Available online at www.sciencedirect.com
ScienceDirect
Intelligence 35 (2007) 427 – 449
INTELLIGENCE

Executive functioning in children, and its relations with reasoning, reading, and arithmetic

Sophie van der Sluis [a,*], Peter F. de Jong [b], Aryan van der Leij [b]

[a] Department of Biological Psychology, Vrije Universiteit, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands
[b] Department of Education, University of Amsterdam, P. O. box 94208, 1090 GE Amsterdam, The Netherlands

Received 26 January 2006; received in revised form 4 September 2006; accepted 4 September 2006
Available online 16 October 2006

Abstract

The aims of this study were to investigate whether the executive functions, inhibition, shifting, and updating, are distinguishable as latent variables (common factors) in children aged 9 to 12, and to examine the relations between these executive functions and reading, arithmetic, and (non)verbal reasoning. Confirmatory factor analysis was used to decompose variance due to the executive and the non-executive processing demands of the executive tasks. A Shifting factor and an Updating factor, but not an Inhibition factor, were distinguishable after controlling for non-executive variance. Updating was related to reading, arithmetic, and (non) verbal reasoning. Shifting was mainly related to non-verbal reasoning and reading. However, in terms of variance explained, arithmetic and reading were primarily related to the non-executive processing demands of the executive measures. The results are discussed in light of the "task impurity problem".

Keywords: Executive functions; Inhibition; Shifting; Updating; Arithmetic; Reading; Reasoning

Executive functions (EFs) are defined as the routines responsible for the monitoring and regulation of cognitive processes during the performance of complex cognitive tasks (e.g., Lindsay, Tomazic, Levine, & Accardo, 1999; Miyake et al., 2000). In neuropsychological settings, executive tasks are often used as diagnostic instruments, and there is abundant evidence that disorders of executive control are associated with damage to the frontal lobes (e.g., Baddeley, 1996; Rabbitt, 1997, but see also Alvarez & Emory, 2006). In children, the majority of studies has been concerned with the comparison of the executive capacity of clinical and non-clinical samples (e.g., Bull & Scerif, 2001; Everatt, Warner, Miles, & Thomson, 1997; Helland & Asbjørnsen, 2002; Sergeant, Geurts, & Oosterlaan, 2002). Therefore, one aim of the current study is to examine the structure of executive functions in normal children.

In general, the study of executive functioning is far from easy. One of the fundamental problems in the measurement of executive functioning is the 'task impurity problem' (e.g., Denckla, 1994; Rabbitt, 1997). Because EFs need a task framework to become manifest, executive tasks always implicate other, non-executive cognitive abilities such as verbal ability, motor speed, or visual–spatial ability. In addition, executive tasks often require more than one EF. Because executive tasks are complex and multi-cognitive in nature (i.e., they are 'impure'), and because they differ greatly in

* Corresponding author.
E-mail address: S.van.der.sluis@psy.vu.nl (S. van der Sluis).

0160-2896/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2006.09.001

their background demands (Burgess, 1997), performance on executive tasks cannot readily be attributed to the absence or presence of a given executive capacity. Likewise, when a relationship is observed between the performance on an executive task and the performance on other cognitive measures, it is unclear whether this relationship is due to the executive or the non-executive processing demands of the executive task.

The impurity and complexity of executive tasks are psychometric problems, which complicate the interpretation of findings, and thus hinder hypothesis testing. Therefore, here we examine the structure of executive functioning in normal children, while addressing the more general problem of task impurity. In addition, as a second aim, the relations of executive and non-executive performance with verbal and non-verbal reasoning ability, reading ability, and arithmetic ability are explored. Below, we first discuss the recent literature on the structure of executive functioning, and the relations of the specific EFs with reasoning ability, reading, and arithmetic ability.

1. Measurement problems and the structure of executive functioning

The term executive functioning pertains to a wide variety of conscious, deliberate, meta-cognitive processes, such as planning, organized search, impulse control, goal directed behavior, set maintenance, flexible strategy employment, selective attention, attentional control, initiation of actions, fluidity, self-evaluation, and dual task performance (e.g., Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Sikora, Haley, Edwards, & Butler, 2002; Wu, Anderson, & Castiello, 2002). Within this profusion of terms, three EFs are generally acknowledged as important, because they are lower-level (i.e., supposedly implicated in performance on complex executive tasks), and relatively well-defined: shifting, inhibition, and updating (e.g., Baddeley, 1996; Miyake et al., 2000; Rabbitt, 1997).

Shifting is defined as the ability to switch between sets, tasks, or strategies, i.e., the disengagement of an irrelevant task set, and the subsequent initiation of a new, more appropriate set. For example, in the Number–Letter task (Miyake et al., 2000), subjects need to switch between judging digits (odd vs. even) and letters (consonant vs. vowel), depending on where these symbols are located on a monitor. Several subtypes of *inhibition* have been distinguished (e.g., Friedman & Miyake, 2004; Nigg, 2000). However, in studies on executive functioning, like the present, the focus is on the ability to deliberately suppress dominant, automatic, or prepotent responses in favor of more goal-appropriate ones.[1] In the Stroop task (Stroop, 1935), for example, subjects are presented with color-words that are printed in incongruent ink colors (e.g., the word 'red' printed in green), and are instructed to name the ink color and to inhibit the automatic tendency to read the word. *Updating* is defined as the ability to monitor and code incoming information, and to update the content of memory by replacing old items with newer, more relevant, information. Updating thus concerns the dynamic, goal directed manipulation of memory content. An example of an updating task is the Digit Monitoring task (Salthouse, Atkinson, & Berish, 2003), where subjects are presented with series of digits, and are asked to respond to each third odd digit by pressing 'Z' on a keyboard, and to all other digits by pressing 'M'.

Whether theoretically distinguishable EFs are actually discernible as distinct factors in factor analysis is an important question. Most studies of executive control in children used exploratory factor analysis (EFA) to address this question (e.g., Anderson, Anderson, Northam, Jacobs, & Catroppa, 2001; Klenberg, Korkman, & Lahti-Nuuttila, 2001; Levin et al., 1991; Welsh, Pennington, & Groisser, 1991). However, the factorial solutions reported differ with respect to both the number and the interpretation of extracted factors. For instance, Levin et al. (1991) reported a solution with three factors, which the authors interpreted as 'semantic association and concept formation', 'freedom from perseveration', and 'planning and strategy'. Klenberg et al. (2001) reported a four-factor solution, with factors interpreted as 'fluency', 'selective visual attention', 'selective auditory attention', and 'simple motor inhibition'.

The variation in the results of these studies is partly due to the use of different test batteries. In addition, the wide age-ranges of the samples used in these EFA studies (e.g., 7–15 years, Levin et al., 1991; 3–12 years, Klenberg et al., 2001; 11–17 years, Anderson et al., 2001) may also constitute problem. However, a specific problem with studies that have used EFA is that the non-executive processing demands of the executive tasks may influence the factor structure. The executive demands of a task refer to the monitoring and regulatory processes that the task requires, while the *non*-executive demands of a task refer to all other abilities that are

[1] Types of inhibition that cannot be considered deliberate, such as negative priming (longer reaction times in response to recently ignored or suppressed stimuli) and reactive inhibition (tendency to suppress previous responses) are usually not regarded as executive in nature (Miyake et al., 2000).

needed to perform the task. For instance, when performing the aforementioned Stroop task, the color naming speed and the speed of word identification constitute the non-executive task demands, while suppression of the tendency to read the words in favor of naming the ink colors constitutes the executive demand. Executive performance thus includes the monitoring and regulation of the non-executive responses. Since executive performance can only be measured indirectly, i.e., within a task framework, executive tasks invariably involve executive as well as non-executive abilities. As a consequence, executive tasks may cluster together in EFA because of their resemblance with respect to non-executive demands. Accordingly, the factor structure will, at least in part, reflect non-executive functioning. The factors distinguished by Klenberg et al. (2001) illustrate this drawback, as these factors mostly reflected the tasks' measurement formats (e.g., the visual attention factor was indicated by the two visual inspection tests, and the auditory attention factor by the two tasks that required listening).

The interpretation of factors extracted by means of EFA is thus hindered by the impurity of executive tasks. As Miyake et al. (2000) put it 'Another related consequence of the unclarity as to what these executive tests really measure is the difficulty of interpreting what construct(s) different factors obtained in many EFA studies of executive functions really represent. Interpretations given to obtained factors often seem quite arbitrary and posthoc' (page 53). Two methods have been proposed to address the problem of task impurity. First, the use of confirmatory factor analysis (CFA) has been suggested, as CFA extracts the variance common to executive tasks that are *presupposed* to call on the same underlying EF (e.g., Rabbitt, 1997). The assumption is that when several tasks are used to measure inhibition, and inevitably a variety of non-executive abilities, the common factor underlying performance on these tasks will be a more pure measure of inhibition than the separate tasks indicating the factor.

Miyake et al. (2000) used CFA to study the distinctiveness of the functions inhibition, shifting, and updating in adults. They arrived at a model with three correlated latent factors representing EFs. Since models with one general executive factor, or two collapsed factors fitted the data significantly worse, the authors concluded that inhibition, shifting, and updating were indeed distinguishable, yet correlated EFs. Similar results with CFA were reported by Lehto et al. (2003) in a sample of children aged 8 to 13. In another study with children aged 6 to 16, Manly et al. (2001) used CFA to distinguish between the factors selective attention, attentional control/switching, and sustained attention. These factors can be considered comparable to inhibition, shifting, and updating, respectively.

The above-mentioned CFA studies seem to suggest that at least some EFs may be viewed as distinct constructs in children as well as adults. Yet, Lehto et al. (2003, p. 75) stressed that the use of different task sets, and the impure nature of executive tasks complicate the comparison of CFA solutions across studies and age groups. In addition, it should be mentioned that in all three studies, the loadings of the executive tasks on the hypothesized constructs were rather low, implying that most of the variance in the executive tasks (in many cases more than 70%) remained unaccounted for.

Furthermore, like in EFA, interpretation of the factors extracted through CFA may not always be straightforward when impure tasks are used. For example, in the study by Lehto et al. (2003), the latent factor Inhibition was indicated by the Tower of London (TOL) and the Matching Familiar Figures Test (MFFT). In the literature, the TOL is viewed as a measure of planning, monitoring, self-regulation, and problem solving (Klenberg et al., 2001), that also calls on visual perception, attention and working memory (Sikora et al., 2002). The MFFT is viewed as a measure of visual search, hypothesis testing, impulse control, and inhibitory processes (Welsh et al., 1991). The proposition that the common factor underlying performance on these tests is indicative of inhibitory ability may therefore not be justified.

As a second method of dealing with the task impurity problem, various authors (e.g., Denckla, 1996; Pennington & Ozonoff, 1996; Scheres et al., 2004; Sergeant et al., 2002) have advocated the use of control tasks. In this design, performance on a control task is compared to the performance on an executive task, which only differs from the control task in its additional call on a given EF. In subsequent analyses, the focus is on the difference in performance on the EF task and its control counterpart (i.e., difference scores), or on the variance in the EF task that could not be explained by its control counterpart (i.e., regression residuals).

The importance of accounting for performance on control tasks was frequently demonstrated. For example, while many authors report greater Stroop-like interference in learning disabled children (e.g., Bull & Scerif, 2001; Everatt et al., 1997), children with ADHD (see Sergeant et al., 2002 for a review), and older people (e.g., Christ, White, Mandernach, & Keys, 2001; West & Baylis, 1998), few found evidence for differential inhibitory ability once initial differences in basic naming speed, as measured by the control task, were controlled for (e.g., Carter, Krener, Chaderjian, Northcutt, & Wolfe, 1995; Graf & Uttl, 1995; Seidman, Biederman, Monuteaux, Weber, & Faraone, 2000; Uttl & Graf, 1997; van der Sluis, de Jong, & van der Leij, 2004; Willcutt et al., 2001). These findings again underline the seriousness of the task impurity problem, as individual differences in performance on executive tasks may be indicative of differences in executive ability, but also of differences in the ability to handle the task's non-executive requirements.

In sum, the structure of executive functioning has been studied in children, but the use of impure tasks complicates the interpretation of the extracted factors. In our view, the issue of task impurity should be addressed explicitly in studies on the factor structure of executive functioning.

2. EFs and their relations with reasoning, reading, and arithmetic

EFs, considered as lower-level abilities, have been hypothesized to underlie a range of higher-order cognitive abilities, such as reasoning, reading, and arithmetic. As mentioned before, in most studies involving children, the executive performance in samples of normal children is compared to that of children in clinical samples, e.g., children with learning deficits in reading and/or arithmetic (e.g., Bull, Johnston, & Roy, 1999; Bull & Scerif, 2001; Sikora et al., 2002; van der Sluis et al., 2004; van der Sluis, van der Leij, & de Jong, 2005), or children with attention deficit/hyperactivity disorder (e.g., Scheres et al., 2004; Shallice et al., 2002; Sonuga-Barke, Dalen, Daley, & Remington, 2002; Stevens, Quittner, Zuckerman, & Moore, 2002; Wu et al., 2002). The general hypothesis of these studies is that defects in executive functioning may be (causally) associated with children's learning or attentional problems. Consequently, measures of executive performance could be used for diagnostic purposes, and to pinpoint the specific cognitive deficits that underlie these problems.

Shifting ability is believed to be involved in arithmetic performance by supporting alternation between arithmetic strategies (e.g., addition, subtraction, multiplication), and arithmetic sub-solutions in multi-step arithmetic problems. Children with arithmetic deficits have been shown to display poorer shifting ability, especially using measures derived from complex shifting tasks (e.g., Bull et al., 1999; Bull & Scerif, 2001; McLean & Hitch, 1999; van der Sluis et al., 2004). Children with reading deficits often display shifting ability similar to that of control children, once differences in various covariates (e.g., IQ, general naming speed, or ADHD symptoms) are taken into account (e.g., Klorman et al., 1999; van der Sluis et al., 2004; Weyandt, Rice, Linterman, Mitzlaff, & Emert, 1998; Willcutt et al., 2001). Likewise, the inhibitory performance of children with reading and/or arithmetic related deficits is often found to be similar to that of control children, once initial differences in covariates (e.g., IQ, general naming speed or numerical skills) are controlled for (e.g., Bull & Scerif, 2001; Everatt et al. 1997; Helland & Asbjørnsen, 2000; Sikora et al., 2002; van der Schoot, Licht, Horsley, & Sergeant, 2000; van der Schoot, Licht, Horsley, & Sergeant, 2002; van der Sluis et al., 2004; Willcutt et al., 2001).

Updating ability has not been studied in relation to reasoning, arithmetic or reading. Studies on working memory capacity (WM) may however be informative about these relations, as WM tasks and updating tasks share the requirement to store information, and to revise the content of memory in the light of new information. In both children and adults, WM correlates strongly to fluid and crystallized intelligence (usually $r>.70$, e.g., Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004; de Jong & Das-Smaal, 1995; Kyllonen, 1993; Kyllonen & Crystal, 1990; Süβ, Oberauer, Wittman, Wilhelm, & Schulze, 2002). Also, WM is believed to support the encoding process of letters, words, and sentences (e.g., de Jong, 1998; Siegel & Ryan, 1989), and WM is thought to support the memorization of numbers and sub-solutions during arithmetic (e.g., Geary, 1993; McLean & Hitch, 1999). In comparison to that of controls, the WM performance of children with reading deficits (e.g., de Jong, 1998; Howes, Bigler, Burlingame, & Lawson, 2003; Howes, Bigler, Lawson, & Burlingame, 1999; Siegel & Ryan, 1989; Swanson, 1999; Swanson & Ashbaker, 2000) or arithmetic deficits (e.g., McLean & Hitch, 1999; Siegel & Ryan, 1989; Swanson, 1994) is often poorer. Yet, again, these differences usually disappear once age, IQ, and the presence of other learning deficits are controlled for (e.g., Bull et al., 1999; van der Sluis et al., 2005; Wimmer & Mayringer, 2002).

The repeated finding that group differences in executive performance are largely diminished once differences in non-executive abilities are controlled, raises questions concerning the discriminant validity of EFs. In this respect a study of Salthouse et al. (2003), with adults, is of interest. Salthouse et al. used CFA to explicitly test the construct validity of inhibition, updating and time-sharing in the context of the well-established constructs fluid intelligence, memory, speed, and vocabulary. It was shown that the executive tasks shared very little residual variance once relations with the non-executive abilities were controlled. The authors concluded that evidence for construct validity of the EFs was "weak to non-existent" (p. 588).

Summarizing, inhibition and shifting, but not updating, have been studied in relation to reading and arithmetic. In these studies, intelligence featured as a covariate but its relation with the individual EFs was not, as in the current study, the topic of investigation. Also, the focus has been on differences in executive performance between clinical and non-clinical samples. The relations of EFs with higher-order cognitive abilities have not been studied in normally functioning children, even though higher-order abilities are considered core abilities, showing considerable individual differences in non-clinical populations. In light of the task impurity issue, the question remains whether the executive or the non-executive aspects of executive measures underlie the relation between performance on executive tasks and higher-order cognitive functioning. The frequent finding that differences in executive functioning between children with learning and/or behavior problems and their normal functioning peers disappear after controlling for differences in non-executive abilities, suggests that the relations emanate from the non-executive rather than the executive processing demands. However, as said, this issue has yet to be addressed in normal children.

3. The present study

The goal of the present study was twofold. First, we wanted to examine whether the three EFs inhibition, shifting, and updating were distinguishable as distinct EFs in normally functioning children, once non-executive variance was controlled for. We argue that specific executive factors merit interpretation as distinct conceptual entities, only if these factors can be detected after individual differences in non-executive processing demands are accounted for. A similar argument was made previously in studies on, for instance, intelligence and working memory (e.g., Beier & Ackerman, 2004; Gustafsson, 1992; Gustafsson & Undheim, 1992; Hertzog & Bleckley, 2001; Oberauer, Süβ, Wilhelm, & Wittman, 2003; Stankov, 2000). Second, we wanted to investigate in normally functioning children whether performance on executive tasks is related to reasoning, reading, and arithmetic through the executive, or through the non-executive task demands.

In the current study, we combined the two methods of dealing with the task impurity problem, i.e., we used control tasks in the context of CFA. That is, CFA was used to differentiate between the non-executive variance, common to both control and executive tasks, and the executive variance, common only to the executive tasks. By combining control tasks and CFA, variance in performance on executive tasks that remains unexplained after accounting for the performance on the control tasks, can be attributed more confidently to the considered EFs.

For the present study, we selected well-known executive tasks (or adaptations thereof) that were amenable to administration in a control condition. The present method of distinguishing between executive and non-executive variance can be likened to a multi-trait multi-method (MTMM) method (Campbell & Fiske, 1959). Specifically, the non-executive processing demands are considered the 'method', and the executive demands the 'traits'. The MTMM method requires tasks that assess similar traits, with different methods, and tasks that involve similar methods, but assess different traits. Here, we applied a multi-trait *mono*-method model, because only one method was common to many well-known executive tasks, i.e., the (rapid) naming design.

Tasks that require the rapid naming of stimuli have proven useful as measures of both inhibition and shifting in previous studies (Anderson et al., 2001; Bull & Scerif, 2001; van der Sluis, et al., 2004). These rapid naming tasks can be viewed as variations of the Stroop task, which is a well-established measure of inhibition (see MacLeod, 1991, for a review). All naming-based tasks consisted of a control task, and a manipulated task that differed from the control task only in its additional appeal to an EF.

This speeded naming design however, is not practicable for measuring updating ability, since failing updating ability is most apparent in an increase in response errors, and not in a decrease in response speed. We therefore did not force updating ability into the rapid naming format, but used adaptations of the updating tasks Keep Track and Letter Memory, previously used by Miyake et al. (2000). As these updating tasks too required the naming of stimuli, the control tasks of the inhibition and shifting measures could be used to account for these non-executive requirements in the updating tasks.

4. Method

4.1. Participants

Parental consent was obtained for 172 children (84 boys, and 88 girls) to participate in this study. The children attended grade 4 ($n=100$) and grade 5 ($n=72$) of six primary schools for regular education in the urban regions of Amsterdam and Amstelveen (the Netherlands).[2] Mean age of the sample was 128.08 months (SD = 8.65).

4.2. Tasks

All children were tested for inhibition, shifting, and updating ability, for verbal and non-verbal reasoning, and for arithmetic and reading ability. Eleven executive tasks were administered: four inhibition[3] tasks, four shifting tasks, and three updating tasks. For all three executive functions, we selected well-known measures, and, where necessary, adapted them to ensure that they were suitable for children. Of the eleven executive tasks, seven had a rapid naming format (3 shifting tasks, and 4 inhibition tasks), one task was speeded, but did not require naming (a shifting task), and three were non-speeded, but did require the naming of stimuli (3 updating tasks).

All seven executive tasks with a rapid naming format had a similar structure. For these tasks, simple rapid naming tasks were administered as control tasks. Manipulated rapid naming tasks, which appealed either to inhibition or shifting, did not differ from these simple tasks in any way, except for the requirement to inhibit or shift. The variance of the manipulated tasks that could not be explained by performance on the control tasks, was taken as indicative of shifting or inhibitory ability.

All naming tasks (control and manipulated) consisted of a card with 40 stimuli (5 rows of 8 stimuli each). In all naming tasks, (combinations of) 4 different stimuli were used, which were presented in random order, and appeared approximately equally often. Participants were instructed to name the stimuli as fast as they could without making errors, but were allowed to correct naming errors, if they noted them. The extra time taken to correct naming errors was not recorded separately. For both the control and the manipulated naming tasks, scores consisted of the number of items named per second (i.e., 40 divided by time on task).

Example items preceded all naming tasks. The examples were used to familiarize the children with the task requirements. All examples consisted of 2 rows of 8 stimuli. The experimenter read the instructions aloud, named the first 4 stimuli of the example, and then asked the child to complete the example. If children were noticeably in any doubt about the instructions, the example was presented again.

All control, inhibition, and shifting tasks are illustrated in Fig. 1.

4.2.1. Inhibition

Four measures of inhibition were administered, which were all based on simple naming tasks. The variance in the inhibition-loaded tasks that could not be explained by the control tasks, was attributed to the additional inhibition requirement, and considered indicative of inhibitory ability.

4.2.1.1. Quantity inhibition. This task was derived from Bull and Scerif (2001). The control task (Quantity) consisted of series of small triangles ranging in number between 1 to 4. The children were required to name the number of triangles in a series. In the manipulated version, the Quantity-I task, children were presented with series of digits instead of triangles. Again, they had to name the number of stimuli within the series. For example, if the series '444' was presented, the correct answer was '3', i.e., the numerical denotation of the digits had to be inhibited in favor of the quantity of digits.

4.2.1.2. Object inhibition. This task was derived from van der Sluis et al. (2004). In the control task (Objects), children were required to rapidly name the geometrical figures circle, square, triangle, and diamond, which were printed in a heavy, black line. In the manipulated version (Objects-I), the same four geometrical objects were presented but now an additional smaller object was placed in varying positions within the larger object. These smaller objects were printed in a light, gray line. Children were instructed to name the smaller, less obtrusive object, i.e., to ignore the larger, prepotent figure in favor of the smaller, less noticeable one.

4.2.1.3. Stroop. The Stroop Color-Word test (Stroop, 1935) was used, and slightly adjusted, so that it was similar in format to the other executive naming tasks (i.e., 5 × 8 stimuli, 2 practice rows). The control task (Color) required the naming of the ink color of colored

[2] No additional information was available on SES, ethnicity, or attendance at learning support classes. However, all children attended schools for regular education in favorable neighbourhoods of Amsterdam and Amstelveen. The present sample is unlikely to include children with multiple learning disabilities or behavioral problems as such characteristics are reason for reference to special education in the Netherlands.

[3] Initially, a fifth inhibition task was included in the study, but this task was excluded from further analysis as the inhibition manipulation of this task failed to bring about the intended experimental effect (i.e., performance on the manipulated task did not differ from the performance on the control task).

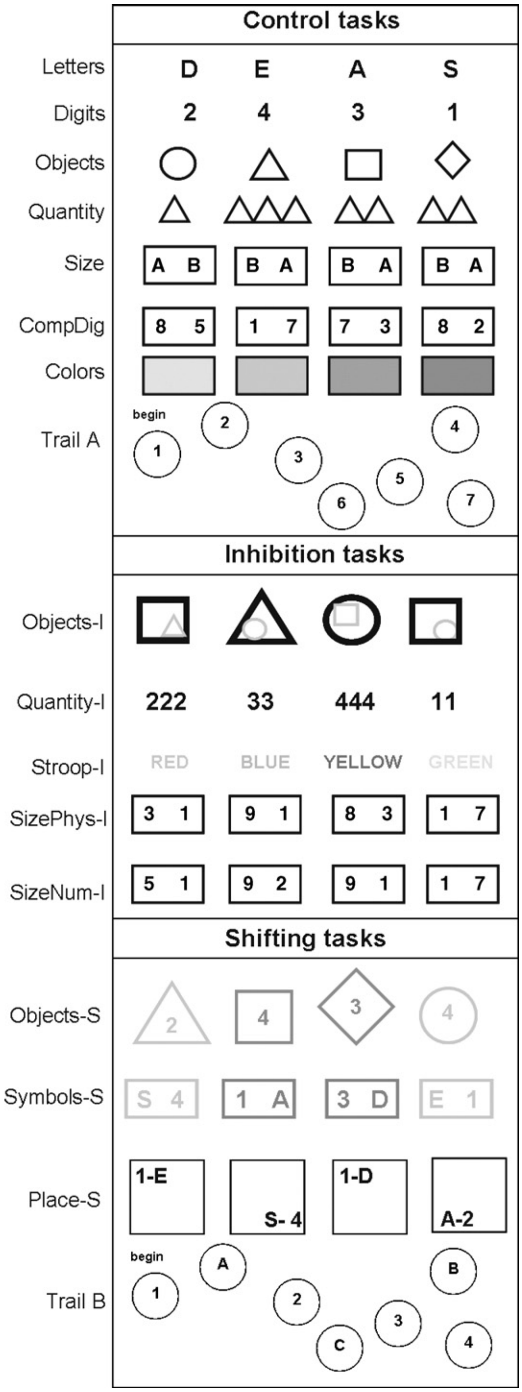

Fig. 1. Examples of the 7 control tasks, 4 inhibition tasks, and 4 shifting tasks. The different shades of gray illustrate the different ink colors in Stroop-I task, Objects-S and Symbols-S.

patches (yellow, red, green, and blue). In the manipulated task (the original color-word task, denoted here as Stroop-I), the color-words ‘red’, ‘blue’, ‘green’, and ‘yellow’ were printed in incongruent ink colors (red, blue, green, and yellow). Children were required to name the color in which these words were printed, and to inhibit the tendency to read the words.

4.2.1.4. Numerical size inhibition. This task was derived from Butterworth (1999). In the control task (CompDig), pairs of single digits (ranging between 1 and 9) were presented. The pairs were enclosed by a rectangle so that each pair was clearly separated from other pairs. The children were instructed to name the numerically largest digit of a pair. In the manipulated task (SizeNum-I), one of the digits was printed in a large font (Courier 28), and one in a smaller font (Courier 20). The children had to name the *numerically* largest digit of a pair, which was always the physically smallest. That is, they had to ignore the larger, more prepotent digit in favor of the smaller one.

4.2.2. Shifting

Three naming-based shifting tasks were administered. The fourth shifting task was speeded, but did not require naming. The variance in the shifting-loaded tasks that could not be explained by the control tasks, was attributed to the additional shifting requirement, and thus considered indicative of shifting ability.

4.2.2.1. Objects shifting. This task was derived from van der Sluis et al. (2004). Two control tasks are administered: the Objects naming control task as described before, and a digit naming control task (Digits), which required naming of numbers (1, 2, 3, and 4). In the manipulated version (Objects-S), a digit (1 to 4) was placed in the center of the four geometrical objects. Depending on the color of the stimulus, children were to name either the object (when blue) or the digit (when yellow).

4.2.2.2. Symbol shifting. The two control tasks included Digits (as described above), and the Letters task, which required the naming of the letters A, E, S, or D. In the manipulated version (Symbols-S), a letter and a digit were presented in pairs, which were enclosed by a rectangle. Depending on the color of the stimulus, children were required to name the letter (when blue) or the digit (when yellow).

4.2.2.3. Place shifting. This task was adapted from the Numbers–Letters task as used by Rogers and Monsell (1995), and Miyake et al. (2000). Like the Symbols-S task, Letters and Digits were used as control tasks. In the manipulated version (Place-S), a letter and a digit were presented in pairs in one of the corners of a square.

Naming depended on the position of the pair in the square; children were to name the letter when the pair was printed in one of the top corners, and the digit when the pair was printed in one of the bottom corners.

4.2.2.4. Making trails task. This task was adopted from the Making Trails task as previously used by, e.g., McLean and Hitch (1999). The control task (Trail A) is the only control task that does not require naming, although it is speeded, like the other control tasks. Children were presented with a card containing 22 numbered circles. The instructions were to start at the circle containing the number '1', and to connect the circles with a pencil in ascending order (i.e., 1-2-3-4-5 etc.) as quickly as possible. In the manipulated version (Trail B), children were presented with a card with 22 circles carrying either a digit or a letter. The children were to start at the circle containing '1', and make a trail, alternately connecting digits and letters in ascending order (i.e., 1-A-2-B-3-C, etc.) as quickly as possible. Both Trail A and Trail B were preceded by an example trail of 8 circles. Scores on the Making Trails tasks consisted of the number of circles connected per second (i.e., 22 divided by the time on task).

4.2.3. Updating

Three tasks were used as indicators of updating ability. These tasks were adaptations of tasks used by Miyake et al. (2000), who in turn adapted their tasks from Morris and Jones (1990), and Yntema (1963).

4.2.3.1. Keep track. In the Keep Track task (Ktrack-U), children were presented with 8 series of 10 cards. On each card, a symbol or picture was shown that belonged to one of the five following categories: Letters (A, E, S, or D), Digits (1, 2, 3, or 4), Objects (circle, square, triangle, or diamond), Animals (cat, dog, bird, or fish), or Vehicles (car, bike, train, or airplane). Before the start of each series, the experimenter designated 3 or 4 'target categories' for that specific series. Each series included 1, 2, or 3 cards from these target categories, supplemented by cards from the other categories. The children were instructed to name all pictures in the series, and to keep track of the target categories. At the end of a series, they were asked to name the last picture that they had seen of each of the target categories. For example, when the target categories were letters and animals, and the series presented was 'E, cat, bike, square, airplane, fish, triangle, A, 1, train', the correct answer was 'A and fish' (the order of recall was free). Of the 8 series, the first 4 required keeping track of 3 target categories, and the last 4 required keeping track of 4 target categories. Thus, in total, 28 pictures had to be recalled. The test score was the proportion of pictures recalled correctly.

Before the beginning of the task, participants were shown all 5 target categories, and the 4 symbols/pictures in each category, to ensure that they were familiar with the depicted objects, and knew to which category they belonged. To familiarize the children with the task, two practice series of 6 pictures and 2 target categories were presented. All symbols/pictures were black and white, and printed in a landscape, A4 size booklet. Within a series, the experimenter turned the pages at a rate of approximately 2 s per page. As a reminder, the target categories remained visible in the left bottom corner of each page within a series.

4.2.3.2. Letter memory. In the Letter Memory task (Letters-U), subjects were presented with series of letters, which varied in length from 3, 5, 7, to 9 letters. Children were asked to constantly recall the last three letters presented to them. That is, they were instructed to update throughout a series, continually adding the last presented letter to a cluster and dropping the earlier fourth letter. For example, when a series included the letters 'R, B, K, A, G', subjects were required to respond 'R', 'RB', 'RBK', 'BKA', and 'KAG'. The length of the series varied unpredictably for the subjects. The test consisted of 8 series total (2 series of each length), resulting in 48 clusters of letters to be recalled. The proportion of clusters recalled correctly, was used as the test score. Before the test began, three practice series (2 of length 3, 1 of length 5) were presented to illustrate the task. The letters used in the Letters-U task were A, O, H, K, B, G, R, T, and S. The letters were printed in Arial 120 in a landscape, A4 size booklet. Within a series, experimenters turned the pages at a rate of approximately 2 s per page.

4.2.3.3. Digit memory. The Digit Memory task (Digits-U) was a numerical version of the Letters-U task. Now, series (length 3, 5, 7, or 9) of digits (1 to 9) were presented, and children were again instructed to update throughout a series. Again, 48 clusters of digits had to be recalled, and the proportion of clusters recalled correctly was used as the test score. The digits were printed in Arial 120 in a landscape, A4 size booklet, and within a series, experimenters turned the pages at a rate of 2 s per page. Like with Letters-U, three practice series were presented (2 series of 3 digits, and 1 series of 5 digits).

4.2.4. Reasoning, arithmetic, and reading

Four standardized tests were administered to measure verbal and non-verbal reasoning ability, arithmetic ability, and reading ability.

4.2.4.1. Verbal reasoning. The Verbal Analogies subtest of the RAKIT, a Dutch intelligence test for children (Bleichrodt, Drenth, Zaal, & Resing, 1987), was used to assess verbal reasoning ability. This test consists of 30 multiple-choice items of the format 'A is to B, as C is to ?', and children were required to select the correct answer from four answer options. Pictures illustrate all words in the test. Completion of the test took 30 min at most.

4.2.4.2. Non-verbal reasoning. The Raven Standard Progressive Matrices (Raven, Court, & Raven, 1979) were used as a measure of non-verbal reasoning ability. The Raven consists of sixty series of abstract patterns from which a piece is missing. Children were required to select the missing piece from a set of 6 to 8 answer options. Completion of this test took 45 min at most.

4.2.4.3. Reading. The One Minute Reading Test (OMRT, Brus & Voeten, 1979) was used as a measure for word reading efficiency. This test is used as a standard measure of early reading acquisition in Dutch education. The OMRT consists of a list of 116 unrelated words of increasing difficulty. Children were instructed to read as many words as they can in 1 min without making errors. The test score consisted of the number of words read correctly in 1 min.

4.2.4.4. Arithmetic. As a measure of arithmetic ability, the Arithmetic Tempo Test (ATT, De Vos, 1992) was administered. Children were presented with three sets of 50 arithmetic problems. The first set requires addition, the second subtraction, and the third multiplication (i.e., three homogeneous sets). All problems consist of 2 numbers, both ranging between 0 to 99. Within a set, problem difficulty increases. The first approximately 20 problems of each set cover arithmetic facts. If arithmetic ability is automatized, these problems thus call mainly on arithmetic fact retrieval. Per set, children were instructed to solve as many problems as possible within 3 min. The test score consisted of the mean number of problems solved correctly over the three subtests.

4.3. Procedure

The first test day, children were tested group-wise for arithmetic ability and verbal reasoning ability, and approximately half of the children were individually tested for reading ability. On the second test day, the test for non-verbal reasoning was administered group-wise, and the remaining children were tested for reading ability. Thereafter, all children were individually tested in three sessions lasting approximately 25 min each. During each session, one updating task was administered, and 6 or 7 naming tasks (control or manipulated). The sessions took place in a period of 3 days to 3 weeks. To prevent order effects, the 7 control tasks and 11 experimental tasks were presented in three different orders, which were composed such that tasks demands were least likely to interfere. Graduate students, who were trained prior to testing, administered the tests.

5. Results

The results are presented in four sections. In the first section, descriptive statistics of all measures are discussed. The second section contains the results of ANOVA's for repeated measures, which were carried out to test the effectiveness of the inhibition and shifting manipulations. Section three includes the results of the confirmatory factor analysis (CFA) using LISREL 8.50 (Jöreskog & Sörbom, 2001). The question whether the executive functions inhibition, shifting, and updating are detectable once non-executive variance is controlled for, is central to this section. In the fourth section, the latent executive factors are related to reading and arithmetic ability, and verbal and non-verbal reasoning ability.

5.1. Descriptive statistics

Of the 172 children tested, 2 had missing values on one measure due to procedural errors (1 on Symbols-S, 1 on Letters). Scores on the naming tasks were recoded as missing if 25% or more of the stimuli were named incorrectly, because naming time can only be considered a valid score if the number of naming errors is limited. Overall, the number of naming errors was small. For most tasks, more than 95% of participants made 2 naming errors or less. For Place-S and Symbols-S, 90% of the participants made 2 naming errors or less. Four children named more than 25% of the stimuli incorrectly on Place-S (3) and SizeNum-I (1). One child named more than 25% of the stimuli incorrectly on both Place-S and Symbols-S. Finally, nine scores were recoded as missing because these were outlying, i.e., more than 3 SDs from the mean. Taken together, this resulted in 17 missing scores, which is less then .5% of the data. Little's MCAR test showed that this missingness could be considered completely at random ($\chi^2(251)=282.95$, ns).

The descriptive statistics for all measures are presented in Table 1. For all speeded tasks, lower scores represent slower performance (i.e., less items named or

Table 1
Descriptive statistics for reading, arithmetic, verbal and non-verbal reasoning, 7 naming speed control tasks, 4 inhibition tasks, 4 shifting tasks, and 3 updating tasks

Test	*M*	SD	Skew.	Kurt.	*n*
Reading	65.81	12.67	−.002	.366	172
Arithmetic	35.98	6.30	−.135	−.261	172
V. reasoning	24.66	3.62	−.867	.362	171
Nonv. reasoning	42.36	7.07	−.530	.203	171
Control tasks					
Letters	2.14	.38	.058	.346	171
Digits	2.27	.45	.281	.482	172
Objects	1.05	.22	.494	.920	172
Quantity	1.64	.27	.374	−.120	172
Color	1.65	.29	−.100	.292	172
Trail A	.98	.36	.661	.027	172
CompDig	1.21	.20	.150	.330	172
Inhibition					
Quantity-I	1.10	.20	.546	.523	172
Objects-I	.85	.17	.121	−.286	172
Stroop-I	.98	.21	.407	.490	172
SizeNum-I	1.16	.27	.787	.526	171
Shifting					
Objects-S	.98	.20	.211	.203	172
Place-S	.81	.17	.334	.026	168
Symbols-S	.84	.16	.436	.066	170
Trail B	.50	.17	.695	.493	169
Updating					
Ktrack-U	70.37	14.63	.163	−.614	172
Letters-U	94.34	6.97	−1.678	2.599	172
Digits-U	96.18	5.36	−2.103	5.154	172

Note. Scores on the control, inhibition, and shifting tasks are number of items named or connected per second. Scores on the updating tasks are proportion of correct answers.

connected per second). Table 2 contains the correlation matrix for all measures in the study. Note that the correlation matrix is based on all 172 cases. Full Information Maximum Likelihood (FIML) estimation was used to calculate the full information covariance matrix from the raw data. The correlation matrix in Table 2 is based on this covariance matrix. The full information covariance matrix was also used for subsequent model fitting in the third and fourth sections.[4]

The correlations in Table 2 are of interest for three reasons. First, all tasks that required the naming of stimuli, either with or without additional executive requirements, appeared considerably interrelated, which was of course due to the communality in measuring method, i.e., naming speed. Although one might expect the manipulated tasks to be especially related to their control counterparts, several manipulated tasks were also considerably correlated to other control tasks. Second, the correlations between the control tasks and the scholastic measures on the one hand, and between the executive tasks and the scholastic measures on the other hand, did not differ much. Third, the *within*-trait correlations, i.e., the correlations between the Inhibition tasks, and the correlations between the Shifting tasks, were approximately of the same order as the *between*-trait correlations. All in all, the correlation matrix did not show much evidence for divergent validity (Campbell & Fiske, 1959; Marsh, 1989; Marsh & Bailey, 1991).

5.2. Effect of the manipulations

For the inhibition and shifting manipulations to be considered effective, performance on the manipulated tasks should first of all be slower than performance on the control tasks. ANOVA's for repeated measures were carried out to examine whether the means on the manipulated tasks were indeed lower than the means on the control tasks. In each of these analyses, a control task and a manipulated task constituted the within-subjects factor. For the manipulated tasks with two control counterparts, two ANOVA's for repeated measures were performed, so that the means of both control tasks were compared separately to the mean of the manipulated task.

An overview of the results of the ANOVA's for repeated measures is presented in Table 3. All repeated measures analyses for the inhibition tasks were significant, meaning that all four inhibition manipulations had slowed down naming speed. Yet, the magnitude of the effects of the manipulations differed considerably between the tasks. The inhibition manipulation in the Objects-I task ($\eta^2=.14$) had slowed down naming speed to a lesser extent than the manipulations in Quantity-I ($\eta^2=.86$), and Stroop-I ($\eta^2=.89$). Also, the mean naming speed of SizeNum-I was much lower than the mean naming speed of the control task Digits ($\eta^2=.86$), but the difference with the control tasks CompDig was less striking ($\eta^2=.04$), although still significant.

All four shifting manipulations had resulted in significant slower naming speed, compared to the respective control tasks. All effect sizes were large, ranging between .54 and .93.

[4] The LISREL program provides only two fit indices when raw data are analyzed (χ^2 and RMSEA). Using the FIML-estimated covariance matrix as input is thus advantageous as more measures of fit become available. Analyses of the full information covariance matrix were deemed justifiable for the present study as there were very few missing values (less than .5% of the data, and mostly 1 data point per cell). Checks were made to ensure that the results based on the estimated covariance matrix were equivalent to the results obtained with raw data analyses.

Table 2
Correlations between 1 reading test, 1 arithmetic test, 1 verbal reasoning task, 1 non-verbal reasoning task, 7 control tasks, 4 inhibition tasks, 4 shifting tasks, and 3 updating tasks ($n=172$)

	Scholastic				Control							Inhibition				Shifting				Updating		
Arithmetic	1.000																					
Reading	.445	1.000																				
Verbal reas.	.141	.198	1.000																			
Non-verbal reas.	.266	.300	.470	1.000																		
Letters	.446	.241	.056	.002	1.000																	
Digits	.414	.225	−.105	−.111	.663	1.000																
Objects	.380	.381	.163	.095	.432	.367	1.000															
Quantity	.433	.465	.198	.189	.562	.495	.597	1.000														
Color	.439	.367	.128	.149	.497	.493	.556	.570	1.000													
Trail A	.277	.337	.164	.167	.325	.309	.287	.456	.351	1.000												
CompDig	.281	.444	.151	.146	.482	.447	.442	.568	.416	.430	1.000											
Quantity-I	.331	.567	.168	.201	.485	.442	.498	.614	.445	.363	.628	1.000										
Objects-I	.263	.248	.060	.028	.368	.324	.607	.478	.457	.201	.362	.409	1.000									
Stroop-I	.446	.455	.137	.219	.479	.447	.629	.583	.604	.343	.490	.601	.435	1.000								
SizeNum-I	.264	.315	.193	.053	.441	.337	.404	.500	.351	.383	.519	.437	.256	.372	1.000							
Objects-S	.355	.408	.155	.099	.419	.458	.549	.577	.477	.449	.445	.573	.419	.577	.409	1.000						
Place-S	.244	.393	.224	.181	.376	.290	.370	.451	.450	.288	.356	.439	.259	.499	.367	.475	1.000					
Symbols-S	.188	.316	.186	.245	.284	.176	.382	.346	.344	.306	.402	.390	.393	.452	.423	.496	.568	1.000				
Trail B	.188	.387	.240	.339	.140	.050	.146	.210	.239	.532	.279	.252	.173	.267	.182	.276	.318	.385	1.000			
Ktrack-U	.236	.208	.351	.269	.109	.058	.035	.179	.151	.187	.186	.177	−.056	.056	.182	.271	.229	.157	.273	1.000		
Letters-U	.281	.210	.157	.231	.097	.073	.189	.201	.177	.255	.009	.122	.050	.228	.138	.248	.218	.108	.271	.387	1.000	
Digits-U	.208	.168	.122	.266	.053	−.040	.102	.091	.131	.225	.019	.062	−.005	.188	.061	.100	.048	.105	.298	.249	.433	1.000

Table 3
Results of the repeated measures analyses for the five Inhibition and four Shifting tasks

Effect	F	df	p	η^2
Inhibition				
Quantity vs. Quantity-I	1065.73	1170	<.001	.86
Objects vs. Objects-I	28.31	1170	<.001	.14
Color vs. Stroop-I	1390.53	1170	<.001	.89
Digits vs. SizeNum-I	1080.32	1170	<.001	.86
CompDig vs. SizeNum-I	6.62	1170	<.01	.04
Shifting				
Trail A vs. Trail B	461.77	1167	<.001	.73
Digits vs. Objects-S	2109.71	1171	<.001	.93
Objects vs. Objects-S	195.27	1170	<.001	.54
Digits vs. Place-S	1896.31	1167	<.001	.92
Letters vs. Place-S	2303.18	1166	<.001	.93
Digits vs. Symbols-S	1718.31	1169	<.001	.91
Letters vs. Symbols-S	2054.18	1168	<.001	.92

Summarizing, all inhibition and shifting manipulations showed the expected decelerating effect on the participants' naming speed, although the magnitude of the effects varied between the manipulations.

5.3. Structure of executive functions

The repeated measures analyses showed that the control tasks and the manipulated tasks differed in their means. Yet, it remains to be seen whether Inhibition, Shifting, and Updating can be distinguished as separate latent common factors, and as such account for variance *in addition to* the variance that is explained by the control tasks. By modeling the control tasks along with the executive tasks, the significance of the executive factors can be established in the actual presence of the control tasks.

All control tasks were modeled as indicators of a general Naming factor. To begin with, the manipulated tasks and the updating tasks were also modeled as indicators of this general Naming factor. Specific relations between manipulated tasks and their individual control counterparts could subsequently be modeled to compensate for additional task specificities, i.e., variability due to the specific type of stimulus that had to be named in certain manipulated tasks.[5] After modeling these non-executive task requirements, we examined whether additional factors for Inhibition, Shifting, and Updating needed to be included in the model in order to arrive at a satisfactory description of the interrelations among the tasks.

To evaluate the fit of the ensuing models to the data, we used the χ^2 statistic, the root mean square error of approximation (RMSEA), and the comparative fit index (CFI) (e.g., Bentler, 1990; Bollen & Long, 1993; Jöreskog, 1993; Schermelleh-Engel, Moosbrugger, & Müller, 2003). The χ^2 statistic is considered a measure of (badness of) fit, with large χ^2 values, relative to the number of degrees of freedom, indicating that the specified model does not adequately account for the observed covariance matrix. As a rule of thumb, χ^2 values smaller than twice the number of degrees of freedom indicate a good fit, and χ^2 values between 2 and 3 times the degrees of freedom indicate adequate fit (Schermelleh-Engel et al., 2003). The RMSEA is a measure of the error of approximation of the covariance (and mean) structures of the specified model to the observed covariance (and mean) structures in the population. As a rule of thumb, Browne and Cudeck (1993) suggested that a RSMEA of .05 or less indicates good approximation, RMSEA between .05 and .07 indicates reasonable approximation, and RMSEA greater than .08 indicates poor approximation. The CFI takes both the parsimony of a model and sample size into account. This fit index ranges from zero to 1.00, and values >.95 are usually taken as indicative of adequate model fit.

These fit indices are informative about the general fit of the model to the data. In addition, for every fixed parameter in the model, the LISREL program provides a Modification Index (MI), the value of which represents the expected drop in overall χ^2 if a parameter is to be freely estimated. We used the MI's to detect local misspecifications in the model. However, given the MI's, parameters were only freely estimated if the additional relations were interpretable (i.e., in line with expectations and former studies), and led to significantly improved model fit ($\alpha=.01$).

The fit of nested models (i.e., if a more restricted model is a subset of a less restricted model) can be compared by subtracting the χ^2 value of the less restricted model (more free parameters) from the χ^2 value of the more restricted model (fewer free parameters). If the resulting χ^2-difference is significant, the less restricted model provides the better fit. Thus, one can test whether freeing parameters results in a significant decrease in overall χ^2. When testing for significance, we used a conservative criterion level of $\alpha=.01$, which was considered reasonable given the complexity of the models, and the number of tests, which are to follow.

[5] Taking the Stroop task as example, the specific requirement to name colors, rather than quantities, alpha-numeric symbols or geometrical objects, might result in specific within-pair common variance that is not explained by the common factor Naming.

First we fitted a 1-factor model to the 7 naming control tasks.[6] The fit indices of this model, Model N_1, are shown in Table 4. The poor fit of Model N_1 was mainly due to an additional relation between Letter and Digit naming, which was not accounted for by the general Naming factor. This relation makes sense as these are the only tasks that require the simple naming of alpha-numeric characters. In Model N_2, the residual terms of Letters and Digits were allowed to correlated. This model fitted well, and the fit of Model N_2 was significantly better than the fit of Model N_1 ($\chi^2_{\text{diff}}(1)=27.55$, $p<.001$). The finding that all control tasks had significant loadings on the general Naming factor indicates that all control tasks appealed to the same underlying ability.

With the control tasks in place, we examined whether the scores of the naming-based Shifting tasks, could be taken as indicative of more than Naming speed alone. Model N_2 was taken as point of departure. We first modeled the shifting tasks as four additional indicators for naming speed by loading them on the Naming factor (Model S_1). This 1-factor model fitted the data poorly (see Table 4): the RMSEA was too high (.11), and the CFI (.87) was too low. However, the loadings of all four shifting tasks on the Naming factor were significant, implying that naming speed did account for a significant amount of the variance in the shifting measures. Naming speed explained 10%, 28%, 34% and 53% of the variance in Trail B, Symbols-S, Place-S and Objects-S, respectively.

Before introducing an additional Shifting factor, we examined whether the fit of Model S_1 could be improved by relaxing specific relations between the shifting tasks and their control counterparts. Such specific relations can be modeled by allowing the residual terms of control and manipulated tasks to correlate. The MI's of Model S_1 showed that the fit of the model would improve greatly if the residual terms of Trail A and Trail B were correlated. In Model S_2, this additional relation between Trail A and Trail B was included. Although Model S_2 fitted the data significantly better than Model S_1 ($\chi^2_{\text{diff}}(1)=35.96$, $p<.001$), the fit was still not satisfactory. No other specific relations between paired tasks could account for this misfit. Therefore, we introduced a common factor for Shifting, on which all four shifting tasks were allowed to load freely. The fit of the resulting model, Model S_3, was good. The loadings of the shifting tasks on the Shifting factor were significant, albeit small. The manipulated naming tasks are thus indicative of more than Naming speed alone, and the additional variance in those four tasks could be accounted for by one factor. The variance of the Shifting factor was small, yet significant, suggesting that the Shifting factor explained variance on top of the variance explained by the Naming factor. Note that, because the Naming and Shifting factors were uncorrelated, the variance that is explained by the Shifting factor can be considered additional to, and independent of, the variance already explained by the Naming factor.

To examine whether the Inhibition tasks were indicative of more than naming speed alone, Model N_2 was again taken as point of departure, and the four Inhibition tasks were initially added as indicators of Naming only. The fit of the resulting 1-factor model, Model I_1, was not satisfactory (see Table 4). However, the loadings of all four inhibition tasks on the Naming factor were significant. Naming speed accounted for 30%, 37%, 53% and 61% of the variance in Objects-I, SizeNum-I, Quantity-I and Stroop-I, respectively.

Again, we first examined whether the fit of Model I_1 could be improved by relaxing specific relations between the residual terms of the inhibition task and their control counterparts. The MI's showed that three specific relations were not accounted for by the general Naming factor: the relation between Objects and Objects-I, the relation between CompDig and SizeNum-I, and the relation between ComDig and Quantity-I. Although ComDig is not the control task of Quantity-I, we deemed this relation acceptable since both tasks require the evaluation of numerical quantity. These three additional relations were incorporated in Model I_2. Model I_2 provided a more accurate description of the data than Model I_1 ($\chi^2_{\text{diff}}(3)=42.40$, $p<.001$). The fit of Model I_2, in which the inhibition tasks were treated as regular naming tasks, was good. Nevertheless, we introduced a common Inhibition factor to examine whether this factor would explain variance in addition to the variance explained by the general Naming factor (Model I_3). The fit of Model I_3 proved however not significantly different from the fit of the model without a common factor for Inhibition ($\chi^2_{\text{diff}}(4)=9.12$, $p=.06$). Besides, all loadings of the inhibition tasks on the Inhibition factor were insignificant and negative, as was the variance of the Inhibition factor. In Model I_4, we examined whether a common factor for Inhibition could be distinguished if only the two most renowned inhibition tasks, Quantity-I and Stroop-I, were used as indicators. Again, the fit of this model was not significantly better than that of the model without an Inhibition factor ($\chi^2_{\text{diff}}(1)=2.27$, ns), and neither the

[6] Note that Trail A is not actually a naming task. Yet, we chose to incorporate Trail A in this general factor because Trail A shares the speed requirement with the other control tasks.

Table 4
Model fit indices

		χ^2	*df*	*p*	RMSEA	CFI
N_1	1 factor, control tasks only	45.38	14	.00	.12	.93
N_2	As N_1, plus additional relation residuals Letters and Digits	17.83	13	.16	.04	.99
S_1	As N_2, plus all shifting tasks as indicators of naming speed	141.39	43	.00	.11	.87
S_2	As S_1, plus additional relation residuals Trail A and Trail B	105.43	42	.00	.10	.92
S_3	As S_2, plus additional factor for Shifting indicated by all shifting tasks	54.32	38	.04	.04	.98
I_1	As N_2, plus all inhibition tasks as indicators of naming speed	93.99	43	.00	.09	.94
I_2	As I_1, plus additional relations between residuals of Objects and Objects-I, CompDig and SizeNum-I, and CompDig and Quantity-I	51.59	40	.10	.04	.99
I_3	As I_2, plus additional factor for Inhibition indicated by all inhibition tasks	42.47	36	.21	.04	.99
I_4	As I_2, plus additional factor for Inhibition indicated only by Quantity-I and Stroop-I	49.32	39	.12	.04	.99
I_5	As I_1, plus additional factor for Inhibition indicated by all inhibition tasks, Objects and CompDig	55.81	37	.02	.05	.98
U_1	As N_2, plus all updating tasks as indicators of naming speed	108.52	34	.00	.11	.86
U_2	As U_1, plus additional factor for Updating indicated by all updating tasks	51.07	31	.01	.05	.96
ISU_1	Three factors: all tasks as indicators of naming speed, Shifting factor indicated by all shifting tasks, and Updating factor indicated by all updating tasks	200.66	123	.00	.05	.94
ISU_2	As ISU_1, plus additional loading of Trail B on Updating factor	190.99	122	.00	.05	.95
ISU_3	As ISU_2, plus correlation between Shifting and Updating factors	190.75	121	.00	.05	.95
$FULL_1$	As ISU_2, plus reading, arithmetic, verbal and non-verbal reasoning	293.66	182	.00	.05	.93
$FULL_2$	As $Full_2$, plus additional relations between Arithmetic and Quantity-I/Trail B, and between verbal reasoning and Ktrack-U	264.53	179	.00	.05	.95

factor loadings, nor the variance of the factor, deviated significantly from zero. Finally, in Model I_5, we abandoned the three correlated residual terms, and examined whether a second common factor could be distinguished if the control tasks Objects and CompDig were allowed to load on this factor, along with the 4 inhibition tasks. Although the overall fit of this model was reasonable, the resulting 'Inhibition' factor was not interpretable as its variance did not deviate significantly from zero, and the factor loadings were both positive and negative. Because the model without a specific Inhibition factor fitted the data adequately, inclusion of this factor did not improve the fit of the model, and the factor itself was not interpretable, we must conclude that a common factor for Inhibition could not be distinguished, and was not necessary to arrive at an adequate

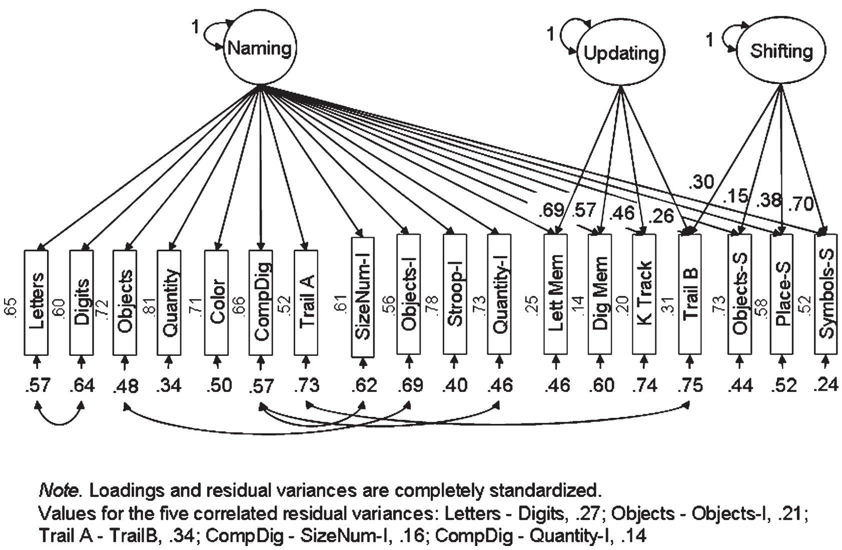

Fig. 2. The estimated three-factor Model ISU_2. The standardized factor loadings for the Naming factor are printed next to the indicators. For Shifting and Updating, the standardized factor loadings are printed next to the single-headed arrows. The numbers at the ends of the smaller arrows are error terms. The curved, double-headed arrows between the error terms represent the specific relations between the indicators that were not explained by the factors.

description of the present data. In the light of these findings, Model I_2 is the preferred model.

Although the three updating tasks did not require speeded rapid naming, subjects were required to name all to-be-recalled stimuli in the updating series aloud. In this respect, naming was involved in the updating tasks. To examine whether the updating tasks were indicative of more than Naming alone, Model N_2 was again taken as point of departure, and the three Updating tasks were initially introduced as indicators of Naming only. The fit of model U_1 was not adequate (see Table 4), but Naming accounted for 2%, 4% and 6% of the variance in Digits-U, Ktrack-U, and Letters-U, respectively, the latter two loadings deviating significantly from zero. The fit of Model U_1 could be not be improved by freeing specific relations between the residual terms of the Updating tasks and the control tasks. A common factor for Updating was therefore introduced in model U_2 on which the three Updating tasks were allowed to load freely. The fit of model U_2 was adequate, and the loadings of the three Updating tasks on the common factor Updating were all significant, as was the variance of the factor. The variance common to the three updating tasks could thus be described by one factor.

Table 5
Completely standardized factor loadings and residual variances for Model Full$_2$

	Factor loadings			Res. variances
	Naming	Shifting	Updating	
Control tasks				
Letters	.66			.56
Digits	.60			.64
Objects	.72			.48
Quantity	.81			.34
Color	.71			.49
Trail A	.52			.73
CompDig	.66			.57
Inhibition				
Quantity-I	.73			–
Objects-I	.55			.69
Stroop-I	.78			.40
SizeNum-I	.61			.62
Shifting				
Objects-S	.73	.14		.45
Place-S	.58	.40		.50
Symbols-S	.53	.64		.31
Trail B	.31	.34	.30	–
Updating				
Ktrack-U	.21		.48	–
Letters-U	.25		.63	.55
Digits-U	.14		.59	.63
Dependent				
Reading	.54	−.17	.25	–
Arithmetic	.55	.06	.16	–
V reasoning	.20	.13	.20	–
Nonv reasoning	.18	.22	.39	–

Note. Values for the five correlated residual variances: Letters–Digits, .26; Objects–Objects-I, .21; Trail A–Trail B, .33; CompDig–SizeNum-I; .16; CompDig–Quantity-I, .12.
Values for the three specific relations: Quantity-I–Arithmetic, .22; Trail B–Arithmetic, .14; Keep Track–Verbal Reasoning, .23.

Table 6
Residual variances and partial correlations for reading, arithmetic, verbal and non-verbal reasoning in Model Full$_2$

	Reading	Arithmetic	V. reasoning	Nonv. reasoning
Reading	.62			
Arithmetic	.22** (.45)	.60		
V. reasoning	.001 (.14)	.04* (.20)	.85	
Nonv. reasoning	.16 (.27)	.13* (.30)	.39 (.47)	.77

Note. On the diagonal are the standardized residual variances of the four dependent factors (i.e., variance not explained by Naming speed, Shifting, Updating, or one of the indicators). Off-diagonals are partial correlations, i.e., correlations between these residuals. In parentheses are the original correlations. The asterisks denote a significant difference between original and partial correlations, *$p<.05$, and **$p<.01$.

In Model ISU$_1$, we modeled all control and manipulated tasks simultaneously, with all tasks as indicators for Naming, the above-described correlated residuals, and the four shifting tasks and three updating tasks as indicators of the common factors Shifting and Updating, respectively. The fit of this model proved adequate (see Table 4), but the MI's showed that the fit of the model would improve if Trail B was allowed to load on the Updating factor as well. This additional loading of Trail B on the Updating factor is justifiable (see the Discussion for a detailed rationale), and was therefore admitted to the model. The resulting Model ISU$_2$ fitted the data better than Model ISU$_1$ ($\chi^2_{\text{diff}}(1)=9.67, p<.01$). In Model ISU$_3$, a correlation between the factors Shifting and Updating was introduced. This correlation was small (.06), and the fit of Model ISU$_3$ was not significantly better than that of Model ISU$_2$ ($\chi^2_{\text{diff}}(1)<1$, ns). The more parsimonious Model ISU$_2$ is therefore the preferred model.[7] Model ISU$_2$ is illustrated in Fig. 2.

[7] Note that the correlation between Shifting and Updating was .18 when Trail B was only allowed to load on the Shifting factor. However, since the communality with Trail B proved the main source of covariance between Updating and Shifting, it is in our view more accurate to acknowledge the multidimensionality of the Trail B tasks, than to suggest the presence of a meaningful correlation between Updating and Shifting. Also, models in which Trail B was allowed to load on only one of the factors, fitted the data significantly worse than the model in which Trail B was allowed to load on both the Shifting and the Updating factor.

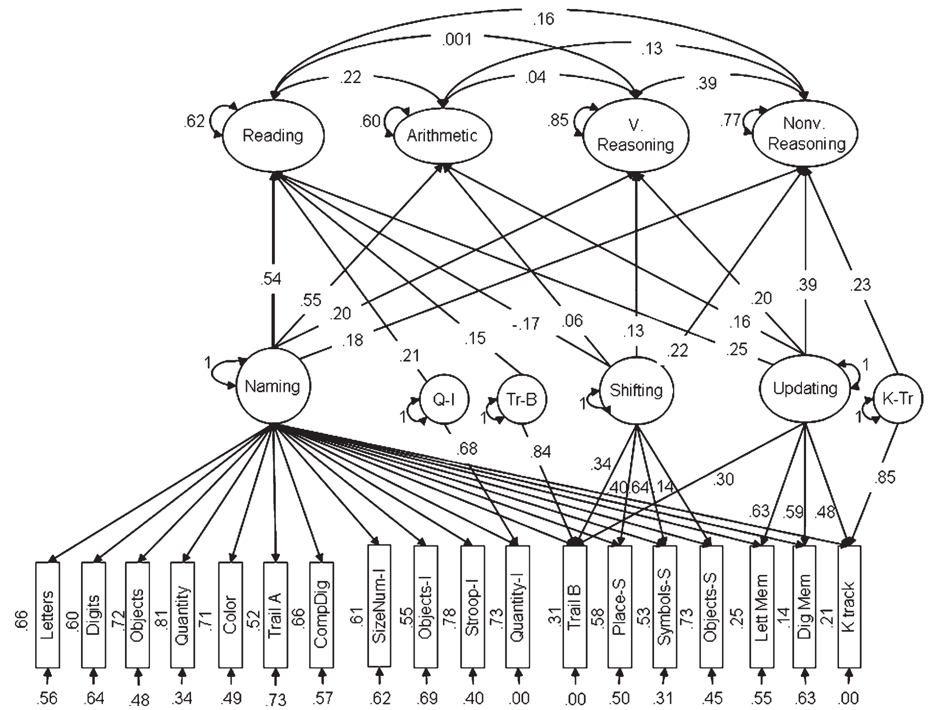

Fig. 3. The estimated structural Model Full$_2$. Note that correlated errors between control tasks and manipulated tasks are not illustrated for reasons of convenience. The standardized factor loadings for the Naming factor are printed next to the indicators.

Finally, it is worth noting that additional confirmatory factor analyses based on a variety of transformed scores all yielded similar results with respect to factor structure and the distinctiveness of the latent factors. Specifically, we analyzed logarithmically transformed scores,[8] difference scores (where the scores on the control tasks are subtracted from the scores on the executive tasks, and CFA is performed on the differences), regression residuals (where scores on the executive tasks are regressed on their control counterparts, and CFA is performed on the regression residuals), and scores corrected for naming accuracy or Raven performance (where naming accuracy and Raven performance were regressed out of both the executive tasks and the control tasks).

5.4. Executive functions, reasoning, arithmetic, and reading

The second objective of the study was to examine whether the executive or the non-executive variance in the executive tasks qualified the relations between executive task performance and higher-level cognitive abilities. To this end, the tasks for verbal and non-verbal reasoning, reading, and arithmetic were introduced in Model ISU$_2$ as four correlated factors with 1 indicator each, resulting in Model Full$_1$.[9] These four factors were regressed on the three common factors for Naming, Shifting, and Updating. As the indices in Table 4 show, the fit of the resulting model was reasonable.

All regression parameters between the Naming factor and the four dependent factors were significant, as were all regression parameters associated with the Updating factor. However, Shifting was only significantly related to non-verbal reasoning and reading.

Inspection of the MI's showed that specific relations existed between the arithmetic test and Quantity-I and Trail B, and between the verbal reasoning task and Ktrack-U. The specific relations of arithmetic ability with Quantity-I and Trail B are in line with the results of a former study (van der Sluis et al., 2004), and both

[8] Log transformations enables one to study proportional change in performance on the manipulated tasks with respect to the control task (Zar, 1999).

[9] For reasons of identification, the errors of those four factors were fixed to zero, the loading of the factors was fixed to 1, and the variances were estimated. Note that these four 'dependent' common factors are actually not latent factors as they have only 1 indicator and zero-error variance. This parameterization thus simply implies that the four indicators were directly regressed on the latent factors for Updating, Shifting, and Naming.

Table 7
Percentages variance explained in reading ability, arithmetic ability, verbal reasoning and non-verbal reasoning by the common factors Naming, Shifting, and Updating, and the three tasks Quantity-I, Trail B, and Ktrack-U in Model Full$_3$

	Naming	Shifting	Updating	Quantity-I	Trail B	Ktrack-U	Total
Reading	29.3	2.7	6.1				38.1
Arithmetic	30.1	.4	2.6	4.7	1.9		39.7
V. reasoning	4.1	1.7	3.9			5.2	14.9
Nonv. Reasoning	3.3	4.6	15.1				23.0

Total = total variance explained in the dependent variables reading, arithmetic, verbal reasoning and non-verbal reasoning.

Ktrack-U and the task for verbal reasoning require knowledge of categories and domains. Inclusion of these three relations was therefore deemed justifiable. Free estimation of these three specific relations in Model Full$_2$ resulted in significant improvement of the fit ($\chi^2_{\mathrm{diff}}(3)=29.13$, $p<.001$). The completely standardized solution of Model Full$_2$ is presented in Tables 5 and 6, and the model is illustrated in Fig. 3.

Together, the uncorrelated factors Naming, Shifting, and Updating, and the three task specific relations, explained 38.1% of the variance in reading ability, 39.7% of the variance in arithmetic ability, 14.9% of the variance in verbal reasoning ability, and 23.0% of the variance in non-verbal reasoning ability. Table 7 shows the partial percentages of variance explained for each predictor separately. Table 6 shows that some of the correlations between the dependent measures had decreased significantly after their relations with naming speed, shifting and updating were accounted for. Interestingly, the correlation between verbal and non-verbal reasoning showed the least decrease.

The executive function Updating proved considerably related to all four dependent variables, but mostly to non-verbal reasoning ability and reading ability, explaining 15.1% and 6.1% of the variances in these measures respectively. The executive function Shifting mostly related to non-verbal reasoning ability, explaining 4.6% of the variance in this measure. The relation with reading was also just significant, with Shifting explaining 2.7% of the variance in reading. The negative sign of this relation implies a trend for greater word reading fluency to coincide with slightly slower shifting performance. It is noteworthy that Naming speed was by far the best predictor of reading and arithmetic ability, explaining 29.3% and 30.1% of the variance in these variables respectively. In contrast, Naming speed explained a relatively small amount of variance in verbal and non-verbal reasoning ability (4.1% and 3.3%, respectively).

Beside the variance that Quantity-I and Trail B already explained through the common factors Naming and Updating, these indicators accounted for an additional 4.7% and 1.9%, respectively, of the variance in arithmetic ability through their specific relations. Likewise, performance on the Keep Track task was not only related to verbal reasoning ability through Naming and Updating, but explained an additional 5.2% of the variance in verbal reasoning ability through its specific relation.

6. Discussion

The objective of the present study was twofold. First, we wanted to establish whether the executive factors Inhibition, Shifting, and Updating were detectable as common factors in children, while controlling for the effects of the non-executive task requirements. Second, we wanted to examine to what extent the relation between performance on executive tasks and higher-order cognitive functioning was attributable to the executive and non-executive aspects of executive measures. Well-established executive measures, which are amenable to administration in a control condition, were selected for this study. The control tasks were used in confirmatory factor analysis (CFA) to separate the executive from the non-executive variance.

CFA showed that all control tasks involved one underlying ability, which we denoted Naming. Former studies on rapid automatized naming report that naming speed partly depends on the type of symbols (e.g., van den Bos, Zijlstra, & Lutje Spelberg, 2002; van der Sluis et al., 2004). In the present study, this domain specificity was evident in the additional correlation between the residuals of the control tasks Letters and Digits. The common factor Naming could be treated as a single source of individual difference in the non-executive processing demands, as the (slightly extended) 1-factor model provided an adequate description of the control data.

The Naming factor explained variance in all executive tasks. However, specific factors for Shifting and Updating were necessary to account for additional relations between the shifting measures and the updating measures, respectively. The finding that performance on these executive tasks could not be entirely explained by performance on the control tasks indicates that these executive tasks had tapped other sources of individual differences, i.e., performance on the shifting and updating tasks was indicative of more than naming speed alone. Although factors for Shifting and Updating were clearly distinguishable, considerable portions of the variance in these executive measures remained unaccounted for. Sizeable residual terms could be the result of random measurement error, i.e., unreliable measures, or simply indicate that the tasks, although reliable in themselves, have little in common.

Interestingly, in the present study, Shifting and Updating were only correlated through their shared non-executive variance, as was captured in the Naming factor. Previous studies reported considerable correlations between Shifting and Updating (e.g., .52, Miyake et al., 2000; .65, Lehto et al., 2003). If none of the shifting and updating tasks in the present study had been controlled for performance on the control tasks, we would have found a correlation of .39. It should be noted that in the analyses that we presented here, the executive measures were corrected for *all* the variance that they had in common with *all* control tasks. So if, for example, an updating measure shared variance with the control task of a shifting measure, this non-executive variance was controlled for. This correction is more thorough than the correction for specific control counterparts only. Nevertheless, the correlation between Shifting and Updating would still have been .34 if the shifting tasks were only corrected for performance on their specific control counterparts. Whether the correlations between Shifting and Updating, as reported in previous studies, were mainly attributable to shared non-executive variance, is an open question.

A factor for Inhibition could not be established in the presence of the Naming factor. All variance that the inhibition tasks had in common was absorbed by the Naming control factor. The difficulty to detect a common factor for inhibition is in line with a study by Shilling, Chetwynd, and Rabbitt (2002), who reported weak and varied correlations (between −.13 and .22) among a set of Stroop-like tasks once performance on the control tasks had been taken into account. Salthouse et al. (2003) also report "…that the relations between the variables postulated to represent inhibition were not significantly different from zero when they were considered in the context of relations to other constructs, suggesting that convergent validity was lacking" (p. 584). In a study of different age groups, Huizinga, Dolan, and van der Molen (2006) also failed to detect a common factor for inhibition after basic speed was controlled for. Miyake et al. (2000) and Friedman and Miyake (2004) did manage to distinguish common factors for inhibition, but the residual variances of the supposed inhibition tasks were large (ranging between .82 and .92, and between .70 and .92, respectively), suggesting that the inhibition tasks had very little variance in common, and that the greater part of the variance in these tasks could not be explained by their models. The present results are therefore consistent with the more general finding that well-established inhibition manipulations fail to tap a common source of individual differences.

The failure to distinguish a common Inhibition factor does not necessarily mean that there is no such thing as 'inhibition' or 'inhibitory ability'. This construct may exist, but might a) not be a large source of reliable individual differences, b) be very difficult to measure reliably, or c) be highly correlated with other constructs (Embretson, 1983). With respect to the first point, we did find that the performance of the children on the inhibition tasks was markedly slower than their performance on the control tasks. The decelerating effect of the inhibition manipulations suggests that the manipulations had broached an additional ability that might be regarded as inhibition. With regard to the last point, the study by Salthouse et al. (2003) showed that inhibition was strongly related to processing speed. Had we modeled Naming and Inhibition as two separate factors, we would have found a correlation of 1.00 between Naming speed and Inhibition. This obviously explains why a common factor for inhibition was not detectable in the present study: all variance common to inhibition and naming speed ended up in the Naming factor, comprising what we called 'non-executive variance'.[10]

The present and previous findings therefore suggest that (well-established) inhibition manipulations show consistent effects on the level of the mean performance (i.e., fixed effects), but fail to tap a common and systematic source of individual differences (i.e., random effects). As a cautionary note to this latter conclusion, we should mention that all these CFA studies on Inhibition involved normally functioning participants. It is conceivable that inhibition is a source of individual differences in clinical samples, or can be used to distinguish between clinical and non-clinical samples. With

[10] Note that it is possible that the correlation between two constructs inflates if these constructs show similar age-related developmental trends (Hofer & Sliwinski, 2001). However, mean naming speed (as calculated over all 7 control tasks) did not differ significantly between the 25% youngest and the 25% oldest children in our sample ($t(84)=1.44$, ns), so it is unlikely that the results in this study were influenced by age-related changes.

respect to the first objective of this study, however, we conclude that factors for Shifting and Updating factors, but not for Inhibition, were detectable in normally functioning children, and that Shifting and Updating should be considered separate executive functions.

The second aim of the study concerned the question to what extent the performance of normal children on executive tasks was related to reasoning, reading, and arithmetic through the executive, and the non-executive task demands. The latter demands were operationalized in the terms of the general Naming factor, which captured the variance common to the control tasks and the EF measures. This non-executive Naming factor was strongly related to reading and arithmetic ability. The relation between reading and the ability to rapidly identify and name stimuli has often been reported in reading disabled subgroups (e.g., de Jong & van der Leij, 2003; van den Bos, 1998; van der Sluis et al., 2004; see Share, 1995; Wolf & Bowers, 1999 for a review), and in non-clinical samples (e.g., de Jong & van der Leij, 2002; de Jong & Wolters, 2002; Manis, Seidenberg, & Doi, 1999; Neuhaus & Swank, 2002; Torgesen, Wagner, Rashotte, Burgess, & Hecht, 1997; van den Bos et al., 2002). This relation is explained by the mutual requirement to quickly retrieve phonological codes from memory (e.g., Manis et al., 1999; Neuhaus & Swank, 2002; Wolf & Bowers, 1999).

The reading and the arithmetic task used in this study are measures of efficiency. Accordingly, high scores on the present reading test indicate fluent, and at least partly automatized, reading ability. Likewise, the arithmetic test reflected the prompt recognition of the visually presented arithmetic problems as familiar, and the fast retrieval of the accompanying answers from memory (i.e., arithmetic fact finding). The finding that the Naming factor was related to both efficiency measures supports the idea that both reading and arithmetic ability depend on the strength and accessibility of representations in memory, and the speed with which information can be retrieved (Geary, 1993, 2004). To some extent, the speeded character of both scholastic measures may have contributed to the strength of their relations with Naming. However, naming speed has previously been found to be related to unspeeded measures of reading and arithmetic as well (e.g., reading accuracy, reading comprehension and orthographic skills, and unspeeded arithmetic performance; Bull & Johnston, 1997; de Jong & Wolters, 2002; Manis et al., 1999; Torgesen et al., 1997).

The non-executive Naming factor was only weakly related to verbal and non-verbal reasoning. This is comprehensible if naming speed is indeed indicative of the speed with which information is retrieved from memory. Reasoning tasks do not measure automatized abilities, and there are no ready answers to reasoning problems stored in long-term memory. Good performance on reasoning tasks is characterized by careful selection, judgment, and evaluation of all possible relations, rather than quick retrieval of information from long-term memory.

We now turn to the results on the relation between higher-order cognitive abilities and the executive processing demands of the executive tasks. Updating ability was positively related to reading, arithmetic, verbal, and non-verbal reasoning, explaining 6.1%, 2.6%, 3.9%, and 15.1% in these measures, respectively. However, these relations were not as strong as previous studies on working memory capacity suggest. Yet, although working memory tasks and updating tasks both require the revision of the content of memory in light of new information, working memory tasks also require the processing of information. It is conceivable that these processing demands, rather than the storage demands, qualify the strong correlations of working memory with higher-order cognitive abilities. Earlier studies already showed that storage per se is hardly related to (non-verbal) reasoning, while storage plus processing is (e.g. Engle, Laughlin, Tuholski, & Conway, 1999; Conway, Cowan, Bunting, Therriault, & Monkoff, 2002). If working memory tasks are indeed related to higher-order cognitive abilities through their processing demands, expectations with respect to updating should be revised.

The weaker relations of updating with the higher-order cognitive abilities could also be due the specific measures used in this study. For example, the arithmetic task required high levels of arithmetic fact finding, rather than the elaborate arithmetical processing, in which updating has been hypothesized to be involved (e.g., Geary, 1993, 2004; McLean & Hitch, 1999). Good updating ability may be more of a prerequisite for successful performance on arithmetic tasks that contain multi-step calculations and word problems (Geary, 2004). Also, some (see Kyllonen, 2002, p.433) have argued that the complex nature of working memory measures accounts for their strong relations with reasoning and intelligence. The updating tasks used in this study were not complex, i.e., did not require reading, arithmetic, counting, or reasoning (see for instance the working memory tasks used by Colom et al., 2004; Conway et al., 2002; Kyllonen & Chrystal, 1990). This relative purity of the present updating measures may have resulted in lower correlations with higher-order cognitive abilities. Finally, we should note that two out of three of our updating tasks showed ceiling effects. This might be another reason for the relatively weak relations with the higher-order cognitive abilities in the present study. These ceiling effects may have resulted in underestimation of the correlations with the respective dependent measures due to restriction of range.

Shifting was only significantly related to non-verbal reasoning and reading, explaining 4.6% and 2.7%, respectively. Unexpectedly, the correlation with reading efficiency was negative. The finding that reading was positively related to Updating, but negatively to Shifting, supports the distinctness of these EFs. The absence of a correlation between Shifting and arithmetic was unexpected. Several authors (e.g., Bull & Colleagues, 1999; Bull & Scerif, 2001; McLean & Hitch, 1999; van der Sluis et al., 2004) have reported associations between arithmetic ability and performance on complex shifting tasks like the Wisconsin Card Sorting Task (WCST) and the Making Trails task. Shifting was therefore believed to support alternation between arithmetic strategies (e.g., addition, subtraction, multiplication), and arithmetic sub-solutions in multi-step arithmetic problems.

The results of the present study suggest that the relationships between shifting and arithmetic ability reported in other studies, may partly have been due to the non-executive requirements of the shifting tasks. Clearly, the relationship between shifting performance and arithmetic performance would have been much stronger, had we not controlled for naming speed. In addition, the role of shifting ability may have been limited in our arithmetic measure as the three arithmetic subtests were homogeneous with respect to the required arithmetical procedure (i.e., they required either addition, *or* subtraction, *or* multiplication). Shifting may have occurred *within* some of the arithmetic problems – such as shifting from arithmetic fact finding strategies to calculation strategies – but the necessity to shift was clearly minimal.

Interestingly, the specific relation between arithmetic ability and performance on the Making Trails task was confirmed here. Van der Sluis et al. (2004) suggested that Making Trails tasks call on other EFs besides shifting. More specifically, they hypothesized that Trail B not only measures shifting (i.e., alternation between letters and digits), but also negative priming (i.e., the response, inhibited on trial $n-1$, is relevant on trial n), and updating (i.e., one needs to keep track of former responses in order to select the correct next response). The present results support this hypothesized multi-componentional character of this measure. Trail B loaded significantly on Updating as well as on Shifting. The current finding that *simple* shifting ability is not related to arithmetic ability, while performance on a multi-componential shifting task is, is therefore in line with the study by van der Sluis et al. (2004). However, we also found that the relation between arithmetic ability and Trail B extended beyond the appeal of Trail B to both shifting and updating. How this additional relation should be explained, is an open question.

Also, the additional specific relationship between the Quantity-I task and arithmetic ability is consistent with previous results (e.g., Bull & Scerif, 2001; van der Sluis et al., 2004). Bull and Scerif suggested that the specificity of this relation might indicate domain-specific problems with inhibition, or a domain-specific problem with strategy generation and evaluation (p. 286).

In sum, the present study demonstrated the usefulness of control tasks in the context of CFA in the study of executive functioning. After controlling for the non-executive task demands, a Shifting and an Updating factor, but not an Inhibition factor, could be distinguished. Previous studies, using different task sets and different age samples, also failed to detect a common factor for Inhibition. It is therefore unlikely that Inhibition failed to appear as a common factor in the present study due to our choice of tasks, or to characteristics of the present sample (i.e., young adolescents). The frequent failure to detect a common factor for inhibition in non-clinical samples suggests that in normal populations, individual differences in inhibitory capacity are small and/or very difficult to measure reliably.

As was to be expected, after accounting for non-executive variance, the correlations of EFs with (non) verbal reasoning, reading, and arithmetic were lower than reported in earlier studies. The non-executive Naming factor explained a considerable amount of variance in the higher-order cognitive abilities, especially in reading and arithmetic. One could argue that these findings were inevitable given the present selection of tasks, i.e., both dependent and independent measures called on speeded naming. However, in our view the present study provides an illustration of how measurement format, in itself nothing but a vehicle, can influence the associations between test scores. The results show clearly that the correlations of executive tasks with other cognitive measures can partly depend on the executive measure's *non*-executive requirements. For instance, performance on a speeded arithmetic task may correlate to performance on a shifting task not because shifting is implicated in arithmetic, but because both tasks require the speeded identification and evaluation of stimuli.

Task impurity may underlie relations between executive measures and other abilities, when these relations are mainly due to non-executive ability. This forms a problem for accurate hypothesis testing and thus theory building. In neuropsychological studies, task impurity may obscure the exact nature of the association between frontal lobe damage and performance on executive tasks. In developmental studies, task impurity raises the question whether improved performance on executive tasks with age (and later decline in old age), is attributable to improved executive or improved non-executive performance. Overall, task impurity poses a real problem, and future progress in the field of executive functioning will benefit from greater awareness of the complications associated with the use of impure measures.

Acknowledgements

We are grateful for the detailed comments and suggestions of three anonymous reviewers. These resulted in many improvements.

We wish to thank Ilja Grauwelman, Annemarie Honig, Marjolein Hulshof, Floris Lemmen, Najat Rabbae, Melle van Tricht and Alexandra Wilson for their assistance with the data collection, and the children and staff of the Roald Dahl school, the Roelof Venemaschool, the Vondelschool, Het Palet, De Pool, and De Westwijzer in Amsterdam and Amstelveen for the time and assistance which they generously contributed to this study.

References

Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review*, *16*(1), 17−42.

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.

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology*, *49A*(1), 5−28.

Beier, M. E., & Ackerman, P. L. (2004). A repraissal of the relationship between span memory and intelligence via “best evidence synthesis”. *Intelligence*, *32*(6), 607−619.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238−246.

Bleichrodt, N., Drenth, P. J. D., Zaal, J. N., & Resing, W. C. M. (1987). *RAKIT; Revisie Amsterdamse Kinder Intelligentie Test.* Lisse: Swets & Zeitlinger Publishers.

Bollen, K., & Long, J. S. (Eds.). (1993). *Testing structural equation models* Newbury Park, CA: Sage.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 136−162). Newbury Park, CA: Sage.

Brus, B. T., & Voeten, M. J. M. (1979). *Een Minuut Test (one minute reading test).* Lisse: Swets & Zeitlinger Publishers.

Bull, R., & Johnston, R. S. (1997). Children’s arithmetic difficulties: Contributions from processing speed, item identification, and short-term memory. *Journal of Experimental Child Psychology*, *65*, 1−24.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children’s mathematics ability: Inhibition, switching and working memory. *Developmental Neuropsychology*, *19*(3), 273−293.

Bull, R., Johnston, R. S., & Roy, J. A. (1999). Exploring the roles of the visual–spatial sketch pad and central executive in children’s arithmetical skills: Views from cognition and developmental neuropsychology. *Developmental Neuropsychology*, *15*(3), 421−442.

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 81−116). Hove, UK: Psychology Press.

Butterworth, B. (1999). *The mathematical brain.* London: Macmillan.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81−104.

Carter, C. S., Krener, P., Chaderjian, M., Northcutt, C., & Wolfe, V. (1995). Abnormal processing of irrelevant information in attention deficit hyperactivity disorder. *Psychiatry Research*, *56*, 59−70.

Christ, S. E., White, D. A., Mandernach, T., & Keys, B. A. (2001). Inhibitory control across the life span. *Developmental Neuropsychology*, *20*(3), 653−669.

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by g. *Intelligence*, *32*, 277−296.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Monkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163−183.

de Jong, P. F. (1998). Working memory deficits of reading disabled children. *Journal of Experimental Child Psychology*, *70*, 75−96.

de Jong, P. F., & Das-Smaal, E. A. (1995). Attention and intelligence: The validity of the Star Counting Test. *Journal of Educational Psychology*, *87*(1), 80−92.

de Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading*, *6*, 51−77.

de Jong, P. F., & van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology*, *95*(1), 22−49.

de Jong, P. F., & Wolters, G. (2002). Fonemisch bewustzijn, benoemsnelheid en leren lezen. *Pedagogische Studiën*, *79*, 53−63.

de Vos, T. (1992). *Tempo Test Rekenen.* Lisse: Swets & Zeitlinger Publishers.

Denckla, M. B. (1994). Measurement of executive function. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 117−142). Baltimore: P.H. Brookes Publishing.

Denckla, M. B. (1996). A theory and model of executive function: A neuropsychological perspective. In G. R. Lyon & N.A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 263−277). Baltimore: MD: Paul H. Brookes.

Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, *93*(1), 179−197.

Engle, R. W., Laughlin, J. E., Tuholski, S. W., & Conway, A. R. A. (1999). Working memory, short term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology. General*, *128*(3), 309−331.

Everatt, J., Warner, J., Miles, T. R., & Thomson, M. E. (1997). The incidence of Stroop interference in dyslexia. *Dyslexia*, *3*, 222−228.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology. General*, *133*(1), 101−135.

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin*, *114* (2), 345−362.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities*, *37*(1), 4−15.

Graf, P., & Uttl, B. (1995). Color- and picture-word Stroop tests: Performance changes in old age. *Journal of Clinical and Experimental Neuropsychology*, *17*, 1−16.

Gustafsson, J. E. (1992). The relevance of factor analysis for the study of group differences. *Multivariate Behavioral Research*, *27*(2), 239−247.

Gustafsson, J. E., & Undheim, J. O. (1992). Stability and change in broad and narrow factors of intelligence from ages 12 to 15 years. *Journal of Educational Psychology*, *84*(2), 141−149.

Helland, T., & Asbjørnsen, A. (2000). Executive functions in dyslexia. *Child Neuropsychology*, *6*(1), 37−48.

Hertzog, C., & Bleckley, M. K. (2001). Age differences in the structure of intelligence: Influences of information processing speed. *Intelligence*, *29*, 191−217.

Hofer, S. M., & Sliwinski, M. J. (2001). Understanding ageing. *Gerontology*, *47*, 341−352.

Howes, N. L., Bigler, E. D., Burlingame, G. M., & Lawson, J. S. (2003). Memory performance of children with dyslexia: A comparative analysis of theoretical perspectives. *Journal of Learning Disabilities*, *36*(3), 230−246.

Howes, N. L., Bigler, E. D., Lawson, J. S., & Burlingame, G. M. (1999). Reading disability subtypes and the test of memory and learning. *Archives of Clinical Neuropsychology*, *14*(3), 317−339.

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

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 294−316). Newbury Park, CA: Sage.

Jöreskog, K. G., & Sörbom, D. (2001). *LISREL 8.50: User's reference guide.* Chicago: Scientific Software International.

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.

Klorman, R., Hazel-Fernandez, L. A., Shaywitz, S. E., Fletcher, J. M., Marchione, K. E., Holahan, J. M., et al. (1999). Executive functioning deficits in attention-deficit/hyperactivity disorder are independent of oppositional defiant or reading disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *38*(9), 1148−1155.

Kyllonen, P. C. (1993). Aptitude testing inspired by information processing: A test of the four-sources model. *The Journal of General Psychology*, *120*(3), 375−405.

Kyllonen, P. C. (2002). g: Knowledge, speeds, strategies, or working-memory capacity? A systems perspective. In R. J. Sternberg & E.L. Grigorenko (Eds.), *The general factor of intelligence — How general is it?* (pp. 415−445). London: Lawrence Erlbaum Associates, Publishers.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity?! *Intelligence*, *14*, 389−433.

Lehto, J. E., Juujärvi, Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*, 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.

Lindsay, R. L., Tomazic, T., Levine, M. D., & Accardo, P. J. (1999). Impact of attentional dysfunction in dyscalculia. *Developmental Medicine and Child Neurology*, *41*, 639−642.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: In integrative review. *Psychological Bulletin*, *109*(2), 163−203.

Manis, F. R., Seidenberg, M. S., & Doi, L. M. (1999). See Dick RAN: Rapid naming and the longitudinal prediction of reading subskills in first and second graders. *Scientific Studies of Reading*, *3*(2), 129−157.

Manly, T., Anderson, V., Nimmo-Smith, I., Turner, A., Watson, P., & Robertson, I. H. (2001). The differential assessment of children's attention: The Test of Everyday Attention for Children (TEA-Ch), normative sample data and ADHD performance. *Journal of Child Psychology and Psychiatry*, *42*(8), 1065−1081.

Marsh, H. W. (1989). Confirmatory factor analysis of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, *13*(4), 335−361.

Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement*, *15*(1), 47−70.

McLean, J. F., & Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning disabilities. *Journal of Experimental Child Psychology*, *74*, 240−260.

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

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, *81*, 111−121.

Neuhaus, G. F., & Swank, P. R. (2002). Understanding the relations between RAN letter subtest components and word reading in first-grade students. *Journal of Learning Disabilities*, *35*(2), 158−174.

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.

Oberauer, K., Süβ, H. -M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, *31*, 167−193.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *37*, 51−87.

Rabbitt, P. (1997). Introduction: Methodologies and models in the study of executive function. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 1−38). Hove, UK: Psychology Press.

Raven, J. C., Court, J. H., & Raven, J. (1979). *Manual for Raven's progressive matrices and vocabulary scales. Section I general overview.* London: H.K. Lewis & Co.

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

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology. General*, *132*(4), 566−594.

Scheres, A., Oosterlaan, J., Geurts, H., Morein-Zamir, S., Meiran, N., Schut, H., et al. (2004). Executive functioning in boys with ADHD: Primarily an inhibition deficit? *Archives of Clinical Neuropsychology*, *19*, 569−594.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, *8*(2), 23−74.

Seidman, L. J., Biederman, J., Monuteaux, M. C., Weber, W., & Faraone, S. V. (2000). Neuropsychological functioning in nonreferred siblings of children with attention deficit/hyperactivity disorder. *Journal of Abnormal Psychology*, *109*(2), 252−265.

Sergeant, J. A., Geurts, H., & Oosterlaan, J. (2002). How specific is a deficit of executive functioning for attention-deficit/hyperactivity disorder? *Behavioural Brain Research*, *130*, 3−28.

Shallice, T., Marzocchi, G. M., Coser, S., Del Savio, M., Meuter, R. F., & Rumiati, R. I. (2002). Executive function profile in children with attention deficit hyperactivity disorder. *Developmental Neuropsychology*, *21*(1), 43−71.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*, 151−218.

Shilling, V. M., Chetwynd, A., & Rabbitt, P. M. A. (2002). Individual inconsistency across measures of inhibition: An investigation in the construct validity of inhibition in older adults. *Neuropsychologia*, *40*, 605−619.

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development*, *60*, 973−980.

Sikora, D. M., Haley, P., Edwards, J., & Butler, R. W. (2002). Tower of London Test performance in children with poor arithmetic skills. *Developmental Neuropsychology*, *21*(3), 243−254.

Sonuga-Barke, E. J. S., Dalen, L., Daley, D., & Remington, B. (2002). Are planning, working memory, and inhibition associated with individual differences in preschool ADHD symptoms? *Developmental Neuropsychology*, *21*(3), 255−272.

Stankov, L. (2000). Complexity, metacognition, and fluid intelligence. *Intelligence*, *28*(2), 121−143.

Stevens, J., Quittner, A. L., Zuckerman, J. B., & Moore, S. (2002). Behavioral inhibition of self-regulation of motivation and working memory in children with attention deficit hyperactivity disorder. *Developmental Neuropsychology*, *21*, 117−140.

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

Süβ, H. -M., Oberauer, K., Wittman, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability — And a little bit more. *Intelligence*, *30*, 261−288.

Swanson, H. L. (1994). The role of working memory and dynamics assessment in the classification of children with learning disabilities. *Learning Disabilities Research*, *9*(4), 190−202.

Swanson, H. L. (1999). Reading comprehension and working memory in learning disabled readers: Is the phonological loop more important than the executive system? *Journal of Experimental Child Psychology*, *72*, 1−31.

Swanson, H. L., & Ashbaker, M. H. (2000). Working memory, short-term memory, speech rate, word recognition and reading comprehension in learning disabled readers: Does the executive system have a role? *Intelligence*, *28*(1), 1−30.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatized naming ability to the growth of word-reading skills in second- to fifth-grade children. *Scientific Studies of Reading*, *1*(2), 161−185.

Uttl, B., & Graf, P. (1997). Color-word Stroop test performance across the adult life span. *Journal of Clinical and Experimental Neuropsychology*, *19*, 405−420.

van den Bos, K. P. (1998). IQ, phonological awareness and continuous-naming speed related to Dutch poor decoding children's performance on two word identification tests. *Dyslexia*, *4*, 73−89.

van den Bos, K. P., Zijlstra, B. H. J., & Lutje Spelberg, H. C. (2002). Life-span data on continuous-naming speeds of numbers, letters, colors, and pictured objects, and word-reading speed. *Scientific Studies of Reading*, *6*(1), 25−49.

van der Schoot, M., Licht, R., Horsley, T. M., & Sergeant, J. A. (2000). Inhibitory deficits in reading disability depend on subtype: Guessers but not spellers. *Child Neuropsychology*, *6*(4), 297−312.

van der Schoot, M., Licht, R., Horsley, T. M., & Sergeant, J. A. (2002). Fronto-central dysfunctions in reading disability depend on subtype: Guessers but not spellers. *Developmental Neuropsychology*, *22*(3), 533−564.

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2004). Inhibition and shifting in children with learning deficits in arithmetic and reading. *Journal of Experimental Child Psychology*, *87*, 239−266.

van der Sluis, S., van der Leij, A., & de Jong, P. F. (2005). Working memory in Dutch children with reading and arithmetic related learning deficits. *Journal of Learning Disabilities*, *38*(3), 207−221.

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.

West, R., & Baylis, G. C. (1998). Effects of increased response dominance and contextual disintegration on the Stroop interference effect in older adults. *Psychology and Aging*, *13*(2), 206−217.

Weyandt, L. L., Rice, J. A., Linterman, I., Mitzlaff, L., & Emert, E. (1998). Neuropsychological performance of a sample of adults with ADHD, developmental reading disorder, and controls. *Developmental Neuropsychology*, *14*, 643−656.

Willcutt, E. G., Pennington, B. F., Boada, R., Ogline, J. S., Tunick, R. A., Chhabildas, N. A., et al. (2001). A comparison of the cognitive deficits in reading disability and attention-deficit/hyperactivity disorder. *Journal of Abnormal Psychology*, *110*(1), 157−172.

Wimmer, H., & Mayringer, H. (2002). Dysfluent reading in the absence of spelling difficulties: A specific disability in regular orthographies. *Journal of Educational Psychology*, *94*(2), 272−277.

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, *91*(3), 415−438.

Wu, K. K., Anderson, V., & Castiello, U. (2002). Neuropsychological evaluation of deficits in executive functioning for ADHD children with and without additional learning disabilities. *Developmental Neuropsychology*, *22*(2), 501−531.

Yntema, D. B. (1963). Keeping track of several things at once. *Human Factors*, *5*, 7−17.

Zar, J. H. (1999). *Biostatistical analysis* (4th ed.). London: Prentice Hall.