Intelligence 90 (2022) 101600

Contents lists available at ScienceDirect

Intelligence

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

The factor structure of executive function in childhood and adolescence

Felien Laureys[a,*,1], Silke De Waelle[a,1], Maria T. Barendse[b], Matthieu Lenoir[a], Frederik J. A. Deconinck[a]

[a] Department of Movement and Sports Sciences, Ghent University, Watersportlaan 2, 9000 Ghent, Belgium
[b] Department of Language and Genetics, Max Planck Institute, Wundtlaan 1, 65256, XD, Nijmegen, the Netherlands

ARTICLE INFO

Keywords:
Executive functions
Cognition
Cross-sectional development
Childhood
Adolescence
Multigroup structural equation modeling

ABSTRACT

Executive functioning (EF) plays a major role in many domains of human behaviour, including self-regulation, academic achievement, and even sports expertise. While a significant proportion of cross-sectional research has focused on the developmental pathways of EF, the existing literature is fractionated due to a wide range of methodologies applied to narrow age ranges, impeding comparison across a broad range of age groups. The current study used a cross-sectional design to investigate the factor structure of EF within late childhood and adolescence. A total of 2166 Flemish children and adolescents completed seven tasks of the Cambridge Brain Sciences test battery. Based on the existing literature, a Confirmatory Factor Analysis was performed, which indicated that a unitary factor model provides the best fit for the youngest age group (7–12 years). For the adolescents (12–18 years), the factor structure consists of four different components, including working memory, shifting, inhibition and planning. With regard to differences between early (12–15 years) and late (15–18 years) adolescents, working memory, inhibition and planning show higher scores for the late adolescents, while there was no difference on shifting. The current study is one of the first to administer the same seven EF tests in a considerably large sample of children and adolescents, and as such contributes to the understanding of the developmental trends in EF. Future studies, especially with longitudinal designs, are encouraged to further increase the knowledge concerning the factor structure of EF, and the development of the different EF components.

1. Introduction

The construct of executive function has sparked the interest of many scientists over the last decades. It has often been defined as a “control mechanism” that modulates the operation of a range of cognitive processes and thereby regulates the dynamics of human cognition and goal directed behavior (Doebel, 2020; Miyake et al., 2000). In this regard, executive functioning plays an important role in many domains such as self-regulation (Dohle, Diel, & Hofmann, 2018; Lang, Stahl, Espie, Treasure, & Tchanturia, 2014), academic achievement (Cragg & Gilmore, 2014), and even sports expertise (Vestberg, Gustafson, Maurex, Ingvar, & Petrovic, 2012). Indeed, it is known that individual variation in EF is an important predictor of developmental outcomes in for example weight status (Liang, Matheson, Kaye, & Boutelle, 2014) or school performance (Mischel, Shoda, & Rodriguez, 1989). This pivotal role of EF in human development has led to much theorizing about the concept, its neural underpinnings, its development, and its relationship with general cognitive ability (or, the g-factor). While traditionally seen as a single construct (Della Sala, Gray, Spinnler, & Trivelli, 1998; Shallice, 1982), EF is now generally considered as multifactorial, encompassing different inter-related and inter-dependent skills that act within an integrated, top-down control system (Alexander & Stuss, 2000; Stuss & Alexander, 2000). The neural basis is within the prefrontal cortex, however, there are multiple connections with many other areas, including the limbic and subcortical regions, supporting its role in various aspects of thought, emotion and action (Stuss & Benson, 1984). According to Zelazo (2015), EF are “attentional skills” that serve to modulate attention in the pursuit of a specific goal, whether simple (e.g. stay focused on what you are reading) or complex (e.g. planning your agenda). In line with the gradually increasing complexity of human behavior, EF improve throughout childhood, adolescence and early adulthood and a pertinent question, in this respect, is to what extent EF

---

[*] Corresponding author at: Ghent University, Watersportlaan 2, 9000 Gent, Belgium.
E-mail addresses: Felien.Laureys@UGent.be (F. Laureys), Silke.DeWaelle@UGent.be (S. De Waelle), Mariska.Barendse@mpi.nl (M.T. Barendse), Matthieu.Lenoir@UGent.be (M. Lenoir), Frederik.Deconinck@UGent.be (F.J.A. Deconinck).
[1] Laureys, F. and De Waelle, S. share first authorship of this manuscript.

https://doi.org/10.1016/j.intell.2021.101600
Received 29 April 2021; Received in revised form 4 November 2021; Accepted 5 November 2021

becomes more specialized and/or fractionated.

In adults, EF is typically categorized and measured according to the “unity and diversity”-theory, by Miyake et al. (2000), which proposes three separate, yet correlated EF components. The first component, shifting, concerns the ability to shift between tasks, operations or mental sets (Monsell, 1996). Updating and monitoring of working memory representations, is proposed as the second EF component in this theory. The third component is inhibition, which involves the ability to deliberately inhibit dominant, automatic, or prepotent responses when necessary (Miyake et al., 2000). This tripartite structure was later replicated on young adults (Friedman et al., 2016) and has been applied frequently and has received substantial support in other studies on adults (Theodoraki, McGeown, Rhodes, & MacPherson, 2019). Despite this widespread support, other factor structures have been suggested. For example, Fournier-Vicente, Larigauderie, and Gaonach (2008) found more fine-grained dissociations within the central EF, with five inter-related factors (a verbal and a spatial coordination function, strategic retrieval, selective attention and shifting). And more recently, Himi, Buhner, and Hilbert (2021) showed that the three-factor model of Miyake et al. (2000) is highly inter-related with other basic cognitive skills (working memory capacity, relational integration, divided attention) and together link to a higher-order factor that would represent general cognitive ability. Note that the above-mentioned factors are also referred to as “cool” EF, i.e. those measured with relatively arbitrary and decontextualized (laboratory) tasks. “Hot” EF, in contrast, are EF that are elicited in situations that evoke emotion, motivation, or tension between instant gratification or long-term reward (Zelazo & Muller, 2002). Deliberate, top-down control of our day-to-day functioning is the result of a combination of cool and hot EF, yet the focus of this article is on cool EF.

Within the different models that have been proposed with regard to EF, planning is generally accepted as part of EF (Anderson, 2002; Miyake & Friedman, 2012; Rommelse et al., 2007). Planning can be defined as the ability to map out a sequence of actions in preparation for a certain task (Morris, Miotto, Feigenbaum, Bullock, & Polkey, 1997), and is usually measured by ‘Tower Tasks’ such as the TOL (Tower of London) (Levin et al., 1996). Numerous studies have included measures of planning abilities when investigating EF (Carlson, Moses, & Claxton, 2004; Hung, Tsai, Chen, Wang, & Chang, 2013; Kenny, Cribb, & Pellicano, 2019). However, the specific relation of planning towards the EF construct seems unclear, as it has seldom been included in latent variable analyses of the EF factor structure. Planning has most frequently been investigated in isolation as a stand-alone EF component in studies that also investigate other EF components such as working memory or inhibition (Albert & Steinberg, 2011; Carlson et al., 2004; Hackman, Gallop, Evans, & Farah, 2015; Hung et al., 2013; Nemati et al., 2017). In a small number of cases, planning has been considered in latent variable analysis, where it has mostly been categorized as part of working memory (Cirino et al., 2018) or inhibition (Lehto, Juujarvi, Kooistra, & Pulkkinen, 2003; Welsh, Satterlee-Cartmell, & Stine, 1999). Both of these approaches seem to be incomplete, as investigating planning as a stand-alone component (i.e., not including it in a model with measures for the other EF-components) ignores the relation between planning and the other EF components, while seeing it as part of working memory or inhibition seems to ignore the complexity and unique characteristics of planning abilities. Therefore, further research is needed to clarify whether planning can be defined as part of inhibition or working memory, or is better classified as a separate component of EF.

With regard to the development of EF, there is converging evidence that the factor structure of the EF construct differs between childhood and adolescence (Karr et al., 2018; Lee, Bull, & Ho, 2013; Xu et al., 2013). In early childhood, a unitary factor model is generally observed between 3 and 5 years old (Brydges, Fox, Reid, & Anderson, 2014). This unitary factor structure has been associated with the structural as well as functional immaturity observed in the prefrontal cortex (Diamond, 2002). Later, during primary school years and along with maturation of the prefrontal cortex, working memory and inhibition seem to emerge as two distinguishable factors, while the three-factor model is often observed during adolescence (Davidson, Amso, Anderson, & Diamond, 2006; Miyake et al., 2000). According to Zelazo (2015), a major driver of these developments is the increased ability to reflect, or iteratively reprocess information, which allows the formulation of more complex rules, and hence better or more efficient control of attention. However, confusion exists as to whether these developments are associated with a change in the observed or measurable fractionation of EF. For example, Prencipe et al. (2011) identified a one-factor model in children (8–11 years old) and adolescents (12 to 15 years old), while others have reported a two-factor model in children, adolescents and even young adults (Huizinga, Dolan and van der Molen, 2006) or have found support for the three factor structure in participants as young as six years old (Brocki & Bohlin, 2004; McAuley & White, 2011).

It should be noted, though, that what we currently know about the developmental changes of EF (and the factor structure thereof) stems predominantly from separate studies, i.e. studies that focus on either childhood or adolescence. The contrasting findings from these separate studies might be the result of great variation in methodology, as well as insensitivities in the tasks or scoring methods in some of these studies (van der Ven, Kroesbergen, Boom, & Leseman, 2013). For example, older children are presented with other, more difficult tasks than younger children to avoid ceiling and floor effects, which complicates the comparison of these two groups (Best, Miller, & Naglieri, 2011). In addition, the majority of the tests used to measure the different EF components address multiple components simultaneously, or they might reflect lower-order underlying processes instead of pure EF performance (Miyake et al., 2000). This issue is also referred to as task impurity and can result in ambiguity about whether changes on test performance are actual changes in EF or rather changes in nonexecutive processing requirements (e.g., language or visuospatial processing; (Miyake et al., 2000)). Related to this, the same task is often classified under a different cognitive component, depending on the study. The Wisconsin Card Sorting Task (WCST), for example, has been classified as an inhibition or shifting task (Garon, Bryson, & Smith, 2008), while it can be argued that a task of such complexity probably taps into multiple EF components simultaneously (Best et al., 2011). Consequently, the current understanding of the EF factor structure in children and (young) adolescents is both fragmented and complicated by methodological issues. Studies using a multidimensional cognitive test battery over a wide age range are therefore necessary to allow for more insights in EF and its factor structure across different ages.

In summary, despite the consensus regarding the dissociation of EF into different inter-related factors, previous studies have not considered planning. Also, there is a dearth of studies that has examined the factor structure across childhood and adolescence using one test battery, which challenges the field to gain a clear developmental perspective on EF. It is apparent that both developmental research and practice would benefit from acquiring a deeper understanding of the factor structure of EF across different developmental periods. To this end, the aim of this study is to investigate the factor structure of EF across different age groups in a large and representative sample of children and adolescents from 7 to 17.99 years of age. A four-factor baseline model, in which the three main EF components (shifting, working memory and inhibition) as well as planning are integrated, is proposed and tested across age groups using confirmatory factor analyses. Furthermore, while investigation of sex differences was not a goal in itself, the influence of sex on the factor structure was explored in a post-hoc analysis, given the current debate in literature (e.g. see Grissom & Reyes, 2019). Likewise, because processing speed has been suggested to explain a significant proportion of EF in childhood (Lee et al., 2013; McAuley & White, 2011) it was deemed necessary to control for the role of processing speed on the factor structure in the analyses.

2. Materials and method

2.1. Participants

Participants in this large-sampled study were Flemish children and adolescents aged between 7 and 17.95 years old, recruited through various channels in order to provide a representative sample of the Flemish youth. The participants were recruited in six elementary schools, seven secondary schools, and over 20 sports clubs from four sports federations located in Flanders, Belgium, between May 2018 and October 2019. The six elementary schools had various types of education (community schools, schools using an alternative pedagogical approach and catholic schools) and the seven secondary schools included a mix of students in general, technical and vocational education. Thereby, a convenient sample of Flemish children and adolescents was gathered, which resulted in a total sample of 2117 participants (1171 boys; 56.8% recruited via schools). Because participants were recruited from both schools and sports clubs, it was deemed important to investigate whether the overall sports participation rate of the current sample was representative for the Flemish population. Of the total sample, 74.5% of the participants practiced in organised sport for at least 1 h per week. Taking into account that there was no data on sports participation available for 13% of the participants, these numbers align with studies indicating that 82% and 85% of 15 to 30 years old Flemish individuals are or have been participating in organised sports at some point (Fransen et al., 2012; Scheerder, Borgers, & Willem, 2015). Consequently, together with the size of the sample and the variety of sampling channels, these numbers speak for the representativeness of this sample. Also, sufficient power is guaranteed since the group sizes are >150 as recommended for complex multi-group factor analysis (Kline, 2015).

Participants were divided into three age groups: childhood, early adolescents and late adolescents, in accordance with Theodoraki et al. (2019) and McAuley and White (2011). Table 1 gives an overview of the mean age and number of participants in each age group.

This project has been conducted in accordance with the Helsinki Declaration and was approved by the local ethics committee of Ghent University. Since all participants were minors, parents or their legal representative gave their informed consent to let their child participate in this study. All data were analyzed confidentially.

2.1. Test battery

The multidimensional cognitive test battery used in this study was the web-based Cambridge Brain Sciences (CBS) cognitive test battery. The cognitive test battery can be personalized to the needs of the study and can contain up to thirteen EF tests, including a wide range of outcome variables. All tests start with the same (low) level of difficulty for each participant, and the complexity increases or decreases depending on the accuracy of response. This implies that the test battery adjusts to the performance level, allowing each participant, independent of age, to execute the tests to their best cognitive abilities. Furthermore, the test battery has an overall adequate test-retest reliability (Hampshire, Highfield, Parkin and Owen, 2012) (for details, see supplementary material S1). Altogether, this makes the CBS a valuable test battery to use in wide age ranges, since it minimizes floor and ceiling effect.

Table 1
Overview of number of participants with the mean, minimum, maximum and numbers (N) per group.

Age Group	Mean Age in years (SD)	Min Age	Max Age	N	(Boys/ Girls)
Childhood	9.92 (1.36)	7.00	11.99	818	(382/436)
Early Adolescents	13.59 (0.90)	12.00	14.99	763	(451/312)
Late Adolescents	16.30 (0.86)	15.00	17.95	536	(338/198)

Seven CBS-tasks that included minimal reading or mathematic abilities were chosen for this study in order to minimize the influence of academic skills on task performance. Additionally, to ensure maximum concentration and motivation during the assessment, it was decided to choose no more than these seven tasks and keep the assessment as short as possible. The selected tasks were Spatial Span, Double Trouble, Token Search, Odd One Out, Spatial Planning, Monkey Ladder and Sustained Attention to Response Task (SART). The tests were always assessed in the same order that is stated above. Screenshots of each test can be found in the online supplementary material (Fig. S2-S8). Through these seven tasks, this computerised assessment covers common EF components, including inhibition, shifting, visuospatial working memory, planning and processing speed.

Spatial Span (SS) is a task based on the Corsi Block Tapping Task (Corsi, 1972) and measures a persons' ability to remember the relations between objects in space. This test consisted of a grid of 4 × 4 boxes, that lit up in a random order on the screen. Participants were instructed to tap the boxes in the same sequence of appearance on the screen. The first trial always had a span length of four blocks. When a trial was executed correctly (correct locations in the correct order) the next trial contained one extra box. An incorrect trial was followed by a trial containing one box less. The test ended after three incorrect responses. Response accuracy (SS RA) was used as performance indicator for the spatial span task, and was calculated as the maximum number of blocks remembered correctly (i.e., span length) for each participant.

Double Trouble (DT) is an adaption of the Stroop test and mainly assesses inhibitory control (Stroop, 1992). Three words are presented to the participant as shown on Fig. S2. Participants were asked to indicate which of two coloured words at the bottom described the colour of the word at the top. The test lasted 90 s in which participants had to give as many correct responses as possible. For this test, three performance indicators were selected. First, total response accuracy (DT RA) was calculated as percentage of correct trials for each participant. Second, mean response time (i.e., the time between the words appearing on screen and the participants tapping on a word) on double incongruent trials (DT RT II) was calculated for each participant. Double incongruent trials were trials where the top word and target word were different and had a different colour. Third, mean response time on double congruent trials (DT RT CC) was calculated for each participant. Double congruent trials were trials where both top word and target word were the same and had the same colour.

Token Search (TS) is a self-guided search task that mainly assesses spatial working memory (Collins, Roberts, Dias, Everitt, & Robbins, 1998). Participants were presented with a number of boxes randomly placed on the screen and were asked to find a token that was hidden underneath the boxes. Each box contained the token only once and the next hiding place was unpredictable. The task requires to hold the selected boxes in memory. Selection of an empty box twice or a box that had previously held the token, resulted in a failure. When a trial was executed correctly (all tokens found without error) the next trial contained one extra box. After an incorrect trial the next trial contained one box less. The test ended after three incorrect responses. Response accuracy (TS RA) was selected as performance indicator for the token search task and was calculated as the maximum number of boxes found without error for each participant.

Odd One Out (OO) is a modern adaptation of classical tests of fluid intelligence (Brenkel, Shulman, Hazan, Herrmann, & Owen, 2017), and mainly assesses deductive reasoning and shifting. This task consists of nine sets of shapes that differ from each other in colour, shape and size. The participant had to point out which shape was the most different from the others. A correct response resulted in the next trial being more complex, while an incorrect trial resulted in the next trial being less complex. The grade of complexity depended on the amount of variance on the three levels (colour, shape, size) within the nine figures. The test lasted 180 s in which participants had to give as many correct responses as possible. Response accuracy as well as response time were selected as performance indicators for this task. Response accuracy for the odd one out task (OO RA) was calculated as the number of correct attempts for each participant (*N attempts – N errors*). For response time (i.e., time between the trial appearing on screen and the participants tapping on a shape), the mean response time per trial was calculated for each participant (OO RT).

Spatial Planning (SP) is an adapted version of the Tower of London Task (Shallice, 1982), which is primarily used to assess planning ability. Participants were asked to sort balls that are positioned on a tree-shaped frame in numerical order in as few moves as possible, by replacing one ball per move (Fig. S5). The problems became progressively more complex to solve as the participant progressed through the task. The test lasted 180 s in which participants had to solve as many problems as possible. Response accuracy was used as a performance indicator for this task and was calculated in two steps. First, trial scores were calculated per trial using the following formula: (*minimum moves required * 2) – moves made*. The total response accuracy (SP RA) was then calculated as the sum of all trial scores for each participant.

Monkey Ladder (ML) is based on a task from the non-human primate literature (Inoue & Matsuzawa, 2007) and mainly assesses visuospatial working memory, or the ability to hold information in memory and to manipulate or update it depending on the purpose or the circumstances. Participants were presented with a number of boxes randomly placed on the screen, with each box containing a number ranging from 1 to the number of boxes. Participants were asked to memorize the numbers appearing in each box and to tap the boxes in numerical order as soon as the numbers disappeared. When a trial was executed correctly, the next trial contained one extra box. After an incorrect trial the next trial contained one box less. The test ended after three incorrect responses. Response accuracy (ML RA) was selected as performance indicator for the monkey ladder task and was calculated as the maximum number of boxes remembered correctly for each participant.

Sustained Attention to Response Task (SART, (Robertson, Ridgeway, Greenfield, & Parr, 1997)) mainly assesses inhibition. Participants were presented with single digits in the centre of the screen, each digit appeared for 250 ms. Participants were asked to respond with a tap on the “GO” button on the screen to each digit (GO) as quickly as possible. However, when the digit “3” appeared on screen (NO GO), participants were asked to withhold a response. Participants had to maintain their attention to this task for four minutes. The response accuracy score (SART RA NG) was calculated as the percentage of correct NO GO trials for each participant.

2.2. Procedure

The cognitive assessments lasted approximately 20–25 min per child, and up to 16 participants could be tested simultaneously. The test battery was administered on a 9.7 in. Apple iPad 2017 (iOS 12.1, Apple Inc., Cupertino, USA) that was held in an upright position. The testing took place in a quiet area, with no influence of parents, teachers/coaches or significant others and under the supervision of qualified and trained researchers. Researchers explained the overall purpose of the game-like cognitive assessment (to get a score as high as possible on each test) and some general rules (e.g., each test works with either a time span or three lives, where you try to get as far as possible). Before each test, researchers explained the test with pictograms, and made sure the participants understood the intention of the test. During the test, researchers supervised and checked if the participants were executing the test correctly. Participants could also see their score, their lives or the time left for each test, motivating them to gain a high score.

2.2. Data analysis

2.2.1. Data processing

For each test, data were checked for impossible values or apparatus malfunctions, which would lead to the deletion of test scores of the given test only. This resulted in the removal of the Double Trouble measurements for two participants, because of unlikely high RTs (24 s and 28 s). Measurements of the Odd One Out test were also removed for two participants, one because of an extremely high RT of 20s, and one because of an unlikely number of attempts and errors (89 attempts/78 errors). For the SP and SAR test, data of respectively 13 and 42 participants was lost due to apparatus malfunction (caused by unstable internet connections). Apparatus malfunction occurred at random and was unrelated to any personal, demographic or other variables in the current dataset. Therefore, data can thus be assumed with reasonable certainty to be missing completely at random (MCAR) (Newsom, 2015). The total number of missing cases adds up to 55 out of 2117 participants, which is 2.6% of the total sample. Means, standard deviations and distributional raw scores per age group per test were also checked (Supplementary Table S9).

2.2.2. Structural equation modeling

2.2.2.1. Modeling steps. Based upon literature and the CBS tests, a four-factor baseline model was proposed, where not only the three main EF components (shifting, working memory and inhibition) were integrated, but planning was also added as a fourth EF component. To address part of the task impurity problem, selected outcome measures from different CBS tests were used as indicators for one EF component. Confirmatory factor analysis (CFA) within the framework of structural equation modelling (SEM) was used to investigate our baseline model. To test this model for measurement invariance across age, a first multigroup CFA model was fit. Next, although a recent meta-analysis indicates that there is little evidence for sex differences in EF during childhood and adolescence, results have been mixed in literature (Grissom & Reyes, 2019). It therefore seems relevant to control for these possible sex differences in our sample, and consequently, a second multigroup CFA model was fitted to explore sex differences on the different EF factors. Lastly, the influence of processing speed on the previously fitted models was also investigated as an additional post-hoc analysis. Although processing speed might play a significant role towards EF, this factor has only seldom been considered in previous literature (for exceptions see Lee et al. (2013), McAuley and White (2011) and Huizinga, Dolan and van der Molen (2006). Brocki and Bohlin (2004) and Fry and Hale (2000) have indicated significant age-related differences in processing speed between the ages of six and twelve years old, coinciding with important developmental changes within EF (Diamond, 2002). Therefore, it was decided to run the two CFA models again with the addition of processing speed to control for the potential influence of processing speed on the factor structure of EF in the different age groups. This leads to the following four models that have been included for analysis, of which the code can be found in the supplementary material (S10):

1. Single group CFA to test the proposed four-factor structure with latent variables for (visual-spatial) working memory, inhibition, shifting and planning (see Fig. 1), and to compare this model to a unitary, two-, and three-factor model (see online supplementary material, Fig. S11).
2. Multigroup CFA to test for measurement invariance across age for the four-factor model.
3. Multigroup CFA to test for measurement invariance across sex for the four-factor model.
4. Addition of processing speed to the previous models. Processing speed was a latent variable that consisted of the response time on the Double Trouble incongruent and congruent trials and the response time on the Odd One Out task (see online supplementary material, Fig. S12).

2.2.2.2. Fit indices. The full information maximum likelihood (ML)

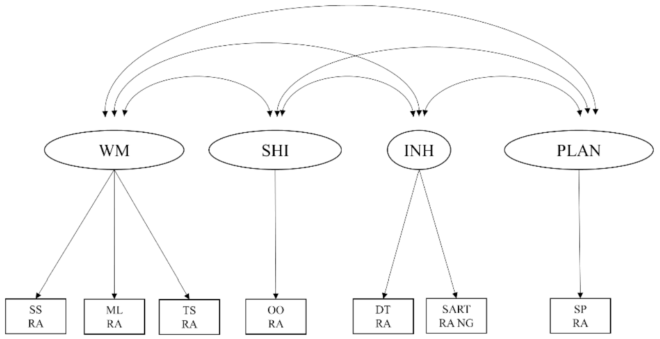

Fig. 1. Representation of the hypothesized four-factor model. The rounded boxes represent the latent EF factors; the square boxes represent the observed variables or test items loading onto these factors. Residuals are not depicted. WM = Working Memory; SHI = Shifting; INH = Inhibition; PLAN = Planning; SS RA = Spatial Span Response Accuracy; ML RA = Monkey Ladder Response Accuracy; TS RA = Token Search Response Accuracy; OO RA = Odd One Out Response Accuracy; DT RA = Double Trouble Response Accuracy; SART RA NG = Sustained Attention To Response Task Response Accuracy No Go; SP RA = Spatial Planning Response Accuracy.

method was applied for handling missing data (Arbuckle, Marcoulides, & Schumacker, 1996; Enders, 2001). To evaluate goodness of fit of these models, the following criteria were used: Model Chi Square (χ^2) (Tabachnick, Fidell, & Ullman, 2007), the Comparable Fit index (CFI) > 0.9 (Hooper, Coughlan, & Mullen, 2008), Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) <0.05 (Hu & Bentler, 1999) and finally the Aikake Information Criterion (AIC; (Kline, 2015)), which was mainly used for model comparison purposes.

2.2.2.3. Measurement invariance tests. Measurement invariance (see multigroup CFA across age and sex) testing consisted of a series of model comparisons, where equality constraints were added to the models consecutively in each step (Meredith, 1993; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000). Achieving measurement invariance is crucial, because if measurement invariance is not achieved, observed differences in test scores do not validly represent actual differences in what the test is supposed to measure. Means of the latent variables can thus only be compared across groups if measurement invariance across groups has been established. The first step is to establish configural invariance, for which the model without constraints was fit across sex and age. Next, to test weak invariance, factor loadings were constrained to be equal across groups, and lastly, to test strong invariance, both factor loadings and intercepts were constrained to be equal across groups. To assess whether a more constrained model (i.e., configural, weak and strong invariance, respectively) would lead to a substantial decrease in model fit, overall model fit was compared and the chi-square difference test was used. If full measurement invariance was established, latent means were compared by setting the means of the comparison group to zero, and freely estimating the means of the other group. Analyses were conducted in R using the lavaan package (lavaan version 0.5–12; (Rosseel, 2012).

3. Results

3.1. Model choice for the factor structure of EF

In the first step of the analyses, the four-factor model (see Fig. 1) as well as a unitary factor, two-factor (working memory + inhibition) and three-factor (working memory + inhibition + shifting) model were applied to the data. All four models consisted of the same items, with the difference being the number of latent factors these items would load on (see online supplementary material, Fig. S11). The unitary factor, two-factor and three-factor model were then compared to the four-factor model to test the hypothesis that a four-factor solution would be the best fit for our data. Results are shown in Table 2A. The unitary factor model as well as the two-factor model, demonstrated a clear inferior model fit compared to the four-factor model (i.e., lower CFI and higher χ^2, SRMR and AIC values). The fit measures of the three-factor and four-factor model were similar, however, taking into account theoretical and practical considerations, it was decided to retain the four-factor model with working memory, inhibition, shifting and planning as our final solution.

3.2. Measurement invariance across age with multigroup CFA

In the second step, to investigate whether the four-factor model holds for all groups, we performed a multigroup analysis across the three different age groups. During these analyses it became apparent that the four-factor model did not provide an adequate fit for the youngest age group, as the model could not converge (i.e., the childhood group). Upon this finding it was decided to split out the analysis and continue with the childhood group and the adolescence group (consisting of early and late adolescents) separately. Table 2B and C shows the results for the childhood and the adolescence group separately. For the childhood group, convergence issues were also apparent for both the two- and three factor model, while the unitary factor model did converge and provided an excellent fit, indicating a unitary construct for EF at that age (Table 2C). In agreement with the full sample, the four-factor model provided the best fit for the adolescence group.

To continue with the second step for the adolescent group only, a multigroup CFA model was tested for measurement invariance across two age groups: early and late adolescents. As shown in Table 3, configural, weak and strong invariance held for the two age groups. After measurement invariance was established in the adolescent group, latent

Table 2
Fit Indices for the different Confirmatory Factor Analysis models for the full sample as well as for the childhood and adolescent groups separately.

N Factors	Factors	df	χ^2	CFI	RMSEA	SRMR	AIC
A. Full Sample							
1	Unitary model	14	52.95	0.975	0.036	0.021	34,044
2	Working Memory - Inhibition	13	52.82	0.975	0.038	0.021	34,046
3	Working Memory – Inhibition-Shifting	12	41.86	0.981	0.034	0.018	34,037
4	Working Memory – Inhibition-Shifting - Planning	10	38.33	0.982	0.037	0.017	34,038
B. Adolescence							
1	Unitary model	14	46.83	0.935	0.042	0.028	21,093
2	Working Memory - Inhibition	13	46.60	0.933	0.045	0.027	21,095
3	Working Memory – Inhibition-Shifting	12	37.00	0.950	0.040	0.024	21,087
4	Working Memory – Inhibition-Shifting - Planning	10	33.75	0.953	0.043	0.023	21,088
C. Childhood							
1	Unitary model	14	17.05	0.994	0.016	0.019	12,229

EF = Executive Function.

Table 3
Measurement invariance tests for age in the adolescent group (early vs late adolescents).

	df	χ^2	CFI	RMSEA	SRMR	AIC	$\Delta\chi^2$
Measurement model	10	33.75	0.953	0.043	0.023		
Invariance (Age)							
Configural	10	47.93	0.936	0.046	0.028	20,934	
Weak (loadings)	23	50.12	0.938	0.043	0.029	20,930	2.19
Strong (loadings, intercepts)	26	58	0.928	0.043	0.32	20,932	7.67
Means (loadings, intercepts, means)	30	172.4	0.676	0.085	0.074	21,039	114.6**

**$p < 0.001$.

means could be compared between groups by setting the means of the reference group to zero and freely estimating the means of the other group. Results demonstrated that late adolescents (LA) have significantly higher scores on inhibition [$\mu_{LA} = 0.068$ (0.008), $p < 0.001$], planning [$\mu_{LA} = 0.264$ (0.050), $p < 0.001$] and working memory [$\mu_{LA} = 0.361$ (0.052), $p < 0.001$] compared to early adolescents as the reference group, but there was no difference between groups with regard to shifting performance.

To give an indication of the age-related differences, z-scores based on the factor scores were plotted for each age in both the childhood and adolescence groups (see Fig. 3 and 4). The factor scores in the childhood group were based on the unitary factor model shown in Fig. 2A. The factor scores for the adolescence group were calculated based on the four-factor model shown in Fig. 2B; for the early and late adolescents together. Factor scores for each participant were calculated. As these factor scores have different scales, they were converted to z-scores to facilitate the comparison of the trends across the different factors. The z-score then gives an indication of the deviation of the individual factor score from the group mean (childhood or adolescence). These z-scores only serve visual presentation purposes, and were not used for any formal analyses.

3.3. Measurement invariance across sex with multigroup CFA

In the third step, measurement invariance across sex (boys and girls) was tested for the childhood and adolescence group separately using two multigroup CFA models. Table 4A shows the measurement invariance results for the unitary factor model in the childhood group. Both configural and weak invariance hold. Strong invariance is not established due to a small difference in intercept for only one of the seven items (SART Response Accuracy No Go). As the overall fit of the model is still acceptable, we continue with a models that contains measurement invariance restrictions to test equality of the means across sex. For the childhood group no differences between boys and girls on the latent factor of EF were found.

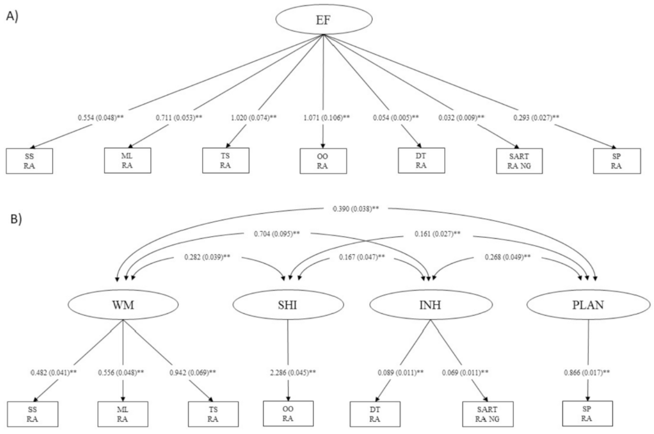

Fig. 2. A) Factor structure for the childhood group; B) Factor structure for the adolescent group. Estimates (standard errors) are displayed (**$p < 0.001$, * $p < 0.05$), error variances and residuals are not displayed. EF = Executive Function; WM = Working Memory; SHI = Shifting; INH = Inhibition; PLAN = Planning; SS RA = Spatial Span Response Accuracy; ML RA = Monkey Ladder Response Accuracy; TS RA = Token Search Response Accuracy; OO RA = Odd One Out Response Accuracy; DT RA = Double Trouble Response Accuracy; SART RA NG = Sustained Attention To Response Task Response Accuracy No Go; SP RA = Spatial Planning Response Accuracy.

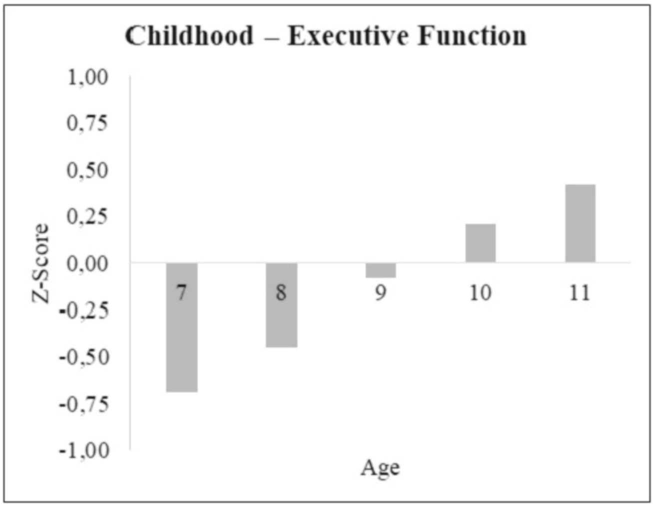

Fig. 3. Representation of the age-related differences in EF during childhood. For each individual a factor score was calculated based on the unitary factor model shown in Fig. 2 A) and then converted into a z-score relative to the group (i.e. childhood) mean. Per age-group the average of the individual z-scores in that age-group is shown [N(7 years) = 106; N(8 years) = 107; N(9 years) = 173; N(10 years) =216; N(11 years) = 216).

Next, measurement invariance across boys and girls was tested for the four-factor model in the adolescence group, Table 4B shows that configural, weak and strong invariance hold for boys and girls. After measurement invariance was established, latent means could be compared between groups. Girls showed significantly higher scores on inhibition [$\mu_{girls} = 0.039$ (0.008, $p < 0.001$], shifting [$\mu_{girls} = 0.896$ (0.126), $p < 0.001$] and working memory [$\mu_{girls} = 0.114$ (0.045), $p = 0.012$] compared to boys, but there was no difference in planning performance between boys and girls.

In addition to step 2 and 3, to further investigate measurement invariance in the adolescence group, factor covariances were also constrained to examine whether these factor covariances differed across groups. Results demonstrated that the covariances between the different latent variables did not significantly differ across different age or sex groups.

3.4. Exploring the role of processing speed

In the fourth step of the analyses, processing speed was added to the model (see online supplementary material, Fig. S12) and a new single group CFA was run for the early childhood group (unitary factor model)

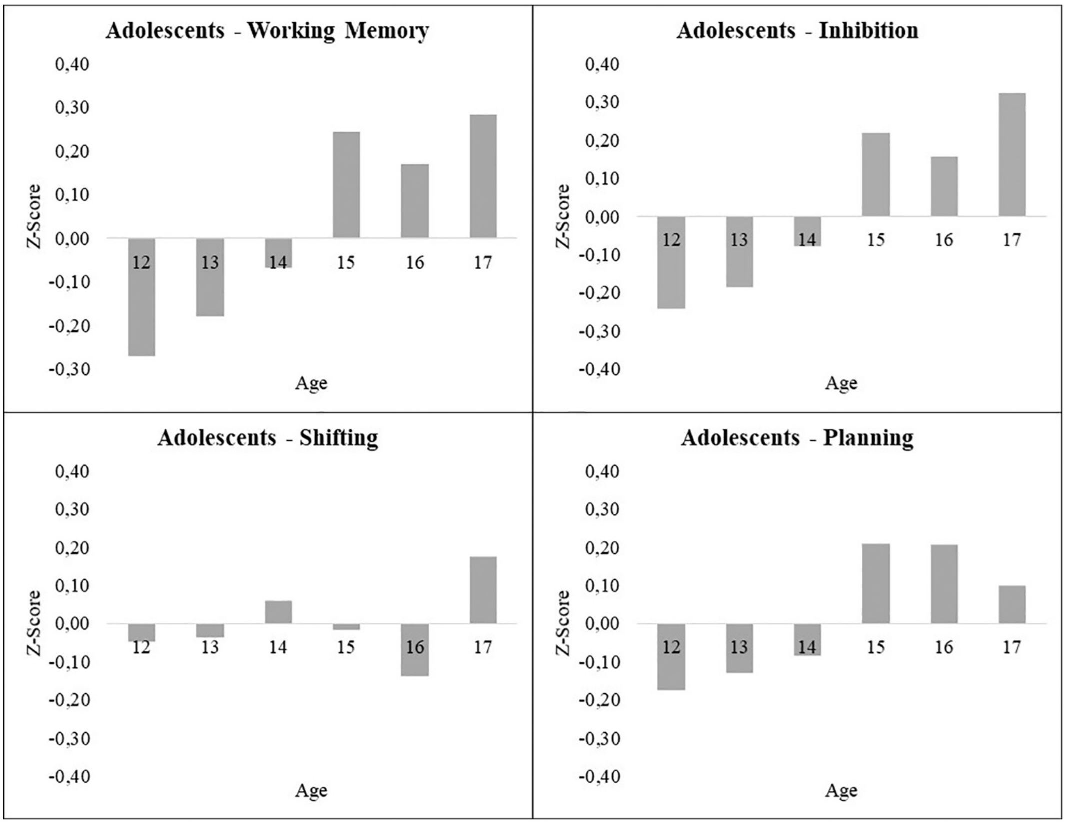

Fig. 4. Representation of the age-related differences in the four factors of EF during adolescence. For each individual the factor score was calculated based on the four-factor model shown in Fig. 2 B) and then converted into a z-score relative to the group (i.e. adolescence) mean. Per age-group the average of the individual z-scores in that age-group is shown [N(12 years) = 241; N(13 years) = 220; N(14 years) = 302; N(15 years) = 239; N(16 years) = 152; N(17 years) = 145).

Table 4
Measurement invariance analyses across boys and girls in the childhood and adolescence age groups.

	df	χ^2	CFI	RMSEA	SRMR	AIC	$\Delta\chi^2$
A. Childhood							
Measurement model	14	17.0	0.994	0.016	0.019	23,392	
Invariance							
Configural	28	28.8	0.998	0.008	0.025	23,357	
Weak (loadings)	34	33.3	1.000	0.000	0.030	23,350	4.59
Strong (loadings, intercepts)	40	61.9	0.953	0.037	0.043	23,367	28.57**
Means (loadings, intercepts, means)	41	72.4	0.932	0.043	0.053	23,375	10.49**
B. Adolescence							
Measurement model	10	33.75	0.953	0.043	0.023		
Invariance (Sex)							
Configural	20	54.72	0.931	0.052	0.030	21,026	
Weak (loadings)	23	63.6	0.919	0.052	0.034	21,029	8.87*
Strong (loadings, intercepts)	26	64.7	0.923	0.048	0.034	21,024	1.11
Means (loadings, intercepts, means)	30	131.9	0.797	0.072	0.055	21,083	67.21**

**$p < 0.001$, *$p < 0.05$.

and the adolescence group (four-factor model) separately to assess the fit of these extended models. As two items from the processing speed latent variable were performance indicators taken from the same test as items from the inhibition latent variables, we added additional residual correlations. A unitary factor model with processing speed as a latent variable in the childhood group did not fit the data. In contrast, for our sample of the adolescence group, the new extended model, i.e. the four-factor model with processing speed, did provide a good fit (CFI = 0.983, df = 25, $\chi^2 = 55.7$, RMSEA = 0.032, SRMR = 0.026).

For the adolescence group, the extended four-factor model with processing speed was tested for measurement invariance across age (early and late adolescence) and sex (boys and girls; see online supplementary material, Table S13). Similar to the results of the baseline four-factor model, measurement invariance was established across sex and age. This again allowed for comparison of the latent means, where the differences between boys and girls and between early and late adolescence on working memory, inhibition, shifting and planning remained similar for the extended model compared to the baseline model. Concerning the new latent variable, results demonstrated no difference in processing speed between boys and girls, while late adolescents showed slightly inferior processing speed than early adolescents. These results indicate that adding processing speed as a fifth factor to the model did not considerably affect the coefficients of the baseline four-factor model.

4. Discussion

The current study aimed to examine the factor structure of EF in childhood and adolescence and explore differences between age groups. Specifically, it was investigated whether the classic tripartite structure, proposed by Miyake et al. (2000), could be expanded with an additional component, i.e. planning. By using the same seven tasks in a large sample across a broad age range, we demonstrate that EF is best described as a unitary EF construct in children between seven and twelve years old, while four separable EF components can be distinguished in adolescents between thirteen and seventeen years old, including inhibition, shifting, working memory, and planning. Comparison between age-groups indicates significantly better performance on planning and working memory in the early versus late adolescents, whereas for inhibition and shifting, the differences are arguably smaller or even absent. Further, post-hoc exploratory analysis indicates no difference in EF between boys and girls in childhood (between 7 and 12 years), whereas in adolescence, sex differences emerge on three of the four EF components in favour of the girls. Finally, processing speed seems to have little effect on the model and the obtained factor structure.

The proposed hypothesis that a four-factor model would best describe EF was supported by the results of the CFA. This four-factor model contains the three EF components proposed by Miyake et al. (2000) (working memory, inhibition and shifting) as well as an additional planning component. Previous studies have included planning as part of working memory (Cirino et al., 2018) or inhibition (Welsh et al., 1999), however, our results suggest that planning can be seen as a fourth EF component that is integrated in the factor structure. Importantly, from a statistical perspective, the three-factor model with planning as part of working memory provided an equally acceptable solution compared to the four-factor model. From a practical perspective, however, a four-factor solution is preferable over a three-factor solution if both are equally supported by the statistical analyses. The fractionation into four EF components is consistent with previous work (e.g. Fournier-Vicente et al., 2008; Himi et al., 2021) and allows for a finer grain level of analysis, where qualitative differences in the four EF components separately, rather than differences in overall EF performance, can be considerably more informative (Hughes & Graham, 2002). The value of this fractionation is also corroborated by the fact that the four factors each show different rates of development during adolescence, as illustrated by the variations in the size of the mean differences between early and late adolescence. Importantly, although four separable components or processes are detected here, there is also a degree of underlying commonality, as demonstrated by the correlations between the components. This is consistent with the “unity-diversity”-theory of Miyake et al. (2000) and provides support for the so-called ‘common executive’, which contributes to the different executive processes (McKenna et al., 2017; Friedman et al., 2008, 2011; Miyake & Friedman, 2012; Himi et al., 2021).

In contrast with the four-factor model in the adolescents, the results of the current study revealed that in childhood EF can be best described as a unitary construct. This is consistent with other studies (e.g., Xu et al., 2013), and theoretical models that relate neural immaturity at the level of the prefrontal cortex, with limited reflective capacity and rule representation (Brydges et al., 2014; Diamond, 2002; Zelazo, 2015). Our findings thus support the conclusion of a recent review by McKenna and colleagues, (2017) that during childhood the cognitive processes (i.e., inhibition, updating and shifting) cannot be distinguished and are all related to a common process. Only at a later age (during adolescence), this common executive specialises into separable processes (McKenna, Rushe, & Woodcock, 2017). Other studies, however, did find multiple factors already in children ranging from 3 to 13 years old depending on the study (Gandolfi, Viterbori, Traverso, & Usai, 2014; Lee et al., 2013; Lehto et al., 2003; Miller, Giesbrecht, Muller, McInerney, & Kerns, 2012). Still, the wide variety of tasks used across studies, in combination with the large differences in age ranges and age groups renders it difficult to uncover whether this discrepancy between studies can be attributed to methodological differences or not (Karr et al., 2018). In this regard, it has been suggested that the identification of a unitary construct in childhood samples is often due to the fact that only a limited number of tasks and performance indicators are included in the analyses (Karr et al., 2018). While this critique might hold for other studies, the current study included seven different tasks. Additionally, the current results align with the findings of Xu et al. (2013), who also used a large number of different tasks that loaded on separate factors in adolescence, but were part of a unitary construct in childhood. Hence, it seems unlikely that the observation of a unitary construct for EF in childhood in the current study was caused by an insufficient number of tests.

In childhood, although statistical limitations prevent cross-sectional comparison between the different age groups, the results seem to indicate better EF in the older compared with the younger children. This is suggestive of a gradual development of EF throughout childhood. By the time children reach adolescence, the unitary factor model seems to have evolved into a four-factor model, a transition that has been associated with myelination and synaptic reorganization in the prefrontal cortex, leading to more efficient functional networks and increased capacity and use increasingly complex rules (Barnea-Goraly et al., 2005; Sowell et al., 2003; Yakovlev, 1967). Inspection of the means of the EF components demonstrated that late adolescents have superior planning and working memory skills, as well as slightly superior inhibition skills compared to early adolescents, suggesting that these EF components are still improving during adolescence, due to neural changes and concurrent improvements in cognitive capacity. In contrast, shifting does not seem to improve throughout adolescence, as no differences in shifting performance were observed between early and late adolescents. These results are in line with previous research, indicating that working memory and inhibition continue to develop into adulthood, while shifting performance levels off around young adolescence (Anderson, Anderson, Northam, Jacobs, & Catroppa, 2001; Huizinga et al., 2006).

In a recent review, Karr et al. (2018) have explicitly described the current methodological issues that EF research suffers from. They argue that the choice of tasks, performance indicators and scoring system greatly influence which factor structure is best supported by the data. However, the tasks used in this study were selected after careful consideration, and were suitable for all ages, with a scoring system that minimized floor and ceiling effects. Furthermore, each task was based on a well-validated task, and the selected performance indicators were also based on previous literature. Nonetheless, while seven different tasks to assess EF is a big step forward from what has generally been done in literature, the current study still has two factors with only one performance indicator (i.e., shifting and planning). Although this is justified from a statistical point of view, loading multiple performance indicators on each EF factor would make the current conclusions even stronger. Still, the current study provides clear support for a unitary construct of EF in childhood, evolving into a four-factor structure in adolescence that includes inhibition, shifting and working memory, but also includes planning as a fourth EF factor.

The current study also explored the influence of sex and processing speed on the EF factor structure. No sex differences were found in the childhood group, while in adolescence, girls outperformed boys on shifting, and also showed slightly better scores on inhibition and working memory. However, consistent with the results of Grissom and colleagues, these differences are small, especially with regard to inhibition and working memory, and their practical relevance probably will be limited (Grissom & Reyes, 2019). In further research, it may therefore be worth considering the mediating role of contextual factors (e.g., motivation, stress, mental state, impulsive behaviour, etc.) rather than biological sex per se (Grissom & Reyes, 2019). Processing speed also does not seem to play an identifiable role towards the unitary construct of EF, as the model for processing speed and EF did not provide an acceptable fit. A plausible explanation could be that processing speed may vary in rate of development compared to EF-components or that the high variability between individual developmental trajectories might prevent the identification of a generalizable model for this age group (Brocki & Bohlin, 2004; Fry & Hale, 2000). In adolescence however, adding processing speed to the four-factor model did provide a good fit, but this did not cause appreciable changes in the model parameters. This finding suggests that processing speed is distinct from the other four processes. Assessments of EF traditionally focus on accuracy, probably due to limitations related to pen and paper tests. With the rise of computerized assessments and given the known trade-off between speed and accuracy, it may be worth considering the additional value of reaction time in future work. Altogether, our results tend to suggest that it seems useful to take into account sex differences and processing speed in EF during adolescence, but additional research is required further clarify the role of sex and processing speed towards EF development.

The present study considerably adds to the EF literature by including multiple EF components across a very broad age range, with a test battery that taps into different EF components and is suited for all ages. It is important to note, however, that the developmental trends discussed in this paper are inferred from cross-sectional comparison of different cohorts of individuals, and longitudinal research would be more appropriate. During puberty especially, follow-up of biological maturation status (i.e., being early, on-time or late maturing compared to same-age, same-sex peers), is also necessary, since a deviation of maturational status could influence EF development (Chaku & Hoyt, 2019; Juraska & Willing, 2017; Laureys et al., 2021; Stumper, Mac Giollabhui, Abramson, & Alloy, 2020). On the other hand, an important asset of this study is the use of a multi-facetted, easily administrable and time-efficient test battery known as the CBS. The CBS is applicable across a broad age range, with minimal floor or ceiling effects, which are known to be problematic in EF literature (Best et al., 2011). The outcome measures of the seven tests were carefully selected, such that different EF components could be revealed, while considering the trade-off between number of tests and number of groups in the factor analysis. Future research may want to include more tests and thus more outcome variables to broaden our understanding of EF measurement and consequently its factor structure and development. In addition, including young adults could also provide further insights, as EF-components such as working memory and inhibition have been suggested to continue to refine into young adulthood (McAuley & White, 2011).

In conclusion, the current study is one of the few studies to include seven different EF tests across a large sample of children and adolescents. Our results show that EF is a unitary construct during childhood and develops into a four-factor construct in adolescence, with factors for working memory, inhibition, shifting and planning. Working memory, inhibition and planning improve with age during adolescence, while shifting seems to reach a plateau. Researchers and practitioners are advised to assess and monitor working memory, inhibition, shifting and planning separately during adolescence, which will allow a more detailed analysis of these EF components and of the overall EF development.

Funding

This work was supported with financial support of the Research Foundation – Flanders (FWO) with grant number FWO_3F0_2018_0031_01.

Declaration of Competing Interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.intell.2021.101600.

References

Albert, D., & Steinberg, L. (2011). Age differences in strategic planning as indexed by the tower of London [Article]. *Child Development, 82*(5), 1501–1517. https://doi.org/10.1111/j.1467-8624.2011.01613.x

Alexander, M. P., & Stuss, D. T. (2000). Disorders of frontal lobe functioning. *Seminars in Neurology, 20*(4), 427–437. https://doi.org/10.1055/s-2000-13175

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8*(2), 71–82.

Anderson, V., Anderson, P., Northam, E., Jacobs, R., & Catroppa, C. (2001). Development of executive functions through late childhood and adolescence in an Australian sample. *Developmental Neuropsychology, 20*(1), 385–406.

Arbuckle, J. L., Marcoulides, G. A., & Schumacker, R. E. (1996). Full information estimation in the presence of incomplete data. *Advanced Structural Equation Modeling: Issues and Techniques, 243*, 277.

Barnea-Goraly, N., Menon, V., Eckert, M., Tamm, L., Bammer, R., Karchemskiy, A., & Reiss, A. L. (2005). White matter development during childhood and adolescence: a cross-sectional diffusion tensor imaging study. *Cerebral Cortex, 15*(12), 1848–1854.

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 [Article]. *Learning and Individual Differences, 21*(4), 327–336. https://doi.org/10.1016/j.lindif.2011.01.007

Brenkel, M., Shulman, K., Hazan, E., Herrmann, N., & Owen, A. M. (2017). Assessing capacity in the elderly: Comparing the MoCA with a novel computerized battery of executive function. *Dementia and geriatric cognitive disorders extra, 7*(2), 249–256.

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology, 26*(2), 571–593.

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

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology, 87*(4), 299–319.

Chaku, N., & Hoyt, L. T. (2019). Developmental Trajectories of Executive Functioning and Puberty in Boys and Girls [Article]. *Journal of Youth and Adolescence, 48*(7), 1365–1378. https://doi.org/10.1007/s10964-019-01021-2

Cirino, P. T., Ahmed, Y., Miciak, J., Taylor, W. P., Gerst, E. H., & Barnes, M. A. (2018). A framework for executive function in the late elementary years [Article]. *Neuropsychology, 32*(2), 176–189. https://doi.org/10.1037/neu0000427

Collins, P., Roberts, A., Dias, R., Everitt, B., & Robbins, T. (1998). Perseveration and strategy in a novel spatial self-ordered sequencing task for nonhuman primates: Effects of excitotoxic lesions and dopamine depletions of the prefrontal cortex. *Journal of Cognitive Neuroscience, 10*(3), 332–354.

Corsi, P. (1972). *Memory and the medial temporal region of the brain*. Unpublished doctoral dissertation). Montreal, QB: McGill University.

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in neuroscience and education, 3*(2), 63–68.

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*(11), 2037–2078.

Della Sala, S., Gray, C., Spinnler, H., & Trivelli, C. (1998). Frontal lobe functioning in man: The riddle revisited. *Archives of Clinical Neuropsychology, 13*(8), 663–682. https://doi.org/10.1016/s0887-6177(97)00093-0

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. *Principles of Frontal Lobe Function*, 466–503.

Doebel, S. (2020). Rethinking executive function and its development [Article]. *Perspectives on Psychological Science, 15*(4), 942–956. Article 1745691620904771 https://doi.org/10.1177/1745691620904771.

Dohle, S., Diel, K., & Hofmann, W. (2018). Executive functions and the self-regulation of eating behavior: A review. *Appetite, 124*, 4–9.

Enders, C. K. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement, 61*(5), 713–740.

Fournier-Vicente, S., Larigauderie, P., & Gaonach, D. (2008). More dissociations and interactions within central executive functioning: A comprehensive latent-variable analysis. *Acta Psychologica, 129*(1), 32–48. https://doi.org/10.1016/j.actpsy.2008.04.004

Fransen, J., Pion, J., Vandendriessche, J., Vandorpe, B., Vaeyens, R., Lenoir, M., & Philippaerts, R. M. (2012). Differences in physical fitness and gross motor coordination in boys aged 6-12 years specializing in one versus sampling more than one sport [Article]. *Journal of Sports Sciences, 30*(4), 379–386. https://doi.org/10.1080/02640414.2011.642808

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. *J. Exp. Psychol. Gen., 137*, 201–225. https://doi.org/10.1037/0096-3445.137.2.201

Friedman, N. P., Miyake, A., Robinson, J. L., & Hewitt, J. K. (2011). Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: a behavioural genetic analysis. *Dev. Psychol., 47*, 1410–1430. https://doi.org/10.1037/a0023750

Friedman, N. P., Miyake, A., Altamirano, L. J., Corley, R. P., Young, S. E., Rhea, S. A., & Hewitt, J. K. (2016). Stability and Change in Executive Function Abilities From Late

Adolescence to Early Adulthood: A Longitudinal Twin Study. *Developmental Psychology, 52*(2), 326–340. https://doi.org/10.1037/dev0000075

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children [Article]. *Biological Psychology, 54*(1–3), 1–34. https://doi.org/10.1016/s0301-0511(00)00051-x

Gandolfi, E., Viterbori, P., Traverso, L., & Usai, M. C. (2014). Inhibitory processes in toddlers: a latent-variable approach. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00381. Article 381.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework [Review]. *Psychological Bulletin, 134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Grissom, N. M., & Reyes, T. M. (2019). Let's call the whole thing off: evaluating gender and sex differences in executive function (vol 44, pg 86, 2019) [Correction]. *Neuropsychopharmacology, 44*(7), 1344. https://doi.org/10.1038/s41386-019-0367-y

Hackman, D. A., Gallop, R., Evans, G. W., & Farah, M. J. (2015). Socioeconomic status and executive function: Developmental trajectories and mediation. *Developmental Science, 18*(5), 686–702. https://doi.org/10.1111/desc.12246

Hampshire, A., Highfield, R. R., Parkin, B. L., & Owen, A. M. (2012). Fractionating human intelligence. *Neuron, 76*(6), 1225–1237.

Himi, S. A., Buhner, M., & Hilbert, S. (2021). Advancing the Understanding of the Factor Structure of Executive Functioning. *Journal of Intelligence, 9*(1). https://doi.org/10.3390/jintelligence9010016. Article 16.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1), 53–60.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Hughes, C., & Graham, A. (2002). Measuring executive functions in childhood: Problems and solutions? *Child and Adolescent Mental Health, 7*(3), 131–142.

Huizinga, M., Dolan, C., & van der Molen, M. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*(11), 2017–2036.

Hung, T. M., Tsai, C. L., Chen, F. T., Wang, C. C., & Chang, Y. K. (2013). The immediate and sustained effects of acute exercise on planning aspect of executive function [Article]. *Psychology of Sport and Exercise, 14*(5), 728–736. https://doi.org/10.1016/j.psychsport.2013.05.004

Inoue, S., & Matsuzawa, T. (2007). Working memory of numerals in chimpanzees. *Current Biology, 17*(23), R1004–R1005.

Juraska, J. M., & Willing, J. (2017). Pubertal onset as a critical transition for neural development and cognition [Review]. *Brain Research, 1654*, 87–94. https://doi.org/10.1016/j.brainres.2016.04.012

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies [Review]. *Psychological Bulletin, 144*(11), 1147–1185. https://doi.org/10.1037/bul0000160

Kenny, L., Cribb, S. J., & Pellicano, E. (2019). Childhood executive function predicts later autistic features and adaptive behavior in young autistic people: A 12-year prospective study. *Journal of Abnormal Child Psychology, 47*(6), 1089–1099. https://doi.org/10.1007/s10802-018-0493-8

Kline, R. B. (2015). *Principles and practice of structural equation modeling.* Guilford Publications.

Lang, K., Stahl, D., Espie, J., Treasure, J., & Tchanturia, K. (2014). Set shifting in children and adolescents with anorexia nervosa: An exploratory systematic review and meta-analysis [Article]. *International Journal of Eating Disorders, 47*(4), 394–399. https://doi.org/10.1002/eat.22235

Laureys, F., Middelbos, L., Rommers, N., De Waelle, S., Coppens, E., Mostaert, M., … Lenoir, M. (2021). The effects of age, biological maturation and sex on the development of executive functions in adolescents. *Frontiers in Physiology, 12*. https://doi.org/10.3389/fphys.2021.703312

Lee, K., Bull, R., & Ho, R. M. (2013). Developmental changes in executive functioning. *Child Development, 84*(6), 1933–1953.

Lehto, J. E., Juujarvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21*, 59–80. https://doi.org/10.1348/026151003321164627

Levin, H. S., Fletcher, J. M., Kufera, J. A., Harward, H., Lilly, M. A., Mendelsohn, D., & Eisenberg, H. M. (1996). Dimensions of cognition measured by the Tower of London and other cognitive tasks in head-injured children and adolescents. *Developmental Neuropsychology, 12*(1), 17–34. https://doi.org/10.1080/87565649609540638

Liang, J., Matheson, B. E., Kaye, W. H., & Boutelle, K. N. (2014). Neurocognitive correlates of obesity and obesity-related behaviors in children and adolescents. *International Journal of Obesity, 38*(4), 494–506. https://doi.org/10.1038/ijo.2013.142

McAuley, T., & White, D. A. (2011). A latent variables examination of processing speed, response inhibition, and working memory during typical development [Article]. *Journal of Experimental Child Psychology, 108*(3), 453–468. https://doi.org/10.1016/j.jecp.2010.08.009

McKenna, Roisin, Rushe, T, & Woodcock, Kate (2017). Informing the Structure of Executive Function in Children: A Meta-Analysis of Functional Neuroimaging Data. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2017.00154

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543.

Miller, M. R., Giesbrecht, G. F., Muller, U., McInerney, R. J., & Kerns, K. A. (2012). A Latent variable approach to determining the structure of executive function in preschool children. *Journal of Cognition and Development, 13*(3), 395–423. https://doi.org/10.1080/15248372.2011.585478

Mischel, W., Shoda, Y., & Rodriguez, M. I. (1989). Delay of gratification in children. *Science, 244*(4907), 933–938.

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*(1), 8–14. https://doi.org/10.1177/0963721411429458

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 [Article; Proceedings Paper]. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Monsell, S. (1996). *Control of mental processes. unsolved mysteries of the mind: Tutorial essays in cognition* (pp. 93–148).

Morris, R. G., Miotto, E. C., Feigenbaum, J. D., Bullock, P., & Polkey, C. E. (1997). The effect of goal-subgoal conflict on planning ability after frontal- and temporal-lobe lesions in humans. *Neuropsychologia, 35*(8), 1147–1157. https://doi.org/10.1016/s0028-3932(97)00009-2

Nemati, P., Schmid, J., Soltanlou, M., Krimly, J.-T., Nuerk, H.-C., & Gawrilow, C. (2017). *Planning and self-control, but not working memory, directly predict multiplication performance in adults*.

Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction.* Routledge.

Prencipe, A., Kesek, A., Cohen, J., Lamm, C., Lewis, M. D., & Zelazo, P. D. (2011). Development of hot and cool executive function during the transition to adolescence [Article]. *Journal of Experimental Child Psychology, 108*(3), 621–637. https://doi.org/10.1016/j.jecp.2010.09.008

Robertson, I. H., Ridgeway, V., Greenfield, E., & Parr, A. (1997). Motor recovery after stroke depends on intact sustained attention: a 2-year follow-up study. *Neuropsychology, 11*(2), 290.

Rommelse, N. N., Altink, M. E., De Sonneville, L. M., Buschgens, C. J., Buitelaar, J., Oosterlaan, J., & Sergeant, J. A. (2007). Are motor inhibition and cognitive flexibility dead ends in ADHD? *Journal of Abnormal Child Psychology, 35*(6), 957–967.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software, 48*(2), 1–36.

Scheerder, J., Borgers, J., & Willem, A. (2015). Sportdeelname in Vlaanderen: trends en profielen [Sport participation in Flanders: trends and profiles]. In *Participatie in Vlaanderen 2: eerste analyses van de participatiesurvey 2014* (pp. 209–250). Acco.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 298*(1089), 199–209.

Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L., & Toga, A. W. (2003). Mapping cortical change across the human life span. *Nature Neuroscience, 6*(3), 309–315.

Steenkamp, J.-B. E., & Baumgartner, H. (1998). *Assessing measurement invariance through multi-sample structural equation modeling*.

Stroop, J. R. (1992). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: General, 121*(1), 15.

Stumper, A., Mac Giollabhui, N., Abramson, L. Y., & Alloy, L. B. (2020). Early Pubertal Timing Mediates the Association between Low Socioeconomic Status and Poor Attention and Executive Functioning in a Diverse Community Sample of Adolescents [Article]. *Journal of Youth and Adolescence, 49*(7), 1420–1432. https://doi.org/10.1007/s10964-020-01198-x

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: a conceptual view. *Psychological Research Psychologische Forschung, 63*(3–4), 289–298. https://doi.org/10.1007/s004269900007

Stuss, D. T., & Benson, D. F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin, 95*(1), 3.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). MA: Pearson Boston.

Theodoraki, T. E., McGeown, S. P., Rhodes, S. M., & MacPherson, S. E. (2019). Developmental changes in executive functions during adolescence: A study of inhibition, shifting, and working memory. *The British Journal of Developmental Psychology, 38*(1), 74–89. https://doi.org/10.1111/bjdp.12307

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research [Review]. *Organizational Research Methods, 3*(1), 4–70. https://doi.org/10.1177/109442810031002

van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2013). The structure of executive functions in children: A closer examination of inhibition, shifting, and updating [Article]. *British Journal of Developmental Psychology, 31*(1), 70–87. https://doi.org/10.1111/j.2044-835X.2012.02079.x

Vestberg, T., Gustafson, R., Maurex, L., Ingvar, M., & Petrovic, P. (2012). Executive functions predict the success of top-soccer players [Article]. *PLoS One, 7(4), 5*. https://doi.org/10.1371/journal.pone.0034731. Article e34731.

Welsh, M. C., Satterlee-Cartmell, T., & Stine, M. (1999). Towers of Hanoi and London: Contribution of working memory and inhibition to performance [Article]. *Brain and Cognition, 41*(2), 231–242. https://doi.org/10.1006/brcg.1999.1123

Xu, F., Han, Y., Sabbagh, M. A., Wang, T. F., Ren, X. Z., & Li, C. H. (2013). Developmental differences in the structure of executive function in middle childhood and adolescence [Article]. *PLoS One, 8(10), 9*. https://doi.org/10.1371/journal.pone.0077770. Article e77770.

Yakovlev, P. (1967). The myelogenetic cycles of regional maturation of the brain. *Regional Development of the Brain in Early Life*, 3–70.

Zelazo, P. D. (2015). Executive function: Reflection, iterative reprocessing, complexity, and the developing brain. *Developmental Review, 38*, 55–68. https://doi.org/10.1016/j.dr.2015.07.001

Zelazo, P. D., & Muller, U. (2002). The balance beam in the balance: Reflections on rules, relational complexity, and developmental processes. *Journal of Experimental Child Psychology, 81*(4), 458–465. https://doi.org/10.1006/jecp.2002.2667