Age-related change in executive function: Developmental trends and a latent variable analysis
Huizinga, M.; Dolan, C.V.; van der Molen, M.W.

Published in:
Neuropsychologia

DOI:
10.1016/j.neuropsychologia.2006.01.010

Link to publication

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Age-related change in executive function: Developmental trends and a latent variable analysis

Mariëtte Huizinga *, Conor V. Dolan, Maurits W. van der Molen

Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands

Received 1 August 2005; received in revised form 17 January 2006; accepted 20 January 2006

Available online 9 March 2006

Abstract

This study examined the developmental trajectories of three frequently postulated executive function (EF) components, Working Memory, Shifting, and Inhibition of responses, and their relation to performance on standard, but complex, neuropsychological EF tasks, the Wisconsin Card Sorting Task (WCST), and the Tower of London (ToL). Participants in four age groups (7-, 11-, 15-, and 21-year olds) carried out nine basic experimental tasks (three tasks for each EF), the WCST, and the ToL. Analyses were done in two steps: (1) analyses of (co)variance to examine developmental trends in individual EF tasks while correcting for basic processing speed, (2) confirmatory factor analysis to extract latent variables from the nine basic EF tasks, and to explain variance in the performance on WCST and ToL, using these latent variables. Analyses of (co)variance revealed a continuation of EF development into adolescence. Confirmatory factor analysis yielded two common factors: Working Memory and Shifting. However, the variables assumed to tap Inhibition proved unrelated. At a latent level, again correcting for basic processing speed, the development of Shifting was seen to continue into adolescence, while Working Memory continued to develop into young-adulthood. Regression analyses revealed that Working Memory contributed most strongly to WCST performance in all age groups. These results suggest that EF component processes develop at different rates, and that it is important to recognize both the unity and diversity of EF component processes in studying the development of EF.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Executive function; Development; Working memory; Shifting; Inhibition; Structural equation modeling; Wisconsin Card Sorting Task; Tower of London

1. Introduction

Across development, children become increasingly more able to control their thoughts and actions (for a review see: Diamond, 2002). This change has been associated with the development of executive function (EF), which is an umbrella term for various cognitive processes that subserve goal-directed behavior (Miller & Cohen, 2001; see also Luria, 1966; Shallice, 1982). EF is especially important in novel or demanding situations (Stuss, 1992), which require a rapid and flexible adjustment of behavior to the changing demands of the environment (Zelazo, Muller, Frye, & Marcoritch, 2003). EF is thought to rely strongly on prefrontal cortex (PFC), as indicated by studies showing that patients with lesions to PFC perform poorly on tasks such as the Wisconsin Card Sorting Task (WCST; Grant & Berg, 1948) and the Tower of London (ToL; Shallice, 1982; for a review see: Stuss & Levine, 2002). On the WCST, which requires flexible switching between sorting rules, PFC patients typically perseverate, i.e., they persist in sorting according to the rule that was previously correct (e.g., Anderson, Damasio, Jones, & Tranel, 1991; Milner, 1963; Nagahama, Okina, Suzuki, Nabatame, & Matsuda, 2005; Stuss et al., 2000). On the ToL, which requires spatial problem solving by moving balls in order to reach a pre-specified goal, PFC patients require more moves to solve the problem (e.g., Andres & Van der Linden, 2001; Carlin et al., 2000; Morris, Ahmed, Syed, & Toone, 1993; Owen, Downes, Sahakian, Polkey, & Robbins, 1990).

Children show a similar pattern as patients with PFC damage; that is, they also perseverate on the WCST and require more moves to solve ToL problems (Anderson, Anderson, & Lajote, 1996; Baker, Segalowitz, & Feltisi, 2001; Chelune & Baer, 1986; Chelune & Thompson, 1987; Heaton, Chelune, Talley, Kay, & Curtis, 1993; Kirk & Kelly, 1986; Lehto, 2004;
A major theoretical issue concerns the organization of EF. It has been suggested that EF is unitary, i.e., it does not include distinct sub-functions or sub-components. This means that the cognitive and behavioral impairments seen after PFC damage can be explained entirely in terms of one dysfunctional system (e.g., Baddeley, 1986; Stuss, Shallice, Alexander, & Picton, 1995; see also Teuber, 1972). The multi-faceted nature of EF is suggested by behavioral studies incorporating batteries of widely used EF tasks. These studies yielded low or non-significant correlations between tasks and exploratory factor analysis yielded multiple factors (Brocki & Bohlin, 2004; Culbertson & Zillmer, 1998; Lehto, 1996; Levin et al., 1996; Pennington, 1997; Robbins et al., 1994; Welsh et al., 1991).

Neuroimaging studies provide evidence in support of the multi-faceted nature of EF, as different components of EF are seen to rely on different parts of PFC. For example, the ability to maintain information in working memory has been found to recruit mostly lateral PFC (Narayanan et al., 2000; Smith & Jonides, 1999). In contrast, others view EF as multi-faceted (non-unitary). These authors argued that EF involves several discrete cognitive processes that have a relatively focal neural representation (e.g., Baddeley, 1986; Stuss, Shallice, Alexander, & Picton, 1995; see also Teuber, 1972). The multi-faceted nature of EF is suggested by behavioral studies incorporating batteries of widely used EF tasks. These studies yielded low or non-significant correlations between tasks and exploratory factor analysis yielded multiple factors (Brocki & Bohlin, 2004; Culbertson & Zillmer, 1998; Lehto, 1996; Levin et al., 1996; Pennington, 1997; Robbins et al., 1994; Welsh et al., 1991).

Developmental studies using standard neuropsychological tests have shown that EF has a protracted course of development, beginning in early childhood and continuing into adolescence. However, these EF tasks are subject to distinct developmental trajectories. For example, on the WCST, analysis of perseverative errors indicates that the performance of children is comparable to that of young-adults by 12 years of age; however, analysis of failure-to-maintain set indicates that children do not reach adult levels of performance until 13–15 years of age (e.g., Chelune & Buer, 1986; Chelune & Thompson, 1987; Levin et al., 1991; Welsh et al., 1991). Similarly, on the ToL task, performance based on errors appears to continuously improve from middle childhood into young-adulthood; however, when performance is based on both errors and time, adult levels of performance may be attained as early as 13 years of age (Baker et al., 2001; see also Levin et al., 1996).

A growing body of research indicating differential trends in the development of EF component processes.

1 A growing body of research appeared recently, focusing on EF in pre-school aged children (e.g., Carlson, Mandell, & Williams, 2004; Diamond, Bratton, Lehto, Juusaaervi, Kooistra, & Puikinen, 2003; Paniak, Miller, Murphy, & Patterson, 1996; Welsh, Pennington, & Greissler, 1991). The slow development of EF has been attributed to the protracted maturation of PFC (e.g., Diamond, 2002). Conclusive evidence about the developmental trajectories of the different EF components in relation to the performance on standard neuropsychological EF tasks has yet to be established. In this study, we examined the development of EF component processes by using a multi-group confirmatory factor analysis. Where we have at our disposal multiple indicators of a given latent variable, this approach has the advantage that it allows us to study performance at the level of the latent variables, according to a pre-specified model of EF.

1.1. Decomposition of executive function

Developmental studies using standard neuropsychological tests have shown that EF has a protracted course of development, beginning in early childhood and continuing into adolescence. However, these EF tasks are subject to distinct developmental trajectories. For example, on the WCST, analysis of perseverative errors indicates that the performance of children is comparable to that of young-adults by 12 years of age; however, analysis of failure-to-maintain set indicates that children do not reach adult levels of performance until 13–15 years of age (e.g., Chelune & Buer, 1986; Chelune & Thompson, 1987; Levin et al., 1991; Welsh et al., 1991). Similarly, on the ToL task, performance based on errors appears to continuously improve from middle childhood into young-adulthood; however, when performance is based on both errors and time, adult levels of performance may be attained as early as 13 years of age (Baker et al., 2001; see also Levin et al., 1996).

1 A growing body of research appeared recently, focusing on EF in pre-school aged children (e.g., Carlson, Mandell, & Williams, 2004; Diamond, Bratton, Lehto, Juusaaervi, Kooistra, & Puikinen, 2003; Paniak, Miller, Murphy, & Patterson, 1996; Welsh, Pennington, & Greissler, 1991). The slow development of EF has been attributed to the protracted maturation of PFC (e.g., Diamond, 2002). Conclusive evidence about the developmental trajectories of the different EF components in relation to the performance on standard neuropsychological EF tasks has yet to be established. In this study, we examined the development of EF component processes by using a multi-group confirmatory factor analysis. Where we have at our disposal multiple indicators of a given latent variable, this approach has the advantage that it allows us to study performance at the level of the latent variables, according to a pre-specified model of EF.

1.1. Decomposition of executive function

Developmental studies using standard neuropsychological tests have shown that EF has a protracted course of development, beginning in early childhood and continuing into adolescence. However, these EF tasks are subject to distinct developmental trajectories. For example, on the WCST, analysis of perseverative errors indicates that the performance of children is comparable to that of young-adults by 12 years of age; however, analysis of failure-to-maintain set indicates that children do not reach adult levels of performance until 13–15 years of age (e.g., Chelune & Buer, 1986; Chelune & Thompson, 1987; Levin et al., 1991; Welsh et al., 1991). Similarly, on the ToL task, performance based on errors appears to continuously improve from middle childhood into young-adulthood; however, when performance is based on both errors and time, adult levels of performance may be attained as early as 13 years of age (Baker et al., 2001; see also Levin et al., 1996).

1 A growing body of research appeared recently, focusing on EF in pre-school aged children (e.g., Carlson, Mandell, & Williams, 2004; Diamond, Bratton, Lehto, Juusaaervi, Kooistra, & Puikinen, 2003; Paniak, Miller, Murphy, & Patterson, 1996; Welsh, Pennington, & Greissler, 1991). The slow development of EF has been attributed to the protracted maturation of PFC (e.g., Diamond, 2002). Conclusive evidence about the developmental trajectories of the different EF components in relation to the performance on standard neuropsychological EF tasks has yet to be established. In this study, we examined the development of EF component processes by using a multi-group confirmatory factor analysis. Where we have at our disposal multiple indicators of a given latent variable, this approach has the advantage that it allows us to study performance at the level of the latent variables, according to a pre-specified model of EF.

1.1. Decomposition of executive function

A major theoretical issue concerns the organization of EF. It has been suggested that EF is unitary, i.e., it does not include distinct sub-functions or sub-components. This means that the cognitive and behavioral impairments seen after PFC damage can be explained entirely in terms of one dysfunctional system (e.g., Baddeley, 1986; Stuss, Shallice, Alexander, & Picton, 1995; see also Teuber, 1972). The multi-faceted nature of EF is suggested by behavioral studies incorporating batteries of widely used EF tasks. These studies yielded low or non-significant correlations between tasks and exploratory factor analysis yielded multiple factors (Brocki & Bohlin, 2004; Culbertson & Zillmer, 1998; Lehto, 1996; Levin et al., 1996; Pennington, 1997; Robbins et al., 1994; Welsh et al., 1991).

Neuroimaging studies provide evidence in support of the multi-faceted nature of EF, as different components of EF are seen to rely on different parts of PFC. For example, the ability to maintain information in working memory has been found to recruit mostly lateral PFC (Narayanan et al., 2000; Smith & Jonides, 1999). In contrast, others view EF as multi-faceted (non-unitary). These authors argued that EF involves several discrete cognitive processes that have a relatively focal neural representation (e.g., Baddeley, 1986; Stuss, Shallice, Alexander, & Picton, 1995; see also Teuber, 1972). The multi-faceted nature of EF is suggested by behavioral studies incorporating batteries of widely used EF tasks. These studies yielded low or non-significant correlations between tasks and exploratory factor analysis yielded multiple factors (Brocki & Bohlin, 2004; Culbertson & Zillmer, 1998; Lehto, 1996; Levin et al., 1996; Pennington, 1997; Robbins et al., 1994; Welsh et al., 1991). The slow development of EF has been attributed to the protracted maturation of PFC (e.g., Diamond, 2002). Conclusive evidence about the developmental trajectories of the different EF components in relation to the performance on standard neuropsychological EF tasks has yet to be established. In this study, we examined the development of EF component processes by using a multi-group confirmatory factor analysis. Where we have at our disposal multiple indicators of a given latent variable, this approach has the advantage that it allows us to study performance at the level of the latent variables, according to a pre-specified model of EF.
These studies, although not entirely unequivocal, show that adult-level performance on different EF tasks is attained at different ages during childhood and adolescence (for reviews see: Diamond, 2002; Welsh, 2002). Working memory capacity has been found to gradually develop throughout childhood and into adolescence (e.g., Beveridge, Jarrold, & Pettit, 2002; Brocki & Bohlin, 2004; DeLuca et al., 2003; Gathercole, Pickering, Ambridge, & Wearing, 2004; Hitch, Halliday, Dodd, & Little, 1989; Luciana, Conklin, Hooper, & Yarger, 2005; Luciana & Nelson, 1998; Luna, Garver, Urban, Lazar, & Sweeney, 2004). In addition, recent studies on the development of task shifting abilities all show that the cost of switching between tasks decreases as children grow older, with adult levels of performance being attained around the age of 12 (Cepeda, Kramer, & Gonzalez de Sather, 2001; Crone, Bunge, Van der Molen, & Ridderinkhof, in press; Huizinga & Van der Molen, in preparation-a; Kray, Eber, & Lindenberger, 2004). Finally, inhibitory control was found to increase throughout childhood (e.g., Klenberg, Korkman, & Laihi Nuuttila, 2001), and to reach adult level of performance in late childhood, around the age of 12 (Bédard et al., 2002; Bunge, Dudukovic, Thomasom, Vaidya, & Gabrieli, 2002; Durston et al., 2002; Ridderinkhof & Van der Molen, 1995; Van den Wildenberg & Van der Molen, 2004), or early adolescence (Williams, Ponesse, Schachar, Logan, & Tamcock, 1999).

A straightforward interpretation of the developmental trends of the EF component processes is hampered by a number of factors. First, different tasks are used across studies to measure the same EF component. For example, the developmental trajectory of Inhibition has been assessed by using a Go/NoGo task (e.g., Durston et al., 2002), the Eriksen Flankers task (e.g., Bunge et al., 2002; Ridderinkhof & Van der Molen, 1995), and the Stop-signal task (e.g., Bédard et al., 2002; Van den Wildenberg & Van der Molen, 2004; Williams et al., 1999). Second, it is unclear whether children at various ages use the same strategy when performing on EF tasks. This issue concerns the question of measurement invariance, i.e., whether we are actually measuring the same construct across age (Meredith, 1993). Third, many developmental studies focused on a single EF component process. This precludes a reliable assessment of developmental change across EF component processes, because differential rates might be due to different samples rather than components. Thus, a reliable assessment of the developmental patterning of EF component processes requires homogeneous age groups and the application of a latent variables approach to extract what the various tasks used to tap EF component processes have in common. This approach has been adopted by Lehto et al. (2003), who were the first to assess the patterning of EF component processes in children by using SEM. Importantly, they observed the same factor structure in a group of 8–13-year old children as previously found by Miyake et al. (2000) in adults.

In the present study, we hope to contribute to these results by adopting a multi-group design, and thus providing a more graded assessment of developmental change in EF component processes.

1.3. The present study

In the present study, we adopted the conceptual framework of Miyake et al. (2000) to assess developmental change in EF. The main goal of this study was to examine age-related changes in the three EF components distinguished by Miyake et al. (2000), i.e., Working Memory, Shifting, and Inhibition. In order to shed some light on the development of these EF component processes, we tested children in three homogeneous age groups (i.e., 7-, 11-, 15-year olds), in addition to a group of young-adults (i.e., 21-year olds). Adult level on EF tasks is typically reached in late childhood or early adolescence (for reviews see: Diamond, 2002; Welsh, 2002). The decision to limit the youngest group to 7-year olds was based on the consideration that the present task battery was probably too difficult for children younger than 7 years of age. The EF components, Working Memory, Shifting, and Inhibition, were indexed by nine experimental tasks, three for each EF component.

Working Memory was defined as the collection of cognitive processes that temporarily retain information in an accessible state, suitable for carrying out any mental task (Cowan, 1998). The essence of this component is the monitoring and coding of incoming information with respect to relevance and replacement of information that is no longer relevant by newly relevant information. Shifting was interpreted as shifting back and forth between multiple tasks (Allport, Styles, & Hsieh, 1994; Monsell, 1996, 2003). When different (usually choice RT) tasks are mixed within blocks, shifting between tasks typically results in an increase in RT and a decrease in accuracy (i.e., shift costs). Inhibition was conceptualized as the ability to deliberately inhibit dominant, automatic, or pre-potent responses (Logan & Cowan, 1984). The essence of this EF component lies in the suppression of a response or in the control of interfering stimuli or competing responses.

We adopted two approaches in analyzing the data. First, we conducted a standard analysis of variance approach. Second, we took a latent variable approach, i.e., multi-group confirmatory factor analyses (Dolan, 2000; Meredith, 1993). We examined (i) the organization of executive function in children and young-adults by investigating whether the indicators of the Working Memory, Shifting, and Inhibition tasks measured the same constructs across age, (ii) whether this organization changes across development, and (iii) how EF component processes contribute to the performance on the WCST and the ToL across age groups, again following Miyake et al. (2000). We included the WCST and the ToL because these tasks have been used previously to study the development of EF (Anderson, Byrd, & Berg, 2005; Baker et al., 2001; Chelune & Baer, 1986; Lehto, 2004; Welsh et al., 1991).
2. Method

2.1. Sample

The present study included four age groups: seventy-one 7-year olds (39 female, M age = 7.2 (age range: 6–8)); M Raven-quartile = 3.6 (S.D. = 0.88); M number of years of education = 0.56 (S.D. = 0.10)); one hundred and eight 11-year olds (62 female, M age = 11.2 (age range: 10–12)); M Raven-quartile = 3.2 (S.D. = 0.93); M number of years of education = 1.92 (S.D. = 0.13)); one hundred and eleven 15-year olds (58 female, M age = 15.5 (age range: 14–16)); M Raven-quartile = 3.1 (S.D. = 0.99); M number of years of education = 7.20 (S.D. = 0.20); and ninety-four 21-year olds (72 female, M age = 20.8 (age range: 18–26)); M Raven-quartile = 3.7 (S.D. = 0.52); M number of years of education = 10.55 (S.D. = 0.31)). Children were recruited from regular local schools; the 21-year olds were university students and they received academic course credit.

Informed consent was obtained. All participants had normal, or corrected-to-normal, vision. The 7- and 11-year olds received a small present for their participation, the 15-year olds received €10, and the 21-year olds received course credit.

To assess intelligence, we used a non-verbal IQ test, the Standard Progressive Matrices (SPM; Raven, Court, & Raven, 1985). Raw scores were converted to quartile scores, following the norms for each age group. Mean Raven-quartile scores differed between groups (F(3, 359) = 17.79, p < 0.001). Positive Bonferroni tests indicated significant differences between the young-adult group and the other groups. Preliminary analyses were performed to control for IQ, and we found the relationship between IQ and the different EF measures absent both within-groups and between-groups. This is consistent with other studies that found no relation between IQ and the EF (e.g., Anderson et al., 1991; Beach & Martin, 2004, but also see Anderson, Anderson, Northam, Jacobs, & Catroppa, 2001; Conway, Kane, & Engle, 2003; Duncan, et al., 1996 for opposite results). The gender distribution across different groups differed significantly, χ^2(2) = 14.64, p < 0.002. This was caused by a relatively larger proportion of women in the young-adult group. The addition of gender as a covariate in the analyses reported below did not change any of the main effects or interactions involving the respective task manipulations. Therefore, Raven IQ scores and gender were not included in the analyses reported below.

2.2. Experimental tasks

The experimental tasks were designed to assess the EF components Working Memory, Shifting, and Inhibition. The task battery also included three complex EF tasks: WCST, the ToL, and the Random Number Generation task. The results relating to the last task are presented elsewhere (Huisings & Van der Molen, in preparation-b). Task administration was computerized, and presented on a Toshiba Satellite 1460 laptop (Intel Celeron 800 MHz processor; 15.6 in. 60 Hz monitor). All tasks required left- and right-hand responses. The response button for the right hand was the ‘‘↑’’ key on the computer keyboard, for the left hand it was the ‘‘↓’’ key (responses were counterbalanced across participants).

The nine EF tasks that we used were speeded choice reaction time (RT) tasks. The WCST and ToL were computerized versions of the standard neuropsychological tests. The participants were coached to balance speed and accuracy when responding on all tasks, except the WCST and ToL. Care was taken to ensure that participants understood the instructions, as indexed by verbal report, response accuracy, and stable RTs.

2.2.1. Working memory

Tic Tac Toe

In the Tic Tac Toe task (adapted from Miller, 1971), participants were required to retain visual information about the orientation of a pattern of figures across their working memory. The task consisted of a memorizing phase and a recognition phase. A pattern consisting of Xs and Os was presented within a 3 × 3 grid during the memorizing phase. Working memory load was varied by using patterns consisting of three versus four letters (i.e., low memory load versus high memory load). The recognition phase was initiated by pressing the space bar. During this phase, the Xs and Os were presented one after another at different positions in the grid, each for a period of 600 ms, in series that varied from four to seven presentations for the low memory load, and from four to nine presentations for the high memory load. As soon as the pattern of Xs and Os included the pre-specified pattern indicated in the memorizing phase, participants were required to press a button. Participants first received a practice block of 3 trials, followed by one block including 15 trials for low memory load and one block including 15 trials for high memory load. The low and high memory load blocks were counterbalanced across participants to control for order effects. Participants had 350 ms to respond, the time interval between trials varied randomly between 900 and 1100 ms (drawn from a uniform distribution). The main dependent variable was the proportion of correct trials.

Tic Tac Toe

Running Memory

Visual search

Visual search

Local Global

Local Global

(i.e., a low memory load versus a medium memory load versus a high memory load) was created by using, respectively, eight pictures of fruit and three of animals; versus eight pictures of fruit and four of animals; versus eight pictures of fruit and five of animals). The participants received one block for each condition, consisting of eight response probes. The order of the low, medium and high memory load blocks was counterbalanced across participants to control for order effects. Participants had 350 ms to respond. The time interval between the stimulus pairs varied randomly between 1000 and 1250 ms (drawn from a uniform distribution). The main dependent variable was the proportion of correct trials.

2.2.2. Shifting

Local-Global (adapted from Miyake et al., 2000). Participants responded to randomly presented rectangular or squares by pressing a left or right response button, respectively. Larger (global) rectangles/squares consist of smaller (local) rectangles or squares. In one block, participants responded to the local figure, in a second block the responded to the global figure (blocks 1 and 2, in a randomized order; 30 practice trials and 50 experimental trials per block), and in the third block they alternated between a series of four “local figures” and a series of four “global” trials (block 3). 90 practice trials, 150 experimental trials). A cue indicated to which dimension (global or local) the participants should respond. Cues that related to the global (local) dimension consisted of a big (small) square, presented at one side of the target stimulus, and a big (small) rectangle, presented at the other side of the target stimulus. The color of the cue and of the target stimulus was 500 ms. The interval between the response
and the presentation of the cue was fixed at 1000 ms. The main dependent variables were the median response latencies on task repetition and task alternation trials.

**Dots–Triangles** (adapted from Rogers & Monsell, 1995). Varying numbers of dots or triangles (from 3 to 6) were presented in a 2 × 2 grid on the screen (i.e., three to eight dots or triangles per half of the grid; equally distributed). During the “dots” task, participants had to decide whether there are more dots in the left or the right part of the screen (block 1: 30 practice trials, 50 experimental trials). During the “triangles” task, participants had to decide whether there are more triangles in the top or in the bottom part of the screen (block 2: 30 practice trials, 50 experimental trials). Blocks 1 and 2 were administered in randomized order. In the third block (30 practice trials, 150 experimental trials), a series of four “dots” trials and a series of four “triangles” trials were alternately presented to the participants. A stimulus remained on the screen until a response was given. Green boxes (> 1000 ms) and red boxes (between 900 and 1100 ms (drawn from a uniform distribution)) were displayed after the presentation of the stimulus. The time interval between the response and the next stimulus was 1000 ms. The main dependent variables were the median response latencies on repetition and alternation trials.

**Smiling Faces** (adapted from Rogers & Monsell, 1995; see also Span, Riddertinkhof, & Van der Molen, 2004). Stimuli were schematic faces (man or woman, smiling or non-smiling) that appeared in a 2 × 2 grid. On each trial a stimulus was presented in the center of one of the four (5 × 5 cm) quadrants. Participants were required to respond to gender if the face was presented in either one of the top two quadrants, or to the expression of the face if presented in one of the two lower quadrants (blocks 1 and 2, in randomized order; 50 practice trials and 50 experimental trials per block). In the third block, the stimulus moved clockwise through the grid and participants had to respond to gender, if the face occurred in one of the top two quadrants, or to the expression of the face, if it occurred in one of the two lower quadrants (blocks 3, 90 practice trials, 150 experimental trials). The stimulus remained on the screen until a response was given. Participants had 1500 ms to respond. The time interval between the response and the stimulus of the subsequent trial varied randomly between 900 and 1100 ms (drawn from a uniform distribution). The main dependent variables were the median response latencies on repetition and alternation trials.

**Stop-signal**. In the present version of the Stop-signal task (adapted from Van Bostel, Van der Molen, Jennings, & Bronia, 2003), participants had to respond as fast as possible to a left or right pointing arrow by a left or right button press. On 25% of the trials, the color of the arrow changed (unpredictably) from green to red, indicating that the response to the arrow stimulus should be inhibited. The time interval between arrow onset and arrow color varied depending on the participant’s performance. A dynamic tracking algorithm was used to ensure that stopping approximated 50% correct inhibited responses. The stimulus remained on the screen until a response was given. Participants had 1250 ms to respond. The time interval between the response and the next arrow onset on the subsequent trial varied randomly between 1650 and 2150 ms (drawn from a uniform distribution). There were 50 practice trials, and 2 blocks of 100 experimental trials. The main dependent variable was the mean stop-signal reaction time (SSRT), reflecting the latency of the internal response to the stop signal (see Logan, 1994).

**Erikson Flankers** (adapted from Riddertinkhof & Van der Molen, 1995). The participant’s task was to respond to a left versus right pointing arrow presented at the center of the screen by pressing a left or right response button. The arrow was flanked by four arrows pointing in the same direction (i.e., → → → → or ← ← ← ← or ← → ← →), or in an opposite direction (i.e., → → ← ← or ← → → ← or ← ← ← →, incongruent condition). The arrow array was presented in a rectangle, which served as the warning stimulus, as well as the response stimulus. The time interval between the onset of the warning stimulus and the presentation of the arrow array was 500 ms. The arrow array remained on the screen until a response was given. Participants had 2500 ms to respond. The time interval between the response and the presentation of the warning stimulus of the subsequent trial was fixed at 1000 ms. There were 50 practice trials and 50 experimental trials (i.e., the congruent trials and 50 incongruent trials, varied pseudo-randomly). The main dependent variables were the median response latencies on congruent and incongruent trials.

**Stroop** (adapted from Stroop, 1935). Stimuli were pictures of “smileys” either with a blue or red color (color task) and either right-side-up or upsidedown oriented (orientation task). On the color task, participants had to respond with their dominant hand to the blue or the red color of the stimulus (pure block 1, 30 practice trials, 85 experimental trials). On the orientation task, participants had to respond to the straight or the upsidedown orientation of the stimulus (pure block 2, 30 practice trials, 85 experimental trials). Blocks 1 and 2 were administered in random order and were presented to establish a strong mapping between stimulus (color or orientation) and response. On a third block (30 practice trials, 100 experimental trials), participants responded to interference trials using the same hand. They responded to one stimulus color (e.g., red), but only when the orientation of stimulus was right-side-up and to the other stimulus color (e.g., blue) but only when the orientation of the stimulus was upsidedown. Stimulus color and orientation corresponded on half of the trials. The stimulus remained on the screen until a response was given. Participants had 900 ms to respond. The time interval between the response and the stimulus of the subsequent trial varied randomly between 600 and 800 ms (drawn from a uniform distribution). The main dependent variable was the median response latency on interference trials. The time interval between the response and the subsequent stimulus of the task varied randomly between 600 and 800 ms (drawn from a uniform distribution). The main dependent variable was the median response latency on interference trials.
moves across each difficulty level), and planning time (i.e., the interval between the occurrence of an item and the first mouse click on one of the balls to be moved).

2.3. Procedure

Testing took place in two sessions. In each session, six tasks were administered. The 7-year olds were tested simultaneously in groups of two. The 11-year olds, 15-year olds, and 21-year olds were tested simultaneously in groups of four. Each test session lasted approximately 1.5 h. The order of tasks was counterbalanced across participants, the WCST and ToL were always administered last, considering the inter-individual variation in total time needed to complete the task. There were 3-min breaks between tasks, and a 10-min break following three tasks. At the end of the test-session, the 15-year olds and 21-year olds completed a paper-and-pencil version of the Raven SPM (Raven et al., 1985); the children completed this task individually in the classroom (with all participants present).

2.4. Outlier analysis

If performance was less than 55% correct on one of the conditions in the Shifting, and Inhibition tasks (except the Stop-signal task), the results from this particular task were coded missing. The same was done for the Stop-signal task in case the proportion of correct inhibits was lower than 20% or higher than 80%. For all nine EF tasks where RT served as a dependent measure, we performed a two-stage trimming procedure. (1) All incorrect responses, as well as responses that were preceded by an incorrect response, or responses with RTs shorter than 120 ms, or with a latency exceeding the mean by more than 2.5 standard deviations (for each participant and task, separately) were excluded from the RT analyses. This amounted to less than 1.5% of all trials. (2) Extreme outliers at group level per condition were identified by SPSS boxplot procedure (SPSS Inc., 2003). Extreme data (i.e., more than three times the inter-quartile range) were scored as missing. For the Working Memory tasks, we performed the abovementioned step 2 on the accuracy data, because differences

### Table 1

<table>
<thead>
<tr>
<th>EF domain</th>
<th>Task</th>
<th>7-year olds</th>
<th>11-year olds</th>
<th>15-year olds</th>
<th>21-year olds</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Working Memory</strong></td>
<td>Tic Tac Toe Low memory load</td>
<td>83.4 (1.10)</td>
<td>91.1 (0.97)</td>
<td>95.1 (0.98)</td>
<td>96.0 (0.98)</td>
</tr>
<tr>
<td></td>
<td>High memory load</td>
<td>51.0 (1.57)</td>
<td>70.7 (1.58)</td>
<td>87.0 (1.66)</td>
<td>90.1 (1.90)</td>
</tr>
<tr>
<td></td>
<td>Mental Counters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 counters and series 5</td>
<td>78.5 (2.30)</td>
<td>84.5 (1.65)</td>
<td>85.0 (1.66)</td>
<td>86.9 (1.68)</td>
</tr>
<tr>
<td></td>
<td>3 counters and series 5</td>
<td>76.2 (2.44)</td>
<td>83.7 (1.83)</td>
<td>85.6 (1.67)</td>
<td>90.4 (1.71)</td>
</tr>
<tr>
<td></td>
<td>2 counters and series 7</td>
<td>61.0 (2.66)</td>
<td>73.4 (2.23)</td>
<td>78.1 (1.94)</td>
<td>85.4 (2.22)</td>
</tr>
<tr>
<td></td>
<td>3 counters and series 7</td>
<td>70.9 (2.86)</td>
<td>73.6 (2.06)</td>
<td>82.4 (2.06)</td>
<td>86.5 (2.04)</td>
</tr>
<tr>
<td><strong>Running Memory</strong></td>
<td>Block 1</td>
<td>76.6 (1.75)</td>
<td>85.6 (1.48)</td>
<td>86.7 (1.49)</td>
<td>89.5 (1.51)</td>
</tr>
<tr>
<td></td>
<td>Block 2</td>
<td>76.7 (2.10)</td>
<td>82.6 (1.64)</td>
<td>87.4 (1.66)</td>
<td>88.2 (1.80)</td>
</tr>
<tr>
<td></td>
<td>Block 3</td>
<td>75.3 (1.91)</td>
<td>84.5 (1.65)</td>
<td>88.0 (1.50)</td>
<td>90.6 (1.73)</td>
</tr>
<tr>
<td><strong>Shifting</strong></td>
<td>Local-Global Repetition</td>
<td>90.3 (0.76)</td>
<td>92.9 (0.58)</td>
<td>93.5 (0.58)</td>
<td>95.1 (0.59)</td>
</tr>
<tr>
<td></td>
<td>Alternation</td>
<td>88.9 (0.75)</td>
<td>92.0 (0.77)</td>
<td>93.5 (0.77)</td>
<td>95.8 (0.78)</td>
</tr>
<tr>
<td></td>
<td>Dots-Triangles Repetition</td>
<td>78.5 (1.06)</td>
<td>80.1 (0.76)</td>
<td>92.7 (0.77)</td>
<td>94.3 (0.97)</td>
</tr>
<tr>
<td></td>
<td>Alternation</td>
<td>71.4 (1.18)</td>
<td>79.7 (1.07)</td>
<td>84.8 (1.11)</td>
<td>87.8 (1.12)</td>
</tr>
<tr>
<td></td>
<td>Smiling Faces Repetition</td>
<td>79.2 (1.25)</td>
<td>85.6 (1.11)</td>
<td>87.8 (1.12)</td>
<td>93.1 (1.35)</td>
</tr>
<tr>
<td></td>
<td>Alternation</td>
<td>73.8 (1.20)</td>
<td>79.9 (1.07)</td>
<td>82.8 (1.09)</td>
<td>88.9 (1.13)</td>
</tr>
<tr>
<td><strong>Inhibition</strong></td>
<td>Stop-signal Correct inhibits</td>
<td>46.0 (0.95)</td>
<td>48.2 (0.69)</td>
<td>50.0 (0.85)</td>
<td>47.9 (0.83)</td>
</tr>
<tr>
<td></td>
<td>Eriksen Flankers Congruent</td>
<td>96.2 (0.39)</td>
<td>97.8 (0.79)</td>
<td>98.6 (0.46)</td>
<td>99.2 (0.40)</td>
</tr>
<tr>
<td></td>
<td>Incongruent</td>
<td>90.8 (0.76)</td>
<td>93.3 (0.77)</td>
<td>93.9 (0.78)</td>
<td>94.1 (0.78)</td>
</tr>
<tr>
<td></td>
<td>Stroop Interference</td>
<td>80.8 (0.90)</td>
<td>90.8 (0.95)</td>
<td>93.1 (0.97)</td>
<td>96.6 (0.98)</td>
</tr>
<tr>
<td><strong>Complex</strong></td>
<td>WCST Categories (%)</td>
<td>2.90 (0.220)</td>
<td>3.51 (0.173)</td>
<td>4.20 (0.146)</td>
<td>5.1 (0.182)</td>
</tr>
<tr>
<td></td>
<td>Perseverative errors (%)</td>
<td>32.20 (1.76)</td>
<td>24.52 (1.38)</td>
<td>22.08 (1.33)</td>
<td>15.56 (1.45)</td>
</tr>
<tr>
<td></td>
<td>Conceptual level (%)</td>
<td>45.27 (2.45)</td>
<td>53.75 (1.93)</td>
<td>60.35 (1.85)</td>
<td>64.26 (2.03)</td>
</tr>
<tr>
<td></td>
<td>ToL Perfect solutions (%)</td>
<td>6.11 (0.10)</td>
<td>8.11 (0.08)</td>
<td>12.56 (0.08)</td>
<td>14.91 (0.09)</td>
</tr>
<tr>
<td></td>
<td>Additional moves (%)</td>
<td>14.08 (0.87)</td>
<td>17.27 (0.69)</td>
<td>12.44 (0.68)</td>
<td>10.80 (0.73)</td>
</tr>
<tr>
<td></td>
<td>Planning time (s)</td>
<td>5.78 (0.49)</td>
<td>4.20 (0.39)</td>
<td>6.22 (0.38)</td>
<td>7.55 (0.41)</td>
</tr>
<tr>
<td><strong>Basic Speed</strong></td>
<td>Stroop Pure blocks</td>
<td>97.8 (0.20)</td>
<td>99.6 (0.20)</td>
<td>99.8 (0.20)</td>
<td>99.8 (0.20)</td>
</tr>
</tbody>
</table>

Standard errors are presented within parentheses.
between conditions were more prominent for accuracy rather than RT. Accuracy scores were transformed by square root. Failure to complete the first category of the WCST was scored as missing. Only correctly solved items of the ToL were analyzed. Finally, failure to attend the second test session and equipment malfunction resulted in the loss of four 11-year olds and five 15-year olds. Missing values amounted to 11% of all observations for the 7-year olds, 6% for the 11-year olds, 7% for the 15-year olds, and 5% for the 21-year olds.

3. Results

We performed three sets of analyses. The first set included analyses of variance to assess developmental trajectories for each task. The second set included multi-group confirmatory factor analysis, to assess when the latent components Working Memory, Shifting, and Inhibition reached adult levels. The third set included regression analyses to assess the contribution of the latent factors to the performance on the WCST and ToL.

3.1. Analysis of variance

Proportions correct and median RTs of the EF tasks in the domains of Working Memory, Shifting and Inhibition, and the performance indices of the WCST and ToL are presented in Tables 1 and 2 for each age group separately. Below, we report the accuracy results for the Working Memory tasks and the RT results for the Shifting and Inhibition tasks, because these were the dependent measures for which Condition and Age effects were most prominent. For the RT effects in the Working Memory tasks and the accuracy effects in the Shifting and Inhibition tasks the Condition × Age interactions were either non-significant or in the same direction. Previous research suggested that developmental differences (i.e., comparing children to adults) in task performance can be accounted for by general speeding (e.g., Corella & Hale, 1994; Kail, 1991; Span et al., 2004). Therefore, in the present study we corrected the RT results for within-group individual differences in basic processing speed. The median RTs obtained from the pure blocks of the Stroop task were taken to provide an estimate of the basic processing speed.

Working Memory: The accuracy scores on the Working Memory tasks were submitted to an ANOVA with Condition (two levels) as within-subjects variable and Age Group (four levels) as between-subjects variable. The ANOVA of the Tic Tac Toe task scores resulted in a main effect of Condition, $F(3, 351) = 300.21, p < 0.001$, $\eta^2_p = 0.46$; Age Group, $F(3, 351) = 116.86, p < 0.001$, $\eta^2_p = 0.50$; and a Condition × Age Group interaction, $F(3, 351) = 44.26, p < 0.001$, $\eta^2_p = 0.27$. Post hoc Bonferroni comparisons for the difference score of the Condition effect showed that the difference between the high and low memory load conditions was larger in 7-year olds than in 11-year olds, and larger in 11-year olds than in 15-year olds. The 15-year olds did not differ from 21-year olds.

A similar ANOVA for the Mental Counters resulted in a main effect of Condition, $F(1, 283) = 67.84, p < 0.001$, $\eta^2_p = 0.20$; Age Group, $F(3, 283) = 21.31, p < 0.001$, $\eta^2_p = 0.22$; and a Condition × Age Group interaction, $F(3, 283) = 6.02, p < 0.001$, $\eta^2_p = 0.18$, and a Condition × Age Group interaction, $F(3, 351) = 21.31, p < 0.001$, $\eta^2_p = 0.34$. Post hoc Bonferroni comparisons for the difference score of the Condition effect showed that performance increased across blocks the difference between series consisting of seven stimuli was larger in 7-year olds than in 11-year olds, and larger in 11-year olds than in 15-year olds. The 15-year olds did not differ from 21-year olds.

Finally, the ANOVA for the Running Memory task only yielded a main effect of Age Group, $F(3, 354) = 16.90, p < 0.001$, $\eta^2_p = 0.19$, showing that performance increased with age. The main effect of Load was not significant, and did not interact with Age Group ($F(3, 354) = 1.13, p = 0.34$). Post hoc Bonferroni comparisons of the accuracy results indicated that accuracy in 7-year olds was lower than in 11-year olds, and 11-year olds did not differ from 15-year olds (who in turn did not differ from 21-year olds).

To summarize, on two of the three Working Memory tasks (the Tic Tac Toe task and the Mental Counters task), adult level of performance was not reached until the age of 15. The Running Memory task showed that accuracy increased with advancing age but in the absence of an interaction with task memory this developmental trend is difficult to interpret.

Shifting: The median RTs on the Shifting tasks were submitted to a MANCOVA with the basic processing speed entered as covariate, Condition (two levels) as within-subjects variable, and Age Group (four levels) as between-subjects variable. The MANCOVA for the Local–Global task resulted in a main effect of Age Group, $F(3, 314) = 30.92, p < 0.001$, $\eta^2_p = 0.17$, and a Condition × Age Group interaction, $F(3, 314) = 7.24, p < 0.001$, $\eta^2_p = 0.07$.

The median RTs on the Stroop task were submitted to a multi-group confirmatory factor analysis (M-FAC), and for the control factor Basic Speed. The median RTs on the Stroop task were submitted to a multi-group confirmatory factor analysis (M-FAC), and for the control factor Basic Speed. The median RTs on the Stroop task were submitted to a multi-group confirmatory factor analysis (M-FAC), and for the control factor Basic Speed. The median RTs on the Stroop task were submitted to a multi-group confirmatory factor analysis (M-FAC), and for the control factor Basic Speed.

Table 2: Median response latencies (ms) for the EF tasks in the domains of Shifting and Inhibition, and for the control factor Basic Speed.

<table>
<thead>
<tr>
<th>EF domain</th>
<th>Task</th>
<th>7-year olds</th>
<th>11-year olds</th>
<th>15-year olds</th>
<th>21-year olds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shifting</td>
<td>Local–Global Repetition</td>
<td>874 (12)</td>
<td>535 (10)</td>
<td>422 (11)</td>
<td>393 (11)</td>
</tr>
<tr>
<td></td>
<td>Alternation</td>
<td>1088 (18)</td>
<td>625 (15)</td>
<td>477 (15)</td>
<td>448 (16)</td>
</tr>
<tr>
<td></td>
<td>Don–Triangle Repetition</td>
<td>1164 (23)</td>
<td>810 (18)</td>
<td>623 (18)</td>
<td>588 (20)</td>
</tr>
<tr>
<td></td>
<td>Alternation</td>
<td>1677 (40)</td>
<td>1200 (31)</td>
<td>819 (30)</td>
<td>817 (33)</td>
</tr>
<tr>
<td>Smiling Faces</td>
<td>Repetition</td>
<td>1071 (23)</td>
<td>726 (19)</td>
<td>591 (19)</td>
<td>518 (21)</td>
</tr>
<tr>
<td></td>
<td>Alternation</td>
<td>1567 (38)</td>
<td>1088 (31)</td>
<td>818 (31)</td>
<td>703 (33)</td>
</tr>
<tr>
<td>Inhibition</td>
<td>Stop-signal RT</td>
<td>269 (8)</td>
<td>238 (7)</td>
<td>217 (7)</td>
<td>201 (8)</td>
</tr>
<tr>
<td></td>
<td>Eriksen Flanker Congruent</td>
<td>815 (10)</td>
<td>523 (9)</td>
<td>415 (9)</td>
<td>394 (9)</td>
</tr>
<tr>
<td></td>
<td>Incongruent</td>
<td>950 (12)</td>
<td>587 (10)</td>
<td>466 (11)</td>
<td>444 (11)</td>
</tr>
<tr>
<td>Stroop</td>
<td>Interference</td>
<td>776 (8)</td>
<td>612 (8)</td>
<td>548 (8)</td>
<td>514 (8)</td>
</tr>
<tr>
<td>Basic Speed</td>
<td>Stroop Pure blocks</td>
<td>579 (6)</td>
<td>424 (6)</td>
<td>356 (6)</td>
<td>341 (6)</td>
</tr>
</tbody>
</table>

Standard errors are presented within parentheses.
The ANCOVA for the Dots-Triangles task resulted in a main effect of Condition, \( F(1, 315) = 4.37, p = 0.037, \) MSE = 20364.45, \( \eta^2_p = 0.02; \) Age Group, \( F(3, 315) = 17.21, \) \( p < 0.001, \) MSE = 10650.38, \( \eta^2_p = 0.14; \) and a Condition × Age Group interaction, \( F(3, 315) = 10.53, p < 0.001, \) MSE = 20364.45, \( \eta^2_p = 0.09. \)

The ANCOVA for the scores on the Smiling Faces task also resulted in a main effect of Age Group, \( F(3, 306) = 11.00, \) \( p < 0.001, \) MSE = 74719.36, \( \eta^2_p = 0.10; \) and a Condition × Age Group interaction, \( F(3, 306) = 4.95, p = 0.002, \) MSE = 21986.26, \( \eta^2_p = 0.05. \) Post hoc Bonferroni comparisons for the shift costs (i.e., the difference in RT between alternation trials and repetition trials in the mixed block) showed that for the Local-Global, the Dots-Triangles, and the Smiling Faces the shift cost was larger in 7-year olds than in 11-year olds, and larger in 11-year olds than in 15-year olds. However, 15-year olds did not differ from 21-year olds.

To summarize, for all three Shifting tasks adult level of performance was not reached until the age of 15.

Inhibition. The ANOVA of the Stop-signal task data resulted in a main effect of Age Group, \( F(3, 283) = 3.94, p = 0.009, \) MSE = 4043.96, \( \eta^2_p = 0.040. \) Post hoc Bonferroni comparisons showed that the SSRT was larger for 7-year olds than for 11-year olds, larger for 11-year olds than for 15-year olds. However, the 15-year olds did not differ from 21-year olds.

The median RTs on the Eriksen Flankers and the Stroop task were submitted to a MANCOVA with the basic processing speed entered as covariate, Condition (two levels) as within-subjects variable, and Age Group (four levels) as between-subjects variable.

The ANCOVA of the Eriksen Flankers task data resulted in a main effect of Condition, \( F(1, 326) = 9.90, p = 0.002, \) MSE = 1306.78, \( \eta^2_p = 0.03; \) Age Group, \( F(3, 326) = 50.70, \) \( p < 0.001, \) MSE = 12400.09, \( \eta^2_p = 0.32; \) and a Condition × Age Group interaction, \( F(3, 326) = 12.73, p < 0.001, \) MSE = 1306.78, \( \eta^2_p = 0.11. \) Post hoc Bonferroni comparisons of the difference score of the Condition effect showed that the difference between congruent and incongruent trials was larger for 7-year olds than for 11-year olds, larger for 11-year olds than for 15-year olds. Again, 15-year olds did not differ from 21-year olds.

Finally, the ANCOVA of the interference RTs on the Stroop task resulted in a main effect of Age Group, \( F(3, 333) = 6.97, \) \( p < 0.001, \) MSE = 2593.43, \( \eta^2_p = 0.06. \) Post hoc Bonferroni comparisons showed that the interference RT was larger in 7-year olds than in 11-year olds, larger in 11-year olds than in 15-year olds, and larger in 15-year olds than in 21-year olds.

To summarize, on two of the Inhibition tasks (the Stop-signal task and the Eriksen Flankers task), adult level of performance was not reached until the age of 15, whereas for the third task (the Stroop task) performance continued to improve into young-adulthood.

Wisconsin Card Sorting Task. The proportion of conceptual level responses, the number of categories completed, and the proportion of perseverative errors on the WCST were submitted to ANOVAs with Age Group as between-subjects variable. All main effects were significant, \( F(3, 346) = 13.96, p < 0.001, \) MSE = 360.48, \( \eta^2_p = 0.11; \) \( F(3, 346) = 23.82, p < 0.001, \) MSE = 2.90, \( \eta^2_p = 0.17; \) and \( F(3, 346) = 18.46, p < 0.001, \) MSE = 185.25, \( \eta^2_p = 0.14, \) respectively. Post hoc Bonferroni comparisons showed that the proportion of conceptual level responses was smaller in 7-year olds than in 11-year olds, and smaller in 11-year olds than in 15-year olds. The 15-year olds, however, did not differ from 21-year olds. The number of categories completed and the proportion of perseverative errors were larger in 7-year olds than in 11-year olds, larger in 11-year olds than in 15-year olds, and larger in 15-year olds than in 21-year olds.

To summarize, for one of the three measures of the WCST (the proportion of conceptual level responses), adult level of performance was not reached until the age of 15, whereas for the other two measures (i.e., the number of categories completed and the proportion of perseverative errors) performance continued to develop into young-adulthood.

Tower of London. The number of perfect solutions on the ToL, the number of additional moves, and planning time were submitted to ANOVAs with Age Group as between-subjects variable. All main effects were significant, \( F(3, 348) = 41.02, p < 0.001, \) MSE = 31.55, \( \eta^2_p = 0.26, \) \( F(3, 348) = 18.21, p < 0.001, \) MSE = 47.38, \( \eta^2_p = 0.14; \) \( F(3, 348) = 12.17, p < 0.001, \) MSE = 14.83, \( \eta^2_p = 0.10. \) Post hoc Bonferroni comparisons showed that the number of perfect solutions was smaller in 7-year olds than in 11-year olds, smaller in 11-year olds than in 15-year olds, and smaller in 15-year olds than for 21-year olds. Additionally, the number of additional moves was larger in 7-year olds than in 11-year olds, larger in 11-year olds than in 15-year olds, while 15-year olds did not differ from 21-year olds. Finally, planning time was longer in 7-year olds than in 11-year olds, longer in 11-year olds than in 15-year olds. The 15-year olds however did not differ from 21-year olds.

To summarize, for two of the three measures of the ToL, (i.e., the number of additional and planning time), adult level of performance was not reached until the age of 15, whereas for the third measure (i.e., the number of perfect solutions) performance continued to develop into young-adulthood.

3.2. Structural equation modeling

3.2.1. Model specification

Structural equation modeling (SEM) was used to formulate and test models of EF explicitly in multi-group analyses. The details of the fitting procedure are described in Appendix A. Working Memory was indexed by the mean accuracy on the Tic Tac Toe and of the Mental Counters tasks. The results of the Running Memory task were not included because of the lack of a task load effect. Shifting was indexed by the median RT on shift trials on the Local-Global task, the Dots-Triangles task, and the Smiling Faces. The three manifest Inhibition variables were the median RT on incongruent trials at the Eriksen Flankers task, the mean latency of the SSRT at the Stop-signal task, and the median RT on interference trials at the Stroop task. Basic processing speed, as indexed by median RT of the pure blocks
of the Stroop task, was used as a control factor. As the three Shifting tasks and the Eriksen Flankers and Stop-signal tasks are RT tasks, these were specified to load on the basic processing speed factor to correct for within-group individual differences in processing speed. Note that the Stop-signal RT task was not specified to load on the basic processing speed factor, because the horse-race model used to estimate SSRT assumes that SSRT and Go RT are independent processes\(^2\) (Logan, 1994).

We first fitted a model to compare EF task performance across groups. We refer to this model as the "partial measurement model\(^3\)\(^3\), as it involved relating the indicators of Working Memory and Shifting to the common Working Memory and Shifting factors. The indicators of Inhibition were not related to a common Inhibition factor, as the correlations (ranging from −0.12 to 0.35) between the Inhibition tasks were such that a common Inhibition factor could not be extracted. Rather than choosing to include one of the Inhibition tasks in the partial measurement model, we decided to include all three tasks as measures of Stop-signal, Eriksen Flankers, and Stroop inhibition, respectively. Thus, the partial measurement model incorporated two latent variables (Working Memory and Shifting), three manifest Inhibition variables, and one control factor.

The model was fitted to the data of all age groups, simultaneously in a multi-group model, to assess whether the indicators of Working Memory and Shifting actually measure the same components across age, and whether observed mean differences are attributable to the common factor mean differences, i.e., we investigated factorial invariance (Meredith, 1993). In fitting the partial measurement model, we introduced various substantive and identifying constraints. We compared the groups by fitting a number of models. First, in order to establish that we measured the same components in all age groups, we defined the factor model M1 within each age group (i.e., we established configural invariance; a path diagram of this model is shown in Fig. 1). In this model, the configuration of factor loadings was identical for all age groups, but parameters were free to vary across groups. This model provided a baseline, by which we evaluated more constrained models. Secondly, we established measurement invariance, that is, we constrained the regression coefficients (factor loadings) of the observed indicators of the common factors to be identical in all age groups. (Note that this does not imply that the group means are identical.) This constrained model was coined M2. Finally, we proceeded to assess the developmental trajectories of EF. Thus, we fitted a model, coined M3, in which the 21-year olds served as a baseline, and we estimated the mean differences of the 7-, 11-, and 15-year olds relative to the 21-year olds. With respect to the Working Memory and Shifting variables, model M3 specified strong factorial invariance (Meredith, 1993). This implies that the observed mean differences between the groups with respect to Working Memory and Shifting are a function of the mean differences of the common factors Working Memory and Shifting. From the perspective of linear regression, model M3 specified that across the groups, the regression coefficients and intercepts in the regression of the observed variables on the factors, were equal.

\(^2\) The data obtained in this study are in accordance with the horse-race model that argues that performance on the Stop-signal task reflects a race between activation and Inhibition processes (see Logan, 1994). An ANCOVA on Stop-signal RT, entering Go RT as a covariate yielded a significant main effect of Age Group \(F(3, 321) = 30.58, p < 0.001\), with Age Group explaining 23% of the variance. As predicted by the horse-race model, it was observed that responses on stop trials that escape Inhibition were executed faster than average Go-signal responses in all age groups, \(F(1, 325) = 825.21, p < 0.001\); mean RT of failed inhibitions was 488 ms (S.D. = 121), and mean Go RTs was 570 ms (S.D. = 144).

\(^3\) Task reliabilities were satisfactory. The reliabilities of the tasks (as indexed by indicators of the partial measurement model) were calculated with Cronbach’s \(\alpha\) (odd–even). Across age groups, the reliabilities ranged from 0.63 to 0.91.
To avoid computational problems that may arise when the variances of the variables differ greatly (as is the case here), we rescaled the measurement variables so that the variances of the young-adult group all equaled about 20 (the choice of the value is obviously arbitrary), and we applied the same linear transformation in the other groups. Rescaling the variables does not alter the group differences in any meaningful way, i.e., the group differences in means and covariance structure are retained, albeit on a linearly transformed scale. Subsequently, the PRELIS program (Jöreskog & Sörbom, 1999) was used to transform the variables by calculating normal scores within each age group. This transformation ensured that skewness and kurtosis of the variables were as close as possible to the values expected under normality (i.e., 0 and 3, respectively). The transformation does not alter the means or the variances of the variables. Because of missingness, we employed raw data likelihood estimation (Schafer & Graham, 2002; Wothke, 2000). That is, rather than fitting the model to the sample summary statistics of the four groups, we fitted the models directly to the data. In doing so, we assumed that the data are missing completely at random (MCAR; see Schafer & Graham, 2002). Judging by the formal test (SPSS Inc., 2003), MCAR appears to hold in the present groups, with the possible exception of the 21-year olds. Here, however, the value of the $\chi^2$ test statistic was hardly cause for alarm (7-year olds: $\chi^2(414) = 406.31$, $p = 0.59$; 11-year olds: $\chi^2(498) = 535.97$, $p = 0.31$; 15-year olds: $\chi^2(473) = 496.63$, $p = 0.22$; 21-year olds: $\chi^2(362) = 419.90$, $p = 0.02$).

3.2.2 Model fitting

Model fitting was done using LISREL (Jöreskog & Sörbom, 1999) and Mx (Neale, Boker, Xie, & Maes, 2003). As we applied raw data likelihood estimation, the usual array of goodness of fit measures are not available. We therefore focused mainly on the $\chi^2$ difference statistics (i.e., loglikelihood ratio tests) to investigate the models directly to the data. In doing so, we assumed that the data are missing completely at random (MCAR; see Schafer & Graham, 2002). First, to establish configural invariance, we fitted the factor model M1. This model served as the baseline model. We introduced (and established) measurement invariance (M2). Following the imposition of the M2 equality constraints, and compared to M1, we observed a $\chi^2$ difference of 51.45 (d.f. = 9; $p < 0.001$). Inspection of the standardized residuals of the covariance structure (between the three oldest age groups), and the standardized residuals for the means between 1.7 and 2.8 in all age groups) suggest that M3 is acceptable. The correlations between the common factors of Working Memory and Smoking explained by the common factors of Working Memory and Shifting and the manifest inhibition measures for each of the four age groups.

Table 4: Correlations of the latent factors Working Memory and Shifting and the manifest inhibition measures for each of the four age groups.

<table>
<thead>
<tr>
<th>Task</th>
<th>7-year olds</th>
<th>11-year olds</th>
<th>15-year olds</th>
<th>21-year olds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working Memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tic Tac Toe (%)</td>
<td>46</td>
<td>7</td>
<td>20</td>
<td>47</td>
</tr>
<tr>
<td>Mental Counters (%)</td>
<td>75</td>
<td>12</td>
<td>25</td>
<td>51</td>
</tr>
<tr>
<td>Shifting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local-Global (%)</td>
<td>28</td>
<td>50</td>
<td>64</td>
<td>67</td>
</tr>
<tr>
<td>Smiling Faces (%)</td>
<td>29</td>
<td>45</td>
<td>41</td>
<td>60</td>
</tr>
<tr>
<td>Basic Speed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroop pure blocks (%)</td>
<td>83</td>
<td>77</td>
<td>86</td>
<td>76</td>
</tr>
</tbody>
</table>

Table 2: Maximum likelihood estimates and standard errors (between parentheses) of differences in latent factor means of Working Memory and Shifting, the three manifest inhibition variables and the control factor Basic Speed.

<table>
<thead>
<tr>
<th></th>
<th>$\omega_1^2$</th>
<th>$\omega_2^2$</th>
<th>$\omega_3^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working Memory</td>
<td>–4.83 (0.72)*</td>
<td>–1.57 (0.28)*</td>
<td>–0.49 (0.19)*</td>
</tr>
<tr>
<td>Shifting</td>
<td>1.76 (0.71)*</td>
<td>1.17 (0.33)*</td>
<td>0.09 (0.10)</td>
</tr>
<tr>
<td>Stop-signal Inhibition</td>
<td>1.54 (0.26)*</td>
<td>0.67 (0.17)*</td>
<td>0.29 (0.17)</td>
</tr>
<tr>
<td>Eriksen Flankers Inhibition</td>
<td>6.81 (0.99)*</td>
<td>1.97 (0.40)*</td>
<td>0.23 (0.21)</td>
</tr>
<tr>
<td>Stroop Inhibition</td>
<td>–1.09 (0.74)</td>
<td>–0.15 (0.31)</td>
<td>0.28 (0.17)</td>
</tr>
<tr>
<td>Basic Speed</td>
<td>7.92 (0.90)*</td>
<td>2.88 (0.37)*</td>
<td>0.62 (0.18)*</td>
</tr>
</tbody>
</table>

* Significant difference relative to 21-year olds.
The maximum likelihood estimates and standard errors of the common factor mean differences (i.e., in the notation of the appendix, the parameters in the vectors $\alpha^*_1$, $\alpha^*_2$, and $\alpha^*_3$) are presented in Table 5. As explained in Appendix A, these parameters represent differences in common factor means relative to the young-adult group. In other words, the observed mean differences between the groups with respect to Working Memory and Shifting are a function of the mean differences of the common factors Working Memory and Shifting. The standard errors of the estimates indicated that the 7-year olds and the 11-year olds differed from the 21-year olds with respect to Working Memory, Shifting, the Stop-signal task, the Eriksen Flankers task, and basic processing speed. The 15-year olds differed from the 21-year olds with respect to Working Memory and basic processing speed. These developmental patterns are apparent in Fig. 2 that provides plots of the decomposition of means of each EF relative to the young-adult group.

3.2.3. Predicting performance on the WCST and the ToL

In a number of separate analyses the complex EF tasks (i.e., the ToL and WCST tasks) were regressed on the latent variables in the partial measurement model. Again we performed 4-group analyses, as we wished to assess possible age-related changes in the regression equations. The analyses were done in three steps. The details of the regression analyses are presented in Appendix B. First, we tested the hypothesis that the regression coefficients differed between age groups (Model A). Second, we tested the hypothesis that the regression coefficients were equal across the groups (Model B). Note that across-group equality of the regression coefficients does not necessarily imply that the amount of explained variance is equal over the age groups. Third, we investigated whether the regression coefficients were significantly greater than zero (Model C).

The results of the regression analyses of the WCST measures indicated that for the number of categories achieved and the proportion of conceptual level responses, Model C did not fit significantly worse than Models B and A. This finding suggests that the common factors Working Memory and Shifting and the manifest Inhibition variables do not contribute substantially to the number of categories and in the proportion of conceptual level responses of the WCST. In the analysis of the proportion of perseverative errors, we found that the difference between Models A and B was not significant ($\chi^2$ difference $= 17.85; \text{d.f.} = 15$), but the difference between Models C and B were ($\chi^2$ difference $= 15.30; \text{d.f.} = 5$, $p < 0.010$), indicating that the regression coefficients were equal across groups, but not zero. The upper panel of Table 6 presents the estimates of the regression coefficients with 95% confidence intervals. It can be seen that in all age groups Working Memory related strongly to the proportion of perseverative errors on the WCST. For the mean number of additional moves and planning time on the ToL, we failed to fit a reasonable model. Regression analysis of the proportion of perfect solutions indicated that the difference between Models A and B was not significant ($\chi^2$ difference $= 16.60; \text{d.f.} = 15$). The difference between Models C and B was significant ($\chi^2$ difference $= 15.30; \text{d.f.} = 5$, $p < 0.010$), indicating that the regression coefficients were equal across groups, but not zero. The upper panel of Table 6 presents the estimates of the regression coefficients with 95% confidence intervals.
and A was significant ($\chi^2$ difference = 21.47, d.f = 5, $p < 0.001$), indicating that the regression coefficients differed across groups and do not equal zero. Calculating the 95% confidence intervals of the regression coefficients in the youngest three age groups proved problematic. In the lower panel of Table 6, it is shown that Stroop performance was a significant predictor of the proportion of perfect solutions. Thus, in 21-year olds, the ability to suppress a pre-potent response, as on the Stroop task, seems to play an important role in the efficiency of solving TOL items.

4. Discussion

In this study, we examined the developmental trajectories of three frequently postulated EF components, Working Memory, Shifting, and Inhibition, in relation to performance on standard neuropsychological EF tasks, the WCST and the TOL. In so doing, we adopted both standard analyses and multi-group latent variable modeling. The latter enabled us to model the structure of the underlying latent variables across age groups, thus going beyond the usual analysis of observed developmental trends in the performance on the task battery.

4.1. Analysis of manifest task performance

First, we analyzed the data using conventional analyses of variance. The developmental trajectories for Working Memory, Shifting and Inhibition that emerged from the task battery were consistent with several developmental studies indicating that developmental trends of different EF component processes are distinct (e.g., Brocki & Bohlin, 2004; Klenberg et al., 2001; Lehto, 2004; Luciana et al., 2005; Luna et al., 2004; Welsh et al., 2003; Gathercole et al., 1999; Luna et al., 2005; Luna et al., 2004). In contrast, the current version of the Stroop task yielded only a weak developmental trend from early childhood into adulthood. Previous studies using a child-
A major aim of the current study was to assess the developmental change in the latent components Working Memory, Shifting, and Inhibition. To address this question, we adopted the SEM approach advanced by Miyake et al. (2000), who established that these EF components are distinct, but moderately correlated, at the latent level. Following this study, we employed multi-group structural equation modeling to evaluate an a priori specified measurement model of EF that included EF components as latent factors underlying age-related changes in EF task performance.

We had to depart from the Miyake et al. approach in that our measurement model included two latent variables, Working Memory and Shifting, three manifest Inhibition variables, and one control factor (basic processing speed). The reason for separating the Inhibition variable into three manifest Inhibition variables was based on preliminary analyses, which revealed low and, in some cases, negative correlations between Inhibition measures. Thus, these measures were retained as separate variables rather than aggregating them in a single common factor (for similar results see Van der Sluis, De Jong, & Van der Leij, 2006). Although several authors have argued that 'inhibition' comes in many varieties depending on the exact paradigm that is used (e.g., Dempster, 1993; Harmishfeger, 1996; Kipp, 2005; Logan, 1994; Nigg, 2000), the current observation of disparate developmental trends in the ability to inhibit speeded motor responses was unexpected. Three paradigms were selected to provide indicators of the ability to inhibit motor responses. In the Eriksen Flankers task, subjects are required to execute a speeded motor response to the central stimulus of the stimulus array and to inhibit the competing response that is activated by the flanker stimuli on incongruent trials. Similarly, in the Stroop task, subjects are required to respond as quickly and accurately as possible to the color of the stimulus, but to inhibit the activated response when the color and orientation of the stimulus does not match on interference trials. Finally, in the Stop-signal task, subjects have to execute a speeded response to the go stimulus, but must inhibit this response on Stop-signal trials. An important similarity is that all three paradigms require a form of inhibition, which concerns the control of competing responses. In this regard, the current finding of low correlations between inhibition measures is difficult to explain. Possibly, subtle differences in task format may have contributed to the dissimilarities across measures. That is, the Eriksen Flankers task is a two-choice task requiring a speeded response on all trials, whereas the Stroop task and the Stop-signal task require a speeded response on some trials, but inhibition of response on other trials. In addition, the Stroop task differs from the Stop-signal task in that a conjunction of stimulus features (color and orientation) determines whether or not a response has to be executed in the Stroop task, whereas the instruction to execute or inhibit a motor response is provided by different signals in the Stop-signal task.

The SEM model presented in Fig. 1 provided an acceptable description of EF task performance. The SEM analyses indicated that the latent factors Working Memory and Shifting, and the manifest Inhibition variables are clearly separable, as their correlations are quite modest. The correlations among the latent and manifest variables constituting the measurement model were found to be rather low. This result provides additional support for the non-unitary nature of EF (Fisk & Sharp, 2004; Miyake et al., 2000; Stuss et al., 1995). Most importantly, however, given the present aims, the current study added a developmental dimension to the EF model fitting. The model presented in Fig. 1 provides an adequate description of EF task performance in each age group. The analyses revealed developmental trajectories that, by and large, are in accord with the ANOVA results. The 7-year olds and 11-year olds differed from 21-year olds in Working Memory and Shifting, and on two of the Inhibition tasks (i.e., the Stop-signal and Eriksen Flankers tasks, but not the Stroop task). Finally, the child groups differed from 21-year olds in basic processing speed. The 15-year olds differed from 21-year olds with respect to Working Memory and basic processing speed. The overall pattern of results suggests that Shifting and performance on the Stop-signal and Eriksen Flankers tasks have reached mature levels by adolescence, while Working Memory and basic processing speed follow a more protracted course of development into young-adulthood. Thus, although analysis of variance of the separate Working Memory tasks suggests that 15-year olds do not differ from 21-year olds, when examining Working Memory at a latent level there is still maturation after the age of 15.

The present modeling approach has two important advantages over the standard analysis of variance approach. First, the analy-
ses of trends were conducted at the level of the latent theoretical variables, in the case of Shifting and Working Memory (but sadly not in the case of Inhibition). This means that the problem of task impurity is at least somewhat reduced. Equally important is the fact that the analyses of trends were conducted subject to strong factorial invariance (see Appendix A). The restrictions associated with this model go a long way to ensuring that we are actually measuring the same hypothetical constructs across different age groups. The standard analysis of variance is based on the tacit assumption that the latent variables are identical across the age groups. However, it remains uncertain whether a given mean age difference is actually attributable to development in the same latent variable. By establishing strong factorial invariance first, we can interpret the observed mean differences in terms of the latent mean differences in the latent variables that the tasks purport to measure.

When regressing the WCST performance (as indexed by perseveration errors) on the latent variables in partial measurement model, we found that in all groups Working Memory related significantly to the proportion of perseverative errors. Moreover, the regression coefficients were equal across groups. These results differ from those of Miyake et al. (2000), who showed that Shifting, rather than Working Memory predicted perseverative errors on the WCST. In earlier studies, it was found that patients who have damage to dorsolateral PFC experience difficulties with both Working Memory (e.g., Bechara, Damasio, Tranel, & Anderson, 1998) and the WCST (e.g., Barceló & Knight, 2002). Thus, it is likely that Working Memory is an important contributor to WCST performance as indexed by perseveration errors. The differences in findings between this study and the study by Miyake et al. (2000) may be associated with differences in task instructions. That is, in the current study we did not give instructions about the switching of sorting rules, as required by the Heaton et al. (1993) manual, whereas Miyake et al. (2000) informed participants that sorting principles would switch from time to time. Recently, Stuss et al. (2000) in a patient study showed that test administration (e.g., in advance verbalizing the sorting rules) played a role in the sensitivity of the WCST.

Finally, when regressing the proportion of perfect solutions on the ToL on the latent variables in the partial measurement model, we found that in the 21-year olds the proportion of perfect solutions on the ToL was significantly predicted by Stroop inhibition. The $\chi^2$ difference indicated that the regression coefficients differed across groups. However, calculating the 95% confidence intervals of the regression coefficients in the youngest three age groups proved computationally problematic, which may be due to missingness (see below). The finding that Stroop performance plays a role in ToL performance is largely consistent with Miyake et al. (2000), who found that Inhibition predicted ToL performance, given “that no specific instructions for strategies were given and many people are likely to use a perceptual strategy to solve the ToL” (cf. Miyake et al., 2000, p. 87). The ability to suppress a pre-potent response (as indicated by Stroop performance) has been reported earlier to be a significant predictor of ToL performance in a young-adult sample (e.g., Welsh, Satterlee-Carmell, & Stine, 1999). The finding that Inhibition (as measured by Stroop performance) did not predict ToL performance in the three child groups suggests that they may have used a different strategy to solve the task than adults did. This should be a focus of research for future studies.

One limitation of the current study concerns missingness. Missing data were treated as missing, and, in contrast to Miyake et al. (2000), were not imputed. Missingness amounted to a loss of 11% of the 7-year olds data, and about 5% in the older age groups (Miyake et al., 2000 reported 2.2% of the data were affected by their trimming procedures). The data were found missing completely at random, allowing us to fit the models directly to the data using raw data likelihood (rather than to the summary statistics; see Schafer & Graham, 2002). This method of estimation does not involve imputation or any arbitrary treatment of missingness, such as pair-wise or list-wise deletion. Thus, from the point of view of estimation, the missingness does not pose a great problem. However, the missingness is known to affect the power adversely (Dolan, Van der Sluis, & Grasman, 2005). With regard to the prediction of performance on the complex tasks, we unexpectedly found that the regression coefficients in all age groups proved to equal zero, when regressing the proportion of conceptual level responses and the number of categories achieved on the WCST on the common factors in the partial measurement model. In addition, for the mean number of additional moves and planning time on the ToL we failed to fit a reasonable model. This may at least in part be due to the missingness. Certainly, the loss of statistical power will have contributed to the failure to detect significant regression on most measures of WCST and ToL performance (Dolan et al., 2005).

A second limitation of the current study refers to the failure to incorporate the Running Memory task as indicator of the common factor of Working Memory. Thus, the Working Memory factor was defined by only two indicators. Generally, it is desirable to have at least three indicators to define a common factor (e.g., Kline, 1998). However, the identification of the Working Memory factor is still statistically feasible given the presence of other (correlated) common factors. That is, while a two indicator common factor model is not identified in isolation, it is identified in the presence of other common factors, provided the two indicator common factor is correlated with the other common factors, as is the case here. Substantively, this common factor can thus still be interpreted as representing the common influence underlying the two indicators of Working Memory.

4.3. Conclusion

The results of the current study provide support for the non-unitary, multi-faceted nature of EF. Two latent factors: “Working Memory” and “Shifting” were identified that correlated only moderately. This finding is consistent with other studies taking a latent variables approach and showing separable EF components (e.g., Fisk & Sharp, 2004; Lehto et al., 2003; Miyake et al., 2000). Importantly, the current study showed different developmental trends in Working Memory and Shifting. Working Memory continued to develop into young-adulthood whereas Shifting attained mature levels during adolescence. The finding of different developmental trends is in accordance with earlier literature.
showing that Working Memory and Shifting mature at different rates (Cepeda et al., 2001; Crane et al., in press; Gathercole et al., 2004; Huizinga & Van der Molen, in preparation; Luciana et al., 2005; Luna et al., 2004). The measurement model that was fitted to the data also included three manifest Inhibition variables that could not be incorporated into a single latent factor. In spite of the conceptual similarity (i.e., all three paradigms required the inhibition of a motor response), the measures derived from the Eriksen flankers, Stop-signal, and Stroop paradigms showed low or negative correlations. The ability to inhibit the response to the Go stimulus on Stop-signal trials of the Stop-signal task, and the ability to control response competition on incongruent trials of the Eriksen Flanker task improved rapidly to the age of 11. These findings are consistent with the results of previous studies using these inhibition paradigms (Bédard et al., 2002; Ridderinkhof & Van der Molen, 1995; Van den Wildenberg & Van der Molen, 2004; Williams et al., 1999). On the Stop-signal task, the speed of responding on interference trials revealed only little developmental change. However, response accuracy increased rapidly during childhood, and continued to develop into young adulthood.

In sum, the current findings correspond well with previous studies showing developmental improvement in the performance on the WCST and the ToL. Adult levels on the WCST and the ToL were attained between 11 and 15 years, although the proportion of perfect solutions on the ToL increased into young adulthood (Anderson et al., 2005; Baker et al., 2001; Chelune & Baer, 1986; Lehto, 2004; Welsh et al., 1991). In all groups Working Memory related equally and significantly to the proportion of perseverative errors. In addition, the current findings are in accord with previous findings demonstrating developmental improvements in EF component processes and recent findings emerging from brain imaging studies of cognitive development (e.g., Bunge et al., 2002; Casey et al., 1997; Crane, Donohue, Wendelken, Honomichl, & Bunge, in preparation; Durston et al., 2002; Klingberg et al., 2002; Kwon et al., 2002; Luna & Sweeney, 2004). Neurophysiological studies showing that the anatomical development of PFC areas only reaches maturity in young-adulthood support this notion (e.g., Caviness, Kennedy, Richelme, Rademacher, & Filipek, 1996; Chugani, Phelps, & Mazziotta, 1987; Hüttenlocher, 1979; Sowell et al., 2004; Yakovlev & Lecours, 1967; for a review see: Casey, Tottenham, Liston, & Durston, 2005). In future research, it would be of interest to evaluate the biological plausibility of EF component processes and their development by examining anatomically and functionally separate systems within PFC in concert with a latent variables approach such as presented in this study.

Acknowledgements

The authors thank Ingmar Visser, Ellen Hamaker, and Verena Schmittmann for valuable comments on an earlier version of this manuscript. Bert van Beek is gratefully acknowledged for programming the tasks. Thanks to Susanne Stuurman, Manon de Beer, and Esther Kooymans for assistance with the data collection.

Appendix A

The partial measurement model incorporated two latent variables, three manifest Inhibition variables, and one control factor (i.e., Basic Speed). The two latent variables represented the EF components Working Memory and Shifting. The mean accuracy of the Tic Tac Toe task, and the Mental Counters task (denoted \( T_1 \) and \( T_2 \)) were the indicators of Working Memory; the median RT on shift trials on the Local–Global task, the Dots–Triangles task, and the Stroop task (denoted \( T_6 \) through \( T_9 \)) were the indicators of Shifting. The three observed Inhibition tasks were the Eriksen Flankers task (median RT on incongruent trials), the Stop-signal task (mean latencies of the Stop-signal RT), and the Stroop task (median RT on interference trials; denoted \( T_6 \) through \( T_9 \)). The control factor represented basic processing speed, as indexed by median RT of the pure blocks of the Stop-signal task (denoted \( T_5 \)). The three Shifting tasks and the Eriksen Flankers and Stroop tasks were specified to load on the basic processing speed factor to correct for within-group individual differences in basic RT.

The partial measurement model was fitted in a number of steps. First, we defined the factor model within each group by specifying the following regressions.

First, we considered the Basic Speed factor:

\[
T_6 = \tau_{6} + \lambda_{9,6} \times BS + \varepsilon_{6}
\]

Note that the index on the factor loading \( \tau \) refers to the position of the matrix of factor loadings shown below.

Secondly, we considered the Working Memory (WM) factor (Tic Tac Toe task, Mental Counters task):

\[
T_1 = \tau_{1} + \lambda_{11} \times WM + \varepsilon_{1}
\]

\[
T_2 = \tau_{2} + \lambda_{12} \times WM + \varepsilon_{2}
\]

Thirdly, we considered the Shifting (S) factor (Local–Global task, the Dots–Triangles task, and the Smiling Faces task):

\[
T_3 = \tau_{3} + \lambda_{31} \times S + \varepsilon_{3}
\]

\[
T_4 = \tau_{4} + \lambda_{32} \times S + \varepsilon_{4}
\]

\[
T_5 = \tau_{5} + \lambda_{33} \times S + \varepsilon_{5}
\]

Finally, we considered the three Inhibition variables (Eriksen flankers (EFL) task, Stop-signal (SS) task, Stroop (STR) task):

\[
T_7 = \tau_{7} + \lambda_{73} \times EFL + \varepsilon_{7}
\]

\[
T_8 = \tau_{8} + \lambda_{83} \times SS + \lambda_{84} \times BS + \varepsilon_{8}
\]

\[
T_9 = \tau_{9} + \lambda_{93} \times STR + \lambda_{94} \times BS + \varepsilon_{9}
\]

In these regression equations, \( \tau \) represents the intercept, \( \lambda \) the regression coefficient, and \( \varepsilon \) is a residual term. The subscript \( i \) denotes age group \((i=1, 2, 3, 4, \) where \( 1 \) = 7-year olds, \( 2 \) = 11-year olds, \( 3 \) = 15-year olds, \( 4 \) = 21-year olds). It is convenient to express these equations in a matrix expression:

\[
T_i = \mathbf{T}_i + \mathbf{A}_i \mathbf{B} + \mathbf{e}_i,
\]

where \( \mathbf{T}_i \) is the random vector of the nine observed variables, \( \mathbf{T}_i \) the vector of intercepts, and \( \mathbf{e}_i \) is the random vector of mutually
uncorrelated residuals. The random vector \( \eta_i \) contains the common factors, WM, S, and RS, and the Inhibition variables EFL, SS, STR. The 9 × 6 matrix \( A_i \) contains the regression coefficients:

\[
A_i = \begin{pmatrix}
\lambda_{11} & 0 & 0 & 0 & 0 & 0 \\
\lambda_{21} & 0 & 0 & 0 & 0 & 0 \\
0 & \lambda_{32} & 0 & 0 & 0 & \lambda_{36} \\
0 & \lambda_{42} & 0 & 0 & \lambda_{46} & 0 \\
0 & \lambda_{52} & 0 & \lambda_{56} & 0 & 0 \\
0 & 0 & \lambda_{65} & \lambda_{66} & 0 & 0 \\
0 & 0 & 0 & \lambda_{76} & 0 & \lambda_{76} \\
0 & 0 & 0 & 0 & \lambda_{86} & \lambda_{86} \\
0 & 0 & 0 & 0 & 0 & \lambda_{96}
\end{pmatrix}
\]

The residuals were constrained to be mutually uncorrelated, so the covariance matrix of the residuals was diagonal. We denote the covariance matrix of the residuals as \( \Psi_i \), where \( \psi_{ij} \) denotes the variance of the residuals. Note that we did not introduce residual variances for the Inhibition variables EFL, SS, STR, as these were not treated as indicators of Inhibition, but rather simply as indicators of themselves. Finally, we considered the covariance matrix of \( \eta_i \), denoted \( \Theta_i \):

\[
\Theta_i = \begin{pmatrix}
\sigma_{11} & \cdots & \sigma_{19} \\
\sigma_{21} & \cdots & \sigma_{29} \\
\sigma_{31} & \cdots & \sigma_{39} \\
\sigma_{41} & \cdots & \sigma_{49} \\
\sigma_{51} & \cdots & \sigma_{59} \\
\sigma_{61} & \cdots & \sigma_{69} \\
\sigma_{71} & \cdots & \sigma_{79} \\
\sigma_{81} & \cdots & \sigma_{89} \\
\sigma_{91} & \cdots & \sigma_{99}
\end{pmatrix}
\]

Note that \( \sigma_{33}^{2} \) was fixed to zero, because it proved impossible to fit the model in all age groups when \( \sigma_{33}^{2} \) was estimated freely. Assuming \( \eta_i \) and \( \tau_i \) are uncorrelated, we expressed the expected covariance matrix in group \( i \), \( \Sigma_i \), as follows:

\[
\Sigma_i = A_i \Theta_i A_i' + \Theta_i
\]

An important aim of the present SEM was to incorporate the means, as this allowed us to study group differences with respect to the common factors, rather than with respect to the observed variables, as in the MANCOVA. To this end, we specified the following expressions for the means, denoted \( \mu_i \), in matrix notation:

\[
\mu_i = \tau_i + A_i \alpha_i
\]

where the six-dimensional vector \( \alpha_i \) equaled \( E[\eta_i] \), i.e., the vector of means of the latent variables \( \eta_i \), the nine-dimensional vector \( \tau_i \) contains the intercepts.

In fitting this model, we introduced various substantive and identifying constraints. We compared the groups by fitting a number of models. In the first model (M1), equality constraints were not imposed. In this model, the configuration of factor loadings was identical over the groups, but no parameters were constrained to be equal over the groups. This model provided a baseline, by which we may evaluate other more constrained models:

\[
\Sigma_i = A_i \Theta_i A_i' + \Theta_i \quad (i = 1, \ldots, 4)
\]

\[
\mu_i = \tau_i \quad (i = 1, \ldots, 4)
\]

In fitting this model, we fixed the variance of \( \eta_i \) to equal 1 in each group (i.e., scaling constraints) and we estimated the regression coefficients in \( A_i \). In addition, we fixed the means \( \alpha_i \) to equal zero, so that there was a one-to-one relationship between the observed means and the intercepts \( \tau_i \). Subsequently, we constrained the regression coefficients to be equal in model 2 (M2):

\[
\Sigma_i = A_i \Theta_i A_i' + \Theta_i \quad (i = 1, \ldots, 4)
\]

\[
\mu_i = \tau_i \quad (i = 1, \ldots, 4)
\]

Given this constraint, we retained the standardization of \( \eta_i \) in group 4, but estimated the variances of \( \eta_i \) freely in the other three groups, i.e., \( \Psi_4 \) is a correlation matrix, while \( \Psi_1, \Psi_2, \) and \( \Psi_3 \) are covariance matrices. This expressed the fact that, across age, the variance of the variables of interest may vary. In model M3, we introduced the means. To this end we fitted the following model:

\[
\Sigma_i = A_i \Theta_i A_i' + \Theta_i \quad (i = 1, \ldots, 3)
\]

\[
\mu_i = \tau_i + \alpha_i \quad (i = 1, \ldots, 3)
\]

\[
\mu_4 = \tau
\]

Note that we equated the means and the intercepts in group 4 (\( \mu_4 = \tau \)), and estimated the parameter vector \( \alpha_i \) in groups 1–3. In this parameterization, group 4 served as a reference group and the parameter vectors \( \alpha_i \) (i = 1–3) represented the mean differences in \( \eta_i \) in the groups 1, 2, and 3 relative to group 4.

Appendix B

We regressed the variables obtained in the complex EF tasks (WCST and Tol tasks) on the latent variables \( \eta_i \). Again we did this in a 4-group analysis, as we wished to investigate possible age-related changes in these regression equations. Let \( Y_i \) denote a given dependent variable, i.e., a WCST or Tol task variable. We specified the following regression equation:

\[
Y_i = B_0 + B_i \times \eta_i + \zeta_i
\]

where \( B_0 \) is the intercept. \( B_i \) contains the regression coefficients and \( \zeta_i \) is a residual (Model A). The mean and variance of \( Y_i \) were modeled as:

\[
\mu_{Y_i} = B_0 + B_i \times \alpha_i
\]

\[
\sigma^2_{Y_i} = \hat{B} \Theta_i \hat{B} + \sigma^2_{\zeta_i}
\]

In this regression model, the variance of the dependent variable (i.e., measures derived from WCST or Tol) is decomposed into a part that is explained by the EF variables, corrected for basic processing speed, and a residual part (\( \sigma^2_{\zeta_i} \)). In this model,
we may test the hypothesis that the regression coefficients are equal. Thus, the group means for each region are not constrained to be equal across the groups. Furthermore, we can investigate whether the regression coefficients are significantly greater than zero (Model C).

References


