Age-related change in executive function: Developmental trends and a latent variable analysis

Huizinga, M.; Dolan, C.V.; van der Molen, M.W.

DOI
10.1016/j.neuropsychologia.2006.01.010

Publication date
2006

Document Version
Final published version

Published in
Neuropsychologia

ELSEVIER

Neuropsychologia 44 (2006) 2017–2036

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

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.

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, & Marcovitch, 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, & Lajoie, 1996; Baker, Segalowitz, & Ferlisi, 2001; Chelune & Baer, 1986; Chelune & Thompson, 1987; Heaton, Chelune, Talley, Kay, & Curtis, 1993; Kirk & Kelly, 1986; Lehto, 2004;

* Corresponding author.
E-mail address: m.huizinga@uva.nl (M. Huizinga).
URL: http://users.fmg.uva.nl/mhuizinga.

0028-3932/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2006.01.010

Lehto, Juujaervi, Kooistra, & Pulkkinen, 2003; Paniak, Miller, Murphy, & Patterson, 1996; Welsh, Pennington, & Groisser, 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., Cohen & Servan-Schreiber, 1992; Duncan, Emslie, Williams, Johnson, & Freer, 1996; Kimberg, D'Esposito, & Farah, 1997). For example, Kimberg et al. (1997) posited that all deficits in PFC function can be attributed to deficits in working memory. 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).

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., 2005; Smith & Jonides, 1999). In contrast, switching between tasks is thought to rely on medial PFC (Crone, Wendelken, Donohue, & Bunge, 2005; Rushworth, Walton, Kennerley, & Bannerman, 2004). Finally, the ability to inhibit responses was found to rely on orbitofrontal cortex (e.g., Aron, Robbins, & Poldrack, 2004; Roberts & Wallis, 2000). Thus, different regions within PFC subserve different components of goal-directed behavior.

At this point, it should be noted that the problem of "task impurity" hinders the interpretation of results reported in behavioral and neuroimaging studies using multiple EF tasks. Task impurity refers to the fact that a single indicator (operationalization) of a given construct (e.g., Working Memory) can rarely, if ever, be viewed as a pure measure of that construct. Most measures are contaminated by random error and systematic error (see Kline, 1998, p. 55). The task impurity problem is highly relevant to EF research, as the manifestation of EF components invariably involves other (non-EF) cognitive processes (e.g., Miyake et al., 2000).

Miyake et al. (2000) presented one way to address the task impurity problem. They proposed using multiple tasks to measure each EF component and adopting a latent variables approach to extract the variance common to those tasks. Latent variables (as incorporated in structural equation models; SEM) refer to what is shared among tasks that are assumed to tap a given EF. The latent variable approach minimizes the task impurity problem, and is therefore especially informative in developmental studies (e.g., Nunally & Bernstein, 1994, p. 85). Using confirmatory factor analysis, Miyake et al. (2000) examined the separability of three frequently postulated EF components: "Working Memory", "Shifting", and "Response Inhibition" (henceforth: Inhibition). Miyake et al. (2000) focused on these three EF components because: (1) they are well-circumscribed, lower-level functions that can be operationalized in a fairly precise manner; (2) they can be studied using commonly used tasks; and (3) they have been implicated in the performance of more complex EF tasks, such as the WCST and ToL. Miyake et al. (2000) tested healthy young-adults on multiple tasks tapping Working Memory, Shifting, and Inhibition, and several standard, but complex, neuropsychological tasks, including the WCST and the Tower of Hanoi (similar to the ToL). The results showed that, although moderately correlated, Working Memory, Shifting, and Inhibition were separable constructs (see also Fisk & Sharp, 2004). Moreover, the EF component processes differentially predicted performance on the complex neuropsychological tasks. For example, Shifting predicted WCST performance, whereas Inhibition predicted ToH performance.

1.2. Development of executive function

Developmental studies using standard neuropsychological tasks 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 & Baer, 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).

There is a growing body of research indicating differential trends in the development of EF component processes.[1]

[1] A growing body of research appeared recently, focusing on EF in pre-school aged children (e.g., Carlson, Mandell, & Williams, 2004; Diamond, Briand,

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, & Littler, 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, & Lahti Nuuttila, 2001), and to reach adult level of performance in late childhood, around the age of 12 (Bédard et al., 2002; Bunge, Dudukovic, Thomason, 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, & Tannock, 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).

Fossella, & Gehlbach, 2004; Diamond, Kirkham, & Amso, 2002; Espy & Bull, 2005; Gerstadt, Hong, & Diamond, 1994; Kirkham, Cruess, & Diamond, 2003; Muller, Zelazo, Hood, Leone, & Rohrer, 2004). These studies, however, applied tasks suitable for very young children that cannot be compared to the tasks used for the age range examined in the current study without violating the assumption of measurement invariance (Meredith, 1993).

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. = 0.13)), 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 = 3.92 (S.D. = 0.13)); one hundred and eleven 15-year olds (58 female, *M* age: 15.3 (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. Teachers assisted in the selection process in order to exclude children with health problems, neurological damage, or psychiatric problems. Similar information was derived from a self-report of the 21-year olds. 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 $F(3, 359) = 11.77$, $p < 0.001$. Post hoc 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 EF (e.g., Anderson et al., 1991; Bechara & Martin, 2004; but see Anderson, Anderson, Northam, Jacobs, & Catroppa, 2001; Conway, Kane, & Engle, 2003; Duncan et al., 1996 for opposite results). The gender distribution across groups differed significantly, $\chi^2(3) = 14.64$, $p = 0.002$. 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 (Huizinga & Van der Molen, in preparation-b). Task administration was computerized, and presented on a Toshiba Satellite 1600 laptop (Intel Celeron 800 mHz processor; 15 in. 60 Hz monitor). All tasks required only 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 "z" 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 Milner, 1971), participants were required to retain visual information about the orientation of a pattern of figures active in 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 3500 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.

Mental Counters. The Mental Counters task (adapted from Larson, Merritt, & Williams, 1988) required participants to retain numerical information active in their working memory. Participants had to keep track of the values of two or three (blocked) independent "counters", which change rapidly and in random order. The counters consisted of a horizontal line, above or below which squares appear. Participants were required to add 1 to the value of the counter, when a square appeared above the line, and to subtract 1, when it appeared below the line. When any counter reached a given criterion value, participants had to press a button. The series of stimuli presented comprised 5 or 7 (chosen randomly and equiprobably); there were two blocks consisting of 15 response probes. Participants had 3500 ms to respond. The interval between consecutive presentations of squares varied randomly between 1000 and 1300 ms (drawn from a uniform distribution). The main dependent variable was the proportion of correct trials.

Running Memory (adapted from Kramer, Larish, & Strayer, 1995). The stimuli in this task were pictures of fruit (apple, banana, lemon, pair, grape) and of animals (cat, dog, bear, frog, mouse, rabbit, horse, monkey), taken from the Snodgrass and Vanderwart (1980) pictures set. The participants were presented with series of fruit-and-animal pairs, and every four to seven presentations (equally distributed within blocks) they were required to indicate by a button press whether or not the current pair matched the last presentation of the pair (e.g., apple–cat?). The participants pressed one button on the computer keyboard if the pair matched the previous presentation of this pair (i.e., apple–cat) and another button if it did not match the last presentation (e.g., apple–frog). Feedback was presented following each response (i.e., '+' for a correct response, '−' for an incorrect response, '×' for an omission), and this information remained on the screen for 400 ms. The association between particular pictures of fruit and animals changed over presentations. Working memory load was varied by increasing the number of animals in a run of trials (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 3500 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 rectangles 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 randomized order; 30 practice trials and 50 experimental trials per block), and in the third block they alternated between a series of four "local" trials 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 cues and target was red. They remained on the screen until a response was given. Participants had 3500 ms to respond. The time interval between presentation 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 (derived from Rogers & Monsell, 1995). Varying numbers of either red dots or green triangles were presented in a 4 × 4 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 (90 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. Participants had 3500 ms to respond. 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, Ridderinkhof, & Van der Molen, 2004). Stimuli were schematic faces (man or woman, smiling or unsmiling) 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; 30 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 (block 3; 90 practice trials, 150 experimental trials). The stimulus remained on the screen until a response was given. Participants had 3500 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.

2.2.3. Inhibition

Stop-signal. In the present version of the Stop-signal task (adapted from Van Boxtel, Van der Molen, Jennings, & Brunia, 2001), 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 stimulus reaction time (SSRT), reflecting the latency of the internal response to the stop signal (see Logan, 1994).

Eriksen Flankers (adapted from Ridderinkhof & 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 ← ← ← ← ←; congruent condition) or by four arrows pointing in the opposite direction (i.e., → → ← → → or ← ← → ← ←; incongruent condition). The arrow array was presented in a rectangle, which served as the warning stimulus, as its onset was followed by the arrow array. 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., 50 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 contour (color task) and either right-side-up or upside-down oriented (orientation task). On the color task, participants had to respond with their dominant hand to the blue or the red contour 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 upside-down 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 responses. 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 upside-down. 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.

2.2.4. Complex EF tasks

Wisconsin Card Sorting Task (WCST). We used a computerized version of the WCST (Somsen, Van der Molen, Jennings, & Van Beek, 2000). Against a light-gray background, four key cards were presented at the top of the screen and were numbered from 1 to 4. The response cards were taken from the original version of the WCST (Grant & Berg, 1948) and were presented one at a time at the bottom of the screen. The task required participants to match the series of response cards with any of four key cards by pressing the number corresponding to that key card. The display remained until a choice was given. Feedback followed the response, and consisted of a "+" sign if the response was correct, or a "−" sign if the response was incorrect.

Response cards could be matched on color (red, green, blue, yellow), shape (triangle, star, cross, circle), or number (1, 2, 3, 4). Once the participant made 10 consecutive correct sorts, the sorting principle changed. The task was terminated either after the participant completed 6 categories (e.g., shape, color, form, color, form, shape), or after the maximum of 128 trials was reached. The order of the sorting principles was randomized, with the constraint that the same sorting principles did not occur consecutively. The test was administered according to the procedure outlined in the Heaton manual (Heaton et al., 1993). The variables of interest were the number of categories achieved, the proportion of perseverative errors (i.e., the total number of errors that occur when a participant is required to switch to another sorting principle, and then persists in responding to the previously correct sorting principle, relative to the number of trials administered, and multiplied by 100), and the proportion of conceptual level responses (i.e., the total number of consecutive correct responses that occur in runs of three or more relative to the number of trials administered, and multiplied by 100).

Tower of London (ToL). The ToL (Shallice, 1982) requires the moving of differently colored balls across three differently sized pegs in order to duplicate a pre-specified target configuration. The smallest peg can hold one ball, the medium-sized two, and the largest peg can hold three balls. Three constraints apply: (1) do not place more than the permitted number of balls on one peg; (2) do not place the balls anywhere other than on a peg; (3) only move one ball at a time. We used a computerized version of the ToL (see for a similar ToL task format: Wiers, Gunning, & Sergeant, 1998). Against a light-gray background, a pegboard with three balls (red, green, blue) positioned on the pegs was presented at the center of the computer screen and a task-specification box (depicting the goal state) in the upper right-hand corner of the computer screen. The balls were repositioned by dragging (with a computer mouse) a ball directly to the position indicated by the task-specification box. The ToL items were taken from Schnirman, Welsh, and Retzlaff (1998). Three difficulty levels were presented: one block of five 4-move trials, one block of five 5-move trials, and one block of five 6-move trials. Performance was scored in terms of the proportion of perfect solutions (i.e., the number of items solved in the minimum number of moves relative to the number of items administered, and multiplied by 100), the mean number of additional moves (i.e., the mean number of moves exceeding the minimum number of 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 box-plot 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
Proportions correct (%) for the EF tasks in the domains of *Working Memory*, *Shifting*, and *Inhibition*, the control factor *Basic Speed*, and performance measures of the WCST and ToL

EF domain	Task	7-year olds	11-year olds	15-year olds	21-year olds
Working Memory	Tic Tac Toe				
	Low memory load	83.4 (1.10)	93.1 (0.97)	95.1 (0.98)	96.0 (0.98)
	High memory load	51.0 (1.57)	767.7 (1.58)	87.0 (1.68)	90.1 (1.90)
	Mental Counters				
	2 counters and series 5	78.5 (2.30)	84.5 (1.65)	85.0 (1.66)	86.9 (1.68)
	3 counters and series 5	76.2 (2.44)	83.7 (1.83)	85.6 (1.67)	90.4 (1.71)
	2 counters and series 7	61.0 (2.66)	73.4 (2.23)	78.1 (1.94)	85.4 (2.22)
	3 counters and series 7	70.9 (2.86)	73.6 (2.06)	82.4 (2.00)	86.3 (2.04)
	Running Memory				
	Block 1	76.6 (1.75)	85.6 (1.48)	86.7 (1.49)	88.9 (1.51)
	Block 2	76.7 (2.10)	82.6 (1.64)	87.4 (1.68)	88.2 (1.88)
	Block 3	75.3 (1.91)	84.5 (1.65)	88.0 (1.50)	90.6 (1.71)
Shifting	Local–Global				
	Repetition	90.3 (0.76)	92.9 (0.58)	93.5 (0.58)	95.1 (0.59)
	Alternation	88.9 (0.75)	92.0 (0.77)	93.5 (0.77)	95.8 (0.78)
	Dots–Triangles				
	Repetition	78.5 (1.06)	89.1 (0.76)	92.7 (0.77)	94.3 (0.97)
	Alternation	71.4 (1.18)	79.7 (1.07)	84.8 (1.11)	87.8 (1.12)
	Smiling Faces				
	Repetition	79.2 (1.25)	85.6 (1.11)	87.8 (1.12)	93.1 (1.35)
	Alternation	73.8 (1.20)	79.9 (1.07)	82.8 (1.09)	88.9 (1.13)
Inhibition	Stop-signal				
	% Correct inhibits	46.0 (0.95)	48.2 (0.69)	50.0 (0.85)	47.9 (0.83)
	Eriksen Flankers				
	Congruent	96.2 (0.39)	97.8 (0.79)	98.6 (0.40)	99.2 (0.40)
	Incongruent	90.8 (0.76)	93.3 (0.77)	93.9 (0.78)	94.1 (0.78)
	Stroop				
	Interference	80.8 (0.90)	90.8 (0.95)	93.1 (0.97)	96.6 (0.98)
Complex	WCST				
	Categories (#)	2.90 (0.220)	3.51 (0.173)	4.20 (0.166)	5.1 (0.182)
	Perseverative errors (%)	32.20 (1.76)	24.52 (1.38)	22.08 (1.33)	15.56 (1.45)
	Conceptual level (%)	45.27 (2.45)	53.75 (1.93)	60.35 (1.85)	64.26 (2.03)
	ToL				
	Perfect solutions (%)	6.11 (0.10)	8.11 (0.08)	12.56 (0.08)	14.91 (0.09)
	Additional moves (#)	14.08 (0.87)	17.27 (0.69)	12.44 (0.68)	10.80 (0.73)
	Planning time (s)	5.78 (0.49)	4.20 (0.39)	6.22 (0.38)	7.55 (0.41)
Basic Speed	Stroop				
	Pure blocks	97.8 (0.20)	99.6 (0.20)	99.8 (0.20)	99.8 (0.20)

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 WSCT 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., Cerella & 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.

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

EF domain	Task	7-year olds	11-year olds	15-year olds	21-year olds
Shifting	Local–Global				
	Repetition	874 (12)	535 (10)	422 (11)	393 (11)
	Alternation	1088 (18)	625 (15)	477 (15)	448 (16)
	Dots–Triangles				
	Repetition	1164 (23)	810 (18)	623 (18)	588 (20)
	Alternation	1677 (40)	1200 (31)	819 (30)	817 (33)
	Smiling Faces				
	Repetition	1071 (23)	726 (19)	591 (19)	518 (21)
	Alternation	1567 (38)	1088 (31)	818 (31)	703 (33)
Inhibition	Stop-signal				
	Stop-signal RT	289 (8)	238 (7)	217 (7)	201 (8)
	Eriksen Flankers				
	Congruent	815 (10)	523 (9)	415 (9)	394 (9)
	Incongruent	950 (12)	587 (10)	466 (11)	444 (11)
	Stroop				
	Interference	776 (8)	612 (8)	548 (8)	514 (8)
Basic Speed	Stroop				
	Pure blocks	579 (6)	424 (6)	356 (6)	341 (6)

Standard errors are presented within parentheses.

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-subject variable. The ANOVA of the Tic Tac Toe task scores resulted in a main effect of Condition, $F(1, 351) = 300.21$, $p < 0.001$, MSE = 0.46, $\eta_p^2 = 0.46$; Age Group, $F(3, 351) = 116.86$, $p < 0.001$, MSE = 0.56, $\eta_p^2 = 0.50$; and a Condition × Age Group interaction, $F(3, 351) = 44.26$, $p < 0.001$, MSE = 0.46, $\eta_p^2 = 0.27$. 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, $F(1, 283) = 67.84$, $p < 0.001$, MSE = 0.34, $\eta_p^2 = 0.19$; Age Group, $F(3, 283) = 21.31$, $p < 0.001$, MSE = 0.72, $\eta_p^2 = 0.18$; and a Condition × Age Group interaction, $F(3, 283) = 6.02$, $p = 0.001$, MSE = 0.34, $\eta_p^2 = 0.06$. Again, post hoc Bonferroni comparisons for the difference score of the Condition effect showed that across blocks the difference between series consisting of five stimuli and 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, $F(3, 354) = 16.90$, $p < 0.001$, MSE = 1.19, $\eta_p^2 = 0.13$, showing that performance increased with age. The main effect of Load was not significant, and did not interact with Age Group (F's < 1). 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 ANCOVA for the Local–Global task resulted in a main effect of Age Group, $F(3, 314) = 30.92$, $p < 0.001$, MSE = 17106.58, $\eta_p^2 = 0.23$; and a Condition × Age Group interaction, $F(3, 314) = 7.24$, $p < 0.001$, MSE = 3330.02, $\eta_p^2 = 0.07$.

The ANCOVA for the Dots–Triangles task resulted in a main effect of Condition, $F(1, 315) = 4.37$, $p = 0.037$, MSE = 20634.45, $\eta_p^2 = 0.02$; Age Group, $F(3, 315) = 17.21$, $p < 0.001$, MSE = 106150.38, $\eta_p^2 = 0.14$; and a Condition × Age Group interaction, $F(3, 315) = 10.53$, $p < 0.001$, MSE = 20634.45, $\eta_p^2 = 0.09$.

The ANCOVA of 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, $\eta_p^2 = 0.10$; and a Condition × Age Group interaction, $F(3, 306) = 4.95$, $p = 0.002$, MSE = 21986.26, $\eta_p^2 = 0.05$. 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 tasks 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, $\eta_p^2 = 0.040$. 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, $\eta_p^2 = 0.03$; Age Group, $F(3, 326) = 50.70$, $p < 0.001$, MSE = 12400.09, $\eta_p^2 = 0.32$; and a Condition × Age Group interaction, $F(3, 326) = 12.73$, $p < 0.001$, MSE = 1306.78, $\eta_p^2 = 0.11$. 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 = 2953.43, $\eta_p^2 = 0.06$. 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, $\eta_p^2 = 0.11$; $F(3, 346) = 23.82$, $p < 0.001$, MSE = 0.2.90, $\eta_p^2 = 0.17$; and $F(3, 346) = 18.46$, $p < 0.001$, MSE = 185.25, $\eta_p^2 = 0.14$, 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, $\eta_p^2 = 0.26$; $F(3, 348) = 18.21$, $p < 0.001$, MSE = 47.38, $\eta_p^2 = 0.14$; $F(3, 348) = 12.17$, $p < 0.001$, MSE = 14.83, $\eta_p^2 = 0.10$. 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

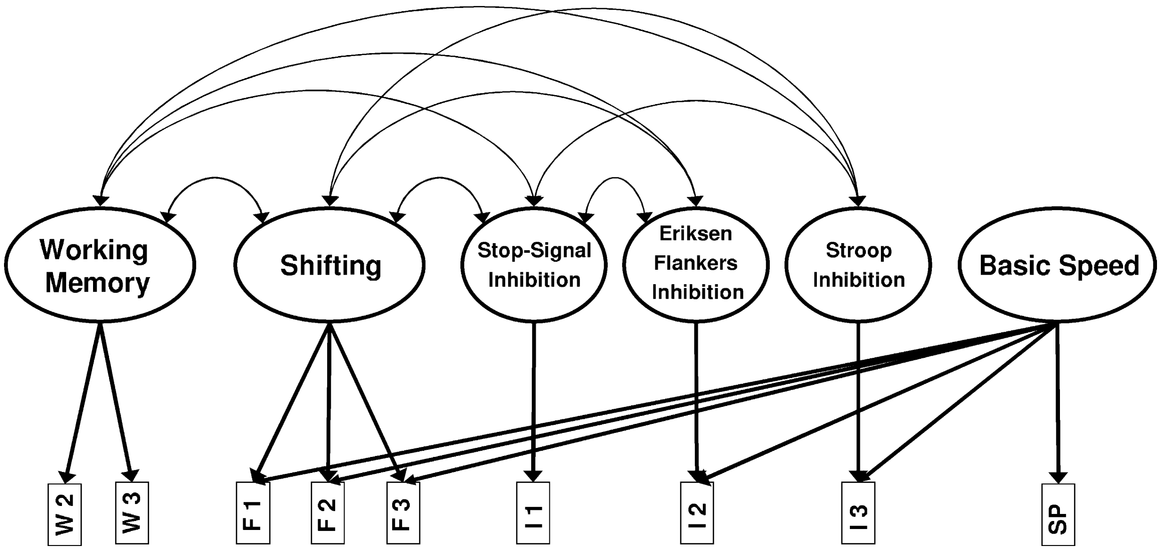

Fig. 1. The partial measurement model used for the confirmatory factor analysis. The ellipses represent the latent factors *Working Memory* and *Shifting*; the manifest *Inhibition* variables and the control factor (*Basic Speed*). The rectangles represent the individual tasks that were chosen to tap the specific EF components. The curved double-headed arrows represent correlations among the latent variables, the Inhibition variables (Stop-signal Inhibition; Eriksen Flanker Inhibition; Stroop Inhibition), and the control factor. *Note*: W2: Tic Tac Toe task; W3: Mental Counters task; S1: Local–Global task; S2: Dots–Triangles task; S3: Smiling Faces task; I1: Stop-signal task; I2: Eriksen Flankers task; I3: Stroop task; SP: Basic Speed.

of the Stroop task, was used as a control factor. As the three Shifting tasks and the Eriksen Flankers and Stroop 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], 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 inhibits 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 α (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 χ^2 test statistic was hardly cause for alarm (7-year olds: $\chi^2(414) = 406.31$, $p = 0.59$; 11-year olds: $\chi^2(498) = 535.97$, $p = 0.11$; 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 χ^2 difference statistics (i.e., loglikelihood ratio tests) to investigate the measurement model and measurement invariance.

First, to establish configural invariance, we fitted the factor model M1. This model served as the baseline model. We found the fit of M1 to be acceptable. The formal goodness of fit, as indexed by the χ^2, equaled 139.34 (d.f. = 67). Next, we introduced (and established) measurement invariance (M2). Following the imposition of the M2 equality constraints, and compared to M1, we observed a χ^2 difference of 17.85 (d.f. = 21, n.s.). The percentages of variances of the data of the five tasks of Working Memory and Shifting explained by the common factors of Working Memory and Shifting are presented in Table 3. To determine the adequacy of the models M1 and M2, we inspected the range of the standardized residuals. These are a function of the difference between the unconstrained covariance matrix and the expected covariance matrix under the specified model (both matrices are estimated using raw data likelihood estimation). In a well fitting model, they are approximately standard normally distributed, and thus should vary between about −3 and 3. In the three oldest age groups, the standardized residuals were acceptable (between −3.6 and 2.8); for the youngest age group only four out of the 45 to be estimated standardized residuals were greater than –3 or greater than 3. These standardized residuals were all related to Mental Counters task performance.

Finally, we performed 4-group confirmatory factor analysis. We therefore introduced the mean structure into the model (M3), and compared to M2, we observed a χ^2 difference of 51.45 (d.f. = 9; $p < 0.001$). Inspection of the standardized residuals of the covariance structure (between −3.6 and 3.1 in the oldest three 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 Shifting, and the three Inhibition tasks are presented in Table 4. Note that the hypothesis of a single common factor in the present normal populations is unlikely, in the light of the generally low correlations between the common factors.

Table 3
Percentage of variance explained by the latent factors *Working Memory* and *Shifting*, and by the control factor *Basic Speed*

Factor	Task	7-year olds	11-year olds	15-year olds	21-year olds
Working Memory	Tic Tac Toe (%)	46	7	20	47
	Mental Counters (%)	75	12	25	51
Shifting	Local–Global (%)	28	50	64	67
	Dots–Triangles (%)	17	41	39	34
	Smiling Faces (%)	29	45	41	60
Basic Speed	Stroop pure blocks (%)	83	77	86	76

Table 4
Correlations of the latent factors *Working Memory* and *Shifting* and the manifest *Inhibition* measures for each of the four age groups

	7-year olds	11-year olds	15-year olds	21-year olds
Working Memory–Shifting	0.40	0.04	−0.08	−0.54*
Working Memory–Stop-signal	0.09	−0.27	−0.42*	−0.55*
Working Memory–Eriksen Fl.	−0.03	−0.11	0.17	−0.21
Working Memory–Stroop	−0.14	−0.17	0.36	−0.20
Shifting–Stop-signal	−0.85*	0.16	0.09	0.27
Shifting–Eriksen Flankers	0.60	0.34*	0.41*	0.35*
Shifting–Stroop	−0.35	0.02	−0.10	0.26*
Stop-signal–Eriksen Flankers	0.12	0.26*	0.09	0.26*
Stop-signal–Stroop	−0.07	0.30*	−0.12	0.19
Eriksen Flankers–Stroop	–	–	–	–

Note: (–) Fixed to zero.
* Significant.

Table 5
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*

	α_1^*	α_2^*	α_3^*
Working Memory	−4.83 (0.72)*	−1.57 (0.28)*	−0.49 (0.19)*
Shifting	1.76 (0.71)*	1.17 (0.33)*	0.09 (0.19)
Stop-signal Inhibition	1.54 (0.26)*	0.67 (0.17)*	0.29 (0.17)
Eriksen Flankers Inhibition	6.81 (0.99)*	1.97 (0.40)*	0.23 (0.21)
Stroop Inhibition	−1.09 (0.74)	−0.15 (0.31)	0.28 (0.17)
Basic Speed	7.92 (0.90)*	2.88 (0.37)*	0.62 (0.18)*

* 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 α_1^*, α_2^*, and α_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 (χ^2 difference = 17.85; d.f. = 15), but the difference between Models C and B were (χ^2 difference = 15.30; 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. Thus, perseveration behavior on the WCST appeared to be strongly related to Working Memory capacity.

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 (χ^2 difference = 16.60; d.f. = 15). The difference between Models C

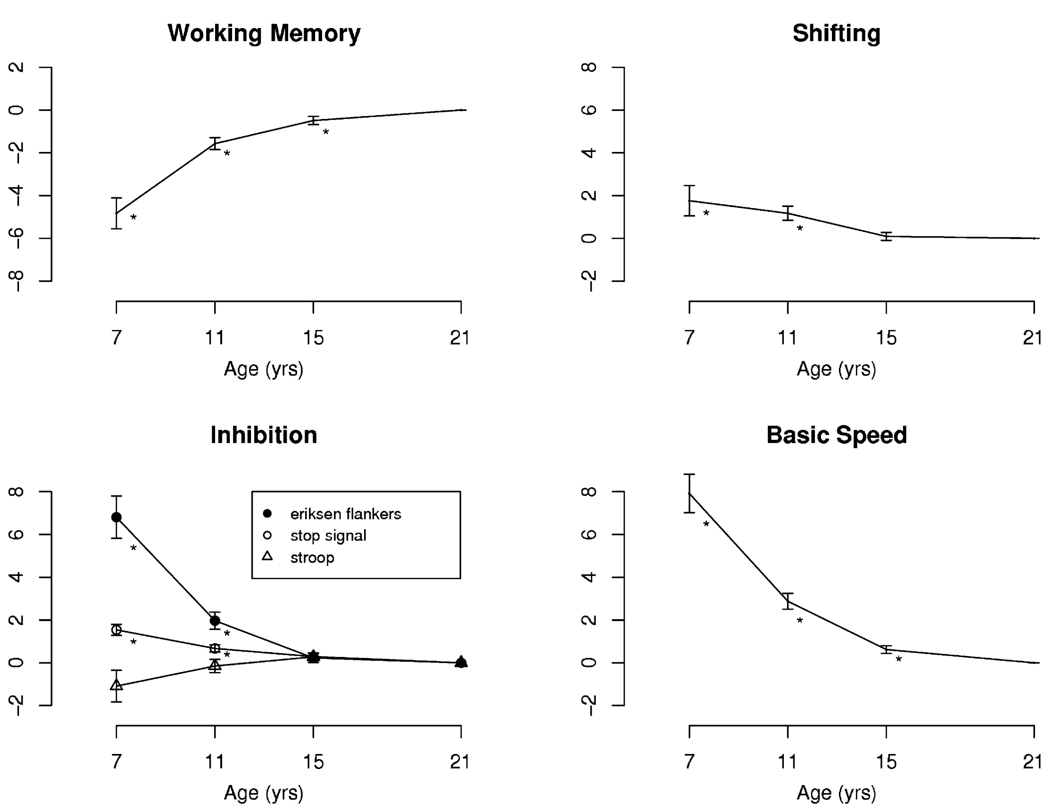

Fig. 2. Maximum likelihood estimates and their standard errors of differences in common factor means of *Working Memory* and *Shifting* (i.e., the observed mean differences between the groups as a function of the mean differences of the common factors), the three manifest *Inhibition* variables and the control factor *Basic Speed*; (*) significant difference relative to 21-year olds.

Table 6
Maximum likelihood estimates of regression coefficients (est), and 95% confidence intervals (cfi) for WCST performance (in all age groups), and ToL performance (in 21-year olds)

		est	95% cfi
WCST: perseverative errors (%)	Working Memory	−0.20*	−0.58 to −0.06
	Shifting	0.04	−0.19 to 0.23
	Stop-signal Inhibition	−0.07	−0.26 to 0.06
	Eriksen Fl. Inhibition	0.03	−0.05 to 0.11
	Stroop Inhibition	0.21	0.04–0.41
ToL: perfect solutions (%)	Working Memory	0.09	−0.40 to 1.06
	Shifting	0.04	−0.40 to 0.66
	Stop-signal Inhibition	0.05	−0.28 to 0.67
	Eriksen Fl. Inhibition	0.10	−0.18 to 0.38
	Stroop Inhibition	−0.44*	−0.78 to −0.14

* Significant estimate.

and A was significant (χ^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., 1991). The correction for within-group individual differences in basic processing speed did not remove the age related trends in task performance. This finding is important, because it indicates that the developmental trends in EF task performance cannot be explained in terms of a global change in basic processing speed (e.g., Cerella & Hale, 1994; Kail, 1991; Span et al., 2004).

Developmental changes in Working Memory are well-established, both at a behavioral level (e.g., Gathercole et al., 2004; Luciana et al., 2005; Luna et al., 2004), and at the level of brain functioning. Neuroimaging studies have shown that development of Working Memory co-occurs with the functional maturation of lateral PFC (e.g., Klingberg, Forssberg, & Westerberg, 2002; Kwon, Reiss, & Menon, 2002). In the current study, Working Memory was tapped using three different tasks, and consistent with prior studies, we found that adult level of performance was not reached before the age of 12 (see Case, Kurland, & Goldberg, 1982; Cohen & Heath, 1990; DeLuca et al., 2003; Gathercole, 1999; Luciana et al., 2005; Luna et al., 2004; Siegel, 1994). The performance on one of the Working Memory tasks (i.e., the Running Memory task), however, failed to reveal a significant effect of task load. The accuracy of performance increased with age but this age-related trend did not interact with task load. The lack of a task load effect was unexpected and, at this point, we do not have a ready explanation for its absence. Possibly, the lack of a task load effect is due to an interference confound. That is, task load increases with longer stimulus pair series but potential interference between stimulus pairs decreases, as the number of possible stimulus pairs increases with series length. Unfortunately, there is not sufficient data to test this interpretation.

Developmental studies on task shifting are still scarce, but studies to date have demonstrated that the cost of shifting between tasks decreases with age (e.g., Cepeda et al., 2001). In the current study, Shifting was assessed using three different tasks that required shifting back and forth between tasks. All three tasks showed that shift costs (i.e., the difference in RT and accuracy on shift trials relative to repeat trials) decreased with age until the age of 15. Thus, the ability to shift task sets does not reach young-adult level of performance until adolescence (Cepeda et al., 2001; Crone et al., in press; Huizinga & Van der Molen, in preparation-a). The age-related change in the ability to shift task sets has been interpreted in terms of two prevailing hypotheses of shift costs. First, it has been argued that the larger shift costs in children reflect immature levels of executive control (Cepeda et al., 2001; Zelazo, Craik, & Booth, 2004). In contrast, Crone, et al. (in press) interpreted larger shift costs in children in terms of memory retrieval of S–R links. Systematic manipulation within the context of the task-shift paradigm should reveal the basic mechanism(s) underlying developmental change in Shifting ability (e.g., Huizinga & Van der Molen, in preparation-a).

Inhibition is the most studied measure of EF in development (e.g., Bjorklund & Harnishfeger, 1990; Dempster, 1992; Nigg, 2000). The current focus was on the inhibition of motor responses, indexed by using three different tasks. These tasks yielded distinct developmental trends. The performance on the Eriksen Flankers task and the Stop-signal task improved rapidly until the age of 11, and at this age performance did not differ from either the 15-year olds or 21-year olds. This developmental trend is consistent with previous studies using the Eriksen Flankers task (e.g., Ridderinkhof & Van der Molen, 1995), and the Stop-signal task (e.g., Van den Wildenberg & Van der Molen, 2004; Williams et al., 1999). In contrast, the current version of the Stroop task yielded only a weak developmental trend from early childhood into adulthood. Previous studies using a childfriendly version of the Stroop task showed a rapid decline in the suppression of pre-potent responses during early childhood (e.g., Wright, Waterman, Prescott, & Murdoch-Eaton, 2003). The current task revealed an early and strong decline in error rate, but not in basic processing speed. The current developmental trends in performance on the Eriksen Flankers task and the Stop-signal task can be interpreted in the light of developmental theories that emphasize the role of Inhibition in cognitive development (e.g., Bjorklund & Harnishfeger, 1990; Dempster, 1992; for a review see: Van der Molen, 2000). These theories assume that the ability to resist interference and inhibit competing responses defines a major developmental dimension promoting performance on a wide array of cognitive tasks (see also Kipp, 2005).

The two standard neuropsychological tasks, the WCST and the ToL, yielded developmental trends that are consistent with the literature. For the WCST, the proportions of perseverative errors and of conceptual level reached young-adult level in 11-year olds, whereas the number of categories completed increased into young-adulthood (Chelune & Baer, 1986; Welsh et al., 1991). For the ToL, planning time and the number of additional moves reached young-adult level in 11-year olds, whereas the proportion of perfect solutions increased into young-adulthood (e.g., Anderson et al., 2005; Baker et al., 2001; Lehto, 2004).

Taken together, the performance analyses revealed that Working Memory, Shifting and Inhibition reached adult levels between 11 and 15 years, whereas some aspects of inhibitory control (Stroop task) did not reach adult levels until after the age of 15. The performance on the WCST and the ToL reached adult levels between 11 and 21 years.

4.2. *Analysis of latent variables*

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; Harnishfeger, 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 analyses 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 χ^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 ToH performance, given "that no specific instructions for strategies were given and many people are likely to use a perceptual strategy to solve the ToH" (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-Cartmell, & 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 retained 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; Crone et al., in press; Gathercole et al., 2004; Huizinga & Van der Molen, in preparation-a; 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 Stroop 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; Crone, 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; Huttenlocher, 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 Smiling Faces task (denoted T_3 through T_5) 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_8). The control factor represented basic processing speed, as indexed by median RT of the pure blocks of the Stroop task (denoted T_9). 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:

$$T_{9i} = \tau_{9i} + \lambda_{9,6i} \times \mathrm{BS}_i + \varepsilon_{4i}$$

Note that the index on the factor loading λ 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):

$$T_{1i} = \tau_{1i} + \lambda_{11i} \times \mathrm{WM}_i + \varepsilon_{1i}$$
$$T_{2i} = \tau_{2i} + \lambda_{12i} \times \mathrm{WM}_i + \varepsilon_{2i}$$

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

$$T_{3i} = \tau_{3i} + \lambda_{32i} \times S_i + \lambda_{36i} \times \mathrm{BS}_i + \varepsilon_{3i}$$
$$T_{4i} = \tau_{4i} + \lambda_{42i} \times S_i + \lambda_{46i} \times \mathrm{BS}_i + \varepsilon_{4i}$$
$$T_{5i} = \tau_{5i} + \lambda_{52i} \times S_i + \lambda_{56i} \times \mathrm{BS}_i + \varepsilon_{5i},$$

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

$$T_{6i} = \tau_{6i} + \lambda_{63i} \times \mathrm{EFL}_i$$
$$T_{7i} = \tau_{7i} + \lambda_{74i} \times \mathrm{SS}_i + \lambda_{76i} \times \mathrm{BS}_i$$
$$T_{8i} = \tau_{8i} + \lambda_{85i} \times \mathrm{STR}_i + \lambda_{86i} \times \mathrm{BS}_i$$

In these regression equations, τ represents the intercept, λ the regression coefficient, and ε 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:

$$\mathbf{T}_i = \boldsymbol{\tau}_i + \boldsymbol{\Lambda}_i \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i, \tag{1}$$

where $\mathbf{T}_i$ is the random vector of the nine observed variables, $\boldsymbol{\tau}_i$ the vector of intercepts, and $\boldsymbol{\varepsilon}_i$ is the random vector of mutually uncorrelated residuals. The random vector $\boldsymbol{\eta}_i$ contains the common factors, WM, S, and BS, and the Inhibition variables EFL, SS, STR. The 9×6 matrix $\boldsymbol{\Lambda}_i$ contains the regression coefficients:

$$\boldsymbol{\Lambda}_i = \begin{matrix} \lambda_{11} & 0 & 0 & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_{32} & 0 & 0 & 0 & \lambda_{36} \\ 0 & \lambda_{42} & 0 & 0 & 0 & \lambda_{46} \\ 0 & \lambda_{52} & 0 & 0 & 0 & \lambda_{56} \\ 0 & 0 & \lambda_{63} & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{73} & 0 & \lambda_{76} \\ 0 & 0 & 0 & 0 & \lambda_{85} & \lambda_{86} \\ 0 & 0 & 0 & 0 & 0 & \lambda_{96} \end{matrix}$$

The residuals were constrained to be mutually uncorrelated, so the covariance matrix of the residuals was diagonal. We denote this covariance matrix $\boldsymbol{\Theta}_i$:

$$\mathrm{diag}(\boldsymbol{\Theta}_i) = [\sigma^2_{\varepsilon 1}\,\sigma^2_{\varepsilon 2}\,\sigma^2_{\varepsilon 3}\,\sigma^2_{\varepsilon 4}\,\sigma^2_{\varepsilon 5}\,0\,0\,0\,\sigma^2_{\varepsilon 9}],$$

where σ^2_{ε} 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 $\boldsymbol{\eta}_i$, denoted $\boldsymbol{\Psi}_i$.

$$\boldsymbol{\Psi}_i = \begin{matrix} \sigma^2_{\mathrm{WM}} \\ \sigma_{\mathrm{S,WM}} & \sigma^2_{\mathrm{Sh}} \\ \sigma_{\mathrm{EFL,WM}} & \sigma_{\mathrm{EFL,S}} & \sigma^2_{\mathrm{EFL}} \\ \sigma_{\mathrm{SS,WM}} & \sigma_{\mathrm{SS,S}} & \sigma^2_{\mathrm{SS,EFL}} & \sigma^2_{\mathrm{SS}} \\ \sigma_{\mathrm{STR,WM}} & \sigma_{\mathrm{STR,S}} & \sigma^2_{\mathrm{STR,EFL}} & \sigma^2_{\mathrm{STR,SS}} & s^2_{\mathrm{STR}} \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{\mathrm{BS}} \end{matrix}$$

Note that $\sigma^2_{\mathrm{STR,SS}}$ was fixed to zero, because it proved impossible to fit the model in all age groups when $\sigma^2_{\mathrm{STR,SS}}$ was estimated freely. Assuming $\boldsymbol{\eta}_i$ and $\boldsymbol{\varepsilon}_i$ are uncorrelated, we expressed the expected covariance matrix in group i, $\boldsymbol{\Sigma}_i$, as follows:

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i \boldsymbol{\Psi}_i \boldsymbol{\Lambda}_i^{\mathrm{t}} + \boldsymbol{\Theta}_i. \tag{2}$$

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 $\boldsymbol{\mu}_i$, in matrix notation:

$$\boldsymbol{\mu}_i = \boldsymbol{\tau}_i + \boldsymbol{\Lambda}_i \boldsymbol{\alpha}_i \tag{3}$$

where the six-dimensional vector $\boldsymbol{\alpha}_i$ equaled E[$\boldsymbol{\eta}_i$], i.e., the vector of means of the latent variables $\boldsymbol{\eta}_i$, the nine-dimension vector $\boldsymbol{\tau}_i$ 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:

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i \boldsymbol{\Psi}_i \boldsymbol{\Lambda}_i^{\mathrm{t}} + \boldsymbol{\Theta}_i \quad (i = 1, \ldots, 4)$$
$$\boldsymbol{\mu}_i = \boldsymbol{\tau}_i \quad (i = 1, \ldots, 4)$$

In fitting this model, we fixed the variance of $\boldsymbol{\eta}_i$ to equal 1 in each group (i.e., scaling constraints) and we estimated the regression coefficients in $\boldsymbol{\Lambda}_i$. In addition, we fixed the means $\boldsymbol{\alpha}_i$ to equal zero, so that there was a one-to-one relationship between the observed means and the intercepts $\boldsymbol{\tau}_i$. Subsequently, we constrained the regression coefficients to be equal in model 2 (M2):

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda} \boldsymbol{\Psi}_i \boldsymbol{\Lambda}^{\mathrm{t}} + \boldsymbol{\Theta}_i \quad (i = 1, \ldots, 4)$$
$$\boldsymbol{\mu}_i = \boldsymbol{\tau}_i \quad (i = 1, \ldots, 4)$$

Given this constraint, we retained the standardization of $\boldsymbol{\eta}$ in group 4, but estimated the variances of $\boldsymbol{\eta}$ freely in the other three groups, i.e., $\boldsymbol{\Psi}_4$ is a correlation matrix, while $\boldsymbol{\Psi}_1$, $\boldsymbol{\Psi}_2$, and $\boldsymbol{\Psi}_3$ 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:

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda} \boldsymbol{\Psi}_i \boldsymbol{\Lambda}^{\mathrm{t}} + \boldsymbol{\Theta}_i.$$
$$\boldsymbol{\mu}_i = \boldsymbol{\tau} + \boldsymbol{\Lambda}_i \boldsymbol{\alpha}_i \quad (i = 1, \ldots, 3)$$
$$\boldsymbol{\mu}_4 = \boldsymbol{\tau}$$

Note that we equated the means and the intercepts in group 4 ($\boldsymbol{\mu}_4 = \boldsymbol{\tau}$), and estimated the parameter vector $\boldsymbol{\alpha}$ in groups 1–3. In this parameterization, group 4 served as a reference group and the parameter vectors $\boldsymbol{\alpha}_{\mathrm{i}}(i = 1\text{–}3)$ represented the mean differences in $\boldsymbol{\eta}$ 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 $\boldsymbol{\eta}$. 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:

$$Y_i = B_{0i} + B_i \times \boldsymbol{\eta}_i + \zeta_i \tag{4}$$

where B_{0i} is the intercept. B_i contains the regression coefficients and ζ_i is a residual (Model A). The mean and variance of Y_i were modeled as:

$$\mu_{Y_i} = B_{0i} + B_i \times \boldsymbol{\alpha}_i \tag{5}$$

$$\sigma^2_{Y_i} = B_i \boldsymbol{\Psi}_i B_i^{\mathrm{t}} + \sigma^2_{\zeta_i} \tag{6}$$

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 over the groups, i.e., $B_1 = B_2 = B_3 = B_4$ (Model B). Note that over-group equality of the regression coefficients does not necessarily imply that the amount of explained variance is equal across the groups (specifically, we do not constrain $\mathbf{\Psi}_i$ or $\sigma^2_{\zeta_i}$ to be equal across the groups). Furthermore, we can investigate whether the regression coefficients are significantly greater than zero (Model C).

References

Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance. XV. Conscious and nonconscious information processing* (pp. 421–452). Cambridge, MA: MIT Press.

Anderson, P., Anderson, V., & Lajoie, G. (1996). The Tower of London Test: Validation and standardization for pediatric populations. *Clinical Neuropsychologist*, *10*(1), 54–65.

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

Anderson, K., Byrd, D., & Berg, W. K. (2005). The development of the contingent negative variation during the Tower of London: Planning from childhood to early adulthood. *Journal of Cognitive Neuroscience*, 59.

Anderson, S. W., Damasio, H., Jones, R. D., & Tranel, D. (1991). Wisconsin Card Sorting Test performance as a measure of frontal lobe damage. *Journal of Clinical and Experimental Neuropsychology*, *13*(6), 909–922.

Andres, P., & Van der Linden, M. (2001). Supervisory attentional system in patients with focal frontal lesions. *Journal of Clinical and Experimental Neuropsychology*, *23*(2), 225–239.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, *8*(4), 170–177.

Baddeley, A. (1986). *Working memory*. New York, NY: Oxford University Press.

Baker, K., Segalowitz, S. J., & Ferlisi, M. C. (2001). The effect of differing scoring methods for the Tower of London task on developmental patterns of performance. *Clinical Neuropsychologist*, *15*(3), 309–313.

Barceló, F., & Knight, R. T. (2002). Both random and perspective errors underlie WCST deficits in prefrontal patients. *Neuropsychologia*, *40*(3), 349–356.

Bechara, A., Damasio, H., Tranel, D., & Anderson, S. W. (1998). Dissociation of working memory from decision making within the human prefrontal cortex. *Journal of Neuroscience*, *18*(1), 428–437.

Bechara, A., & Martin, E. M. (2004). Impaired decision making related to working memory deficits in individuals with substance addictions. *Neuropsychology*, *18*(1), 152–162.

Bédard, A. C., Nichols, S., Barbosa, J. A., Schachar, R., Logan, G. D., & Tannock, R. (2002). The development of selective inhibitory control across the life span. *Developmental Neuropsychology*, *21*(1), 93–111.

Beveridge, M., Jarrold, C., & Pettit, E. (2002). An experimental approach to executive fingerprinting in young children. *Infant and Child Development*, *11*(2), 107–123.

Bjorklund, D. F., & Harnishfeger, K. K. (1990). The resources construct in cognitive development: Diverse sources of evidence and a theory of inefficient inhibition. *Developmental Review*, *10*(1), 48–71.

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

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. E. (2002). Development of frontal lobe contributions to cognitive control in children: Evidence from fMRI. *Neuron*, *33*, 301–311.

Carlin, D., Bonerba, J., Phipps, M., Alexander, G., Shapiro, M., & Grafman, J. (2000). Planning impairments in frontal lobe dementia and frontal lobe lesion patients. *Neuropsychologia*, *38*(5), 655–665.

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: Stability and prediction from ages 2 to 3. *Developmental Psychology*, *40*(6), 1105–1122.

Case, R., Kurland, D., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*(3), 386–404.

Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: What have we learned about cognitive development? *Trends in Cognitive Sciences*, *9*(3), 104–110.

Casey, B. J., Trainor, R. J., Orendi, J. L., Schubert, A. B., Nystrom, L. E., Giedd, J. N., et al. (1997). A developmental functional MRI study of prefrontal activation during performance of a Go-No-Go task. *Journal of Cognitive Neuroscience*, *9*(6), 835–847.

Caviness, V. S., Kennedy, D. N., Richelme, C., Rademacher, J., & Filipek, P. A. (1996). The human brain age 7–11 years: A volumetric analysis based on magnetic resonance images. *Cerebral Cortex*, *6*(5), 726–736.

Cepeda, N. J., Kramer, A. F., & Gonzalez de Sather, J. C. M. (2001). Changes in executive control across the life span: Examination of task-switching performance. *Developmental Psychology*, *37*(5), 715–730.

Cerella, J., & Hale, S. (1994). The rise and fall in information-processing rates over the life-span. *Acta Psychologica*, *86*(2–3), 109–197.

Chelune, G. J., & Baer, R. A. (1986). Developmental norms for the Wisconsin Card Sorting Test. *Journal of Clinical and Experimental Neuropsychology*, *8*(3), 219–228.

Chugani, H. T., Phelps, M. E., & Mazziotta, J. C. (1987). Positron emission tomography study of human brain functional development. *Annals of Neurology*, *22*(4), 487–497.

Chelune, G. J., & Thompson, L. L. (1987). Evaluation of the general sensitivity of the Wisconsin Card Sorting Test among younger and older children. *Developmental Neuropsychology*, *3*(1), 81–89.

Cohen, R. L., & Heath, M. (1990). The development of serial short-term-memory and the articulatory loop hypothesis. *Intelligence*, *14*(2), 151–171.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, *99*(1), 45–77.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552.

Cowan, N. (1998). Visual and auditory working memory capacity. *Trends in Cognitive Sciences*, *2*(3), 77–78.

Crone, E. A., Bunge, S. A., Van der Molen, M. W., & Ridderinkhof, K. R. (in press). Switching between tasks and responses: A developmental study. *Developmental Science*.

Crone, E. A., Donohue, S. E., Wendelken, C., Honomichl, R., & Bunge, S. A. (in preparation). Prefrontal brain regions contributing to developmental changes in rule use.

Crone, E. A., Wendelken, C., Donohue, S. E., & Bunge, S. A. (2005). Neural evidence for dissociable components of task-switching. *Cerebral Cortex* [July 6 Epub ahead of print].

Culbertson, W. C., & Zillmer, E. A. (1998). The construct validity of the Tower of London (DX) as a measure of the executive functioning of ADHD children. *Assessment*, *5*(3), 215–226.

DeLuca, C. R., Wood, S. J., Anderson, V., Bucanan, J., Proffitt, T. M., Mahony, K., et al. (2003). Normative data from the Cantab: I. Development of executive function over the lifespan. *Journal of Clinical and Experimental Neuropsychology*, *25*(2), 242–254.

Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review*, *12*(1), 45–75.

Dempster, F. N. (1993). Resistance to interference: Developmental changes in a basic processing mechanism. In M. L. Howe & R. Pasnak (Eds.), *Emerging themes in cognitive development, Vol. I. Foundations* (pp. 3–27). New York: Springer.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 466–503). New York: Oxford University Press.

Diamond, A., Briand, L., Fossella, J., & Gehlbach, L. (2004). Genetic and neurochemical modulation of prefrontal cognitive functions in children. *American Journal of Psychiatry*, *161*(1), 125–132.

Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology*, *38*(3), 352–362.

Dolan, C. V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research*, *35*(1), 21–50.

Dolan, C. V., Van der Sluis, S., & Grasman, R. P. P. (2005). A note on normal theory power calculation in SEM with data missing completely at random. *Structural Equation Modeling—A Multidisciplinary Journal*, *12*(2), 245–262.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology*, *30*(3), 257–303.

Durston, S., Thomas, K. M., Yang, Y. H., Ulug, A. M., Zimmerman, R. D., & Casey, B. J. (2002). A neural basis for the development of inhibitory control. *Developmental Science*, *5*(4), F9–F16.

Espy, K. A., & Bull, R. (2005). Inhibitory processes in young children and individual variation in short-term memory. *Developmental Neuropsychology*, *28*(2), 669–688.

Fisk, J. E., & Sharp, C. A. (2004). Age-related impairment in executive functioning: Updating, inhibition, shifting, and access. *Journal of Clinical and Experimental Neuropsychology*, *26*(7), 874–890.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, *3*(11), 410–419.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177–190.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action—Performance of children 31/2–7 years old on a Stroop-like Day–Night Test. *Cognition*, *53*(2), 129–153.

Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *Journal of Experimental Psychology*, *34*, 404–411.

Harnishfeger, K. K. (1996). Inhibitory processes in childhood development. *International Journal of Psychology*, *31*(3–4), 2665.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtis, G. (1993). *Wisconsin Card Sorting Test manual, revised and expended*. Odessa: Psychological Assessment Resources.

Hitch, G. J., Halliday, M. S., Dodd, A., & Littler, J. E. (1989). Development of rehearsal in short-term memory: Differences between pictorial and spoken stimuli. *British Journal of Developmental Psychology*, *7*(4), 347–362.

Huizinga, M., & Van der Molen, M. W. (in preparation-a). Developmental change in switching from color to shape, and from stopping to going.

Huizinga, M., & Van der Molen, M. W. (in preparation-b). Development of random number generation from childhood to young-adulthood.

Huttenlocher, P. R. (1979). Synaptic density in human frontal cortex—Developmental changes and effects of aging. *Brain Research Bulletin*, *163*(2), 195–205.

Jöreskog, K. G., & Sörbom, D. (1999). *LISREL 8*. Chicago: Scientific Software International.

Kail, R. (1991). Processing time declines exponentially during childhood and adolescence. *Developmental Psychology*, *27*(2), 259–266.

Kimberg, D. Y., D'Esposito, M., & Farah, M. J. (1997). Cognitive functions in the prefrontal cortex—Working memory and executive control. *Current Directions in Psychological Science*, *6*(6), 185–192.

Kipp, K. (2005). A developmental perspective on the measurement of cognitive deficits in attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *57*(11), 1256–1260.

Kirk, U., & Kelly, M. A. (1986). Children's differential performance on selected dorsolateral prefrontal and posterior cortical functions: A developmental perspective. *Journal of Clinical and Experimental Neuropsychology*, *7*, 604.

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science*, *6*(5), 449–467.

Klenberg, L., Korkman, M., & Lahti Nuuttila, P. (2001). Differential development of attention and executive functions in 3- to 12-year-old Finnish children. *Developmental Neuropsychology*, *20*(1), 407–428.

Kline, R. T. (1998). *Principles and practice of structural equation modeling*. New York: Guilford.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. *Journal of Cognitive Neuroscience*, *14*(1), 1–10.

Kramer, A. F., Larish, J. F., & Strayer, D. L. (1995). Training for attentional control in dual task settings: A comparison of young and old adults. *Journal of Experimental Psychology: Applied*, *1*(1), 50–76.

Kray, J., Eber, J., & Lindenberger, U. (2004). Age differences in executive functioning across the lifespan: The role of verbalization in task preparation. *Acta Psychologica*, *115*(2–3), 143–165.

Kwon, H., Reiss, A. L., & Menon, V. (2002). Neural basis of protracted developmental changes in visuo-spatial working memory. *Proceedings of the National Academy of Sciences*, *99*(20), 13336–13341.

Larson, G. E., Merritt, C. R., & Williams, S. E. (1988). Information processing and intelligence: Some implications of task complexity. *Intelligence*, *12*(2), 131–147.

Lehto, J. H. (1996). Are executive function tests dependent on working memory capacity? *Quarterly Journal of Experimental Psychology Section a. Human Experimental Psychology*, *49*(1), 29–50.

Lehto, J. H. (2004). A test for children's goal-directed behavior: A pilot study. *Perceptual and Motor Skills*, *98*(1), 223–236.

Lehto, J. H., Juujaervi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*(1), 59–80.

Levin, H. S., Culhane, K. A., Hartmann, J., Evankovich, K., Mattson, A. J., Harward, H., et al. (1991). Developmental changes in performance on tests of purported frontal-lobe functioning. *Developmental Neuropsychology*, *7*(3), 377–395.

Levin, H. S., Fletcher, J. M., Kufera, J. A., Harward, H., Lilly, M. A., Mendelsohn, D., et al. (1996). Dimensions of cognition measured by the Tower of London and other cognitive tasks in head-injured children and adolescents. *Developmental Neuropsychology*, *12*(1), 17–34.

Logan, G. D. (1994). *On the ability to inhibit thought and action: A users' guide to the stop signal paradigm*. San Diego, CA: Academic Press.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*(3), 295–327.

Luciana, M., Conklin, H. M., Hooper, C. J., & Yarger, R. S. (2005). The development of nonverbal working memory and executive control processes in adolescents. *Child Development*, *76*(3), 697–712.

Luciana, M., & Nelson, C. A. (1998). The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologia*, *36*(3), 273–293.

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, *75*(5), 1357–1372.

Luna, B., & Sweeney, J. A. (2004). The emergence of collaborative brain function—fMRI studies of the development of response inhibition. *Adolescent Brain Development: Vulnerabilities and Opportunities*, *1021*, 296–309.

Luria, A. R. (1966). *Higher cortical function in man*. Oxford, England: Basic Books.

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

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

Milner, B. (1963). Effects of different brain lesions on card sorting. *Archives of Neurology*, *9*, 100–110.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Monsell, S. (1996). Control of mental processes. In V. Bruce (Ed.), *Unsolved mysteries of the mind: Tutorial essays in cognition* (pp. 93–148). New York: Erlbaum.

Monsell, S. (2003). Task-set reconfiguration processes do not imply a control homunuculus: Reply to Altmann. *Trends in Cognitive Sciences*, *7*(8), 341–342.

Morris, R. G., Ahmed, S., Syed, G. M., & Toone, B. K. (1993). Neural correlates of planning ability—Frontal lobe activation during the Tower of London test. *Neuropsychologia*, *31*(12), 1367-&.

Muller, U., Zelazo, P. D., Hood, S., Leone, T., & Rohrer, L. (2004). Interference control in a new rule use task: Age-related changes, labeling, and attention. *Child Development*, *75*(5), 1594–1609.

Nagahama, Y., Okina, T., Suzuki, N., Nabatame, H., & Matsuda, M. (2005). The cerebral correlates of different types of perseveration in the Wisconsin Card Sorting Test. *Journal of Neurology Neurosurgery and Psychiatry*, *76*(2), 169–175.

Narayanan, N. S., Prabhakaran, V., Bunge, S. A., Christoff, K., Fine, E. M., & Gabrieli, J. D. E. (2005). The role of the prefrontal cortex in the maintenance of verbal working memory: An event-related fMRI analysis. *Neuropsychology*, *19*(2), 223–232.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical Modeling*. (6th ed.). VCU Box 900126, Richmond, VA 23298: Department of Psychiatry.

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, *126*(2), 220–246.

Nunally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw Hill.

Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal-lobe lesions in man. *Neuropsychologia*, *28*(10), 1021–1034.

Paniak, C., Miller, H. B., Murphy, D., & Patterson, L. (1996). Canadian developmental norms for 9 to 14 year-olds on the Wisconsin Card Sorting Test. *Canadian Journal of Rehabilitation*, *9*, 233–237.

Pennington, B. F. (1997). Dimensions of executive functions in normal and abnormal development. In G. R. Lyon & N. A. Krasnegor (Eds.), *Development of the prefrontal cortex: Evolution, neurobiology, and behavior* (pp. 265–281). Baltimore, MD: Paul H. Brookes.

Raven, J. C., Court, J. H., & Raven, J. (1985). *Raven's progressive matrices*. London: J.C. Raven.

Ridderinkhof, K. R., & Van der Molen, M. W. (1995). A psychophysiological analysis of developmental differences in the ability to resist interference. *Child Development*, *66*(4), 1040–1056.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., McInnes, L., & Rabbitt, P. (1994). Cambridge Neuropsychological Test Automated Battery (Cantab)—A factor-analytic study of a large-sample of normal elderly volunteers. *Dementia*, *5*(5), 266–281.

Roberts, A. C., & Wallis, J. D. (2000). Inhibitory control and affective processing in the prefrontal cortex: Neuropsychological studies in the common marmoset. *Cerebral Cortex*, *10*(3), 252–262.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207–231.

Rushworth, M., Walton, M., Kennerley, S., & Bannerman, D. (2004). The role of medial frontal cortex in making decisions and changing tasks. *International Journal of Psychology*, *39*(5–6), 514.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177.

Schnirman, G. M., Welsh, M. C., & Retzlaff, P. D. (1998). Development of the Tower of London—Revised. *Assessment*, *5*(4), 355–360.

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

Siegel, L. S. (1994). Working memory and reading—A life-span perspective. *International Journal of Behavioral Development*, *17*(1), 109–124.

Smith, E. E., & Jonides, J. (1999). Neuroscience—Storage and executive processes in the frontal lobes. *Science*, *283*(5408), 1657–1661.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215.

Somsen, R. J. M., Van der Molen, M. W., Jennings, J. R., & Van Beek, B. (2000). Wisconsin Card Sorting in adolescents: Analysis of performance, response times and heart rate. *Acta Psychologica*, *104*(2), 227–257.

Sowell, E. R., Thompson, P. M., Leonard, C. M., Welcome, S. E., Kan, E., & Toga, A. W. (2004). Longitudinal mapping of cortical thickness and brain growth in normal children. *Journal of Neuroscience*, *24*(38), 8223–8231.

Span, M. M., Ridderinkhof, K. R., & Van der Molen, M. W. (2004). Age-related changes in the efficiency of cognitive processing across the life span. *Acta Psychologica*, *117*(2), 155–183.

SPSS Inc. (2003). *SPSS Base 12.0 for Windows User's Guide*. SPSS Inc., Chicago, IL.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.

Stuss, D. T. (1992). Biological and psychological development of executive functions. *Brain and Cognition*, *20*(1), 8–23.

Stuss, D. T., & Levine, B. (2002). Adult clinical neuropsychology: Lessons from studies of the frontal lobes. *Annual Review of Psychology*, *53*(1), 401–433.

Stuss, D. T., Levine, B., Alexander, M. P., Hong, J., Palumbo, C., Hamer, L., et al. (2000). Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: Effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia*, *38*(4), 388–402.

Stuss, D. T., Shallice, T., Alexander, M. P., & Picton, T. W. (1995). A multidisciplinary approach to anterior attentional functions. In J. Grafman & K. J. Holyoak (Eds.), *Structure and functions of the human prefrontal cortex. Annals of the New York Academy of Sciences* (pp. 191–211). New York, NY: New York Academy of Sciences.

Teuber, H.-L. (1972). Unity and diversity of frontal lobe functions. *Acta Neurobiologiae Experimentalis*, *32*, 615–656.

Van Boxtel, G. J. M., Van der Molen, M. W., Jennings, J. R., & Brunia, C. H. M. (2001). A psychophysiological analysis of inhibitory motor control in the stop-signal paradigm. *Biological Psychology*, *58*(3), 229–262.

Van den Wildenberg, W. P. M., & Van der Molen, M. W. (2004). Developmental trends in simple and selective inhibition of compatible and incompatible responses. *Journal of Experimental Child Psychology*, *87*(3), 201–220.

Van der Molen, M. W. (2000). Developmental changes in inhibitory processing: Evidence from psychophysiological measures. *Biological Psychology*, *54*(1–3), 207–239.

Van der Sluis, S., De Jong, P. F., & Van der Leij, A. (2006). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. Submitted for publication.

Welsh, M. C. (2002). Developmental and clinical variations in executive functions. In D. L. Molfese & V. J. Molfese (Eds.), *Developmental variations in learning: Applications to social, executive function, language, and reading skills* (pp. 139–185). Mahwah: Lawrence Erlbaum.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology*, *7*(2), 131–149.

Welsh, M. C., Satterlee-Cartmell, T., & Stine, M. (1999). Towers of Hanoi and London: Contribution of working memory and inhibition to performance. *Brain and Cognition*, *41*(2), 231–242.

Wiers, R. W., Gunning, W. B., & Sergeant, J. A. (1998). Is a mild deficit in executive functions in boys related to childhood ADHD or to parental multigenerational alcoholism? *Journal of Abnormal Child Psychology*, *26*(6), 415–430.

Williams, B. R., Ponesse, J. S., Schachar, R. J., Logan, G. D., & Tannock, R. (1999). Development of inhibitory control across the life span. *Developmental Psychology*, *35*(1), 205–213.

Wothke, W. (2000). Longitudinal and multi-group modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wright, I., Waterman, M., Prescott, H., & Murdoch-Eaton, D. (2003). A new Stroop-like measure of inhibitory function development: Typical developmental trends. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *44*(4), 561–575.

Yakovlev, P., & Lecours, A. R. (1967). The myelogenetic cycles of regional maturation of the brain. In A. Minkovsky (Ed.), *Regional development of the brain in early life* (pp. 3–70). Oxford: Blackwell.

Zelazo, P. D., Craik, F. I. M., & Booth, L. (2004). Executive function across the life span. *Acta Psychologica*, *115*(2–3), 167–183.

Zelazo, P. D., Muller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function: Cognitive complexity and control-revised. *Monographs of the Society for Research in Child Development*, *68*(3), 93–119.