sibling.

	Main effect age	Main effect age^2	Factor loading		Loading on residual factor	
			Main effect	Linear age moderation	Main effect	Linear age moderation
Inhibition	.21*** (.21)	−.01 (−.02)	.15*** (.16)	−.01 (−.02)	.12 (.07)	.01 (.10)
Switching	.21*** (.20)	−.01 (−.03)	.33*** (.27)	−.01 (−.04)	.11 (.12)	−.13*** (−.10)
Updating	.28*** (.28)	−.04*** (−.04)	.59*** (.62)	.01 (.01)	.00 (.17)	.00 (.02)
Working memory	.22*** (.21)	−.02** (−.02)	.37*** (.35)	−.01 (−.01)	.12 (.12)	−.00 (−.00)
Stroop			1 (1)	–	.89*** (.81)	−.13*** (−.19)
Mickey			.82*** (.63)	–	.94*** (.96)	−.08** (−.09)
Stop Signal auditory			.24 (.35)	–	.99*** (.99)	–
Stop Signal visual			.49* (.58)	–	.98*** (.97)	–
Trail Making			1.16*** (1.37)	– (.12)	.76*** (.78)	–
Plus–Minus			.38*** (.74)	– (−.02)	.84*** (.81)	−.23*** (−.23)
Local–Global			1.07*** (1.30)	– (.06)	.81*** (.79)	–
Cognitive Flexibility			1 (1)	–	.84*** (.87)	–
Running Memory			1 (1)	–	.64***	–
Keeping Track			.79*** (.76)	–	.79***	–
2-Back			.85*** (.84)	.14** (.08)	.74***	–
N-Back			.78*** (.83)	–	.79***	–
Symmetry Span			1.33*** (1.29)	–	.70*** (.75)	–
Listening Recall			1.49*** (1.52)	–	.59*** (.62)	.03 (.03)
Digit Span Backward			1 (1)	–	.84*** (.85)	– (.03)

Note. Estimates from the full sample are in parentheses.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

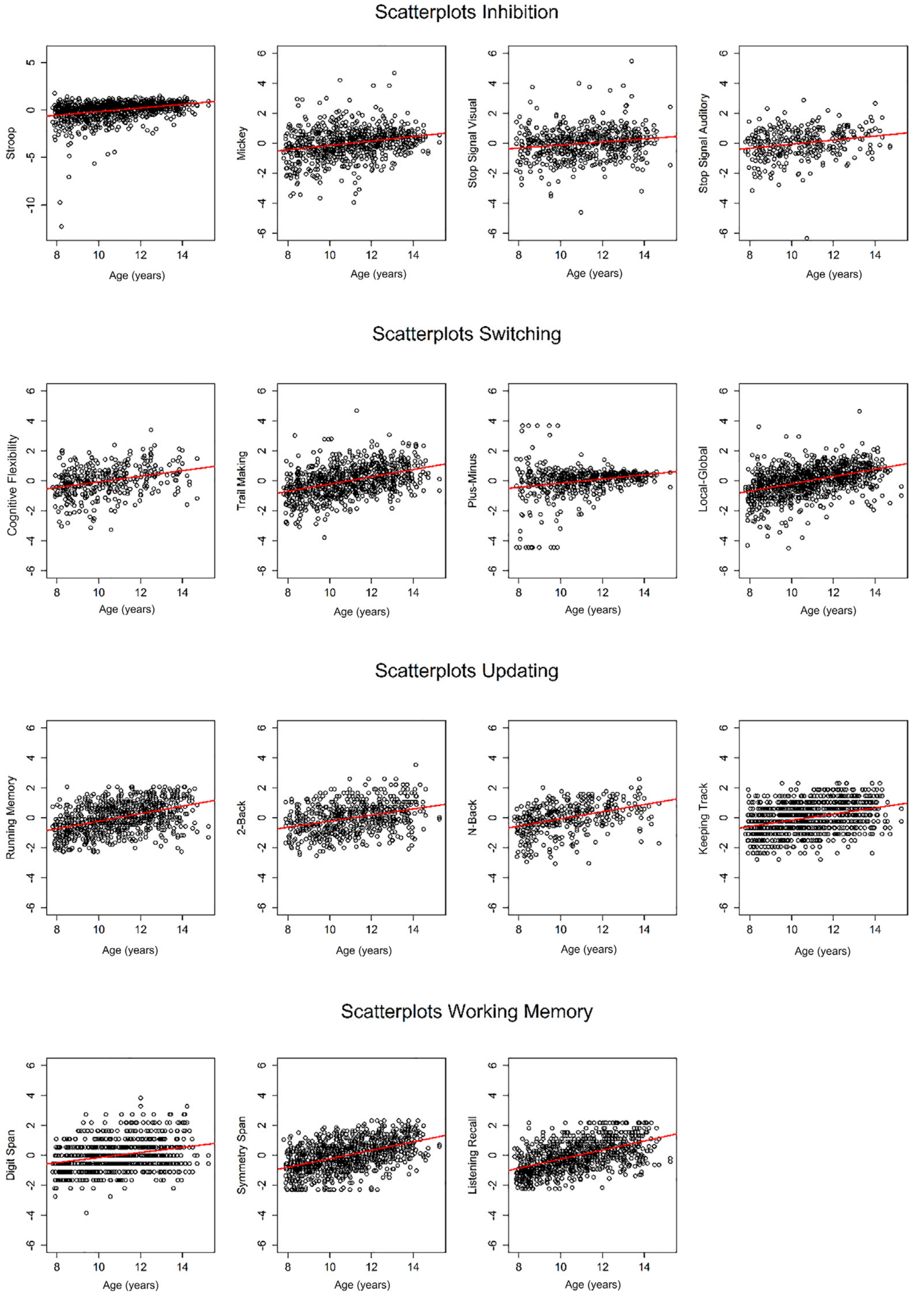

Fig. A1. Scatterplots of the scores in the single tasks across age.

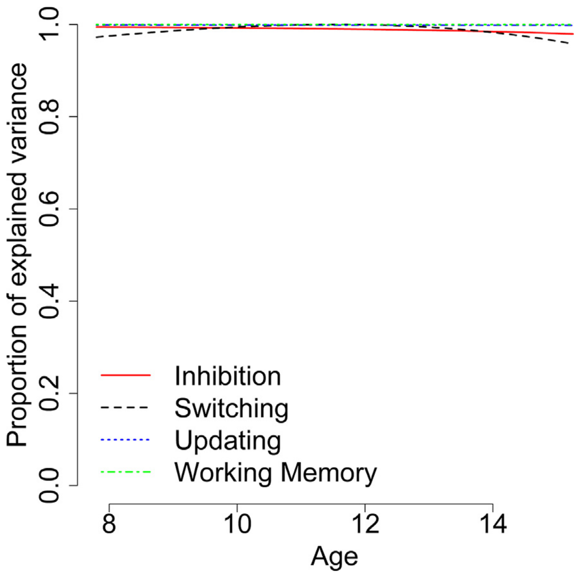

Fig. A2. Proportion of variance in the first-order factors explained by the general executive function factor across age, as implied by moderated factor analysis in the subsample with only one sibling per family.

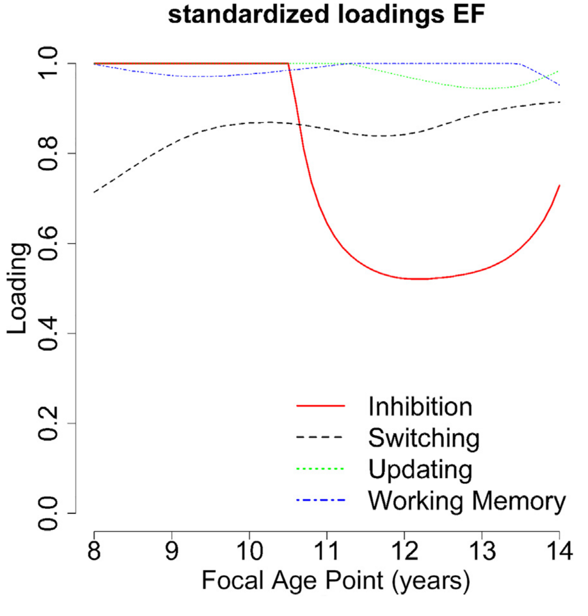

Fig. A3. Standardized second-order factor loadings across age from local structural equation modeling in the subsample with only one sibling per family. EF, executive function.

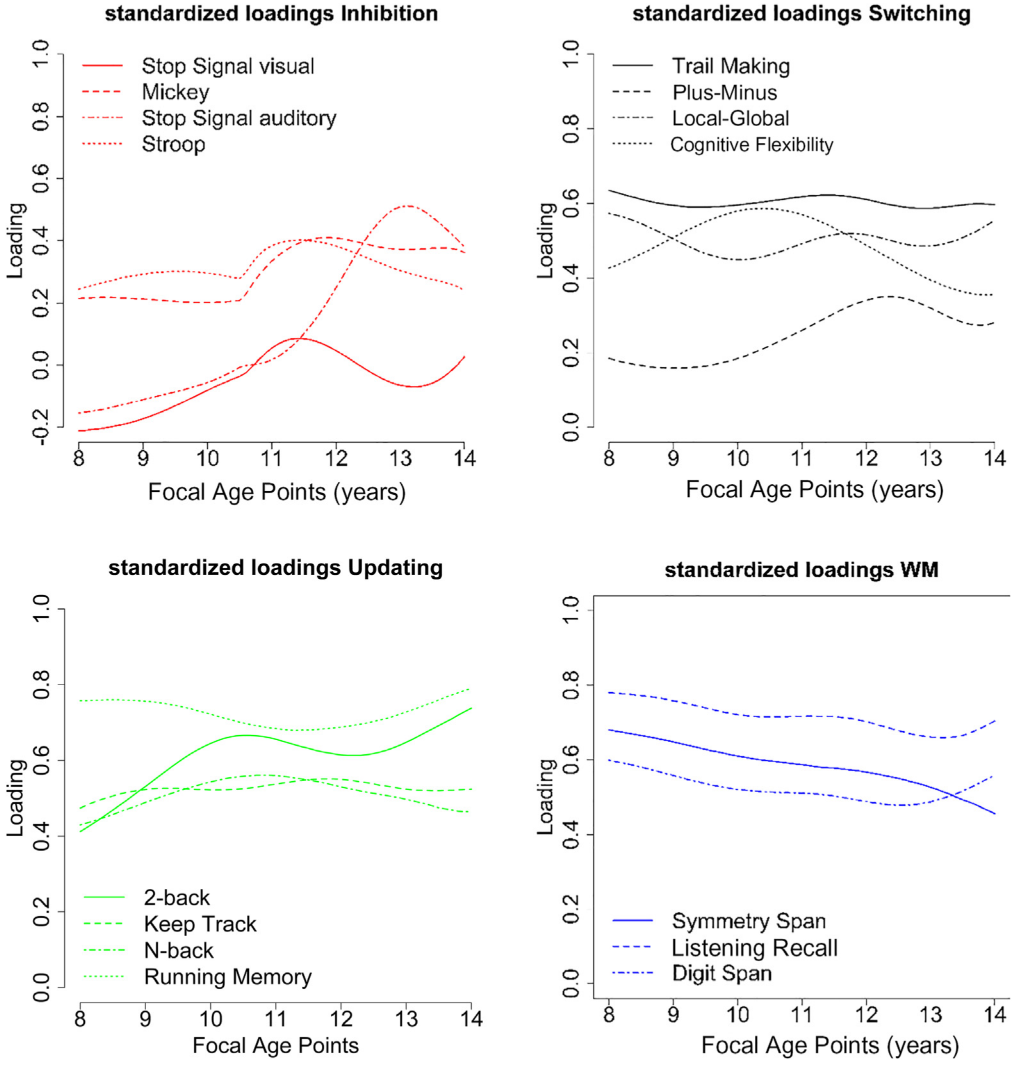

Fig. A4. Standardized first-order loadings on first-order factor loadings from local structural equation modeling in the subsample.

Fig. A5 plots the centrality indices by age from the subsample. There were no major differences in the network analysis of the complete sample.

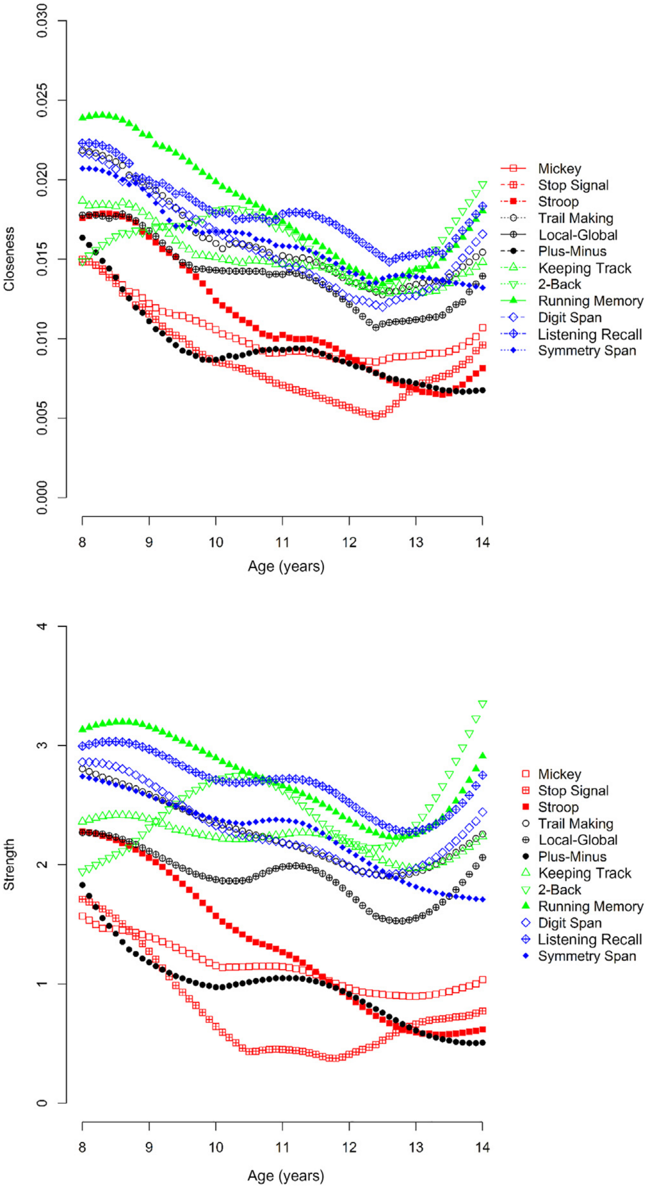

Fig. A5. Closeness and strength of the 12 nodes in the network models across age in the subsample.

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2019.104681.

References

Agostino, A., Johnson, J., & Pascual-Leone, J. (2010). Executive functions underlying multiplicative reasoning: Problem type matters. *Journal of Experimental Child Psychology, 105*, 286–305.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development, 81*, 1641–1660.

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review, 29*, 180–200.

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences, 21*, 327–336.

Bishop, D. V. M., Nation, K., & Patterson, K. (2014). When words fail us: Insights into language processing from developmental and acquired disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*, 20120403.

Briley, D. A., Harden, K. P., & Tucker-Drob, E. M. (2015). Genotype × Cohort interaction on completed fertility and age at first birth. *Journal of Organizational Behavior, 16*, 201–213.

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods, 42*, 563–570.

Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 361–379). New York: Guilford.

Chein, J. M., Moore, A. B., & Conway, A. R. (2011). Domain-general mechanisms of complex working memory span. *NeuroImage, 54*, 550–559.

Cheung, A., Harden, K. P., & Tucker-Drob, E. M. (2015). From specialist to generalist: Developmental transformation in the genetic structure of early child abilities. *Developmental Psychobiology, 57*, 566–583.

Congdon, E., Mumford, J. A., Cohen, J. R., Galvan, A., Aron, A. R., Xue, G., ... Poldrack, R. A. (2010). Engagement of large-scale networks is related to individual differences in inhibitory control. *NeuroImage, 53*, 653–663.

Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we?. *Psychological Bulletin, 74*, 68–80.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*, 450–466.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*, 2037–2078.

Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence, 23*, 105–132.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168.

Ecker, U. K. H., Lewandowsky, S., & Oberauer, K. (2014). Removal of information from working memory: A specific updating process. *Journal of Memory and Language, 74*, 77–90.

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. E. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 170–189.

Engelhardt, L. E., Briley, D. A., Mann, F. D., Harden, K. P., & Tucker-Drob, E. M. (2015). Genes unite executive functions in childhood. *Psychological Science, 26*, 1151–1163.

Engelhardt, L. E., Mann, F. D., Briley, D. A., Church, J. A., Harden, K. P., & Tucker-Drob, E. M. (2016). Strong genetic overlap between executive functions and intelligence. *Journal of Experimental Psychology: General, 145*, 1141–1159.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19–23.

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science, 13*, 190–193.

Epskamp, S., Rhemtulla, M. T., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika, 82*, 904–927.

Filella, J. F. (1960). Educational and sex differences in the organization of abilities in technical and academic students in Colombia, South America. *Genetic Psychology Monographs, 21*, 157–174.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137*, 201–225.

Garrett, H. E. (1946). A developmental theory of intelligence. *American Psychologist, 1*, 372–378.

Hallquist, M. (2016). MplusAutomation: Automating Mplus model estimation and interpretation Retrieved from https://cran.r-project.org/package=MplusAutomation.

Harden, K. P., Tucker-Drob, E. M., & Tackett, J. L. (2013). The Texas twin project. *Twin Research and Human Genetics, 16*, 385–390.

Hartmann, P. (2006). Spearman's law of diminishing returns. *Journal of Individual Differences, 27*, 199–207.

Heaton, R. K. (1993). *Wisconsin card sorting test: Computer version 2*. Odessa, FL: Psychological Assessment Resources.

Hildebrandt, A., Lüdtke, O., Robitzsch, A., Sommer, C., & Wilhelm, O. (2016). Exploring factor model parameters across continuous variables with local structural equation models. *Multivariate Behavioral Research, 51*, 257–258.

Hildebrandt, A., Wilhelm, O., & Robitzsch, A. (2009). Complementary and competing factor analytic approaches for the investigation of measurement invariance. *Review of Psychology, 16*, 87–102.

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

Huizinga, M., & van der Molen, M. W. (2007). Age-group differences in set-switching and set-maintenance on the Wisconsin Card Sorting Task. *Developmental Neuropsychology, 31*, 193–215.

Hülür, G., Wilhelm, O., & Robitzsch, A. (2011). Intelligence differentiation in early childhood. *Journal of Individual Differences, 32*, 170–179.
Isquith, P. K., Gioia, G. A., & Espy, K. A. (2004). Executive function in preschool children: Examination through everyday behavior. *Developmental Neuropsychology, 26*, 403–422.
Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between *n*-back performance and matrix reasoning—Implications for training and transfer. *Intelligence, 38*, 625–635.
Juan-Espinosa, M., García, L. F., Escorial, S., Rebollo, I., Colom, R., & Abad, F. J. (2002). Age dedifferentiation hypothesis: Evidence from the WAIS III. *Intelligence, 30*, 395–408.
Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189–217.
Lee, K., Bull, R., & Ho, R. M. (2013). Developmental changes in executive functioning. *Child Development, 84*, 1933–1953.
Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives, 7*, 199–204.
Logan, G. D., Schachar, R. J., & Tannock, R. (1997). Impulsivity and inhibitory control. *Psychological Science, 8*, 60–64.
MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40.
McAuley, T., & White, D. A. (2011). A latent variables examination of processing speed, response inhibition, and working memory during typical development. *Journal of Experimental Child Psychology, 108*, 453–468.
McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology, 24*, 222–243.
Messer, D., Bernardi, M., Botting, N., Hill, E. L., Nash, G., Leonard, H. C., & Henry, L. A. (2018). An exploration of the factor structure of executive functioning in children. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01179.
Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*, 8–14.
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.
Molenaar, D., Dolan, C. V., Wicherts, J. M., & van der Maas, H. L. (2010). Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence, 38*, 611–624.
Muthén, L. K., & Muthén, B. O. ((1998–2012)). *Mplus user's guide* (7th ed.). Los Angeles: Muthén & Muthén.
Olaru, G., Schroeders, U., Hartung, J., & Wilhelm, O. (2019). Ant colony optimization and local weighted structural equation modeling: A tutorial on novel item and person sampling procedures for personality research. *European Journal of Personality, 33*, 400–419.
Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks, 32*, 245–251.
R Development Core Team (2001). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
Salthouse, T. A. (2011). What cognitive abilities are involved in trail-making performance? *Intelligence, 39*, 222–232.
Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General, 132*, 566–594.
Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1089–1096.
Shilling, V. M., Chetwynd, A., & Rabbitt, P. M. A. (2002). Individual inconsistency across measures of inhibition: An investigation of the construct validity of inhibition in older adults. *Neuropsychologia, 40*, 605–619.
Shing, Y. L., Lindenberger, U., Diamond, A., Li, S.-C., & Davidson, M. C. (2010). Memory maintenance and inhibitory control differentiate from early childhood to adolescence. *Developmental Neuropsychology, 35*, 679–697.
Snyder, H. R., Miyake, A., & Hankin, B. L. (2015). Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and cognitive approaches. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00328.
St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology, 59*, 745–759.
Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology, 45*, 1097–1118.
Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychological bulletin, 133*(6), 1038–1066. https://doi.org/10.1037/0033-2909.133.6.1038.
van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence, 35*, 427–449.
Vandierendonck, A. (2012). Role of working memory in task switching. *Psychologica Belgica, 52*, 229–253.
Verbruggen, F., Logan, G. D., & Stevens, M. A. (2008). STOP-IT: Windows executable software for the stop-signal paradigm. *Behavior Research Methods, 40*, 479–483.
Wechsler, D. (2003). *Wechsler intelligence scale for children–fourth edition, technical*. San Antonio, TX: Psychological Corporation.
Wechsler, D. (2008). *Wechsler adult intelligence scale-fourth edition (WAIS–IV)*. Bloomington, MN: Pearson.
Wechsler, D. (2011). *Wechsler abbreviated scale of intelligence* (2nd ed.). Bloomington, MN: Pearson.
Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology, 44*, 575–587.
Willoughby, M. T., Wirth, R. J., & Blair, C. B. (2012). Executive function in early childhood: Longitudinal measurement invariance and developmental change. *Psychological Assessment, 24*, 418–431.

Wright, I., Waterman, M., Prescott, H., & Murdoch-Eaton, D. (2003). A new Stroop-like measure of inhibitory function development: Typical developmental trends. *Journal of Child Psychology and Psychiatry, 44*, 561–575.
Wu, K. K., Chan, S. K., Leung, P. W., Liu, W.-S., Leung, F. L., & Ng, R. (2011). Components and developmental differences of executive functioning for school-aged children. *Developmental Neuropsychology, 36*, 319–337.
Xu, F., Han, Y., Sabbagh, M. A., Wang, T., Ren, X., & Li, C. (2013). Developmental differences in the structure of executive function in middle childhood and adolescence. *PLoS One, 8*(10) e77770.