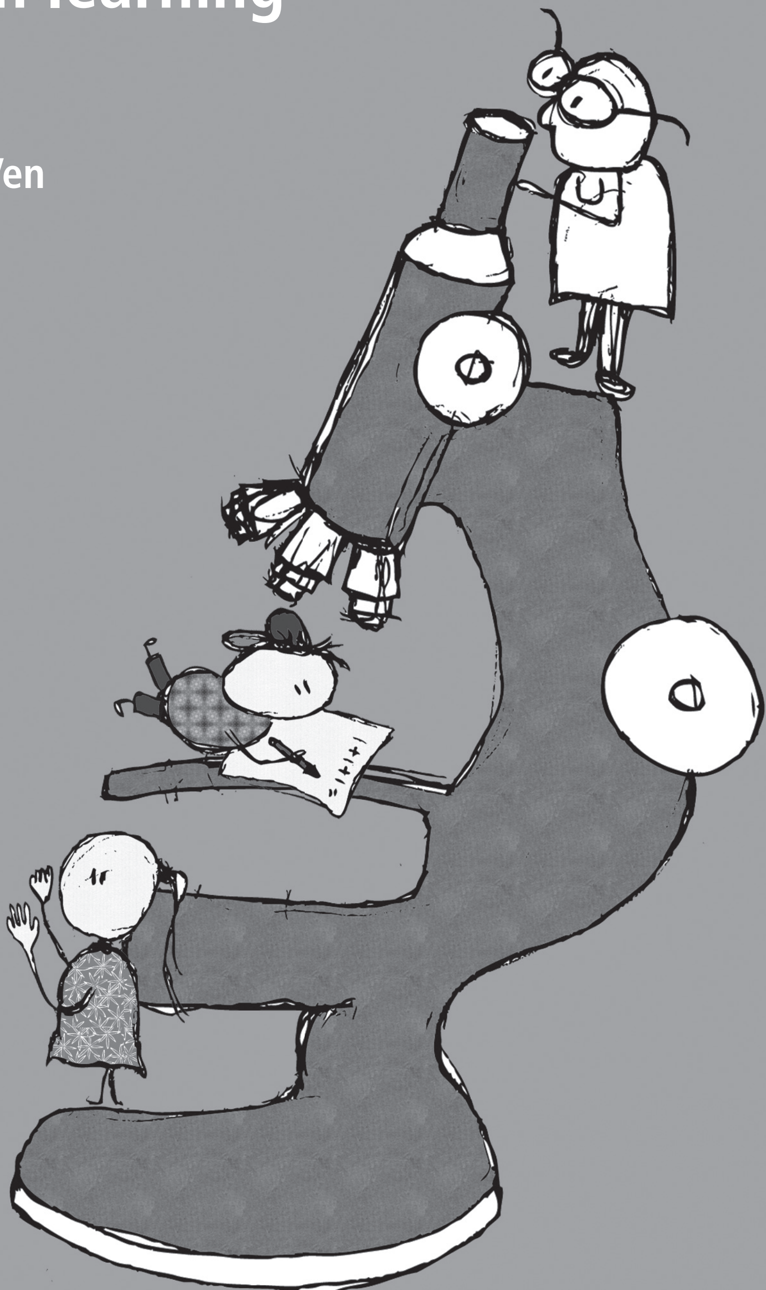
The structure of executive functions and relations with early math learning

Sanne van der Ven



Cover design Kitty van der Veer, Proefschrift.nu
Layout Renate Siebes, Proefschrift.nu
Printed by Labor Grafimedia BV, Utrecht
ISBN 978-90-393-5586-2

The structure of executive functions and relations with early math learning

De structuur van executieve functies en verbanden met leren rekenen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht op gezag van de rector magnificus,
prof.dr. G.J. van der Zwaan, ingevolge het besluit van het
college voor promoties in het openbaar te verdedigen
op donderdag 7 juli 2011 des middags te 4.15 uur

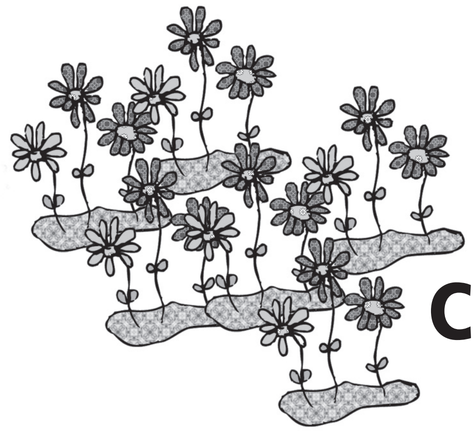
door

Sanne Helena Gerarda van der Ven

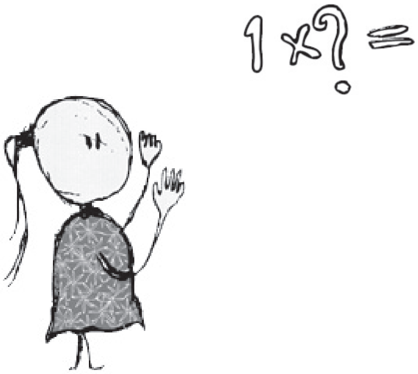
geboren op 16 maart 1982
te Waalre

Promotor: Prof.dr. P.P.M. Leseman

Co-promotoren: Dr. E.H. Kroesbergen
Dr. J. Boom



Contents



1

General introduction

Executive functions and mathematics: a short introduction

Every day, we make many decisions. We set goals, from long-term career plans to the breakfast we make in the morning. In order to do this, we have several abilities. We are able to control our impulses when this is necessary. We remember the goals that we have set, and we are able to monitor the progress of achieving our goals. Flexibility is crucial: when our progress is not according to plan, we are able to make adaptations on the spot, such as when we sense a burning smell and switch off the toaster. Without the ability to behave flexibly and adaptively we would be helpless in the ever-changing environment of everyday life. Perhaps this ability is the essence of what sets us human beings apart from computers and robots – at least, with the current technological state of the art. The ability to plan ahead and adjust our behavior flexibly is called executive functioning (Garon, Bryson, & Smith, 2008; Jurado & Rosselli, 2007). Because this is so important in everyday life, executive functioning has become an ever-growing area of research in the past decades. This research centers around five main themes, which are all interrelated.

One theme is the identification of the problems that patients with damaged executive functioning, caused by accidents or stroke, face in everyday life. This theme has dominated executive functioning research for a long time. It is investigated to what extent recovery in these patients is possible and how this recovery can be enhanced. Nevertheless, although in the past executive functioning was regarded as a unitary system (Baddeley & Hitch, 1974; Norman & Shallice, 1986), not every patient faces the same problems. While some are unable to adapt their behavior adequately because they cannot inhibit their impulses anymore, other patients have lost this ability because they suffer from severe memory loss and are therefore unable to remember what goals they have set. A second line of research therefore investigates the structure of executive functioning: can we identify different executive functions or processes, and how are these related to each other?

The third main theme is the identification of the brain structures that are involved in executive functioning. Since many patients scoring low on traditional executive function (EF) tasks such as the Wisconsin Card Sorting Task (Grant & Berg, 1948) and the Tower of London (Shallice, 1982) had suffered damage to the frontal lobes, it was long thought that the (pre)frontal cortex is the localization of all executive functions, and the term ‘frontal syndrome’ was even considered synonymous with impaired executive functioning (Godefroy, 2003). While it is still acknowledged that prefrontal areas are involved in executive control (Crone, Donohue, Honomichl, Wendelken, & Bunge, 2006), we now know that more areas in the brain are involved (Andres, 2003; Bunge & Wright, 2007; Collette et al., 2005).

A fourth strand of research deals with the development of executive functions in children. We all know the toddler with a temper tantrum in the middle of the supermarket because he doesn’t get the candy he wants. Compare this toddler to an adult that does not get the new job that he or she desired. It is far less acceptable and also highly unlikely for this adult to have a temper tantrum. This illustrates that young children are not as good as adults in controlling their behavior: their executive functions are still underdeveloped. Nevertheless, even infants already have some degree of executive functioning (Diamond & Goldman-Rakic, 1989): they are, for example, capable of controlling their gaze. There are still many open questions about how executive functions develop during childhood. While young children may superficially resemble patients with damaged executive functioning, the difficulties they face and the processes in the brain that are involved are by no means identical. Adults with brain injuries suffered more or less local damage to a fully developed system. This damage disturbs networks in the brain locally, and after initial recovery has taken place their situation is more or less static. In children, on the other hand, these neural networks are still being formed, based on input from the environment, on gene expressions, and on interactions between these two (Karmiloff-Smith, 2009). This development is a dynamic process, since the factors influencing this development change continuously. Processes that may be crucial for healthy development at some time point may not be so important earlier or later. This means that knowledge obtained from the challenges that adult patients face may serve as clues for developmental research, but can never be generalized to the developing brain.

Because executive functions are responsible for flexible and adaptive behavior, they are often considered as predictive of and responsible for other skills, for which this flexibility is important. The fifth research theme is therefore dedicated to the relationship between executive functions and these abilities in other domains. One of these domains is academic skills. In school, children learn a variety of new skills, and are therefore in a constant need to devise strategies to apply their newly learned knowledge to new situations, to monitor successes and failures of their new approaches and to change their strategies when necessary. In other words, they need their executive functions. Individual differences in executive functioning have often been found predictive of (later) scholastic achievement, such as reading and mathematics (Bull, Espy, & Wiebe, 2008; Gathercole & Pickering, 2000b; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; St Clair-Thompson & Gathercole, 2006); possibly they even play a larger role than intelligence (Bull & Scerif, 2001; Kroesbergen, Van de Rijt, & Van Luit, 2007; Lee, Ng, Ng, & Lim, 2004).

In this dissertation, the relationship between executive functions and mathematics in children in the first two years of primary education (age 6–8 years) was scrutinized. To this end, the structure of executive functions was analyzed. It was investigated whether it is possible to distinguish several more or less independent executive functions. The development of these executive functions was analyzed and relations with development in mathematics were investigated. In order to do this in detail, math development itself was also investigated. This means that in this dissertation the second, fourth and fifth main themes of research as delineated above were combined.

Structure of executive functions

Although executive functioning was treated as a unitary system in the past, nowadays the dominant view is that there are several executive functions that are interrelated but separate cognitive processes (Baddeley, 1996; Burgess, Alderman, Evans, Emslie, & Wilson, 1998; Miyake et al., 2000). Less is known, however, about the nature of these processes and the degree to which they are interrelated. Different distinctions have been proposed: examples are a distinction in inhibition, intentionality and executive memory (Burgess et al., 1998), and a distinction in coordinating performance on two separate tasks, selective attention, switching retrieval strategies, and the capacity to hold and manipulate information in long-term memory (Baddeley, 1996). The final three components of the latter distinction have been used in many studies, in which they have been renamed as inhibition, shifting, and updating respectively. Inhibition is usually defined as the ability to suppress a dominant, prepotent response deliberately in favor of another response or no response at all. Shifting, sometimes called switching or attentional flexibility, is the ability to switch between mental sets, rules or tasks, such as alternating between sorting objects according to color and shape. Updating, finally, is the ability to code incoming information for relevance, to store this information in working memory and to update it with newer, more relevant information when necessary.

Some studies found evidence favoring the validity of this theoretical distinction, both in young college students (Miyake et al., 2000) and in the elderly (Fisk & Sharp, 2004). Additional evidence has been found in a PET study, in which the distinction in three independent but related executive functions was confirmed: some brain areas were always active, but each executive function was also associated with uniquely activated areas (Collette et al., 2005). Nevertheless, the validity of this division in three executive functions remains controversial, as it could not always be found, neither statistically (Salthouse, Atkinson, & Berish, 2003) nor in terms of neural correlates (Tamnes et al., 2010).

There is some conceptual confusion considering the term updating. Some studies used the term working memory instead, and the difference between the two is not always clearly defined. There is a slight difference: whereas updating is the ability to change temporarily stored information in the light of incoming information, working memory is referred to as the ability to store and process information simultaneously. Updating is therefore typically measured with tasks in which the to-be-remembered information changes continuously, such as a string of letters of unknown length, of which the final three letters must be remembered: upon each new letter that is presented, the identity of the final three letters changes. Working memory, on the other hand, is measured with complex span tasks, in which a list of items must be remembered while another task must be performed simultaneously. An example is the listening span task, in which a series of sentences (such as “tomatoes play football”) must be judged for validity, but also the final word of each sentence must be recalled. However, since both types of tasks have been shown to load on the same statistical factor (St Clair-Thompson & Gathercole, 2006), the empirical distinction between the two concepts seems small, despite the conceptual distinction.

Measuring executive functions

Studies investigating executive functions are complicated because they face two challenges: dealing with the impurity problem in an appropriate way and finding the right method to score a task. The impurity problem refers to the fact that executive functions are defined as regulating other cognitive processes, such as visuospatial or language processing (Miyake et al., 2000). This means that tasks measuring executive functions are by definition impure: they also measure these other processes. Low performance on an EF task therefore does not necessarily imply that executive functioning is impaired, as the impairment may also be in these lower-level processes. The impurity problem can be addressed by assessing executive functions with multiple tasks that differ in lower-level processes that are involved, and subsequently extracting their shared variance with exploratory or confirmatory factor analysis. Care must be taken that these tasks differ as much as possible with regard to all other cognitive processes that are involved. When tasks are chosen carefully, the latent factors that arise may serve as purer measures of the executive function constructs they represent. Nevertheless, the literature relating executive functions to academic achievements still abounds with studies that use only one task to assess a particular executive function. It must, however, be noted that even when multiple tasks are used, the factors that emerge are only as pure as their tasks are. The variance that is shared by the indicator tasks is supposed to be limited entirely to the executive function the tasks were designed to measure. Since it is not possible to be certain whether this requirement is met, factor analysis can never completely eliminate the possibility that a particular factor structure arises because of secondary processes.

Another approach that has been used to circumvent the impurity problem, is the use of control tasks. This control task taps the same skills as the EF task: stimuli and response format are similar, but the executive function load is absent. For example, when a shifting task requires fast and flexible shifting between sorting according to color and according to shape, the control task requires fast sorting according to only one dimension, e.g., color. It is then assumed that the extra time needed for the executive condition compared to the control condition, or the larger number of mistakes that is made, can be attributed to the executive function process. This way, theoretically one EF task combined with a control task could suffice to measure a particular executive function.

Unfortunately, this method has severe drawbacks, leading to the second challenge in executive function research: scoring methods. When the reaction times of two conditions are compared, the difference can be obtained in different ways, that may potentially lead to very different conclusions (Lord, 1967). The most straightforward way is to take the absolute difference between the two conditions, but this means that measurement error increases (Linn & Slinde, 1977), and, in addition, this method fails to take into account the overadditivity effect: participants that are slow on a control task, slow down disproportionally on any more difficult task (Faust, Balota, Spieler, & Ferraro, 1999). Another option is a ratio score, which is obtained by dividing the EF score by the control score, but this method also has its drawbacks (Wainer, 1991). The best method is probably somewhere in-between taking the ratio and the difference, but a score can only be obtained when many conditions are presented that gradually differ in executive function requirement (Faust et al., 1999), so with the use of linear regression for each participant a line can be estimated through the averages of all these conditions. Since this method requires many different conditions, it takes a long time to administer and is therefore not feasible to use in young children.

Another problem is that performance is not only characterized by speed, but also by accuracy: strong participants are both fast and accurate, and weak participants are slow and inaccurate, but there is a speed-accuracy trade-off in each individual: how does a participant that is slow but accurate compare to another participant that is fast but makes many mistakes? In addition, there is a speed-accuracy trade-off between conditions: one participant may slow down more in the executive condition while maintaining equally high accuracy, on the other extreme there may be a participant that responds just as fast but makes more mistakes, and everything in-between is also possible. When speed and accuracy are considered separately, this trade-off is not taken into account (Wagenmakers, Van der Maas, & Grasman, 2007). These issues may have posed a threat to the validity of previous studies. While it is not easy to find a final solution in which both issues are addressed adequately, in this dissertation an attempt was made in *Chapter 3* by including baseline speed measures, executive speed measures and accuracy in one structural equation model.

Executive functions in children

In the past, it was often assumed that young children did not have executive functions yet. The immature prefrontal cortex, largely responsible for executive functioning, was even said to be 'functionally silent'. Now, we know that even in their first year of life, infants are capable of exerting executive control (Diamond & Goldman-Rakic, 1989). Nevertheless, executive functions develop strongly during childhood (Garon et al., 2008), and are not mature until puberty or even early adulthood (Huizinga, Dolan, & Van der Molen, 2006).

Still an open question on which no consensus has been reached yet, is whether the structure of executive functions in children is the same as in adults, i.e., whether executive functions in children can be divided in inhibition, shifting, and updating (Espy, 2004). This distinction has been confirmed in children in some studies (Espy et al., 2004; Hughes, 1998; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003), but not or only partially in others (Huizinga et al., 2006; Van der Sluis, De Jong, & Van der Leij, 2007; Wiebe, Espy, & Charak, 2008). These contradictory results might reflect age differences in the structure of executive functions: it has been suggested that executive functioning is a hierarchical system, with more complex functions arising later during childhood (Garon et al., 2008). Statistically this may yield different factor structures at different ages. However, this is not a sufficient explanation, as evidence both in favor and against the three-factor structure was found in the same wide age range: from preschoolers to adolescence. An alternative explanation may be that the factor structures that were found in these studies were influenced unintentionally by task-specific characteristics. As described before, when tasks are highly similar in non-executive aspects, a statistical factor that is based on these tasks will also partially reflect these unintended constructs. Since the tasks that were used differed between studies, differences in obtained factor structures might be due to differences in these shared non-executive characteristics of the tasks, rather than to differences in the structure of executive functions in different populations.

Some studies indeed used highly similar tasks to measure the same construct, suggesting that the impurity problem may be a real threat to the validity of the factor structure in these studies. Others were not so careful in how scores were obtained from the tasks. A carefully designed study is therefore needed to address this issue. In this dissertation an attempt was made to do this, by designing a test battery with tasks that differed as much as possible from each other in non-executive aspects, in order to handle the task impurity problem as well as possible. This test battery was administered four times with six-month intervals, in order to follow the development of the children.

Learning mathematics

When six-year-old children enter the first year of primary education, they have already acquired preparatory mathematical skills, such as counting, predominantly in informal situations. These skills are important, as they are predictive of later mathematical skills (Krajewski & Schneider, 2009a; Toll, Van der Ven, & Kroesbergen, 2010). Nevertheless, systematic formal instruction in mathematics starts in grade 1. In the first years of primary schooling, children learn mathematical operations: addition and subtraction in grade 1, and in grade 2 also multiplication. In the Netherlands, all mathematical teaching methods nowadays use a so-called 'realistic mathematics' approach (Freudenthal, 1991), which is based on constructivism theories: rather than drilling mathematical facts, such as the tables of multiplication, these math problems are embedded in a meaningful context. For example, instead of memorizing that 4 x 6 = 24, this problem is presented in a little story: "John is carrying 4 bags, each of which contains 6 books. How many books is John carrying?" Learning mathematics this way rather than memorizing answers is thought to encourage deeper understanding. One of the aims of early mathematics education is still memorization of basic number facts, such as addition under 20 and single-digit multiplication, but it is believed that this aim is also achieved by sufficient exposure to these problems in a meaningful way: each correct solution of a mathematical procedure strengthens the connection between problem and answer, until this connection is strong enough to retrieve the answer without mathematical procedures (Siegler, 1988). Fast and accurate problem solving strengthens the connections between problem and answer and increases the chances of finding the correct answer the next time the same problem is presented (Geary, Brown, & Samaranayake, 1991).

When children solve mathematical problems of which they have not memorized the answers yet, they can use several different computational strategies. For example, when children solve multiplication problems, they can make a drawing of the problem and count the items in their drawing. They may also count on their fingers, or use repeated addition, i.e., they may change the multiplication problem into an addition problem: 3 x 7 = 7 + 7 + 7. They may also use derived facts or shortcuts to reach the answer: if a child already knows that 3 x 4 = 12, this knowledge can be used to solve the unknown problem of 6 x 4, as 6 x 4 = 2 x (3 x 4) = 2 x 12. This order of strategies reflects increasing maturity: each consecutive strategy requires fewer steps and is therefore potentially less error-prone than the previous strategy, but the demands on a knowledge base that is already present also increase. Counting does not require much mathematical knowledge; repeated addition requires a certain proficiency in addition and derived facts can only be used when the child already knows the answer to another relevant math problem. Since children are adaptive in their strategy choice (Torbeyns, Verschaffel, & Ghesquière, 2002, 2004), i.e., they tend to apply the most mature strategy they are able to perform, it can be expected that the strategies that children use change over time. This is precisely what has been found, and the strategy maturity of their choice is also predictive of children's mathematical successes (Imbo & Vandierendonck, 2007; Lemaire, 2010). Siegler (1996) suggested a model of strategy development that can be applied to all cognitive skills: the Overlapping Waves model. According to this model, children use increasingly more mature strategies, but they do not do so in an all-or-none fashion. Instead, children have a strategy repertoire that slowly changes towards more mature strategies. If the frequencies with which the different strategies are used are plotted over time, the shapes of these curves resemble overlapping waves. If this model is applied to learning single-digit multiplication, it can be predicted that children start using predominantly counting strategies, but later, as their experience with the problems and their computational skills grow, slowly shift towards repeated addition, which is in turn gradually replaced by derived facts and eventually by retrieval only. In this dissertation this hypothesis was tested by using categorical growth models. This technique also served as a statistical test of Siegler's (1996) Overlapping Waves model, which thus far has not been formally tested, but has mainly served as a metaphor. This was done by means of a microgenetic study. In microgenetic studies, the development of a skill is investigated in depth, by means of dense, in-depth observations during a relatively short period of time, while rapid developmental changes are taking place (Siegler & Svetina, 2002). The method is time consuming, but can yield deeper insight in the processes that underlie development than single measurements.

Executive functions and mathematics

Executive functions are important for the acquisition of academic skills, as confirmed by numerous studies that found a relationship between mathematics and executive functions. Especially the evidence for a relationship with updating (or working memory) is strong (De Smedt et al., 2009; Lee et al., 2011; Passolunghi, Mammarella, & Altoè, 2008; Swanson & Kim, 2007). Relations with inhibition and shifting have been found in some studies (Blair & Razza, 2007; Bull et al., 2008; Van der Sluis et al., 2007) but these were not or only partially found in other studies (Censabella & Noël, 2008; Espy et al., 2004; Lee et al., 2011; Rasmussen & Bisanz, 2005).

Nevertheless, the question why exactly these relations are expected to exist in the first place, has received remarkably little scientific attention. Exactly which cognitive demands do mathematical problems impose, how are executive functions used to meet these demands and at which stage(s) of the mathematical problem solving process do they play a role? This is left implicit in almost all studies investigating executive functions and mathematics. Thus far, many studies have established significant relationships or sometimes failed to establish these, but possible underlying mechanisms have not been investigated.

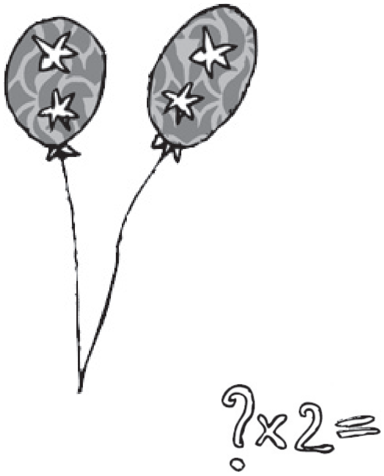
It might be possible that inhibition is necessary to inhibit immature strategies, such as finger counting, once a child has learned more mature strategies. Inhibition may also be needed to suppress the use of prepotent but inappropriate strategies, such as the well-practiced addition of two numbers when they should be multiplied instead. Inhibition may also be important to prevent irrelevant information in a context story from taking up valuable working memory resources. Shifting could be important to switch between different strategies, and between steps in a multi-step problem. Updating, finally, may be necessary to store and retrieve partial results, and to remember the important information that was presented in a complex math problem.

The role of these three executive functions may also be dynamic and change during development. Perhaps inhibition and shifting are not very important during early stages of mathematics learning, when the child does not know many prepotent but inappropriate strategies yet, and math problems are still relatively easy. Working memory or updating is probably most important when a child uses immature strategies, such as counting and repeated addition, to solve a problem. These strategies require many intermediate steps, which must all be remembered. Good working memory skills may enhance the formation of strong neural connections between math problem and answer, allowing the child to progress to more mature strategies faster. Once the child has memorized the answer to a certain math problem, executive functions are probably barely needed to retrieve the answer, as it is then retrieved automatically from long-term memory. These hypotheses need further confirmation in empirical studies. The microgenetic study was designed as a first step to do this.

Outline of this dissertation

This dissertation contributes to the growing body of knowledge of executive functions, as it has two main aims that are centered around three of the five research themes mentioned before. The first aim of the dissertation was to unravel the structure and development of executive functions in children. The second aim was to investigate how executive functions are related to mathematics learning in children. These aims were first addressed in a pilot study. Because the results of this study were promising, a longitudinal study was designed in which the aspects of the pilot study were investigated in more depth. In this longitudinal study, over 200 children were followed during the first two years of primary education. A test battery designed to measure the executive functions of inhibition, shifting, and updating was administered four times, with six-month intervals. The factor structure of the test results was examined and related to mathematical performance. Approximately half of the children participating in the longitudinal study also took part in a microgenetic study, in which mathematical development was investigated in depth.

In *Chapter 2* the results of the pilot study are reported. In this study, children in the first grade of primary education were followed for four weeks. Twice a week, their addition skills were assessed. Different aspects of their progress were related to their executive functions, of which the structure was analyzed using exploratory factor analysis. Although the sample was small, the results looked promising and gave rise to the larger longitudinal study, presented in the later chapters. In *Chapter 3*, the factor structure of executive functions was assessed. To this end, confirmatory factor analysis was performed on the data of the first and last measurement: at the beginning of grade 1 and at the end of grade 2. It was investigated whether the distinction in inhibition, shifting, and updating was tenable in this age group, with carefully chosen tasks to minimize the possible effects of task impurity. *Chapter 4* concerns the longitudinal analyses. Latent growth modeling techniques were used to model the development of both mathematics and executive functions, using the data from all four time points. In order to unravel the longitudinal relationship between mathematics and executive functions, these latent growth models were also related to each other. In *Chapter 5*, the results of the microgenetic study are reported, in which the development of a mathematical skill, single-digit multiplication, was investigated in depth. On a weekly basis during eight weeks, the children's problem solving strategies on various single-digit multiplication problems were assessed. Their strategy selection and accuracy were then modeled in a categorical growth model and related to executive functions. Since by the time of this study previous longitudinal results had showed that only updating was significantly related to mathematics, only this part of the executive function battery was included in the chapter. Because the term updating is usually used in combination with inhibition and shifting, in this chapter the term working memory is used instead; although there are some slight differences between the two concepts, they are closely related. In *Chapter 6*, the general discussion, the results of the four empirical chapters are discussed and a theoretical framework is introduced that may provide a better account of the empirical findings.



2

Executive functions and learning mathematics in seven-year-old children: A pilot study

An abbreviated and translated version of this chapter has been published:
Kroesbergen, E. H., Van der Ven, S. H. G., Kolkman, M. E., Van Luit, J. E. H., & Leseman, P. P. M. (2009). Executieve functies en de ontwikkeling van (voorbereidende) rekenvaardigheid. *Pedagogische Studiën*, 86, 334-349.

Abstract

Children differ greatly in their mathematical proficiency. In this study the role of executive functions in the explanation of these differences was investigated in twenty-six normally developing children in first grade who were learning addition. The children were followed for four weeks, during which they learned addition over ten. They all also completed a test battery of executive functions and fluid intelligence. The validity of the distinction in three executive functions: inhibition, shifting, and updating was assessed. These executive functions were related to different aspects of the mathematics learning process, such as initial level and improvement. Especially updating and shifting were found to be related to mathematical development; inhibition only played a minor role. These results indicate that executive functions are important in mathematical development and that the role of each executive function is different.

Introduction

Already at a young age, children show large individual differences in mathematical proficiency and around 6–7% even encounter serious difficulties in math acquisition (Geary, 2004). Although intelligence is related to mathematical performance (Resing, Ruijssenaars, & Bosma, 2002), it only explains 9–25% of the variance and many children with math learning difficulties have an IQ in the normal range. Moreover, intelligence is a broad measure of cognitive processing that does not allow a detailed analysis of the processes underlying academic achievements. When investigating mathematical performance, an increasing number of studies therefore took a different approach: executive functions as a predictor of mathematical performance.

Executive functions are the set of cognitive skills that are necessary for goal-directed behavior, such as planning, monitoring, and inhibition (Garon, Bryson, & Smith, 2008). When learning new skills these abilities are therefore very important. The relationship between executive functions and learning academic skills such as mathematics is therefore now a topic of study receiving increased attention (e.g., Bull & Scerif, 2001; Passolunghi, Mammarella, & Altoè, 2008; Swanson, 2006a). However, the majority of these studies investigated mathematical knowledge on only one or a few occasions and administered relatively easy mathematical problems. Therefore the role of executive functions during the learning process itself, when they are presumably most important, remains largely unrevealed.

In addition, only few studies investigated the role of different executive functions, even though the diversity of executive functioning is now well-established (Hughes, 1998; Miyake et al., 2000). The precise distinction is still a topic of debate. This study was an attempt to fill this gap: children in the first year of primary education were tested repeatedly on their addition skills, while they were learning addition. Different aspects of their mathematical development were related to three different executive functions: inhibition, shifting, and updating. The validity of this distinction in three executive functions was also scrutinized. This approach yields more precise information of the relationship between executive functions and learning mathematics, which is a step forward in the predictions of future difficulties in mathematics in young children and in the development of individual intervention programs that take into account a child's specific strengths and weaknesses.

The structure of executive functions

Many studies investigating executive functions use Baddeley and Hitch's (1974) model of working memory, consisting of a central executive and three slave systems: the phonological loop, the visuospatial sketchpad and the later added episodic buffer (Baddeley, 2000). The phonological loop and visuospatial sketchpad are temporary stores for phonological and visuospatial information and are measured with simple span tasks, such as digit span forward or Corsi blocks. The episodic buffer also stores information for a short period of time, but binds information from different modalities, including long term memory. The central executive, finally, serves as an attentional control mechanism that coordinates the information stored in the two slave systems (Baddeley, 2000; Baddeley & Hitch, 1974).

Although the central executive was originally regarded as a unitary system, it is now commonly seen as consisting of a variety of skills that together constitute the executive functions (Baddeley, 1996), although there is no consensus yet about the exact number and nature of executive functions. Research trying to identify the nature of these different executive functions is hampered by the impurity problem. This refers to the fact that executive functions control other cognitive processes, which makes them impossible to be measured in isolation (Miyake et al., 2000). A possible solution to overcome this problem is a latent variable analysis. By using this method, Miyake et al. (2000) confirmed the existence of three distinct executive functions in healthy adults: inhibition, shifting, and updating. Inhibition involves the conscious suppression of a dominant response in favor of another response or no response at all. Shifting involves switching back and forth from one response set to another. Updating accounts for storage of intermediate results and revising this information based on new input. The three factors of inhibition, shifting, and updating were found to be distinct but also moderately correlated (Miyake et al., 2000). A similar pattern of both activation common to all three brain areas and unique activation areas for each executive function has been found in the brain (Collette et al., 2005).

Executive functions are in development until late adolescence or even young adulthood (Gathercole, Pickering, Ambridge, & Wearing, 2004; Huizinga, Dolan, & Van der Molen, 2006; Luciana & Nelson, 1998). Whether these differences are only quantitative or qualitative, i.e., whether executive functions are also organized differently in the child brain is still unclear. Research regarding the structure of executive functions in children has yielded conflicting results. Some studies found factor structures that were similar to the structure of inhibition, shifting, and updating in children aged 6–13 (Brocki & Bohlin, 2004; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003), and even as young as 2–5 years (Espy et al., 2004; Hughes, 1998), but other studies could not replicate this factor structure (Huizinga et al., 2006; St Clair-Thompson & Gathercole, 2006; Van der Sluis, De Jong, & Van der Leij, 2007; Wiebe, Espy, & Charak, 2008). Especially the inhibition factor was often not found. The latter results confirm the possibility that in young children executive functions are not as differentiated as in adults, and each executive function may have its own maturational path (Bull, Espy, & Wiebe, 2008; Bull & Scerif, 2001; Van der Sluis et al., 2007). An alternative explanation, however, is that the factor structure that arises is a reflection of the similarity of the tasks that were used: the aim is to construct purer measures of executive functions, but if tasks show too much overlap in non-executive processes, the factors that arise are still impure. Altogether, these results suggest that in young children executive functions are immature, yet present and probably similar (though not necessarily identical) in structure as in adults.

Executive functions and learning mathematics

When solving math problems, all three executive functions of inhibition, shifting, and updating are expected to be involved. While the exact nature of the relationship between executive functions and mathematics is left implicit in many studies, inhibition may be hypothesized to be required when irrelevant information in a math problem must be suppressed and when the use of an old and well-practiced strategy, such as finger counting, must be inhibited in favor of a newly acquired, more efficient strategy. Children with low inhibitory skills may also have difficulty refraining from stating the first and probably incorrect number that comes to their mind, leading to a high error percentage and a slower automatization process. Shifting proficiency might be expected to be important when solving multiple-step problems, and to enable flexible strategy switching between problems. A child with low shifting abilities that has solved some subtraction problems may fail to solve the problem 201 – 199 = ... by means of indirect addition (199 + 2 = 201). Updating, finally, is supposed to be responsible for the storage and retrieval of intermediate results in working memory. A child with low updating abilities may have forgotten the results of intermediate steps or make more procedural errors. A problem like 6 x 7 can be solved by changing the problem into an addition problem (7 + 7 + 7 + 7 + 7 + 7), but a child with low updating abilities may lose track of how many 7s have been added.

A relationship between executive functions and mathematical skills has been found, both in studies that compared a group of mathematically disabled children to a normally developing group and in studies investigating these relationships in the normal range of academic achievers. Relationships have been found between math and inhibition (Passolunghi et al., 2008; St Clair-Thompson & Gathercole, 2006; Swanson & Kim, 2007), shifting (Censabella & Noël, 2008; McLean & Hitch, 1999; Rasmussen & Bisanz, 2005) and updating (De Smedt et al., 2009; Passolunghi et al., 2008; Van der Sluis, De Jong, & Van der Leij, 2004). However, there are also studies that failed to find a relationship between executive functions and mathematical skills, especially inhibition (Censabella & Noël, 2008; McLean & Hitch, 1999; Rasmussen & Bisanz, 2005; Van der Sluis et al., 2004) and shifting (Espy et al., 2004; Van der Sluis et al., 2004).

Studies investigating the slave systems of Baddeley's working memory model also found mixed results. The phonological loop has been found to be related to mathematical achievements (Panaoura & Philippoua, 2007) but other studies did not find this relationship (Passolunghi, Vercelloni, & Schadee, 2007). Similarly, children with mathematical difficulties have been found to perform poorer on phonological loop tasks (D'Amico & Guarnera, 2005), but this could not be confirmed in other studies (Geary, Hoard, & Hamson, 1999; McLean & Hitch, 1999; Wilson & Swanson, 2001). Children with mathematical difficulties have also been found to perform lower on visuospatial sketchpad measures in some studies (McLean & Hitch, 1999; Wilson & Swanson, 2001) but others found no relation between the visuospatial sketchpad and mathematical performance (Bull, Johnston, & Roy, 1999; Swanson, 2006a).

There are several possible explanations for these contradictory results. First, all studies measured mathematical abilities only once, often only measuring already established math knowledge: the math problems were often relatively easy and answers can be expected to be retrieved directly from longterm memory. Executive functions, however, are expected to be most important when skills are still being learned and flexible strategies must be employed to obtain the answer to a math problem. Therefore it is likely that these studies underestimate the influence of executive functions on mathematical development. On the other hand, impurity of the executive tasks, especially the presence of mathematical elements in these tasks, might lead to spurious relationships that are caused by non-executive elements of the tasks.

The present study

The aim of this study was to investigate the role of the executive functions of inhibition, shifting and updating, as well as the visuospatial sketchpad, phonological loop and fluid intelligence, in the development of mathematical knowledge in the first year of primary education. A mathematical skill was chosen that the children had not learned yet at the beginning of the study: addition over ten. In order to maximize the chances of finding significant development during these four weeks, only tie problems were used: problems in which both addends are the same. These problems are solved faster and more accurately than non-ties (Groen & Parkman, 1972).

The development of tie problem solving was investigated using some of the characteristics of a microgenetic study design. In this design the development of a particular skill is studied intensively during a relatively short period of time in which a developmental transition is expected (Siegler, 1996). In a mathematical study this implies that extensive testing takes place during the period in which mathematical skills are being learned: a period when executive functions are expected to play a large role and retrieval is expected to be of lesser importance. Based on Siegler's dimensions of development which had been proven useful in other studies (e.g., Siegler & Svetina, 2002), different aspects of the children's mathematical development during this period were investigated: the initial math level, improvement during the testing period, path of improvement during the four weeks of the study, and the generalization to different but related math problems: tie subtraction problems, such as 16 – 8. Although children were familiar with the principles of subtraction, these particular problems were not taught explicitly during the study. However, when one knows that 8 + 8 = 16, it is easier to solve 8 + . = 16, and, if a child realizes this, 16 – 8 = . is easier too. Differences between children on each aspect identified by Siegler, i.e., initial level, improvement, path and generalization, and one additional aspect, namely error making, were correlated with the different executive functions.

Because of the intensive data collection and the pioneer nature of this type of study, the sample size of this study is small and results should therefore be interpreted cautiously. Nevertheless, we were able to reveal numerous significant relationships, indicating that the study had sufficient power and this approach is a fruitful direction for future studies.

Method

Participants

Twenty-seven first grade Dutch children, all from the same class, were recruited from a Dutch primary school in a middle-sized village. Due to repetitive non-compliance to the task requirements, one boy's results were removed. This left a total of 26 children aged 6;5 to 7;8 years (M = 7;0 years, SD = 4.0 months; 13 boys, 13 girls). During the time of study (spring), the children learned tie problems over 10 as a part of the normal curriculum. At the beginning the children were familiar with basic single digit addition and subtraction, but did not have much experience with problems exceeding 10.0.

Procedure

Twice a week, during four weeks, the children were given a three-minute mathematical test, administered by the teacher in the classroom to all children simultaneously. Five children were absent on one mathematics testing session and another four missed two sessions. In the same period the children were tested individually on the slave systems, intelligence and executive function tasks, in two thirty-minute sessions on different days. In the first session the children completed Digit Span Forward, Digit Span Backward, Symbol Shifting, Keep Track, and Expressive Attention. In the second session, children were tested on Dot Matrix, Children's Color Trail Test, Tower of London and Raven's Coloured Progressive Matrices. All children completed this test battery.

Instruments

Executive functions

For each of the executive functions of inhibition, shifting, and updating, two tasks were selected that were used in previous studies but that differed as much as possible in all non-executive aspects. Whenever possible, control tasks were used in addition to the executive tasks. Care was also taken that the tests did not contain components related to mathematics. All tasks were preceded by some practice items.

Inhibition

Expressive Attention. The Expressive Attention task comes from the Cognitive Assessment System (Naglieri & Das, 1997). The task requires the child to name the real life size (big or small) of an animal, while ignoring the size of the drawing (see Figure 2.1A). In the control condition, a page with 40 of the animal pictures, all drawn equally large, was shown and the real life sizes of all the animals on the page had to be named as rapidly as possible. In the interference condition most big animals were drawn small and vice versa. Each task was preceded by ten practice items. Number of mistakes and time needed to complete the task were recorded. An interference score was obtained by adding three extra seconds for each mistake to the total time obtained in the interference condition, and by dividing this time by the time obtained in the control condition. This is not the standard procedure of the CAS, but the standard procedure did not allow for a score combining the three vital elements of the task: speed in the control condition and speed and number of mistakes in the inhibition condition.

Tower of London. The NEPSY version (Korkman, Kirk, & Kemp, 1998) was administered on a computer. The task involves three differently colored balls that are placed on three pegs of unequal sizes. The child was shown a goal configuration of the balls and had to copy this configuration within a time limit by moving the balls one by one

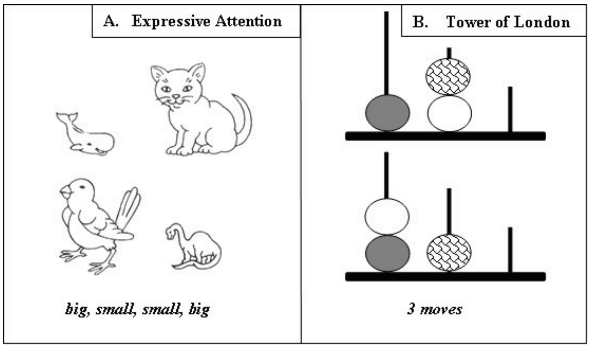


Figure 2.1 Example items of the two inhibition tasks.

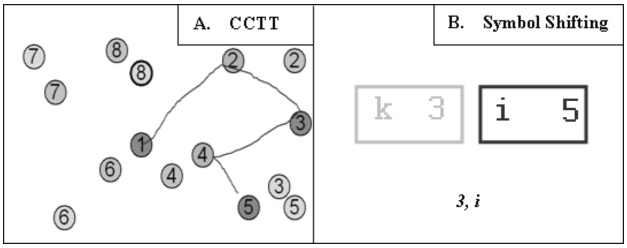


Figure 2.2 Example items of the two shifting tasks.

to other pegs in a minimum number of moves (see Figure 2.1B). There was one practice trial and twenty test trials of increasing difficulty. Testing was discontinued when on four subsequent items the minimum number of steps or the time limit was exceeded. This task is often considered a planning task, but since it requires repressing making intuitive but wrong moves impulsively, it has also been shown to require inhibition (Welsh, Satterlee-Cartmell, & Stine, 1999; Zook, Davalos, DeLosh, & Davis, 2004).

Shifting

Children's Color Trail Test. A computerized adaptation of the Children's Color Trail Test (Llorente, Williams, Satz, & D'Elia, 2003) was used. This version consisted of a screen filled with circles in two different colors. Each circle contained a number, such that the numbers 1 to 15 were present in both colors. The circles had to be connected while alternating colors: from blue 1 to green 2 to blue 3, and so forth (see Figure 2.2A). When a mistake was made, the mouse pointer returned to the previous, correct circle, from where the child could continue. Two control conditions were presented first, in which the children also connected the circles, but each number was present only once. A final score was obtained by dividing the time needed for the shifting condition by the average time of the two control conditions.

Symbol Shifting. This task was adapted from the Symbol Shifting task developed by Van der Sluis et al. (2007). A sheet with blue and yellow colored boxes was shown, each box containing a pair of a same-colored letter (a, i, s, or k) and digit (2, 3, 4, or 5; see Figure 2.2B). The child had to name the letter when the color was blue and the digit when the color was yellow. There were 8 practice items, followed by 40 test items. For each error five extra seconds were added to the total time (a higher number than in Expressive Attention, as the total time was longer in this task). Prior to the task, the child named two control sheets, one with 40 digits and one with 40 letters. As a final score, the ratio between time taken for the shifting task and the average time for the two control tasks was taken. Sometimes a child reversed the naming rules, naming the letter when the color was yellow and the digit when the color was blue. Scoring all items as incorrect would yield a score that is too low, since the child did show flexible rule switching. When a child reversed the naming criteria for more than four subsequent items, only ten extra seconds were added to the total time and from then on false answers were considered correct and vice versa.

Updating

Keep Track. A computerized version of the Keep Track task was created. This task was adapted from the task used by Van der Sluis et al. (2007). The child was shown pictures that belonged to five categories: letters (a, i, s, k), numbers (1, 2, 3, 4), shapes (square, triangle, circle, star), animals (dog, cat, fish, bird) and vehicles (train, car, bike, airplane). The pictures were shown in series of ten; each picture was displayed for three seconds (see Figure 2.3). The child named each picture and tried to recall the last item of certain before-mentioned categories after each series. During the series, small black-and-white pictures symbolizing the to-be-remembered categories were shown in the upper left corner. The number of to-be-remembered categories increased from one to three, with two series of each difficulty level, yielding a total of six series. Each correct answer was noted, yielding a maximum score of twelve correct responses. The total

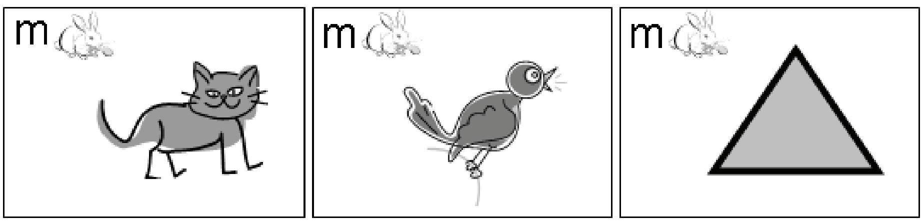


Figure 2.3 Keep Track. The child had to name all pictures and report the last items of the categories depicted in the upper left corner. In this series first cat must be remembered but after the bird is presented, bird must be remembered and cat can be forgotten. The triangle must be named but does not have to be recalled.

number of correct responses was used as a final score. Prior to testing, the child was familiarized with the pictures belonging to each category, and the test was preceded by two practice items.

Digit Span Backward. The Dutch version of Digit Span Backward from the Automated Working Memory Assessment test battery (Alloway, 2007) was administered. Starting with a sequence of two digits, the sequence length increased after four trials of the same length had been remembered correctly. The task was discontinued when three trials of the same length had been reproduced incorrectly. For each completed length six points were given, plus one additional point for each completed sequence of the length that was not finished.

Visuospatial sketchpad

Dot Matrix. An adapted version of the dot matrix task from the Automated Working Memory Assessment computerized test battery (Alloway, 2007) was used. The child was shown a 4 x 4 grid of empty white squares. In one square a big red dot appeared. After disappearing, the dot reappeared in another white square. The child had to reproduce the sequence of the dot by mouse-clicking the same sequence in an empty grid. The number of dots in a sequence gradually increased. The child progressed to a longer sequence when four trials of the same length had been remembered correctly. The task was discontinued when three trials of the same length had been reproduced incorrectly. For each completed length six points were given, plus one additional point for each completed sequence of the length that was not finished.

Phonological loop

Digit Span Forward. A Dutch version of the Digit Span Forward task, part of the Automated Working Memory Assessment test battery (Alloway, 2007), was used. The child heard a sequence of digits and had to repeat the sequence. Continuation and scoring were similar to Dot Matrix.

Fluid intelligence

Raven's Coloured Progressive Matrices (Raven's CPM). A computerized version of the Raven's Coloured Progressive Matrices (Raven, 1962) was administered. A colored pattern was shown with a missing piece. Below the pattern, six pieces, all fitting in the blank but with different patterns, were shown. The child had to select the piece that fitted in the pattern above. The total number of correct answers was recorded. There were 5 practice and 31 test items.

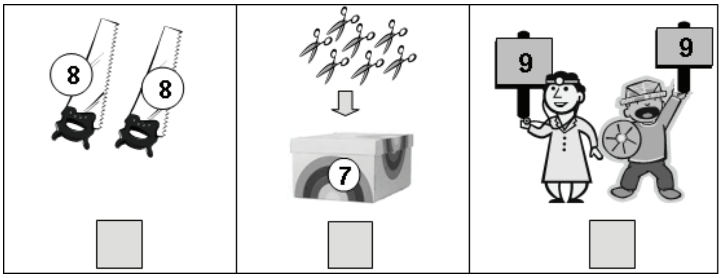


Figure 2.4 Example items of mathematical problems.

Basic processing speed

No separate tasks measuring basic processing speed were performed, but a measure of speed was computed by taking the average of the normalized scores of the control conditions of the three speeded executive tasks: Expressive Attention, Symbol Shifting and Children's Color Trail Test. The basic processing speed measure was used as a control measure.

Mathematics

Addition. This test, administered eight times, contained the four tie problems over 10, i. e., 6 + 6, 7 + 7, 8 + 8, and 9 + 9, presented as pictorial math problems related to everyday life. Some example items are shown in Figure 2.4. The children solved as many problems as they could within three minutes. Each test was five pages long. Each page contained three tie problems and one simple addition problem under 10, preventing children from memorizing the answers. Children could not look back to answers given on previous pages. On some pages both addends were digits; on others the children counted one addend. The booklets always contained the same problems, but in each session the order of the problems varied and the lay-out was different. Different aspects of math development were investigated: initial level, improvement, error making and path of development. These analyses were based on the number of problems solved correctly per minute unless specified otherwise.

Subtraction. In order to investigate generalization to unpracticed but related math problems, a subtraction test was administered in the first and last testing session only. It contained the same problems as in the tie condition, but now presented as subtraction problems: 12 – 6, 14 – 7, 16 – 8, and 18 – 9. There were five pages, each containing three tie subtraction problems and one simple subtraction problem below 10. Presentation of the items was similar to the addition problems. Again, the number of problems solved correctly per minute was used in analyses, unless specified otherwise.

Table 2.1 Descriptive statistics of all tasks

Task	*M*	*SD*
Inhibition		
Expressive Attention	1.31[a]	0.17
Tower of London	10.58	2.04
Shifting		
Children's Color Trail Test	2.21[a]	0.54
Symbol Shifting	3.26[a]	0.83
Updating		
Digit Span Backward	8.26	3.27
Keep Track	8.08	1.98
Visuospatial sketchpad		
Dot Matrix	19.04	3.42
Phonological loop		
Digit Span Forward	21.88	3.23
Fluid intelligence		
Raven's Coloured Progressive Matrices	26.19	5.31
Basic processing speed		
Speed	0.00[a]	0.77

[a] high performance is indicated by low scores

Results

Structure of executive functions

First, the structure of the executive functions was investigated. Descriptive statistics of performance on the executive and related tasks are presented in Table 2.1. The distributions are all normal or approximately normal.

Pearson's correlations were computed between the executive tasks, the phonological loop and visuospatial sketchpad tasks, Raven's CPM and basic processing speed. Tasks were recoded when necessary: low scores always indicate low performance. All correlational tests were one-tailed, as correlations were expected to be positive between all tasks and the sample size is small, which increases the risk of a type II error. A correlation matrix is shown in Table 2.2. Significant correlations with basic processing speed were present, especially with Keep Track. Therefore partial correlations, corrected for speed were also computed and presented in Table 2.2.

When basic processing speed was not partialed out, as expected, the correlations between the two shifting tasks and the two updating tasks were significant. The correlation between the two inhibition tasks was marginally significant ($p = .06$). A few other significant

Table 2.2 Pearson's correlation coefficients between the executive functions and control tasks (below diagonal) and partial correlation coefficients controlling for basic processing speed (above diagonal)

	Task	1	2	3	4	5	6	7	8	9
1	Expressive Attention	-	.30	.03	-.14	.01	-.18	.09	-.19	-.42*
2	Tower of London	.32	-	.32	.20	.32	.42*	.05	.24	.04
3	CCTT[a]	.02	.31	-	.42*	.14	.17	.06	.38*	.08
4	Symbol Shifting	-.15	.17	.42*	-	.16	.51**	.31	.19	-.10
5	Digit Span Backward	-.01	.29	.15	.07	-	.40*	.30	.30	.17
6	Keep Track	.06	.22	.16	.46**	.40*	-	.43*	.18	.00
7	Dot Matrix	.13	-.02	.07	.33*	.33	.53**	-	.06	.42*
8	Digit Span Forward	-.21	.20	.38*	.21	.32	.24	.11	-	.41*
9	Raven's CPM[b]	.43*	-.01	.09	-.08	.19	.14	.46**	.43*	-
10	Speed	-.12	-.18	.05	.10	.15	.61**	.34	.16	.23

[a] Children's Color Trail Test
[b] Coloured Progressive Matrices

Table 2.3 Factor structure of executive functions with Direct Oblimin rotation

Task	**Factor 1 Updating**	**Factor 2 Inhibition**	**Factor 3 Shifting**
Inhibition			
Expressive Attention	-.02	**.86**	-.10
Tower of London	.35	**.72**	.39
Shifting			
Children's Color Trail Test	.17	.20	**.86**
Symbol Shifting	.46	-.19	**.79**
Updating			
Digit Span Backward	**.87**	.14	.11
Keep Track	**.82**	.04	.36

correlations were also found, which can be found in Table 2.2. All significant correlations were in the expected direction, except for the correlation between Raven's CPM and Expressive Attention: children with a higher score on Raven's CPM obtained lower scores on Expressive Attention. Correction for basic processing speed changed two significant results: Symbol Shifting and Dot Matrix did not correlate significantly anymore, while the correlation between Keep Track and the Tower of London now reached significance. The correlation between the two inhibition tasks was still marginally significant (p = .07).

An exploratory principal component analysis was applied on the tests measuring executive functions. Three factors were requested. Direct Oblimin rotation was applied, as the factors had been found to correlate in previous research (Espy et al., 2004; Miyake et al., 2000). The structure matrix is shown in Table 2.3. Although the sample size is small, the expected factor structure was found. The first factor was the updating factor, explaining 35.2% of the variance. The second factor was the inhibition factor, explaining 21.2% and the last factor was the shifting factor, accounting for 16.3% of the variance. All tasks loaded higher than .7 on their expected factor and lower than .5 on the other factors, indicating that the factor structure of inhibition, shifting and updating was present in the children in this study. The factor scores obtained in the factor analysis were used for further analyses. The updating factor correlated significantly with basic processing speed ($r = .40$, $p < .05$). Therefore all further correlations are partial, correcting for basic processing speed.

Mathematical development

As a measure of mathematical performance, number of addition problems solved correctly per minute was used, unless otherwise specified. The average scores on the addition problems are shown in Figure 2.5. The group developed steadily, with quite some individual variation. The error percentages on these problems also varied, without a clear trend, though children made more mistakes on odd session numbers than on even session numbers.

Executive functions and mathematics

In the final step, executive functions were related to the five aspects of mathematical performance as mentioned before: initial level, improvement, path, generalization and error making.

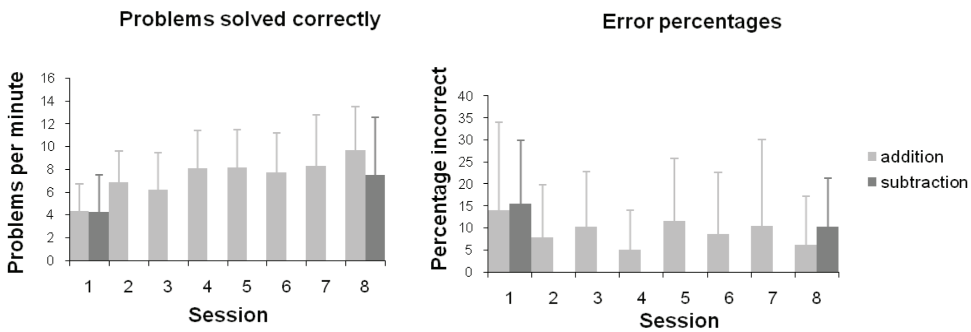


Figure 2.5 Average performance on the math problems: the number of problems solved correctly per minute and error percentages in each session.

Initial level

Because the group scored lower on odd sessions than on even sessions, scores of the first two sessions were standardized and these standardized scores were averaged. When a child missed one session, the standardized score of the other session was used. This initial level score was correlated with the factor scores of updating, inhibition and shifting, created in the factor analysis, as well as with the other measures: phonological loop, visuospatial sketchpad and intelligence. These correlations are shown in Table 2.4.

Table 2.4 Partial correlations between the executive functions and mathematics initial level and improvement, partialing out basic processing speed

	Initial level	Improvement
Inhibition	.04	-.08
Shifting	.44*	.21
Updating	.37*	.41*
Dot Matrix	.47**	.23
Digit Span Forward	.48*	.35*
Raven's CPM[a]	.37*	.36*

Note. improvement score is defined by the difference score between first two and final two measurements.
[a] Coloured Progressive Matrices
* $p < .05$, one-tailed, ** $p < .01$, one-tailed

Table 2.5 Regression analysis of executive functions and mathematical performance

		Initial level			Improvement		
		β	r^2	Δr^2	β	r^2	Δr^2
1	**EF**		.35			.23	
	Inhibition	-.14			-.15		
	Shifting	.28			.10		
	Updating	.43*			.43*		
2a	**Intelligence**		.17			.15	
	Raven's CPM[a]	.41*			.39*		
2b	**Intelligence and EF**		.44	.27		.31	.16
	Raven's CPM[a]	.32†			.30		
	Inhibition	-.06			-.07		
	Shifting	.28			.10		
	Updating	.37*			.37†		

[a] Coloured Progressive Matrices
* $p < .05$, † $p = .07$

All measures, with the exception of inhibition, correlated significantly with initial level, with correlation coefficients around .40.

Next, a regression analysis was carried out. As the group is small, there was not sufficient power to enter all variables as predictors. Therefore the first analysis was limited to executive functions. In a second, stepwise analysis the additional value of executive functions over the predictive value of intelligence was analyzed. The results of these analyses are shown in the left part of Table 2.5.

Improvement

In order to create a more reliable score, a measure of improvement was obtained by subtracting the average score of the first two sessions from the average score of the final two sessions. This difference score was correlated with the executive functions, again controlling for basic processing speed. The results can be seen in the rightmost column of Table 2.4. It was found that only updating, the phonological loop, and intelligence correlated significantly with improvement. The significant correlations were again of the same order of magnitude. Next, regression analyses were carried out again, following the same procedure as for initial level. The results of these analyses are presented on the right half of Table 2.5. The results were similar to the results related to initial score: of the three executive functions, updating was the only significant predictor of improvement. Intelligence was also a significant predictor of improvement, but the explained variance increased greatly when the three executive functions were added, again only due to a significant predictive value of updating. When all three executive functions and intelligence were included in the same analysis, intelligence was even no longer significant.

Path

First, children making a leap in their development were identified. A leap was qualified by an improvement of at least 5 problems per minute which was not followed by a regression of more than 2 problems per minute in the following sessions. However, none of the children met these criteria. Then it was investigated how well each child's improvement followed a straight line. Therefore Pearson's r^2 over 8 testing sessions was calculated for each child: children whose development is in a linear fashion show a higher r^2 than children whose development shows more fluctuations. Mean r^2 was .54, SD = .30. Path did not correlate with the measure of initial level (r = .06, p = .79) , but it did correlate significantly with final level (r = .45, p = .02) and especially with improvement (r = .74, p < .01). Thus, children with a more linear development improved more and ended on a higher level than children with a less linear path of development. The linearity measure

Table 2.6 Partial correlations between subtraction problems and executive functions, controlling for basic processing speed

	First session	Last session	Improvement
Inhibition	.13	.09	.02
Shifting	.30	.40*	.01
Updating	.57**	.69**	.38*
Dot Matrix	.46*	.67**	.08
Digit Span Forward	.23	.43*	.62**
Raven's CPM[a]	.38*	.51*	.19

[a] Coloured Progressive Matrices
* $p < .05$, one-tailed, ** $p < .01$, one-tailed

was then correlated with the EF and control measures, but no significant correlations were found.

Generalization

Scores on the subtraction problems served as an indication of generalization of obtained knowledge. An improvement score was obtained, calculating the difference between the initial and final level. The group improved on solving these problems from an average number of 4.3 problems per minute (SD = 3.3) to an average of 7.5 problems per minute (SD = 5.0). Partial correlations correcting for basic processing speed were computed for the first and second measurement separately, and for the improvement score (see Table 2.6). Results are similar to the addition problems: the individual measurements correlated significantly with updating and to a lesser degree also with shifting. The VSP and Raven's CPM also correlated significantly, while the phonological loop only did so with the final subtraction session. Considering the improvement score only updating correlated significantly, together with the phonological loop.

Error percentages

An additional analysis of the error percentages was carried out. Within each testing session there was a substantial number of children who did not make mistakes. In order to obtain a measure of error making with sufficient variance, the proportion of errors made was averaged over all sessions. The average percentage of errors made was 10.0% (SD = 10.3, range 0.6–45.9%). First, the average error percentage was correlated with initial level, final level and improvement. Correlations with initial level, $r = -.70$, $p < .01$, and final level, $r = -.56$, $p < .01$ were significant, with a higher level indicating a lower error

Table 2.7 Partial correlations between average error percentage and executive functions, correcting for basic processing speed and for initial level

	Errors (corrected for speed only)	Errors (corrected for speed & initial level)
Inhibition	-.29	-.35*
Shifting	-.28	.03
Updating	-.14	.13
Dot Matrix	-.35*	.08
Digit Span Forward	-.48**	.25
Raven's CPM[a]	-.58**	-.46*

[a] Coloured Progressive Matrices
* $p < .05$, one-tailed, ** $p < .01$, one-tailed

percentage, but the correlation with improvement was not significant, $r = -.25$, $p = .21$. Correlations with the executive tasks were also computed, both correcting for basic processing speed only and for initial level, as the latter was found to correlate highly with error percentage. Correlations were one-tailed, because high performance on executive measures was expected to correlate negatively with error making. The results are shown in Table 2.7. When controlling for basic processing speed only, none of the correlations with the executive measures were significant, but correlations with both slave systems and fluid intelligence were: high performance on the slave systems and Raven's CPM indicated a low error percentage. When initial level was also controlled for, the correlation between Raven's CPM and error making was still significant, but correlations with the two slave systems were no longer significant. Inhibition, however, now did correlate significantly: children scoring low on inhibition made more errors than children scoring high on inhibition, when initial level was controlled.

Discussion

Structure of executive functions

The factor structure of inhibition, shifting, and updating found in previous research (Espy et al., 2004; Hughes, 1998; Miyake et al., 2000) could be replicated in this study. One of the shifting tasks, Symbol Shifting, also showed a loading of .46 on the updating factor, which probably reflects the notion that children found it difficult to remember the two sorting rules. This was supported by the fact that some children reversed the sorting rules during the task. Together, these findings support the notion that even in young children these three executive functions are present, also when tasks are selected that are as dissimilar as possible in all non-executive aspects. However, as the sample size is small, these results cannot be generalized.

In this study the central executive factors were not completely independent from the slave systems of visuospatial sketchpad and phonological loop: Dot Matrix, a measure of the visuospatial sketchpad, correlated significantly with updating. This correlation has been reported before (Gathercole & Pickering, 2000a). Digit Span Forward correlated significantly with the shifting factor. Correlations with the separate tasks showed that Digit Span Forward correlated significantly only with Children's Color Trail Test. A possibility therefore is that the involvement of a number sequence in both tasks led to this correlation.

None of the three executive functions correlated significantly with fluid intelligence when basic processing speed was controlled for. This is an indication that the correlations between executive functions and addition skills are independent from and additional to correlations between intelligence and addition skills. Without controlling for basic processing speed there was a significant correlation between Raven's CPM and Keep Track, one of the updating tasks, so possibly the relationship sometimes reported between intelligence and executive functions is mediated by processing speed.

Executive functions and mathematics

As expected, many significant relations between executive functions and the mathematical tasks have been found. These are discussed according to the different aspects of mathematical development that were distinguished in this study.

Initial level

This aspect is comparable to studies that used a single measurement of mathematical skill. Significant relationships between shifting and updating and initial level were found, confirming previous research (Barouillet & Lépine, 2005; Bull & Scerif, 2001; Rasmussen & Bisanz, 2005; Van der Sluis et al., 2007). Fluid intelligence and the phonological loop were also significantly related to initial level of performance. In a regression analysis, updating proved a better predictor than intelligence. The relationship with updating can be explained from the fact that solving the problems required storage of intermediate results. The relationship with shifting may relate to the fact that the same problems were presented several times, requiring shifting between the problems and strategies. Moreover, shifting proficiency may be useful when a problem such as 7 + 7 is broken down into 7 + 3 + 4 = 10 + 4 = 14. There was no role for inhibitory skills. This might be due to the immature mathematical proficiency of the children at this age: their mathematical knowledge is limited, so there is not much interfering knowledge that needs to be suppressed.

Improvement

Only updating was significantly related to the improvement score. This lack of more significant relations might be due to the fact that only a relatively short period of time was investigated. In addition, correlations between executive functions and improvement were lower than correlations with initial level. The lack of a significant correlation with shifting might be due to the low power of the study, as the group of children is small. Another option is that shifting is of lesser importance in this type of problems, as the problems were relatively simple and did not need many shifts in order to be solved. The relation with inhibition, however, was again negligible and not likely to be affected by the low power of the study.

Path

No significant correlations with path of development were found. However, this does not necessarily imply that executive functions do not play a role in the path of mathematics learning. Although the children had not learned addition over 10 in the normal curriculum, by the time of the first measurement many children were already rather proficient at solving the mathematical problems. Therefore part of the path of development is not known and possible leaps in development remained unrevealed. Moreover, strategy use of the children is not known. Knowledge of strategy use would enrich understanding of the path of development and show differences in trajectories that remain unrevealed when only accuracy is assessed.

Generalization

Even though the children had quite some practice in tie problems, both by completing the eight booklets and having additional practice in class during the weeks of testing, their progress on the subtraction problems was virtually similar and the same executive functions were of importance. This suggests that generalization has taken place in most children in this period: either they recognized the subtraction problems as tie problems and were able to retrieve the answer, or they developed more advanced, faster mathematical procedures that they were able to apply in both addition and subtraction problems. Canobi (2004) showed that older children do not have better conceptual knowledge of part-whole relations, i.e. that knowledge of $8 + 8 = 16$ can be used in the problem $16 - 8 = 8$, but older children do have better procedural knowledge than younger children. It is therefore likely that progress in subtraction problems in this study is mostly due to faster procedural skills developed during the period of testing. Updating was the only executive function that correlated significantly with the first measure and the improvement score of the subtraction problems. Shifting also correlated significantly with the final measure. The latter finding suggests that children who are proficient at solving the subtraction problems are quicker at shifting from the subtraction problem to the corresponding addition problem. However, as the strategy that the children used is not known, this possibility awaits confirmation in future studies.

Error percentages

A correlation between error making and inhibition arose when the initial level, which correlated highly with error making, was controlled. This suggests that errors are mostly related to the mathematical proficiency of the children, but that inhibition also plays a minor role: it influences error making within a certain proficiency range. The role of inhibition may increase in later school years when problems are more difficult and more erroneous answers can be retrieved – such as when children learn multiplication and confuse it with addition. The lack of relationship with shifting and especially updating is unexpected.

General discussion

Altogether, this study revealed two main things. First, the study showed that each executive function has its own contribution to early mathematics learning. The roles of updating and shifting were most pronounced, while inhibition played only a minor role. This lack of significant findings might be attributed to the small sample size, but might also reflect that this study took great caution in the prevention of spurious results caused by other, non-executive aspects. As mentioned before, there is now a growing body of studies that doubt the roles of inhibition (Censabella & Noël, 2008; Rasmussen & Bisanz, 2005) and shifting (Espy et al., 2004; Van der Sluis et al., 2004) in math learning.

The second finding is that the fractionation of mathematics learning was fruitful, as different aspects of mathematics learning were influenced differentially by the three executive functions: (improvement in) number of items answered correctly was most strongly related to updating and, to a lesser degree, shifting, but error making was more strongly related to inhibition. This contribution to more precise knowledge of the relations between different executive functions and different aspects of mathematics learning may in the future be useful for diagnostic purposes. Assessing the cognitive profiles of individual children in an early stage of the educational process will give insight in the pattern of strengths and weaknesses in mathematics that this child is likely to experience. This clarifies specific individual educational needs that children with different executive profiles have and opens possibilities for future individual educational methods and early interventions.

Even though the sample size in this study is small, rather strong and significant relationships were found, indicating that lack of power was not a large problem in this study. This shows that the approach of this study is a fruitful one which can be elaborated upon in the future. It is therefore recommended to continue this type of research, incorporating larger samples and more detailed observations of the children's mathematical problem solving, and including more precise measures of reaction times and strategy use.

3

The structure of executive functions in children: A closer examination of inhibition, shifting, and updating

S.H.G. Van der Ven
E.H. Kroesbergen
J. Boom
P.P.M. Leseman

Manuscript submitted for publication.

Abstract

An increasing number of studies tried to investigate the latent factor structure of executive functions. Some studies found a distinction in inhibition, shifting, and updating, but others failed to find this factor structure. We hypothesized that the task choices and scoring methods might be responsible for these contradictory findings. Therefore we selected tasks in which input modality was varied, tasks controlling for baseline speed were used and scores were obtained in various ways in order to investigate whether the distinction in inhibition, shifting, and updating could be replicated. In a group of 211 children, who were tested both at the beginning of grade 1, at approximately 6 years of age, and 18 months later, it was found that the best fitting factor model did not contain the three expected executive function factors, but instead an updating factor and a combined inhibition and shifting factor, besides two baseline speed factors (verbal and motor). We argue that these results might indicate that the structural organization of executive functions might be different in children than in adults, but that there is also an alternative explanation: the aforementioned distinction in executive functions might not accurately represent cognitive structures but instead be a methodological artifact.

Introduction

In the past years, an increasing number of studies have concentrated on executive functions, or the set of cognitive processes that regulate behavior, as predictors of the development of academic and other cognitive skills. This approach has been fruitful: many studies found relationships between executive functions and scholastic achievement in the normal population, and children with learning disabilities score significantly lower on many tests measuring executive functions (e.g., Barouillet & Lépine, 2005; Blair & Razza, 2007; Bull, Espy, & Wiebe, 2008; Bull, Johnston, & Roy, 1999; Bull & Scerif, 2001; D'Amico & Guarnera, 2005; McLean & Hitch, 1999; Passolunghi & Siegel, 2001; St Clair-Thompson & Gathercole, 2006; Swanson & Beebe-Frankenberger, 2004; Swanson & Sachse-Lee, 2001; Van der Sluis, De Jong, & Van der Leij, 2007). However, the number of executive functions and the degree to which they are separable, especially in children, are still a topic of debate. Further clarification of these issues is essential for developing a better understanding of the cognitive processes underlying scholastic achievement and a precise and individual identification of the causes of learning disabilities.

Most researchers agree that executive functions are at least partially independent cognitive processes that are all involved in cognitive control. The number and nature of these different processes, however, are still largely unknown. Baddeley (1996) suggested a distinction in four different processes: (1) coordinating performance on two separate tasks, (2) selective attention, (3) switching between retrieval strategies, and (4) the capacity to hold and manipulate information in memory. The latter three are often used as a starting point in executive function research, and are referred to as inhibition, shifting, and updating, although the exact definition of these terms differs somewhat from Baddeley's distinction. Inhibition is usually defined as the ability to suppress a dominant, prepotent response deliberately in favor of another response or no response at all. A task that is often used to measure inhibition is the famous Stroop task (Stroop, 1935), in which a color word (e.g., red) is printed in another ink color (e.g., blue) and the ink color must be named. Shifting, sometimes called switching or attentional flexibility, is the ability to switch between mental sets, rules or tasks, such as alternating between sorting objects according to color and shape. Updating, finally, is the ability to code incoming information for relevance, to store this information in working memory and to update it with newer, more relevant information when necessary. While all three executive functions are supposed to be parts of the central executive, all updating seems closest related to the central executive in the working memory model by Baddeley and Hitch (1974). Tasks that have been used to measure updating and working memory are similar, though not identical to traditional working memory tasks (tasks that require continuous updating of information for updating and complex span tasks for working memory) and the abilities that are measured have been shown to load on the same factor in a factor analysis (St Clair-Thompson & Gathercole, 2006). Some studies prefer use of the term working memory rather than updating (e.g., Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003). In the present study, tasks from both research traditions were used and the term updating in this article may also be interpreted as the more general term working memory.

Whether these three different executive functions represent meaningful dissociable cognitive constructs, is still a topic of debate. Executive functions by definition regulate other cognitive processes, such as verbal or visuospatial processing of a stimulus. An executive function task score will therefore never be a pure measure of executive functioning, as low scores can also be due to these other processes. This phenomenon is referred to as the impurity problem (Miyake et al., 2000). Several studies tried to overcome the impurity problem by using factor analysis, attempting to extract only the common executive component of a variety of tasks designed to measure the same executive function. As one of the first studies using this approach, Miyake et al. (2000) found that a three-factor model of inhibition, shifting, and updating, with moderately correlating factors fitted their data best in college students. The authors concluded that inhibition, shifting and updating are distinct executive functions that partially use the same resources. They hypothesize that the cause of these correlations might be the requirement of some degree of inhibitory skills in all tasks. These results were replicated in adults, including the elderly (Fisk & Sharp, 2004). Also studies investigating children found results similar to Miyake et al. (2000): the same three-factor structure with correlating factors has been found in 2-5-year-olds (Espy et al., 2004) and in 8-13-year-olds (Lehto et al., 2003), and, with some reservations, even in 3- and 4-year-olds (Hughes, 1998). Another study, which did not include shifting measures, found an inhibition and a working memory factor in addition to a speed/arousal factor (Brocki & Bohlin, 2004). Nevertheless, the distinction of inhibition, shifting, and updating is not uncontroversial: there are also studies that failed to replicate this factor structure. Wiebe, Espy and Charak (2008), who investigated working memory and inhibition only, found that in children aged 2-6 years a model with only one factor fitted the data just as well. St Clair-Thompson and Gathercole (2006) found an inhibition and an updating factor but no shifting factor in 11-year-olds. Two other studies, one with children between 7 and 15 and young adults, the other with children aged 9-11, did not identify a separate inhibition factor (Huizinga, Dolan, & Van der Molen, 2006; Van der Sluis et al., 2007). Finally, also in adults one study showed that updating and inhibition cannot be distinguished as separate factors (Salthouse, Atkinson, & Berish, 2003).

There are different possible explanations for these conflicting results. The first is related to the developmental aspect: the structure of executive functions might vary in the course of child development. This, however, does not seem a satisfactory explanation, since the studies concerning children mentioned before did not obtain converging results. A second explanation, also acknowledged by Miyake et al. (2000), relates to the choice of tests supposed to measure executive functions. While the use of factor analysis is a good way to circumvent the impurity problem, a drawback of the method is that it is never certain what the common latent factor entails exactly. Thus far, consensus has not been reached with regard to which tasks are appropriate: each task is rarely used in more than one of the studies discussed above. Task selection is, however, of the greatest importance: when tasks are selected that are similar in any non-executive aspect, the resulting latent factors will still be impure. A closer examination of the tasks used in the studies that confirmed the division of inhibition, shifting, and updating, yields some sources of doubt. In the study by Espy et al. (2004), for instance, the shifting factor comprised two tasks that were highly similar in aspects not related to shifting: in both tasks two wells were presented to a child and a reward was located in one of them in an alternating fashion. In the study by Hughes (1998) both inhibitory tasks involved making hand movements. And two of the inhibition tasks and all three shifting tasks in the study by Van der Sluis et al. (2007) consisted of pairs of stimuli of which one should be named. All these similarities between tasks potentially threaten the interpretation of study outcomes, since the latent factors that are supposed to be pure measures of executive functions might inadvertently be contaminated with nonexecutive measures to a certain degree.

The scoring systems of executive tasks might also influence the final factor structure. Scoring methods also differ greatly between studies and a rationale for the scoring method used is hardly ever provided. Whereas updating outcome measures are generally straightforward (number of items answered correctly), for inhibition and shifting measures this is not the case: here efficient executive functions are reflected by both fast and accurate performance. Nevertheless, most scoring methods take only one of these two aspects into account. Outcome measures sometimes involve the number of mistakes (Espy et al., 2004; Hughes, 1998; Lehto et al., 2003; Miyake et al., 2000; St Clair-Thompson & Gathercole, 2006), but for other tasks measures are created by using the time needed to complete the task, often controlling for baseline speed, e.g., by subtracting or dividing by the time needed to complete a control task (Lehto et al., 2003; Miyake et al., 2000; St Clair-Thompson & Gathercole, 2006). However, the use of difference scores compromises task reliability, as it may increase measurement error. A better solution is to include a latent speed factor in the model, as has been done in some studies (Huizinga et al., 2006; Van der Sluis et al., 2007), but in these studies accuracy was not included in the model. Others used a combination of speed and accuracy, such as the number of obtained sets divided by errors (e.g., Espy et al., 2004), but the relationship between speed and accuracy is still unknown.

The selection of proper executive tasks that show as little overlap as possible, as well as the method of scoring may thus far not have received the attention they deserve. Therefore the present study was set up to address these issues. It was investigated whether the factor structure of inhibition, shifting and updating could be found when the tasks were selected according to a number of criteria described below.

While in many studies only speed or only accuracy was used in the analyses, good performance on inhibition and shifting tasks is reflected by both fast and accurate responses. Therefore the inhibition and shifting factors received loadings from both speed and accuracy measures. However, speed measures are also influenced by baseline speed abilities of the participant, notably verbal speed (i.e. word formation and preparing the articulation of the response) and motor speed. This issue is sometimes addressed by using the difference between these two conditions, but this method increases error variance (Linn & Slinde, 1977) and fails to take into account the overadditivity effect: participants that are slow on a control task, slow down disproportionally on any more difficult task (Faust, Balota, Spieler, & Ferraro, 1999). Therefore a different solution was chosen: baseline speed factors (a verbal and a motor factor) were constructed. These factors received loadings from all speeded tasks: control tasks and executive tasks. This procedure ensured that speeded loadings on the executive tasks were corrected for baseline speed.

Since a longitudinal design was used, it was also investigated whether the factor structure changed during the first two years of primary education (grade 1 and 2). The study started with children that had just started formal education, around six years of age. This age group is especially interesting because of the developmental perspective: since executive functions are increasingly considered as predictors for learning disabilities, it is important to unravel the structure of executive functions when formal education starts and learning difficulties starts to surface. In addition, at the start of the study the children's executive functions could not have been influenced to a large degree by formal education yet. In order to test the robustness of the results, the study was repeated after 18 months with the same children and the same tasks.

The tasks were selected according to a number of criteria. First of all, preferably they had been used successfully in previous research. Second, in order to prevent the creation of impure latent factors, they were as different as possible from each other in non-executive aspects: the modality of input (auditory or visual) and the way of responding (verbal or motor) varied. Third, whenever possible a control task was administered that was similar to the executive task in all non-executive aspects. Because of the nature of the tasks this was possible for inhibition and shifting tasks, but not for updating tasks. Fourth, as the children had only just started reading education, none of the tasks involved reading. And finally, the tasks should be possible to administer on a computer, allowing measurement of response times for each trial separately. With these carefully selected tasks, it was investigated whether the factor structure of inhibition, shifting, and updating, as found in previous research, could also be found in our study.

Method

Participants

At the beginning of the study, 227 children (120 boys, 107 girls) with a mean age of 6;5 years (*SD* = 4.3 months; range 5;9 – 7;7 years) took part. Children came from 18 classes in 10 schools. Parental consent was obtained from all participating children. School choice was based on two criteria: a low number of children not speaking Dutch at home and diversity in SES (i.e. schools with high and low numbers of parents that had completed less than 2 years of secondary education). As the aim was to obtain a representative sample of children following regular education, there were no stringent exclusion criteria; however, three children were excluded because of failure to understand the task instructions (one child with Down Syndrome and two refugee children with insufficient mastery of the Dutch language).

During the course of the study, thirteen more children (5.8%) dropped out, due to moving (seven children), grade retention (three children), and accelerating a grade (three children). All analyses were performed with the 211 remaining children (110 boys, 101 girls, mean age = 6;5 years, *SD* = 4.4 months). In the final wave, 209 children completed the Raven's Standard Progressive Matrices (Raven, 1958). Scores were slightly below average: 66 children (31.6%) obtained scores in the lowest quartile, 69 children (33.0%) in the second quartile, 60 children (28.7%) in the third and 14 children (6.7%) in the highest quartile.

Procedure

The children were tested individually in a separate room by trained research assistants, who administered the tasks on a laptop. The tasks were divided into three sessions, lasting 35, 25 and 15 minutes respectively. Approximately 75% of the children completed session 3 in pairs, each having their own laptop. The task order was fixed for all children:

Session 1: Animal Stroop, Keep Track, Trail Making Test in Colors, Digit Span Backwards.
Session 2: Animal shifting, Odd One Out, Local Global
Session 3: Simon Task, Sorting Task

Instruments

Based on the criteria mentioned in the introduction, three tasks per executive function were selected, that can be found in Table 3.1. All tasks were administered on a laptop, and except for the Trail Making Test in Colors, all were programmed in E-Prime (Psychological Software Tools, http://www.pstnet.com). Computerized assessment allowed for detailed data: accuracy and response latency were obtained for each stimulus separately. Several tasks required a quick verbal response of the child. Because the use of voice keys is unreliable in the noisy environment of a school, the experimenter indicated the start of the response by pressing the spacebar. After this, the experimenter indicated the accuracy of the response. Reaction time and accuracy were always based on the first response, even when self-corrections were made and also when the first response wasn't vocalized completely (e.g., c…duck).). In order to ensure identical testing conditions for each child, stimuli were always presented in the same, balanced order. Example items of each executive function test are

Table 3.1 Characteristics of selected tasks

Task	Input	Response	Control task
Inhibition			
Animal Stroop	visual	verbal	yes
Local Global	visuospatial	verbal	yes
Simon Task	visuospatial	motor	yes
Shifting			
Animal Shifting	visual	verbal	yes
Trail Making Test in Colors	visuospatial	motor	yes
Sorting Task	visual	motor	yes
Updating			
Digit Span Backwards	verbal	verbal	no
Odd One Out	visuospatial	motor	no
Keep Track	visual	verbal	no

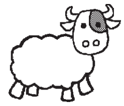
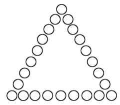
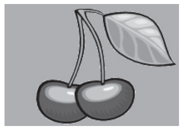
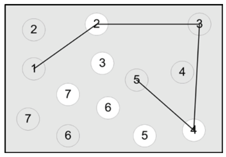
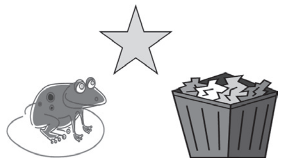
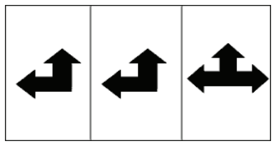
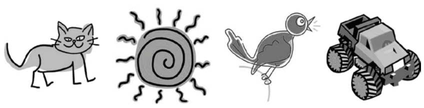
Task name	**Instruction**	**Example item**
Animal Stroop	"Name the animal body"	
Local Global	"Name the smaller shape"	
Simon Task	"If you see a mouse, press the left button, if you see a dragon, press the right button"	 (presented on right or left of computer screen)
Animal Shifting	"If the screen is yellow, name the fruit; if the screen is purple, name the animal"	  (varying background color)
Trail Making in Colours	"Draw a line from 1 to 2 to 3 …, while alternating the colours."	
Sorting Task	"If you see a dog, give it the stimulus if it's blue; throw it away if it's orange. If you see a frog, give it the stimulus if it's a star; throw it away if it's a square."	
Digit Span Backwards	"Repeat the sequence that you hear backwards"	"3 5 1 6"
Odd One Out	"Point at the shape that is different and remember its location."	
Keep Track	"Name each picture you see and remember the last animal."	 (sequence of 10 successive stimuli)

Figure 3.1 Example items of the executive function tasks.

presented in Figure 3.1. For the inhibition and shifting tasks, speeded measures of control condition and executive condition were obtained, and accuracy of the executive condition. Because of ceiling effects, accuracy of the control condition was not usable.

Inhibition

Animal Stroop. An adaptation of the Animal Stroop task, created by Wright, Waterman, Prescott, and Murdoch-Eaton (2003) was used. In this task animal drawings are presented that are composed of the body of one animal and the head of another. The participant has to name the animal body rather than the more salient animal head. The task consisted of two blocks. In each block animal stimuli were presented one at a time, preceded by a 400 ms fixation cross. The stimuli remained on the screen until the child responded. The child named each animal as quickly as possible. The first block consisted of 4 normal animal stimuli with congruent heads (cow, sheep, duck or pig), each presented 12 times, yielding a total of 48 trials. The second block was a mixed block of control stimuli with human heads and incongruent stimuli with incorrect animal heads. Twelve different stimuli per condition were each presented four times, yielding 48 control and 48 incongruent trials. Reaction time of the trials with normal animals (control measure) and reaction time and accuracy of the trials with incorrect animal heads (inhibition measure) were used.

Local Global.[1] In this task identical small geometrical shapes (circle, triangle, square) that together constituted a larger, different geometrical shape were presented. Sometimes the large, global shape had to be named, sometimes the small, local shapes. In general, people show a global preference: responses are faster when the large image has to be named rather than the local shapes (Navon, 1977). The task consisted of three blocks. In the first block 48 single, small geometrical shapes were presented as control stimuli. The second and third block each consisted of 48 larger stimuli that were constructed from the stimuli presented in the first block. All stimuli in these blocks were incongruent: the shape of the larger image was always different from the elements from which it was built. In the second, global block, the large stimulus had to be named, thus strengthening the global preference, while in the third, local block the small image had to be named. Trials were preceded by a 400 ms fixation cross. Reaction time of the first block (control measure) and reaction time and accuracy of the third block (inhibition measure) were used.

Simon Task. The Simon effect (Simon & Berbaum, 1990; Simon & Rudell, 1967) can be elicited with tasks where stimuli are presented at different horizontal locations on a

1 In some studies, e.g. by Miyake et al. (2001), this task was used as a shifting task instead of an inhibition task. However, the task requirements were different: when used as a shifting task, participants had to alternate between naming the large and the small shape within one block. This flexible alternation is the key element of shifting tasks and this element was not present in the current study. Key to an inhibition task is the presence of a dominant tendency to be suppressed; a dominant tendency to name the large shape has been demonstrated convincingly by Navon (1977).

screen, while this spatial aspect must be ignored. A response (a left or right button press) is based on some non-spatial aspect of the stimulus, such as the identity. Nevertheless, the spatial aspect is difficult to ignore: when a correct response entails pressing the button on the side where the stimulus is presented, this response is generally faster and more accurate than a response on the other side. The task was presented as a zoo game with animals (mice and dragons) that had escaped. The child was instructed to capture the escaped animals as quickly and accurately as possible. On a computer screen, a picture of a mouse or a dragon appeared. When a mouse appeared, it had to be captured by a left key press (A); a dragon was captured by pressing a right key (L). The task started with a block with 40 control items, in which the stimuli appeared in the centre of the screen. After the control block, a Simon block appeared, where the animals were presented on the left or right side of the screen. There were 40 congruent (same-side button press) and 40 incongruent (other-side button press) trials. Both types of trials were intermixed. Stimuli were present until the child pressed a button. When the button press was correct, feedback was provided: a cage appeared around the animal. Nothing happened upon a false button press. Trials were preceded by a 500 ms fixation cross. The stimuli were randomized once and then presented in the same, fixed order to each child. Reaction time of the first block with stimuli in the middle (control measure) and reaction time and accuracy of the incongruent items of the second block (inhibition measure) were used.

Shifting

Animal Shifting. In this task, two stimuli were presented simultaneously on the computer screen (an animal and a piece of fruit), and the participant had to switch between naming the animal and the fruit. The task consisted of two blocks. In the control block, 40 stimuli were presented one at a time, each belonging to one of two categories: fruit (strawberry, pear, cherry, banana) or animal (cat, dog, bird, fish). The child named each stimulus as quickly as possible. In the shifting block, two stimuli were presented together: one from the fruit category and one animal. The child had to name only one of them, depending on the color of the screen background: the fruit when it was yellow, the animal when it was purple. Again, 40 stimuli were presented. All stimuli were preceded by a 700 ms fixation cross. The stimuli were randomized once and presented in a fixed order. Reaction time of the first block (control measure) and reaction time and accuracy of the second block (shifting measure) were used.[2]

2 It is also possible to contrast switch trials (in which the rule is different from the previous trial) with nonswitch trials (in which the rule stays the same). However, this measure has been proven a less sensitive measure of executive functioning (Dibbets & Jolles, 2006; Kray & Lindenberger, 2000) and was therefore not used in this study.

Trail Making Test in Colours. This test is based on the Children Colored Trail Test (Llorente, Williams, Satz, & D'Elia, 2003), in which coloured circles with a number in the middle must be connected in the correct order. The task consisted of a control task and a shifting task. In the control task, 20 circles numbered 1-20 were presented on the screen. Half of the circles were orange and the other half were blue, but this information was not relevant. The child had to connect the circles 1-10 by clicking them with the mouse. When a mistake was made, the mouse pointer returned to the previous circle. The tracks followed a relatively easy path over the screen, thus reducing the visual search component of the task. In the first shifting task, the configuration of the circles was the same as in the first control task, but mirrored vertically. Now the 20 circles were numbered 1-10, each number being present in both blue and orange. The child had to connect the circles while alternating the colors. Before each task, a small practice item was administered. This version of the task differed in two main aspects from the task it was based on. In the original task, in the control condition all numbers are presented only once. This decreases the number of circles by a half and therefore also decreases the visual search load in the control condition. In addition, in the original task the order of the circles is different in each condition, which also changes the visual search load. This was also controlled for in our task, by mirroring the image. Reaction time of the first (control measure) and reaction time and accuracy of the second (shifting measure) task were used.

Sorting Task. In this task, inspired by Zelazo and colleagues (Zelazo, Müller, Frye, & Marcovitch, 2003) the child had to alternate between two sorting rules: according to colour and according to shape. The task was presented as a game in which a dog that likes blue and a frog that likes stars were introduced to the child. The child had to give the animal the stimuli that it liked and throw away the stimuli the animal did not like. The task contained two control blocks and a shifting block. In each block, the stimuli were 40 orange and blue stars and squares. In the first control task, the child was introduced to the dog. It was told that this dog loved blue but hated orange. The dog was shown on the lower left side of the screen, and a waste bin was shown on the lower right side of the screen. After 700 ms a stimulus appeared. When the stimulus was blue, the child 'gave' it to the dog by pressing the A button on the left side; when it was orange, the child 'threw it away' by pressing the L on the right side. No feedback was provided. The second control task was similar to the first control task, but now the dog was replaced by a frog, and the child was told that the frog loved stars and hated squares. The same stimuli were presented again, and this time the child sorted the items according to shape. The shifting task was a mixed block in which sometimes a dog and sometimes a frog appeared; the same 40 items were shown again. The stimuli were randomized once and presented in a fixed order. Reaction time on the two control conditions and reaction time and accuracy on the shifting condition were used.

Updating

Digit Span Backwards. An adaptation of the Dutch version of Digit Span Backwards from the Automated Working Memory Assessment test battery (Alloway, 2007) was administered. The child heard a digit sequence and had to repeat it backwards, starting with a two-digit sequence. After three correct answers, the sequence increased by one. When two mistakes were made in trials of the same length, the task was discontinued. The number of correct answers was used as a final score.

Odd One Out. This task was also adapted from the Automated Working Memory Assessment (Alloway, 2007). Three boxes with shapes were presented next to each other. One of the shapes was different from the other two. The child pointed at the different shape. Then three new shapes appeared. At the end of each trial three empty boxes appeared and the child had to point at the locations of the previously-shown different shapes in the same order in which they appeared. An answer was considered correct if each location was recalled correctly in the right order. The task started with only one item; after three correct answers of the same length the sequence increased by one. When two mistakes were made in trials of the same length, the task was discontinued. The number of correct responses was used as a final score.

Keep Track. A computerized version of the Keep Track task was created. This task was adapted from the task used by Van der Sluis et al. (2007). The child was shown pictures, each of which belonged to one of the following five categories: the sky (sun, moon, stars, cloud), fruit (strawberry, pear, cherry, banana), shapes (square, triangle, circle, heart), animals (dog, cat, fish, bird) and toys (teddy bear, scooter, LEGO®, car). The pictures were shown in series of ten, and each picture was displayed for 3.5 s. During the series, the child's task was to name each picture. At the end of the series the child had to recall the last item of certain categories. In each series one, two, three items of each designated category were presented. The remainder consisted of items from other categories. During the series, small white pictures symbolizing the to-be-remembered categories were shown in the bottom of the screen, serving as a reminder. The number of to-be-remembered categories increased from one to four. There were two series in each difficulty level, yielding a total of eight series. Each correct answer was noted. There could thus be one, two, or three correct answers in each trial, yielding a maximum score of twenty correct responses. The total number of correct responses was recorded. Prior to testing, the child was familiarized with the pictures and the categories. The test was preceded by a practice item that was repeated when necessary.

Analysis decisions

Many tasks involved response times. Generally, response times are not normally distributed, but have a rather large right tail, or an ex-Gaussian distribution, where the cases in the right tail are most likely due to lapses in attention (Whelan, 2008). Our data are no exception. This reduces the power of many analyses. Possible solutions to this problem involve the use of cut-offs or data transformation (Whelan, 2008). In order to achieve a better normal distribution and maintain higher power, in this study first cut-offs were made, after which the data were inverse-transformed. In other words, response time was transformed into speed: the number of responses per second. The cut-offs were based on visual inspection of each task separately and were chosen rather lenient (i.e., only really extreme times were removed before transformation). Extremely fast answers were also removed, as these are likely to reflect fast guessing (Whelan, 2008). For inhibition, the cut-offs were 400 and 5000 ms for Animal Stroop, 400 and 6000 ms for Local Global, and 200 and 3500 ms for the Simon Task. For shifting tasks, cut-off values were 500 and 8000 ms for Animal Shifting, and 400 and 6000 ms for the Sorting Task. Response times for false responses and response times directly after a false response were also removed, following Huizinga et al. (2006).

Of all reaction time trials, 11% were removed by following these procedures. Note that for each child an average speed score could be calculated without these removed trials, so the procedure did not lead to any final missing data. However, 0.24% of the data was missing, and, in addition, when a child's accuracy for a certain condition was not significantly higher than what could be expected based on chance (assuming a binomial distribution), the speed measure for this condition was deemed unreliable and was removed for this child. This led to removal of 0.99% of the data. Finally, for all variables, including non-speeded tasks, values that deviated more than 3 *SD*s from the mean were replaced by values deviating exactly 3 *SD*s from the mean. This happened to 0.84% of the cases.

Analyses were performed for both waves separately. Within each wave, first the factor structure of the control tasks was investigated. It was tested whether one factor labelled 'latent baseline speed' fitted the control data. In the next step, the accuracy and speed measures of the executive conditions of the tasks were included and it was tested whether the best model fit was obtained when the full three-factor structure of executive functions was added, or whether a better fit would be provided by fewer than three factors (by combining two or even three factors into one). In a final step, the best fitting models from both waves were compared in a multigroup comparison.

Confirmatory factor analyses were carried out by using structural equation modeling. As is common in structural equation modeling, overall model fit is indicated by several fit indices, which each evaluate different aspects of the model. Here, the fit indices recommended by Blunch (2008) are reported: chi square (χ^2) with its *p*-value, CFI, RMSEA with its P_{close}-value and AIC. Chi square is a discrepancy measure; between the current model and the saturated model, and should therefore be as low as possible, with a *p*-value that is as high as possible. CFI compares the fit of the model to the independence model and is good if CFI > .95 and acceptable if CFI >.90. RMSEA is a parsimony measure, favoring simpler models, and is good if < .05 and acceptable if < .08. Its P_{close}-value should be high. AIC, finally, is not an absolute fit index, but is used to compare models: the lowest AIC value is preferred.

Results

For each task, descriptive measures and test-retest correlations after transformation and removal of extreme values are presented in Table 3.2. Skewness of all measures was below 1.5 and kurtosis was 2.0 or lower, indicating a good normal distribution. For each variable a high score also means high performance, since the speeded measures represent the number of correct answers given per second (rather than reaction time in milliseconds). Correlations between the different measures within each wave are given in Table 3.3.

First, it was tested whether the experimental manipulations had worked, i.e., whether response speed was slower in the executive condition than in the control condition of each task. Repeated measures ANOVAs showed that all differences between control and executive condition of the same task were significant, with one unexpected effect: for the TMT-C the effect was in the opposite direction in wave 2, where on average the shifting task was performed faster than the control task (see Table 3.4). Given this finding, together with the finding that although in wave 1 the mean effect was in the expected direction, the effect size is small compared to the other tasks and 96 of the 211 children also showed the opposite pattern in wave 1, it was decided to remove this task from all further analyses.

Next, confirmatory factor analyses were carried out for both waves separately. Since results were highly similar for both waves, only the results of wave 1 are displayed graphically in Figure 3.2. When results did differ between the waves, relevant coefficients are mentioned in the text. Fit indices are given for each wave separately in Table 3.5.

First, baseline speed was modelled, by creating a latent speed factor that received loadings from all speeded tasks: the five control tasks and the five inhibition and shifting tasks. The errors of the control and executive condition from each task were allowed to covary, in order to account for task-specific effects. This model did not provide a good fit to the data (see Table 3.5). When response mode was taken into account, however, by creating a motor speed factor and a verbal speed factor, the fit was reasonably good, as is shown in Table 3.5. This model is a part of the final model shown in Figure 3.2: it included the motor speed and verbal speed factors with their indicators.

This baseline speed model was expanded with the accuracy measures of the executive conditions. Following the procedure by Miyake et al. (2000), it was investigated whether the executive tasks were best represented by one, two or three additional latent factors, representing the three proposed executive functions. The model with three executive factors included factors for inhibition, shifting, and updating respectively; in

Table 3.2 Descriptive statistics and test-retest correlation coefficients for all variables, after removal of extreme values and after transformations. Accuracy measures reflect the proportion of correct answers; speed measures are indicated by the number of responses per second.

Task	Wave 1			Wave 2			
	N	*M*	*SD*	*N*	*M*	*SD*	*r*
Inhibition							
Animal Stroop							
Accuracy incongruent	210	.93	0.05	210	.95	.04	.28**
Speed congruent	211	0.90	0.13	210	0.94	0.14	.33**
Speed incongruent	210	0.65	0.10	210	0.79	0.12	.30**
Local global							
Accuracy local	210	.93	0.06	210	.95	.05	.17*
Speed control	211	0.81	0.13	210	0.93	0.21	.26**
Speed local	210	0.61	0.10	210	0.72	0.18	.30**
Simon Task							
Accuracy incongruent	208	.91	0.08	208	.93	.06	.15*
Speed control	211	1.46	0.18	210	1.66	0.21	.54**
Speed incongruent	208	1.05	0.14	207	1.26	0.18	.54**
Shifting							
Animal Shifting							
Accuracy shifting	201	.92	0.07	208	.95	.05	.32**
Speed control	211	0.86	0.13	210	0.88	0.19	.37**
Speed shifting	201	0.48	0.08	208	0.55	0.10	.35**
Trail Making Test in Colours							
Accuracy shifting	211	.85	0.18	211	.98	.05	.01
Speed control	211	0.19	0.06	211	0.33	0.09	.43**
Speed shifting	211	0.16	0.06	211	0.35	0.09	.48**
Sorting Task							
Accuracy shifting	208	.89	0.08	208	.92	.06	.24**
Speed control	211	1.13	0.17	209	1.33	0.21	.44**
Speed shifting	208	0.70	0.17	208	0.92	0.22	.39**
Updating							
Digit Span Backwards	211	3.75	1.67	211	5.71	1.81	.37**
Odd One Out	211	6.76	2.40	210	9.39	2.68	.25**
Keep Track	211	11.73	3.02	186	14.91	2.67	.36**

* $p < .05$, ** $p < .01$.

Table 3.3 Correlation matrix

	Task		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	AS	ACC inh	-	-.09	-.06	.26**	.01	.04	.28**	-.05	.03	.16*	-.01	-.03	-.05	.07	-.04	.17*	.01	-.02	.00	.04	.04
2	AS	SP control	.00	-	.75**	-.05	.34**	.31**	.07	.38**	.36**	-.06	.59**	.45**	.07	.17*	.09*	-.01	.31**	.19**	.08	.13	.30**
3	AS	SP inh	.02	.72**	-	-.02	.47**	.44**	.06	.41**	.46**	-.12	.64**	.54**	.08	.20**	.23**	-.03	.42**	.28**	.07	.12	.36**
4	LG	ACC inh	.27**	.02	.11	-	.00	.06	.22**	-.11	.02	.32**	-.01	-.07	.03	.17*	.13	.23**	.03	-.01	.14*	.18*	.08
5	LG	SP control	-.07	.56**	.55**	-.03	-	.74**	.06	.32**	.36**	-.09	.65**	.58**	.09	.17	.10	-.06	.37**	.21**	.10	.09	.23**
6	LG	SP inh	-.08	.48**	.51**	-.03	.67**	-	.03	.29**	.38**	-.10	.57**	.57**	-.01	.15*	.11	-.12	.46**	.35**	.11	.13	.23**
7	SI	ACC inh	.26**	-.01	-.04	.37**	-.03	-.09	-	-.03	-.05	.21**	.08	.02	.07	.10	.08	.35**	-.05	-.12	.02	.17*	.11
8	SI	SP control	-.02	.28**	.41**	.08	.39**	.42**	-.15	-	.73**	.05	.40**	.45**	.06	.30**	.29**	-.01	.66**	.54**	.08	.07	.37**
9	SI	SP inh	.09	.29**	.40**	.07	.34**	.37**	-.13	.70**	-	.05	.43**	.48**	.02	.32**	.27**	-.09	.73**	.58**	.17*	.16*	.41**
10	AN	ACC shifting	.33**	-.03	.00	.36**	-.08	-.13	.27**	.08	.13	-	-.08	-.07	-.02	.05	.06	.34**	.03	-.01	.09	.11	.10
11	AN	SP control	-.01	.65**	.66**	.01	.65**	.53**	-.01	.27**	.33**	-.12	-	.67**	.12	.14	.17*	.01	.38**	.22**	.15*	.12	.32**
12	AN	SP shifting	.02	.48**	.56**	.00	.58**	.58**	-.01	.43**	.44**	.01	.61**	-	.07	.28**	.23**	.00	.41**	.34**	.16*	.20**	.40**
13	TM	ACC shifting	.22**	.16*	.15*	.23**	.02	.01	.13	.11	.16*	.19**	.11	.07	-	.05	.35**	.11	.04	-.08	.01	.10	.09
14	TM	SP control	.16*	.22**	.32**	.26**	.15*	.14*	.13	.28**	.37**	.15*	.13	.17*	.28**	-	.47**	.02	.25**	.23**	.13	.25**	.24**
15	TM	SP shifting	.18**	.23**	.29**	.21**	.19**	.22**	.03	.30**	.42**	.19**	.19**	.29**	.52**	.48**	-	.04	.26**	.18**	.14*	.19**	.32**
16	ST	ACC shifting	.19**	-.01	-.06	.35**	-.05	-.10	.24**	.00	-.03	.38**	-.06	.01	.21**	.08	.16*	-	-.14*	-.22**	.06	.13	.12
17	ST	SP control	.03	.26**	.35**	-.03	.36**	.42**	-.22**	.64**	.70**	.05	.31**	.44**	.08	.30**	.32**	-.18*	-	.74**	.15*	.16*	.35**
18	ST	SP shifting	-.08	.14*	.28**	-.18*	.31**	.39**	-.28**	.48**	.52*	-.05	.25**	.42**	-.01	.09	.20**	-.28**	.67**	-	.16*	.12	.31**
19	DB	Total score	.12	.28**	.20**	.20**	.14*	.06	.11	.09	.16*	.16*	.20**	.23**	.25**	.14*	.28**	.06	.17*	.09	-	.24**	.15*
20	OO	Total score	.05	.19**	.15*	.13	.04	.07	.05	.17*	.22**	.16*	.06	.10	.24**	.16*	.27**	.04	.16*	.08	.32**	-	.34**
21	KT	Total score	.21**	.11	.19**	.23**	.06	.10	.10	.10	.19**	.16*	.10	.19**	.17*	.11	.33**	.46	.20**	.10	.26**	.24**	-

Note. AS = Animal Stroop, LG = Local Global, SI = Simon Task, AN = Animal Shifting, TM = Trail Making Test in Colours, ST = Sorting Task, DB = Digit Span Backwards, OO = Odd One Out, KT = Keep Track, ACC = accuracy, SP = speed, inh = inhibition, * $p < .05$, ** $p < .01$. Below diagonal: Wave 1. Above diagonal: Wave 2.

Table 3.4 Repeated Measures ANOVAs for all speeded tasks

Task	Wave 1				Wave 2			
	F	*df*	*p*	η^2	*F*	*df*	*p*	η^2
Inhibition								
Animal Stroop	1540.72	1, 209	<.001	.88	552.44	1, 209	<.001	.73
Local global	842.103	1, 209	<.001	.80	455.82	1, 209	<.001	.69
Simon Task	1901.20	1, 207	<.001	.90	1475.95	1, 206	<.001	.88
Shifting								
Animal Shifting	2536.76	1, 200	<.001	.92	1141.08	1, 207	<.001	.85
TMT-C[a]	64.92	1, 210	<.001	.24	6.707	1, 210	<.001	.03
Sorting Task	2006.91	1, 207	<.001	.91	1462.74	1, 207	<.001	.88

[a] Trail Making Test in Colours

the three two-factor models two of these factors were merged into one factor and in the one-factor model all executive tasks loaded on the same latent factor. The inhibition and shifting factors received loadings from both accuracy and speed measures (see lower part of Figure 3.2). Covariances between all latent factors (both speed and executive) were allowed but were removed when not significant to obtain optimal fit of each model. Table 3.5 shows that in both waves the best fit was provided by two models: the three-factor model and the two-factor model with inhibition and shifting combined into one factor. The more parsimonious two-factor model is therefore preferred, as is also reflected in the lower AIC value of this model. Moreover, in the model with three executive function factors the estimated correlation between inhibition and shifting is very high, wave 1: $r = .93$, $p < .001$; wave 2: $r = .85$, $p < .001$, which makes it questionable whether these two factors can be considered to be independent.

The fit of both models was best when the updating factor was allowed to covary with both speed factors, while there was no covariation between inhibition/shifting and speed. The covariation between updating and speed probably reflects that fast baseline speed enables an efficient processing of information, which is likely to aid performance on memory tasks, especially the Keep Track task, as this was a speeded task. The lack of relation with inhibition/shifting was also expected, as the baseline speed requirements of these tasks were already accounted for by the two latent speed factors.

A multigroup comparison[3] of the models in both waves was performed to see if the factor structure was not significantly different in both waves. Due to identification

3 Conventionally, multigroup analyses are performed with multiple independent groups. Here, the groups consisted of the same members; the difference is the 18-month time interval. While for many analyses this requires a correction for autocorrelation, here the scores on the two occasions are never directly compared to each other. Rather, the two models are analyzed independently; the only comparison that is made is the creation of shared restrictions on the two models to test the validity of the assumption that parameter values have not changed.

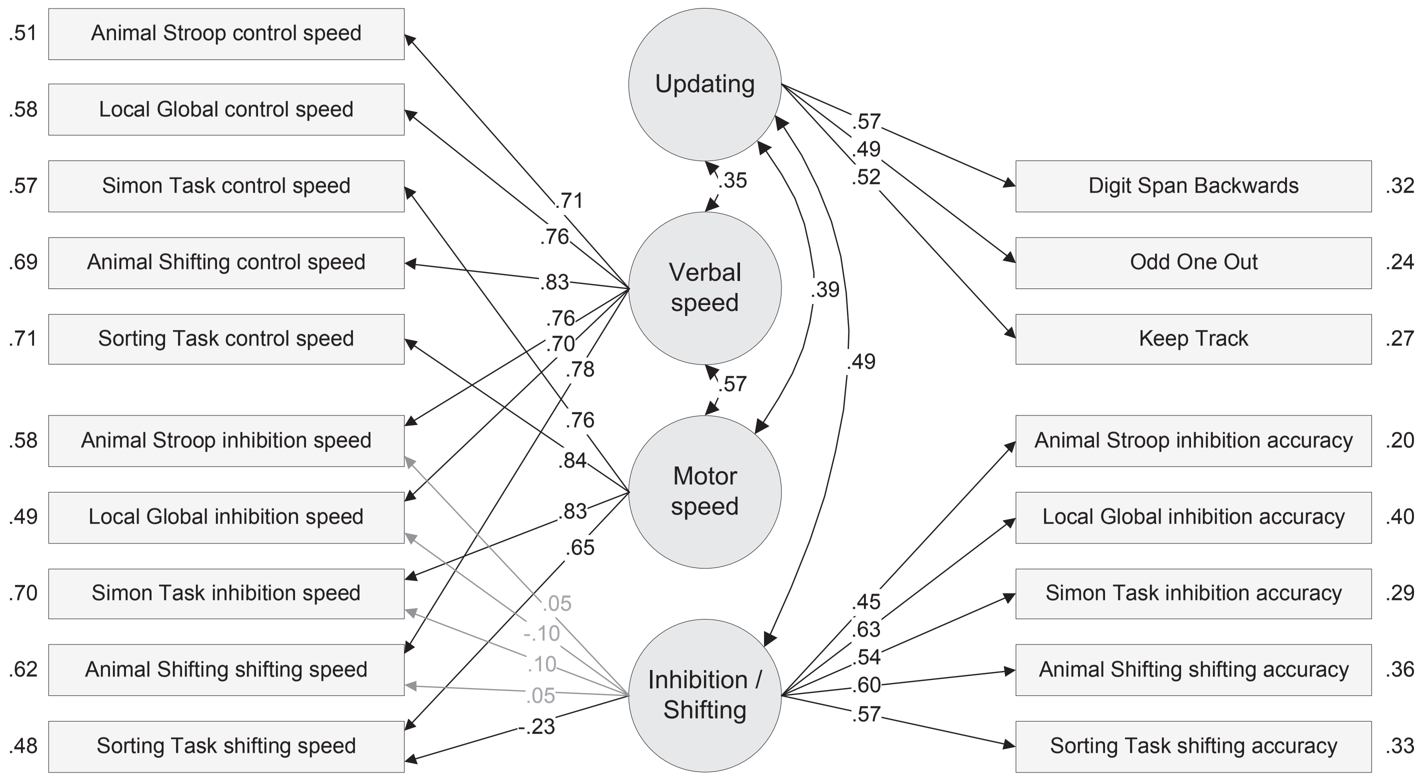


Figure 3.2 Best fitting latent factor model in wave 1. Proportion explained variance is displayed next to each observed variable. Black arrows represent significant coefficients ($p < .01$ in all cases); gray arrows represent coefficients that were not significant ($p > .05$). The five correlations between the error terms of the control tasks and their corresponding executive tasks, e.g., between Animal Stroop control speed and Animal Stroop executive speed, were also included but not presented here to enhance visibility. From top to bottom, their sizes were .38, .29, .17, -.11, and .29 respectively.

Table 3.5 Model fit indices of confirmatory factor analyses with best fitting models in italics

	χ^2	df	*p*	CFI	RMSEA	P_{close}	AIC
			Wave 1				
Baseline speed							
1 latent factor	168.28	30	<.001	.88	.15	<.001	238.28
2 latent factors	61.01	29	<.001	.97	.07	.07	133.01
3 EF	168.03	118	<.001	.97	.05	.70	310.03
2 EF							
I/S and U	173.43	121	.001	.96	.05	.68	309.43
I/U and S	211.84	122	<.001	.94	.06	.13	345.83
S/U and I	211.81	122	<.001	.94	.06	.13	345.80
1 EF	219.09	124	<.001	.93	.06	.10	349.09
Scoring based	245.04	127	<.001	.92	.07	.02	369.04
			Wave 2				
Baseline speed							
1 latent factor	175.14	30	<.001	.89	.15	<.001	245.14
2 latent factors	64.87	29	<.001	.97	.08	.04	136.87
3 EF	164.07	119	<.01	.97	.04	.78	304.07
2 EF							
I/S and U	165.57	121	<.01	.97	.04	.80	301.57
I/U and S	202.843	121	<.001	.94	.06	.20	338.84
S/U and I	220.98	121	<.001	.93	.06	.06	356.98
1 EF	235.56	124	<.001	.92	.07	.03	365.56
Scoring based	240.16	129	<.001	.93	.06	.04	360.16

Note. I = Inhibition, S = Shifting, U = Updating.

problems of the model, the five covariances between control and executive speeded measures had to be removed. However, these covariances improved the fit of the model but did not exert much influence on the rest of the model. The analysis showed indeed a large degree of measurement invariance: the fit was best for the measurement weights model, in which the factor loadings are constrained to be equal in both waves but the covariations between the latent factors are allowed to vary, $\chi^2(264) = 415.79$, $p < .01$, CFI = .95, RMSEA = .04, $P_{close} = 1.00$. Although the pattern of covariations between the latent factors was the same in both waves, compared to wave 1 (see Figure 3.2) the covariations between updating and both speed measures were stronger in wave 2, $r_{updating\text{-}motor\ speed} = .57$, $r_{updating\text{-}verbal\ speed} = .55$, while the covariation between updating and inhibition/shifting was somewhat lower in wave 2, $r = .40$.

Striking in the final model is the fact that most of the loadings of the speed measures of the executive tasks on the latent combined inhibition/shifting factor are very low. This factor therefore predominantly represents accuracy. Since all updating tasks were also measures of accuracy, finally a model was tested with only three latent factors: the same two verbal and motor baseline speed factors as in the previous models, with loadings of all respective verbal and motor speed measures of control, inhibition and shifting tasks, and one accuracy factor with loadings of the accuracy measures of all inhibition, shifting, and updating tasks. The latent factors were allowed to covary. If this model fits the data well, it implies that the factors merely capture the scoring method. However, this was not the case, as the fit of these models was not good (see Table 3.5: Scoring based).

To summarize, while we could replicate the distinct executive function factor of updating, we did not find strong evidence in favor of a distinction in inhibition and shifting as separable executive functions. Instead, in both waves the final model that is accepted is a model with only two executive functions: one for updating, and one combined factor for the inhibition and shifting tasks, with a moderate correlation between the two. In addition, the relation between response speed on simple tasks and response speed on executive tasks is strong: variance in response speed on most of the executive tasks was already accurately captured by the two simple speed factors.

Discussion

In this study we tried to replicate the factor structure of inhibition, shifting, and updating, as previously found in several studies (Espy et al., 2004; Hughes, 1998; Lehto et al., 2003; Miyake et al., 2000) in children, with tasks that varied as much as possible in non-executive aspects. In our data, however, a two-factor model in which inhibition and shifting were combined fitted the data just as well as the three-factor model. Moreover, in the three-factor model the correlation between the inhibition and shifting factors was so high that it is questionable whether these factors should be interpreted as representing different constructs. An interpretation of this finding may be that both shifting and inhibition require the resolution of a conflicting stimulus in which two responses are possible: in the inhibition tasks the most salient response must be inhibited in favor of another response, and in the shifting tasks two possible responses compete. Therefore this combined factor could be interpreted as the ability to process conflicting information and select the proper response (cf. Garon, Bryson, & Smith, 2008).

Moreover, the combined inhibition/shifting factor was dominated by accuracy loadings. The speed loadings were very low, but these measures loaded high on the baseline speed factors. This suggests that the speed measures of the inhibition and shifting tasks do not measure something fundamentally different from baseline speed. This does not mean that the measures are not executive: even in a simple naming task, executive skills such as decision making are probably involved (Szmalec, Vandierendonck, & Kemps, 2005). However, it does mean that administering complex speeded tasks does not have an additional value over simple speeded response tasks. This confirms the findings by

Van der Sluis et al (2007), who found that speeded inhibition tasks loaded on a baseline speed factor while addition of an inhibition factor did not improve the model.

The fact that the two-factor model as a best solution with low speed loadings on the executive factors was found twice, with measurements eighteen months apart, strengthen our findings and confirm that this result is not coincidental, specific to a certain age or merely caused by the notorious unreliability of executive function tasks. Rather, it shows that the often-used distinction in inhibition and shifting needs serious reconsideration, although replication of this finding in a different sample is desirable. It must, however, be noted that highly similar results have already been obtained in a study employing similar measures (Lee et al., 2011). In addition, the results illustrate that the decision of which measure to use to obtain an executive function task score is not trivial: when speed scores are used without a control task, contamination with baseline speed is strong, but when difference scores are used, the amount of measurement error is large.

One limitation of the study that needs to be mentioned is the low number of mistakes that were made. A ceiling effect in accuracy of the control task prevented controlling for baseline accuracy. In addition, one might argue that ceiling effects in executive accuracy measures indicated that these executive tasks were too easy and merely measured baseline speed. That would also explain the high correlations between the speed measures of these tasks and the control tasks. There are two reasons, however, that make us believe that a potential lack of difficulty of the tasks cannot fully account for the findings. First, there was a significant slow-down on the executive condition compared to the control condition on each of the five tasks that were used in the CFA. This means that the executive condition was significantly more difficult. In addition, the test-retest correlation of the accuracy measures was always significant, even though there was a time interval of 18 months, meaning that although the number of mistakes was not large, there is stability in the variation in the number of mistakes that are made. However, the speed loadings on the inhibition/shifting factor were low, and the test-retest correlations of the accuracy measures were sometimes somewhat low too. Given the difficulties that reaction time measures yielded, it is therefore recommended to use tasks that provide more variance in accuracy in future studies, so only accuracy can be used as a final score. Especially for inhibition skills one might argue that accuracy is more important than speed, since fast performance might also be an indication of impulsivity. However, it must be taken into account that in this study baseline accuracy was not included in the model, so interpretations on this account must be cautious.

Another possible weakness of the present study is the indication of response time by the research assistants, inducing another possible source of measurement error. Although we admit that this method might not be ideal – although the use of voice keys is not ideal either in the noisy environment of a school – it is not a severe threat to the validity of the study, as also here the test-retest correlations are rather strong, even though no child was tested by the same assistant in both waves. Moreover, another study showed that the correlation between voice key response and button press response is very high, $r = .90$ (Wright et al., 2003).

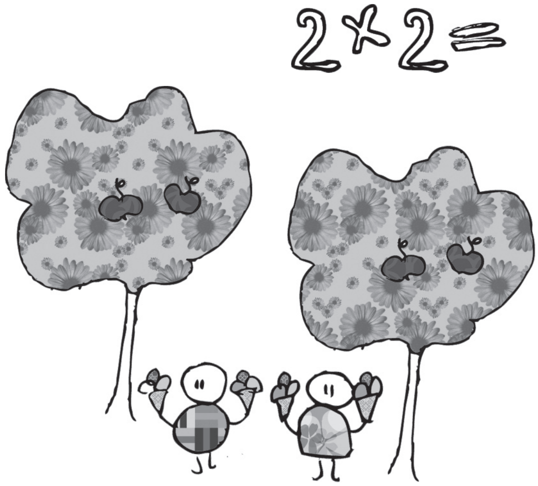
Thus, our results present strong evidence against the dissociability of inhibition and shifting, at least in children. This, however, does not explain the reason why we failed to distinguish separate inhibition and shifting factors. A possible reason might be that the organization of executive functions is different in children than in adults (St Clair-Thompson & Gathercole, 2006). More studies that failed to find the three-factor structure in children might at first sight also support this view (Huizinga et al., 2006; St Clair-Thompson & Gathercole, 2006; Van der Sluis et al., 2007; Wiebe et al., 2008). However, while these studies all failed to replicate the three-factor model, each found a different acceptable solution. This lack of consensus with regard to the factor structure of executive functions in children might reveal a more fundamental problem: statistical factors and correlations that are found need not reflect real cognitive processes, but might also be caused by other processes (see also Salthouse et al., 2003). Instead, they might be mere statistical by-products, caused by a variety of lower-order processes. The concept of higher-order functions in itself is problematic, as these functions still include a homunculus in disguise (Zelazo et al., 2003). Therefore it has been argued that the fractionation of the central executive into very broad executive functions is not the right direction and that instead it is important to focus on lower-level processes and interactions (Fournier-Vicente, Larigauderie, & Gaonac'h, 2008).

It has already been described how the coordination of resolution of conflict between task goal representations in working memory and processes that are triggered automatically can lead to executive control (Vandierendonck, Szmalec, Deschuyteneer, & Depoorter, 2007). This dynamics of lower-order processes may eventually even yield one statistical factor, possibly leading to the false conclusions that this factor represents one coherent and static skill. This case has already been convincingly made for intelligence: although a general 'g'-factor is found consistently, this 'g' need not necessarily reflect one underlying biological factor (Van der Maas et al., 2006).

The same psychometric one-factor structure also arises when it is assumed that several cognitive processes each influence each other during development: a mutualism model. This mutualism model is theoretically more plausible, as thus far no single process has ever been identified to be responsible for intelligence. It seems likely that a similar case can be made for executive functioning: that task performance on executive function tasks is also influenced by a variety of cognitive processes that develop in a mutualistic manner. In the future, it will be important to unravel these lower-order processes and to discover how these processes may mutually influence each other. A starting point may be a model in which the ability to sustain the activation of representations in working memory on the one hand, and the ability to prevent activation of unwanted representations, such as dominant responses, by reallocating attentional resources. Since it is expected that processes influence each other over time, longitudinal studies are recommended, starting as early in childhood as possible and using very precise measuring instruments. Baseline speed, a measure that seems somewhat neglected in the literature concerning executive functions may also be one of these important lower order processes, as it is shown in our study to be related to updating skills and other studies have shown it to be a strong predictor of cognitive skills in a wide age range, from children born prematurely (Rose, Feldman, Jankowski, & Van Rossem, 2005) to cognitive decline in the elderly (Salthouse et al., 2003).

Summarizing, two main conclusions can be drawn from this study. The first is that the distinction of executive functions in inhibition, shifting, and updating is not tenable, at least not in this age group. Just like 'g' in the last century and the 'central executive' of the past decades were oversimplifications, so is also the division of this central executive in only a handful executive functions. It will be more fruitful to abandon the top-down approach of looking for higher-order meaningful constructs and to focus instead on unraveling the many lower-order processes that may all contribute to proper executive functioning. Second, extreme care must be taken with regard to the choice of tests and scoring methods measuring executive functions. When scoring methods are based on reaction times, especially differences in reaction times, unwanted effects can appear that seriously hamper the interpretation of test scores.

4



The development of executive functions and early mathematics: A dynamic relationship

S.H.G. Van der Ven
E.H. Kroesbergen
J. Boom
P.P.M. Leseman

British Journal of Educational Psychology (2011),
doi: 10.1111/j.2044-8279.2011.02035.x

Abstract

Background. The relationship between executive functions and mathematical skills has been studied extensively, but results are inconclusive and how this relationship evolves longitudinally is largely unknown.

Aim. The aim was to investigate the factor structure of executive functions in inhibition, shifting, and updating, the longitudinal development of executive functions and mathematics and the relation between them.

Sample. 211 children in grade 2 (7–8 years old) from 10 schools in the Netherlands.

Method. Children were followed in grade 1 and 2 of primary education. Executive functions and mathematics were measured four times. The test battery contained multiple tasks for each executive function: Animal Stroop, Local Global and Simon task for inhibition, Animal Shifting, Trail Making Test in Colours, and Sorting Task for shifting, and Digit Span Backwards, Odd One Out, and Keep Track for updating. The factor structure of executive functions was assessed and relations with mathematics were investigated using growth modeling.

Results. CFA showed that inhibition and shifting could not be distinguished from each other. Updating was a separate factor, and its development was strongly related to mathematical development while inhibition and shifting did not predict mathematics in the presence of the updating factor.

Conclusions. The strong relationship between updating and mathematics suggest that updating skills play a key role in the maths learning process. This makes updating a promising target for future intervention studies.

Introduction

Executive functions are often studied as predictors of mathematical skills. While a universally agreed definition of the term executive functions is still lacking, it is commonly agreed that executive functions is an umbrella term for the different cognitive skills that are needed to behave flexibly and adaptively in new situations. Since learning involves being presented with new situations, it is assumed that executive functions play an important role in the maths learning process. Many processes that are executive in nature are assumed to be important when maths problems are being solved: partial results must be stored in working memory and retrieved or replaced when necessary, strategies must be chosen flexibly, old and immature strategies or possible dominant but false answers must be inhibited, and the solution process must be monitored and adapted when necessary. Indeed, many studies have found relationships between tasks measuring executive functions and (preparatory) maths skills (e.g., Barouillet & Lépine, 2005; Blair & Razza, 2007; Bull & Scerif, 2001; D'Amico & Guarnera, 2005; Espy et al., 2004; Gathercole & Pickering, 2000b; Geary, Hoard, Byrd-Craven, & DeSoto, 2004; Van der Sluis, De Jong, & Van der Leij, 2007). However, results are still inconclusive, as other studies did not find some of these relationships: especially relationships considering the executive functions of inhibition and shifting (more elaborated upon in the next section) are not always found (Censabella & Noël, 2008; Rasmussen & Bisanz, 2005; Van der Sluis, De Jong, & Van der Leij, 2004). We believe that the causes for these contradictory results may lie in the difficulties of measuring executive functions. In order to advance the field of investigating this complicated relationship between executive functions and mathematics further, we identified three issues that we believe are important and that we address in the present study: (1) the existence of multiple interrelated but distinct executive functions, (2) the 'impurity problem', and (3) developmental aspects.

First of all, while executive functions were treated as a unitary system in the past, now many researchers believe that there are several different executive functions that are distinct but interrelated (Miyake et al., 2000). A distinction between inhibition, shifting, and updating based on a proposal by Baddeley (1996) is often used in this respect. Inhibition entails the suppression of a dominant or pre-potent response in favour of another response or no response at all. Shifting refers to the ability to switch flexibly between rules or tasks. Updating, finally, is the ability to manipulate and update information temporarily stored in working memory. Studies that related these executive functions to maths found a relationship between maths and updating almost invariably (De Smedt et al., 2009; Lee et al., 2011; Passolunghi, Mammarella, & Altoè, 2008; St Clair-Thompson & Gathercole, 2006; Swanson & Jerman, 2006; Swanson & Kim, 2007; Wu et al., 2008), but results regarding shifting and inhibition are less clear, with some studies finding a relationships between tasks measuring inhibition and/or shifting abilities (Blair & Razza, 2007; Bull, Espy, & Wiebe, 2008; Van der Sluis et al., 2007), while others did not or only partially support these findings (Censabella & Noël, 2008; Espy et al., 2004; Lee et al., 2011; Rasmussen & Bisanz, 2005; Van der Sluis et al., 2004). These inconclusive findings might be explained by the fact that some of these studies did not control for other executive functions, even though they are interrelated (Miyake et al., 2000). However, failure to tackle the impurity problem may be a more important issue.

The impurity problem refers to the fact that executive functions regulate other cognitive functions, which means that every single executive task also measures some non-executive skills, such as verbal speed or visual search efficiency (Hughes, 1998; Miyake et al., 2000). Because of the impurity problem, executive functions cannot be measured in isolation by a single task: a single score is inevitably contaminated with variance in non-executive skills. This is especially problematic when tasks are used that contain contents related to the criterion variable: in this case mathematical or numerical contents. The frequently used counting span task illustrates this issue: performance depends not only on updating skills, but also on counting speed (Case, Kurland, & Goldberg, 1982), and mathematically disabled children are slower counters (Hitch & McAuley, 1991). The impurity problem can be addressed by considering the results of a confirmatory factor analysis, rather than individual task scores, as is often done. In studies that applied factor analysis, even the distinction of executive functions in inhibition, shifting, and updating proved questionable, with some studies finding evidence in favour of this distinction (Espy et al., 2004; Hughes, 1998; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Miyake et al., 2000) while others, including a previous study in our lab (Van der Ven, Kroesbergen, Boom, & Leseman, 2011a)[1] did not (Huizinga, Dolan, & Van der Molen, 2006; Lee et al., 2011; St Clair-Thompson & Gathercole, 2006; Van der Sluis et al., 2007; Wiebe, Espy, & Charak, 2008). An updating or working memory factor was found in most of the latter studies, but whether inhibition and shifting can be seen as separable constructs is less clear. In addition, correlations between the factors were substantial: e.g., around .40 (Huizinga et al., 2006), or even around .50 (Miyake et al., 2000) to .65 (Lehto et al., 2003). Some of the studies that incorporated factor analysis also related the obtained factors to maths performance (Espy et al., 2004; Lee et al., 2011; St Clair-Thompson & Gathercole, 2006; Van der Sluis et al., 2007). Results were inconclusive: the only commonality between these three studies was the updating factor, which was invariantly present as a separable factor and always significantly related to mathematics. When separate inhibition and shifting factors were identified, these were not always significantly related to mathematics. These studies suggest that relationships between executive functions and mathematics may not always be present and positive findings might be caused by imperfections in the research design.

The third issue relates to the development of maths. Both executive functions and mathematical skills develop strongly during childhood; executive functions are therefore supposed to be especially important in the learning process (Holmes, Gathercole, & Dunning, 2009; Posner & Rothbart, 2005) and executive and other cognitive skills are likely to influence each other mutually during development (Jones, Gobet, & Pine, 2007; Messer, Leseman, Boom, & Mayo, 2010; Ottem, Lian, & Karlsen, 2007; Van der Maas et al., 2006). A longitudinal design can capture more of these complex relationships, as it can capture the relationship between growth in executive functioning and growth in mathematical skills, even though the exact direction of these hypothesized mutually causal relationships is difficult to determine.

Longitudinal studies are still scarce and inconclusive. To date, the available studies did not investigate different executive functions and most did not use factor analysis. In addition, most longitudinal studies used executive function measures at one time as predictors for mathematical skills at a later time. This approach does not allow the investigation of growth itself. However, these studies did show that tasks measuring one or more executive functions were predictive of maths performance up to three years later (Bull et al., 2008; Clark, Pritchard, & Woodward, 2010; De Smedt et al., 2009; Jenks, De

1 Chapter 3 of this dissertation.

Moor, & Van Lieshout, 2009; Krajewski & Schneider, 2009b; Welsh, Nix, Blair, Bierman, & Nelson, 2010). Passolunghi and colleagues (2008) even showed relationships between cognitive skills and maths that differed over time: in first grade, the phonological loop, performance IQ and updating all contributed to maths performance, whereas in second grade only updating was a predictor of math. This pattern that changes over time stresses again that cognitive skills in development are complex and dynamic and that preferably also the development of these skills should be investigated. Some studies did investigate growth: one found that both cross-sectional initial differences and individual growth differences in updating predicted mathematical problem solving one year later (Swanson, 2006b), another that maths was predicted by both inhibition and short term memory (Panaoura & Philippoua, 2007). Moreover, in the latter study inhibition was also significantly related to growth in maths skills.

Studies are needed that tackle all three issues raised before simultaneously: studies that incorporate multiple executive functions, with multiple measures for each executive function in order to address the impurity problem, using a longitudinal design. The current study was designed as a first effort to take up these challenges. During the first two years of primary education, children's development of executive functions was followed and related to the development of mathematical skills. The use of multiple growth models enabled a detailed assessment of the development of executive function and mathematical skills.

The first aim of the study was to investigate for each wave separately whether it is possible to identify the three executive functions of inhibition, shifting, and updating by means of confirmatory factor analyses. Stability of this factor structure was examined by repeating the procedure at each wave. The obtained factors were related to concurrent mathematics scores. The design enabled a comparison of two possible competing hypotheses. On the one hand, there are theoretically sound arguments, that are strikingly left implicit in most studies, that argue for a relationship between each executive function and mathematics separately: inhibition because a child has to inhibit possible prepotent but false answers and the use of old and well-practiced but inefficient strategies, such as finger counting; shifting may facilitate switching between different maths strategies; and updating enables the storage and updating of relevant aspects of the problem and partial results in working memory while solving a problem. On the other hand, there is the alternative hypothesis that positive relationships found in the past might have been caused by the two aforementioned methodological issues: failing to address the impurity problem by only using single tests for an executive function, or assessing only a single executive function although they are interrelated. The fact that an increasing number of studies has not replicated the three-factor structure argues for this interpretation.

The next aim was to investigate these relationships longitudinally. Latent Variable Growth Curve Models were built to model the development of each skill separately. These models were related to each other to investigate how these skills develop mutually. We expected that the relationship between executive functions and mathematics is dynamic: both are expected to influence each other's development.

Method

Participants

At the beginning of the study, 227 children (120 boys, 107 girls) with a mean age of 6 years 5 months (*SD* = 4.3 months; range 5;9–7;7 years) took part. Children came from 18 classrooms in 10 schools. Parental consent was obtained from all participating children. School choice was based on three criteria: (1) a low number of children not speaking Dutch at home, (2) diversity in SES (i.e., schools with high and low numbers of parents that had completed less than 2 years of secondary education) and (3) use of the same mathematics teaching method. As the aim was to obtain a representative sample of children following regular education, there were no stringent exclusion criteria; however, three children were excluded because of failure to understand the task instructions (one child with Down syndrome and two refugee children with insufficient mastery of the Dutch language). During the course of the study, thirteen more children (5.8%) dropped out, due to moving (seven children), grade retention (three children) and accelerating a grade (three children). All analyses were performed with the 211 remaining children (110 boys, 101 girls, mean age = 6;5 years, *SD* = 4.4 months). In the final wave, 209 children completed the Raven's Standard Progressive Matrices (Raven, 1958). Scores were slightly below average: 66 children (31.6%) obtained scores in the lowest quartile, 69 children (33.0%) in the second quartile, 60 children (28.7%) in the third and 14 children (6.7%) in the highest quartile.

From 130 children (62%), the educational background of at least one of the parents was available and rated according to the International Standard Classification of Education (UNESCO Institute for Statistics, 1999). Of these parents, 22.3% had finished only lower secondary education or less (ISCED level 1 or 2), 31.9% had completed upper secondary education or vocational training (ISCED level 3 or 4) and 45.8% had completed higher vocational training or university (ISCED level 5 or 6). The children whose parental education data were available did not differ significantly from the 38% whose data were missing on executive function or maths measures, $F\,(40, 136) = 1.33$, $p = .12$, and relations between parental education and executive function and maths measures were weak (range: $r = .01$ to $r = .17$). Therefore this measure is not used in further analyses.

Procedure

Testing of executive functions took place in four measurement waves (from now on referred to as W1–W4) with six-month intervals: in fall (W1) and spring (W2) of first grade and fall (W3) and spring (W4) of second grade. The children were tested individually in a separate room by trained research assistants. The tasks were divided over three sessions, lasting 35, 25, and 15 minutes respectively. Maths test results were obtained from the school; maths tests were administered approximately three months after each wave.

Instruments

Multiple tasks per executive function were selected. The tasks met several criteria: (1) whenever possible, they had been used in previous research or were based on such tasks, (2) they differed as much as possible in all non-executive aspects, such as input modality (visual or verbal) and response mode (verbal or motor), and (3) they did not contain mathematical contents. All tasks were administered on a laptop, and except for the Trail Making Test in Colours, all were programmed in E-Prime (Psychological Software Tools, http://www.pstnet.com). Computerized assessment allowed for detailed data: accuracy and response latency were obtained for each stimulus separately. Several tasks required a quick verbal response of the child. The experimenter indicated the start of the response by pressing the spacebar and rated the accuracy of the response after each stimulus. Reaction time and accuracy were always based on the first response, even when self-corrections were made and also when the first response wasn't vocalized completely (e.g., c…duck for cow…duck). In order to ensure identical testing conditions for each child, stimuli were always presented in the same, balanced order. Example items of each executive function task are given in Figure 4.1.

Inhibition

Animal Stroop. An adaptation of the Animal Stroop task, created by Wright, Waterman, Prescott, and Murdoch-Eaton (2003) was used. In this task, animals are presented that are composed of the body of one animal and the head of another. The participant has to name the animal body rather than the more salient animal head. The task consisted of two blocks. In each block animal stimuli were presented one at a time, preceded by a 400 ms fixation cross. The stimuli remained on the screen until the child responded. The child named each animal as quickly as possible. The first block consisted of 4 normal animal stimuli with congruent heads (cow, sheep, duck or pig), each presented 12 times, yielding a total of 48 items. The second block was a mixed block of control stimuli with human heads and incongruent stimuli with incorrect animal heads. Twelve different stimuli per condition were each presented four times, yielding a total of 48 control and 48 incongruent items. The number of inhibition items answered correctly per second was used as a final score. The observed range was 0.33–1.00 items per second.

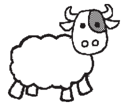
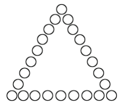
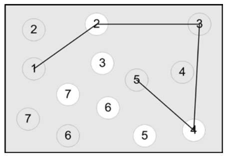
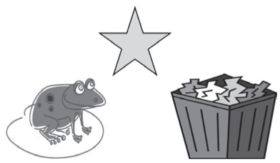
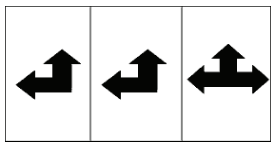
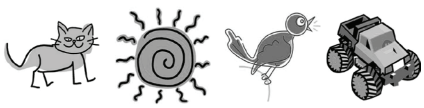
Task name	Instruction	Example item
Animal Stroop	"Name the animal body"	
Local Global	"Name the smaller shape"	
Simon Task	"If you see a mouse, press the left button, if you see a dragon, press the right button"	 (presented on right or left of computer screen)
Animal Shifting	"If the screen is yellow, name the fruit; if the screen is purple, name the animal"	 (varying background color)
Trail Making in Colours	"Draw a line from 1 to 2 to 3 …, while alternating the colours."	
Sorting Task	"If you see a dog, give it the stimulus if it's blue; throw it away if it's orange. If you see a frog, give it the stimulus if it's a star; throw it away if it's a square."	
Digit Span Backwards	"Repeat the sequence that you hear backwards"	"3 5 1 6"
Odd One Out	"Point at the shape that is different and remember its location."	
Keep Track	"Name each picture you see and remember the last animal."	 (sequence of 10 successive stimuli)

Figure 4.1 Example items of the executive function tasks.

Local Global. In this task identical small geometrical shapes (circle, triangle, square) that together constituted a larger, different geometrical shape were presented. Sometimes the large, global shape had to be named, sometimes the small, local shapes. In general, people show a global preference: responses are faster when the large image has to be named rather than the local shapes (Navon, 1977). The task consisted of three blocks. In the first block, 48 single, small geometrical shapes were presented as control stimuli. The second and third block each consisted of 48 larger stimuli that were constructed from the stimuli presented in the first block. All stimuli in these blocks were incongruent: the shape of the larger image was always different from the elements from which it was built. In the second block, the large stimulus had to be named, while in the third block the small image had to be named. Trials were preceded by a 400 ms fixation cross. The number of items answered correctly per second in the third block was used as a final score. The observed range was 0.37–0.98 items per second.

Simon Task. The Simon effect can be elicited with tasks where stimuli are presented at different horizontal locations on a screen, while this spatial aspect must be ignored. A response (a left or right button press) is based on some non-spatial aspect of the stimulus, such as the identity. Nevertheless, the spatial aspect is difficult to ignore: when a correct response entails pressing the button on the side where the stimulus is presented, this response is generally faster and more accurate than a response on the other side (Simon & Berbaum, 1990; Simon & Rudell, 1967). The task was presented as a zoo game with animals (mice and dragons) that had escaped. The child was instructed to capture the escaped animals as quickly and accurately as possible. On a computer screen, a picture of a mouse or a dragon appeared. When a mouse appeared, it had to be captured by a left key press (A); a dragon was captured by pressing a right key (L). Both keys were marked with a green sticker. The task started with a block with 40 control items, in which the stimuli appeared in the centre of the screen. After the control block, a Simon block appeared, where the animals were presented on the left or right side of the screen. There were 40 congruent (same-side button press) and 40 incongruent (other-side button press) trials. Both types of trials were intermixed. Stimuli were present until the child pressed a button. When the button press was correct, feedback was provided: a cage appeared around the animal. Nothing happened upon a false button press. Trials were preceded by a 500 ms fixation cross. The number of incongruent items answered correctly per second was used as a final score. The observed range was 0.48–1.71 items per second.

Shifting

Animal Shifting. In this task, the participant had to name stimuli that were presented on the computer screen as quickly as possible. The stimuli came from two categories: fruit (strawberry, pear, cherry, banana) or animal (cat, dog, bird, fish). The task consisted of two blocks that both contained 40 items, each preceded by a 700 ms fixation cross. In the first block, the child named each stimulus as quickly as possible. In the second block a piece of fruit and an animal appeared together. Depending on the background colour of the screen, the child named only one: the piece of fruit when it was yellow, the animal when it was purple. The number of shifting items answered correctly per second was used as a final score. The observed range was 0.09–0.75 items per second.

Trail Making Test in Colours. This test is based on the Children Colored Trail Test (Llorente, Williams, Satz, & D'Elia, 2003), in which coloured circles with a number in the middle must be connected in the correct order. The task consisted of a control task and a shifting task. In the control task, 20 circles numbered 1–20 were presented on the screen. Half of the circles were orange and the other half were blue, but this information was not relevant. The child had to connect the circles 1–10 by clicking them with the mouse. When a mistake was made, the mouse pointer returned to the previous circle. The tracks followed a relatively easy path over the screen, thus reducing the visual search component of the task. In the first shifting task, the configuration of the circles was the same as in the first control task, but mirrored vertically. Now the 20 circles were numbered 1–10, each number being present in both blue and orange. The child had to connect the circles while alternating the colours. Before each task, a short practice item was administered. The number of shifting items answered correctly per second was used as a final score. The observed range was 0.03–0.71 items per second.

Sorting Task. In this task, inspired by Zelazo and colleagues (Zelazo, Müller, Frye, & Marcovitch, 2003) the child had to alternate between two sorting rules: according to colour and according to shape. The task was presented as a game in which a dog that likes blue and a frog that likes stars were introduced to the child. The child had to give the animal the stimuli that it liked and throw away the stimuli the animal did not like. The task contained two control blocks and a shifting block. In each block, the stimuli were 40 orange and blue stars and squares. In the first control task, the child was introduced to the dog. The child was told that this dog loved blue but hated orange. The dog was shown on the lower left side of the screen, and a waste bin was shown on the lower right side of the screen. After 700 ms a stimulus appeared. When the stimulus was blue, the child 'gave' it to the dog by pressing the A button on the left side; when it was orange, the child 'threw it away' by pressing the L on the right side. No feedback was provided. The second control task was similar to the first control task, but now the dog was replaced by a frog, and the child was told that the frog loved stars and hated squares. The same stimuli were presented again, and this time the child sorted the items according to shape. The shifting task was a mixed block in which sometimes a dog and sometimes a frog appeared; the same 40 items were shown again. The number of shifting items answered correctly per second was used as a final score. The observed range was 0.29–1.82 items per second.

Updating

Digit Span Backwards. An adaptation of the Dutch version of Digit Span Backwards from the Automated Updating Assessment test battery (Alloway, 2007) was administered. The child heard a digit sequence and tried to repeat it backwards, starting with a two-digit sequence. After three correct answers, the sequence increased by one digit. When two mistakes were made in trials of the same length, the task was discontinued. The number of correct answers was used as a final score. The observed range was 0–12 correct trials.

Odd One Out. This task was also adapted from the Automated Updating Assessment (Alloway, 2007). A series of stimuli was shown consecutively. Each stimulus consisted of three boxes with shapes presented next to each other. One of the shapes differed from the other two. The child was asked to point at the deviant shape. Then the next stimulus was presented. At the end of each trial three empty boxes appeared; the child had to point at the locations of the previously shown deviant shapes in the same order in which they had appeared. An answer was correct if each location was recalled correctly in the right order. The task started with only one item; after three correct answers of the same length, the sequence increased by one. When two mistakes were made in trials of the same length, the task was discontinued. The number of correct sequences was used as a final score. The observed range was 3–16 correct items.

Keep Track. A computerized version of the Keep Track task was created. This task was adapted from Van der Sluis et al. (2007). The child was shown pictures, each of which belonged to one of the following five categories: sky (sun, moon, stars, cloud), fruit (strawberry, pear, cherry, banana), shapes (square, triangle, circle, heart), animals (dog, cat, fish, bird), and toys (teddy bear, scooter, 'LEGO®', car). The pictures were shown in series of ten; each picture was displayed for 3.5 seconds. The child was asked beforehand to pay special attention to one or more designated categories. During the series, the child had to name each picture. At the end, the child had to recall the last item of the designated categories. In each series, one, two, or three items of each designated category were presented. The remainder consisted of filler items from other categories. During the series, small pictures symbolizing the to-be-remembered categories were shown in the bottom of the screen, serving as a reminder. The number of to-be-remembered categories increased from one to four. There were two series of each difficulty level, yielding a total of eight series. Each correct answer was noted. Depending on the length of the sequence, there could thus be a maximum of one, two, three, or four correct answers in a trial, yielding a maximum possible score of twenty correct responses. The total number of correct responses was recorded. Prior to testing, the child was familiarized with the pictures and the categories. The test was preceded by a practice item that was repeated when necessary. The observed range was 1–20 correct items.

Mathematics

Cito Mathematics. All schools administered standardized Dutch maths tests twice a year. The tests that were used were the mathematical part of Cito tests (Janssen, Scheltens, & Kraemer, 2005). These are national Dutch tests with good psychometric properties that are commonly used in Dutch schools to monitor the progress of primary school children. For each grade, there are two level-appropriate tests: one is administered in January, the other in June (halfway and end of year). In grades 1 and 2, five main domains are covered: (1) numbers and number relations, covering the structure of the number line and relations between numbers, (2) simple addition and subtraction, (3) simple multiplication and division, (4) complex maths applications, often involving multiple mathematical manipulations and (5) measuring (e.g., weight and length). However, this division is not clear-cut: many problems tap knowledge from more than one domain and the number and relative difficulty of problems per domain varies per wave, because each test reflects the skills that have been learned in that specific period. Therefore it is not possible to investigate sub-domains of mathematics; instead, only the final score can be used as an indication of broad mathematical skills.

The problems are presented in a booklet. Each problem comes with a picture that sometimes, but not always, contains necessary information. The test is administered group-wise by the teacher, who reads the problems out loud. The tests contain 50 (grade 1), 52 (halfway grade 2) or 54 (end grade 2) items that are administered on two separate days. Raw scores are converted into competence scores that increase throughout primary school, enabling the comparison of results of different tests. The observed range was a score of 0–109.

Analyses

The analyses were performed using Latent Variable Growth Curve Models, a technique that uses Structural Equation Modelling. As is common in Structural Equation Modeling, overall model fit is indicated by several fit indices, which each evaluate different aspects of the model. Here, the fit indices recommended by Blunch (2008) are reported: chi square (χ^2) with its *p*-value, CFI, RMSEA with its P_{close}-value and AIC. Chi square is a discrepancy measure; between the current model and the saturated model. Since this discrepancy should be small, chi square should be as low as possible. The *p*-value indicates the probability of obtaining discrepancies with the observed data that are at least as extreme, given that the model is correct. Therefore, *p* should be as high as possible. However, larger sample sizes increase the chances of obtaining a significant *p*-value. Therefore other fit indices are also necessary. CFI compares the fit of the model to the independence model, in which all covariances are constrained to zero. CFI values of > .95 can be considered a good fit, and > .90 an acceptable fit. RMSEA estimates the amount of error of approximation per degree of freedom, thus favoring simpler models. RMSEA is good if < .05 and acceptable if < .08. P_{close} tests the null-hypothesis that the population RMSEA < .05 and should therefore be high. AIC, finally, cannot be used as an absolute fit index, but is used to compare models: the lowest AIC value is preferred.

Results

Means and standard deviations of the final scores of each task at each wave are given in Table 4.1. Correlations between the tasks, averaged over the four measurement waves, are given in Table 4.2. Correlations at each wave separately do not differ much from the averaged correlations as mentioned in the table.

Concurrent analyses

First, a confirmatory factor analysis was performed on the executive function tasks. Five alternative models were tested for each wave separately. The first model was the classical model as proposed by Miyake et al. (2000), with three latent factors: inhibition, shifting, and updating; each latent factor having three indicators. In three alternative models, two of these factors were collapsed, and finally a model was tested with all executive tasks

Table 4.1 Descriptive statistics

	Wave1		Wave 2		Wave 3		Wave 4	
Task	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Mathematics								
Cito Mathematics	35.92	13.57	45.99	14.05	53.17	14.57	63.12	15.01
Inhibition								
Animal Stroop	0.61	0.10	0.62	0.10	0.62	0.10	0.62	0.10
Local Global	0.57	0.10	0.59	0.11	0.60	0.12	0.67	0.17
Simon Task	0.95	0.15	1.04	0.19	1.09	0.17	1.16	0.19
Shifting								
Animal Shifting	0.43	0.09	0.45	0.09	0.47	0.09	0.52	0.10
TMT-C[a]	0.16	0.06	0.24	0.08	0.27	0.08	0.35	0.09
Sorting Task	0.62	0.14	0.69	0.19	0.76	0.21	0.82	0.22
Updating								
Digit Span Backwards	3.75	1.68	4.68	1.56	4.98	1.58	5.72	1.82
Odd One Out	6.77	2.44	7.75	2.64	8.41	2.63	9.40	2.68
Keep Track	11.73	3.02	12.99	2.70	13.97	2.90	14.91	2.67

[a]Trail Making Test in Colours

Table 4.2 Correlations between the tasks, averaged over the four waves

Task	1	2	3	4	5	6	7	8	9
1. Cito Mathematics	-								
2. Animal Stroop	.18	-							
3. Local Global	.18	.32	-						
4. Simon Task	.29	.36	.42	-					
5. Animal Shifting	.20	.30	.47	.43	-				
6. TMT-C[a]	.31	.27	.19	.36	.28	-			
7. Sorting Task	.17	.26	.33	.44	.35	.25	-		
8. Digit Span Backwards	.22	.13	.15	.18	.21	.19	.12	-	
9. Odd One Out	.34	.21	.15	.27	.12	.26	.16	.23	-
10. Keep Track	.34	.28	.22	.33	.29	.33	.24	.22	.33

[a]Trail Making Test in Colours
Note. Since correlations were averaged over the four waves, *p*-values are not meaningful.

loading on a single factor. These models were compared and the best fitting model was chosen for further analyses. From Table 4.3 it can be concluded that a two-factor model, with one factor for inhibition and shifting combined and another for updating always provided the best fit. This result was also found when alternative scoring methods were used: for a detailed description see Van der Ven et al. (2011a). Therefore this two-factor model was used in further analyses.

In the next step, mathematics scores were added to the models described in the previous section. Correlations between the two latent factors and maths were estimated for each wave separately. These were always significant, as was the correlation between the two latent factors (see Table 4.4: Correlation). In addition, the unique contribution of each factor to maths performance was investigated, by using these latent factors (including their correlation) as predictors for maths performance. Updating was always a significant predictor, but in the presence of updating, inhibition/shifting never was (see Table 4.4: Regression).

Longitudinal analyses

The next step was to investigate the development of both executive functions and maths, first univariately. This was done by means of Latent Variable Growth Curve Models. With this analysis the development of a skill is modelled by estimating the best growth curve for each child. While a growth curve can theoretically take any shape, in this study linear growth provided the best fit and only these results are presented here.

A linear growth curve can be described by a linear equation: $y = a*t + b$, in which a represents the slope and b the intercept of the line. Time is represented by t: $t = 0$ at the first wave and 1, 2, and 3 at subsequent waves. The intercept reflects the height of

Table 4.3 Fit indices for models in all combinations of possible factor structures at each wave, with the best fitting models in italics

	χ^2	*df*	*p*	*CFI*	*RMSEA*	P_{close}	*AIC*
			Wave 1				
3 factors	N.A.						
2 factors							
(1) Inh & Shift; (2) Upd	*45.24*	*26*	*.01*	*.94*	*.06*	*.27*	*101.24*
(1) Inh & Upd; (2) Shift	N.A.						
(1) Shift & Upd; (2) Inh	66.42	26	< .001	.87	.09	.01	122.42
1 factor	66.84	27	< .001	.87	.08	.02	120.84
			Wave 2				
3 factors	N.A.						
2 factors							
(1) Inh & Shift; (2) Upd	*45.34*	*26*	*.01*	*.95*	*.06*	*.27*	*101.34*
(1) Inh & Upd; (2) Shift	N.A.						
(1) Shift & Upd; (2) Inh	68.23	26	< .001	.88	.09	.01	124.23
1 factor	70.01	27	< .001	.88	.09	.01	124.01
			Wave 3				
3 factors	N.A.						
2 factors							
(1) Inh & Shift; (2) Upd	*41.66*	*26*	*.03*	*.96*	*.05*	*.39*	*97.66*
(1) Inh & Upd; (2) Shift	N.A.						
(1) Shift & Upd; (2) Inh	N.A.						
1 factor	N.A.						
			Wave 4				
3 factors	N.A.						
2 factors							
(1) Inh & Shift; (2) Upd	*35.13*	*26*	*.11*	*.97*	*.04*	*.65*	*91.13*
(1) Inh & Upd; (2) Shift	N.A.						
(1) Shift & Upd; (2) Inh	N.A.						
1 factor	42.28	27	.03	.95	.05	.73	96.28

Note. CFI = Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation, AIC = Akaike's Information Criterion, Inh = Inhibition, Shift = Shifting, Upd = Updating, N.A. = not admissible.

Table 4.4 Estimated relations between latent executive function factors and mathematics

	Wave 1		**Wave 2**		**Wave 3**		**Wave 4**	
	Maths	Upd	Maths	Upd	Maths	Upd	Maths	Upd
Correlation (*r*)								
Inh & Shift	.37**	.57**	.35*	.63**	.41**	.74**	.41*	.79**
Upd	.55**		.49*		.61**		.55**	
Regression (β)								
Inh & Shift	.08		.02		-.12		-.30	
Upd	.51**		.44**		.68**		.84**	

Note. In the regression analyses, a correlation between Inh & Shift and Upd was allowed; the size was the same as in the correlational analyses. * $p < .05$, ** $p < .01$. Inh = Inhibition, Shift = Shifting, Upd = Updating.

the line, or, in terms of cognitive ability, the general ability of the child throughout the study. The slope reflects the average growth between measurements. Both the intercept and the slope are estimated as latent variables, with the scores at each wave as observed indicators. The intercept does not change over time. Therefore each indicator loads equally strongly on the intercept: loadings are fixed at 1. The slope exerts an increasingly larger influence on task scores at later waves, and therefore these loadings also increase: 0 for the first wave, and 1, 2, and 3 for successive waves. The linearity of this increase in loadings reflects linear growth.

For both the intercept and the slope, means and variances are estimated, reflecting the average intercept and slope and individual differences respectively. A significant covariance between intercept and slope indicates that children with a higher level develop faster if the covariance is positive or the reverse if it is negative.

First, separate growth models were created for mathematics, updating, and inhibition/shifting (univariate growth). The fit indices of the three separate growth models are listed in Table 4.5 under univariate growth. The fit for the mathematics growth model was excellent. There was no significant covariance between the intercept and slope, and therefore the covariance was fixed to 0, as is commonly done with non-significant relations in Structural Equation Modelling. The maths growth model is depicted on the right side of Figure 4.2. The explained variance shows that a linear growth model captures most of the variance of the mathematics tasks. The mean and variance of the intercept were both significant, indicating that children have an average level that is different from zero, and they differ significantly in their average level. The mean and variance of the slope were also significant, which means that children developed significantly during the time of the study and also differed in this development.

The growth model for updating was more complex, as it consisted of three different tasks, each with a different scale. Common methods to model this growth are factor-of-curves models and curve-of-factors models (Duncan, Duncan, & Strycker, 2006). We created a factor-of-curves model. In this model, separate growth models are created for each task. Two higher-order factors are then added: a higher-order intercept, on which

Table 4.5 Growth model fit indices

	χ^2	*df*	*p*	*CFI*	*RMSEA*	p_{close}
Univariate growth						
Mathematics	7.99	6	.24	1.00	.04	.53
Updating	66.01	62	.34	.99	.02	.97
Inhibition/Shifting	1579.72	277	< .001	.62	.15	< .001
Multivariate growth						
Updating and Mathematics	117.604	114	.39	1.00	.01	1.00

Note. CFI = Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation.

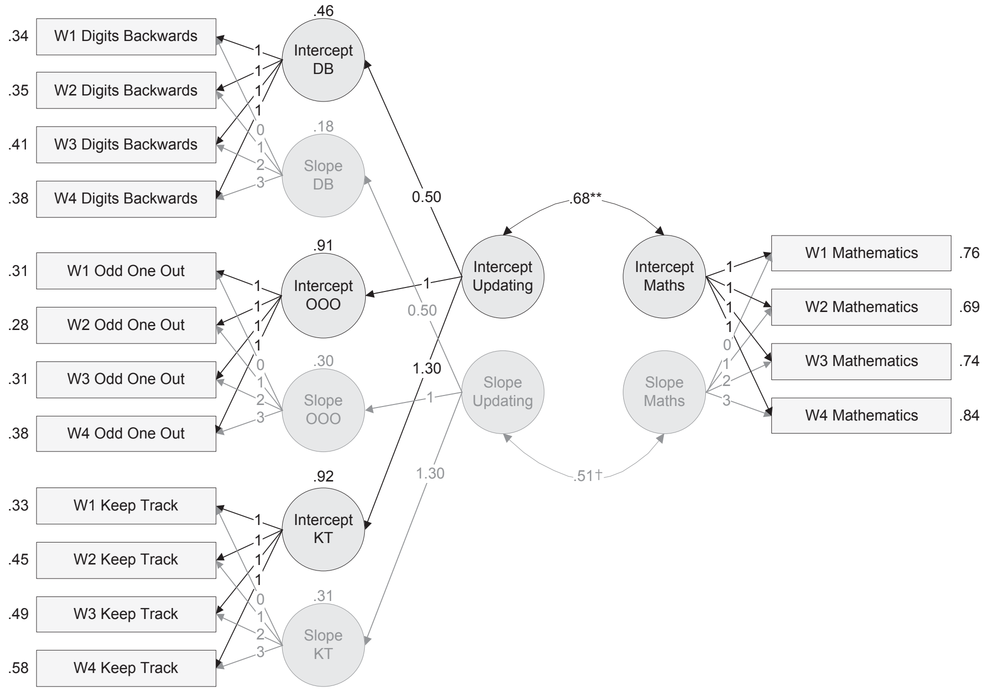


Figure 4.2 Growth model of updating (left part), mathematics (right part) and the correlation between the two (middle part). Factor loadings (above arrows) are unstandardized; explained variance (next to each rectangle and above the task-specific growth factors) and correlations between latent factors are standardized. To aid visibility, error terms are not displayed and all intercept loadings are printed in gray. * $p < .05$, ** $p < .001$, † $p = .12$.

the three task-specific intercepts load, and a higher-order slope, on which the three slopes load (see the left side of Figure 4.2). The intercept and slope loadings of each task were constrained to be equal. A model with a very good fit was obtained (see Table 4.5). The fit of the model was best with a covariance of 0 between slope and intercept. This growth model is shown on the left side of Figure 4.2. The model fitted the data well, but the explained variance is somewhat low. This, however, was expected and is not problematic, because the three tasks do not exclusively tap the same construct: all executive function tasks contain task-specific and possibly non-executive elements that are not captured by a common latent factor. The variance of this common factor for the intercepts is significant: children differed in their average updating proficiency. The variance of the common factor for the slopes was not significant and far smaller than the intercept variance. Nevertheless, it is unlikely that children do not differ in improvement. Moreover, overall fit was not improved by constraining this variance to zero. Therefore it was retained in the model.

For inhibition/shifting, a procedure similar to the updating factor was followed, but despite several different attempts to model its growth, it proved impossible to create a growth model with a satisfactory fit. This indicates that there was no consistent growth of this latent factor. Fit indices are again presented in Table 4.5.

Finally, the relationship between the growth models of executive functions and growth in mathematical performance was investigated in a combined SEM model: multivariate growth. The combined shifting/inhibition function was excluded, as it proved impossible to model the growth of this factor and, moreover, its value in the prediction of maths in the presence of the updating factor was limited, judging from the abovementioned concurrent analyses. The relationship between updating and maths was investigated further by connecting the intercepts and slopes of the two growth models to each other. Covariances rather than regressions were modelled, as it is reasonable to assume that executive functions and mathematics influence each other mutually (Welsh et al., 2010). The relations between the two intercepts and between the two slopes were strong. However, it must be noted that the slope-slope correlation was not significant, despite its estimated large size. This means that there probably is a strong relationship between the two slopes, but its size was difficult to determine. This is also reflected in the large 95% confidence interval (-.20–1.21) of the slope-slope correlation. The 95% confidence interval of the correlation between the two intercepts was far smaller (.58–.77): this relationship could be estimated with more precision.

The intercept-slope and slope-intercept relationships were negligible and non-significant and were therefore fixed at 0. The fit of this model was good (see Table 4.5: Multivariate Growth). These results suggest that both the overall level of the child and growth during the two years of study in both updating and mathematical performance were strongly related to each other.

Discussion

In this study, we investigated relationships between different executive functions and mathematics, using a longitudinal design. We set ourselves three goals: (1) a comprehensive test battery that included tasks measuring the three different executive functions of inhibition, shifting, and updating, with (2) multiple tasks for each executive function in order to address the impurity problem. In addition, (3) the study was longitudinal in order to investigate developmental processes. With this approach we came to several conclusions.

We found that the three-factor structure of inhibition, shifting, and updating, could not be replicated. Instead, the inhibition and shifting tasks loaded on one factor only. The underlying reason for the shifting and inhibition tasks loading on one factor may be that both inhibition and shifting tasks require making a fast and accurate choice between two competing responses (a dominant and a non-dominant response in the case of inhibition; two equivalent responses of which only one is correct for shifting tasks). Since this factor structure was found in each wave and using various scoring methods of the executive tasks, this is evidence against the dissociability of inhibition and shifting.

We also found that both executive function factors were significantly related to mathematics at each wave. However, when the predictive value of both was assessed simultaneously, only updating was a significant predictor. In other words, the shared variance of both executive factors fully accounted for the relationship between the inhibition/shifting factor and mathematics. Miyake and colleagues (2000) suggested that the correlation between executive functions could be explained by the inhibitory requirements that all executive function tasks share. Our findings suggest that the opposite might also be true: all executive function tasks, including inhibition and shifting tasks, require updating skills. Updating is likely to be necessary to maintain representations of the inhibition and shifting task requirements and, in the case of shifting tasks, of the sets between which shifting is necessary (Garon, Bryson, & Smith, 2008). These updating requirements might also be responsible for the relationships between inhibition, shifting, and maths that were found in previous studies (e.g., Blair & Razza, 2007; McLean & Hitch, 1999; St Clair-Thompson & Gathercole, 2006). These findings highlight again that, because of the strong correlations between the different executive functions, it is important to include more in one study.

In order to investigate development, we created growth models of both executive functioning and maths. While linear growth models fitted both the maths tests and the updating tasks well, this was not true for the inhibition/shifting tasks: the common ability tapped by the latter tasks did not show consistent growth. This lack of consistent growth is further evidence against the distinction of the different executive functions. While updating is clearly a separate factor, the existence of inhibition and shifting is unclear, as is the interpretation of the combined factor. The separability of inhibition and shifting and possible alternative distinctions deserve more attention in future studies, given the fact that many studies using factor analysis could not confirm this distinction (Huizinga et al., 2006; St Clair-Thompson & Gathercole, 2006; Van der Sluis et al., 2007; Wiebe et al., 2008), but nevertheless it is used without examining its validity in many other studies concerning the relationship between executive functions and mathematics (e.g., Blair & Razza, 2007; Bull & Scerif, 2001; Jenks et al., 2009).

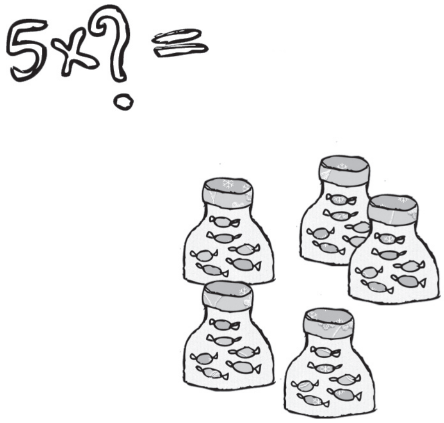
Finally, we found that the intercept of updating correlated strongly with the intercept of math, and both slopes were also strongly correlated. Especially this slope-slope effect suggests that updating is an important factor during the maths learning process and that updating and maths influence each other mutually. The capacity to store and manipulate information seems a vital process in maths learning: the information in a maths problem and the necessary operations must be remembered and partial results must be stored.

Updating capacity in its turn might also increase because children practice these skills during maths problem solving. Children with lower updating capacities are likely to be slower and more error-prone in their mathematical procedures, leading to a slower development.

These results reflect a strong relationship between updating and mathematics. Nevertheless, some limitations of this study need mentioning. First of all, the test used to measure mathematics does not allow a detailed analysis of specific subcomponents of mathematics. In addition, the age range is limited, and results cannot readily be generalized. Moreover, the aim was to investigate several executive functions and relate these to mathematics, but eventually only updating was significantly related to mathematics. However, we believe that it is also important to report the lack of findings considering inhibition and shifting: the fact that our tasks were selected carefully and confirmatory factor analysis was performed argues that previous studies may have found spurious results. Nevertheless, the possibility remains that at other ages this relationship might be stronger: in younger children because their inhibitory skills are still relatively immature, and in older children because they have already learned more mathematical skills, meaning that there is also more that must potentially be inhibited. This strengthens the need for further longitudinal studies that cover a longer period. Considering the growth models of updating and mathematics, it must be mentioned that although the relationship between the two slopes was estimated to be strong, the confidence interval was large, meaning that the exact size of the relationship is difficult to estimate. In addition, the task loadings on the slope factors were rather low. That implies that the slopes carry only a small part of the total variance, and therefore even a strong relationship explains only a small proportion of the total variance. This might be due to the slope coefficients. A starting point of 0 for the first wave (and subsequent coefficients of 1, 2, 3, …) is conventional, but also suggests that the first wave is a true zero point. This assumption is unlikely to be valid, as children did not enter the study at the start of their development in maths or updating skills. The true zero point is likely to lie further in the past, and better coefficients should therefore be higher, for instance 11, 12, 13, and 14, although their true value is unknown. The current coefficients caused part of the slope variance to be included in the intercept variance. This might also inflate the intercept-intercept coefficient and underestimate the slope-slope effect. In addition, tasks were administered only once at each wave, leading to measurement uncertainty that affects the slope estimates worse than it affects the intercept. The latter effect also serves as a warning against single measurements. Almost invariably, studies investigating executive functions use designs that lack a longitudinal component and also administer each task only once. This approach is likely to underestimate the size of the relationships that are found. Our longitudinal approach at least increased the reliability of the intercept, since this was based on four measurements, not one. We also found a stronger relationship between updating and maths than is found in most studies, which confirms the importance of multiple measurements.

Together, these results partially confirmed our hypotheses. We found no distinction between inhibition and shifting and also no relationship between inhibition/shifting and math in presence of the updating factor. Our results did clearly show a strong and dynamic relationship between mathematics and updating. One of the next questions is how this knowledge can be used to help children with mathematical disabilities or children at risk of developing these disabilities. Our results suggest that there is a mutual, continuing relationship between the two. This may yield training opportunities for children with weak executive function and especially weak updating skills. It is therefore recommended to pursue this type of study with intervention studies. These also form a powerful tool to establish the causality of these relationships. The first, promising steps have already been taken (Holmes et al., 2009; Kroesbergen, Van 't Noordende, Kolkman, & Huiting, 2010): in these studies an intervention targeting updating/working memory skills also significantly improved children's (preparatory) maths skills. Another question that still remains is to what degree the relations between updating and maths are domain-specific. While in this study we made an effort to fractionate different executive functions, the mathematics test that was used did not allow to investigate different components of mathematical skills, enabling only general conclusions. Future studies should try and disentangle specific mathematical skills in relation to executive functions. Another limitation of the mathematics test is that only accuracy scores are obtained. It has been hypothesized that children with strong executive skills use more mature mathematical strategies and apply these skills more flexibly, but this requires closer examination in future research.

5



The development of strategy selection in learning single-digit multiplication and working memory: A microgenetic study

S.H.G. Van der Ven
J. Boom
E.H. Kroesbergen
P.P.M. Leseman

Manuscript submitted for publication.

Abstract

Variability in strategy selection is an important characteristic of learning new skills, such as mathematical skills: strategies gradually come and go during this development. Siegler (1996) described this phenomenon as Overlapping Waves. In this microgenetic study it was attempted to model these overlapping waves statistically. In addition, it was investigated whether development in strategy selection was related to development in accuracy. We also investigated the role of working memory, as we expected that the strategies children select and their success in the execution of the selected strategy are partially determined by working memory: children with poor working memory are limited in their possibilities to make the connections that are necessary to progress to more mature strategies. This would explain the often-found relationship between working memory and mathematical abilities.

To address this aim, the strategy selection and accuracy of 98 children who were learning single-digit multiplication was assessed eight times on a weekly basis. Using latent growth modeling for categorical data, we confirmed Siegler's hypothesis that strategies come and go like overlapping waves during development. In addition, strategy selection and accuracy were found to be strongly interrelated: children who used mature strategies also solved more problems correctly, and progress during the study in strategy selection and accuracy were also interrelated. Finally, working memory predicted both strategy selection and accuracy, confirming that children with poor working memory suffer from a double deficit: they choose immature strategies that are error-prone and make more procedural mistakes carrying out these strategies.

Introduction

A relationship between working memory and mathematical abilities in children has been established in many studies, both in the normally developing population (Bull & Scerif, 2001; De Smedt et al., 2009; Espy et al., 2004; Lee, Ng, Ng, & Lim, 2004; St Clair-Thompson & Gathercole, 2006; Swanson & Kim, 2007; Van der Sluis, De Jong, & Van der Leij, 2007; Van der Ven, Kroesbergen, Boom, & Leseman, 2011b),[1] and in children with mathematical disabilities (Gathercole & Pickering, 2000b; Passolunghi & Siegel, 2004; Swanson & Beebe-Frankenberger, 2004; Wu et al., 2008). The subsequent question, which thus far has received only little attention, is through which mechanisms these relations evolve: what exactly do children do when they learn mathematics, and why does that require working memory? In order to unravel the dynamics of these relations, in-depth studies are needed that go beyond relating working memory scores to the proportion of items answered correctly on a math task. Therefore we carried out a microgenetic study in which we investigated which strategies children employ when solving math problems and

1 Chapter 4 of this dissertation.

how well they execute these strategies. We modeled development in strategy choice and accuracy over time and related working memory to this development. We hypothesized that children's selection of strategies determines their success in mathematics and their speed of mathematics learning, and we expected that working memory partially determines the strategies children are able to execute.

A number of studies have already investigated strategy use in mathematics learning, and it was found that strategy selection is important in determining mathematical performance and that the strategies that are chosen change during the course of development (Imbo & Vandierendonck, 2007; Lemaire, 2010; Lemaire & Siegler, 1995). A certain degree of variability in strategy selection is an important characteristic of the learning process (Flynn & Siegler, 2007): greater initial variability even predicts greater learning (Siegler, 2007).

The Overlapping Waves model, presented by Siegler (1996), poses that solution strategies resemble waves that rise and fall during development: after a certain strategy is discovered, the frequency with which it is used initially increases but decreases later, when the child discovers more advanced strategies that slowly replace the older, less mature strategy. An important feature of the model is that the strategies that are used do not replace each other in an all-or-none fashion: rather, at each time point children have a repertoire of strategies, of which the relative use slowly changes. The model can be applied to mathematical problem solving: initially, typically developing children predominantly rely on simple counting strategies (Geary, 2004). As their mathematical skills improve, they build a knowledge base that enables them to use progressively more advanced strategies. For example, once a child is sufficiently proficient in addition, repeated addition can be applied to solve multiplication problems (3 x 9 = 9 + 9 + 9 = 27). Since this strategy requires fewer steps than counting, it can be considered a more mature strategy, which poses a lower load on working memory. In the end, children rely most on memory retrieval (Lemaire & Siegler, 1995; Mabbott & Bisanz, 2003) and are sometimes able to use memorized answers as a shortcut to problems they haven't memorized yet: a child that knows that 10 x 3 = 30 can use this knowledge to solve 9 x 3, as 9 x 3 = 30 – 3. It should be noted that strategy selection in mathematical problem solving relies heavily on earlier acquired knowledge and is also problem-specific, especially for young children whose exposure to multiplication problems is limited (Mabbott & Bisanz, 2003): a child's ability to solve 3 x 2 by means of retrieval does not automatically imply that this child will solve 8 x 7 by means of retrieval too, as retrieval is only an efficient strategy if the child is confident of the answer, i.e., if there is sufficient associative strength between a specific math problem and its answer (Siegler, 1988).

Initial evidence indicating that the Overlapping Waves model can be applied to multiplication learning was obtained in a study that investigated children's development three times in the year after children were first exposed to multiplication problems

(Lemaire & Siegler, 1995). During this year, the frequency of strategy repeated addition decreased while the frequency of retrieval increased during the year, consistent with the hypothesis of rising and falling waves. Nevertheless, at each measurement, children used a mixture of strategies at the individual level, confirming the overlap in these waves.

There are also interindividual differences in strategy use, which are reflected in mathematical abilities. Children with mathematical difficulties generally have the same strategy repertoire as typically developing children, but tend to use immature strategies more often: they persist in relying heavily on counting strategies, while also making more procedural errors than normal achievers (for a review, see Geary, 2004). The question arises why children with mathematical disabilities use more immature strategies than typically developing children. One important factor may be working memory: the ability to store and process information temporarily (Baddeley, 1992; Baddeley & Hitch, 1974). Mathematical problem solving often requires multiple steps, such as multiple additions when repeated addition is used to solve a multiplication problem. The child must keep track of the steps that have been taken and the steps that still must be carried out. In addition, the intermediate outcomes of these steps must be remembered. Both procedures likely require working memory. Empirical evidence supports this theory. A causal relation between working memory and mathematics was found in a study in which children solved simple addition problems, both with and without a secondary task that posed an extra working memory load. This extra working memory load decreased the solution efficiency of addition problems for all types of strategies used, but it did not change the selection of the strategy itself (Imbo & Vandierendonck, 2007). Another study showed that increasing the working memory load of a math problem by adding an illustration worsened math performance in children (Berends & Van Lieshout, 2009). This decrement in performance was even worse in children with poor working memory. This suggests that working memory is needed during mathematical problem solving, but the degree to which it is needed depends on the strategy that is chosen to solve the problem: more mature strategies require fewer working memory resources (Imbo & Vandierendonck, 2007). Especially the less mature counting strategy poses a high load on working memory (Cragg & Gilmore, 2011; Hecht, 2002): counting involves a large number of steps to keep track of, especially when the operands are large. Children with poor working memory are therefore expected to have problems carrying out these counting strategies, particularly on math problems with large operands. Indeed, low working memory scores in young primary school children are associated with more counting errors when counting strategies are chosen to solve simple addition problems (Geary, Brown, & Samaranayake, 1991). Solving math problems fast and accurately is a prerequisite to form strong neural connections between the problem and the correct answer (Geary et al., 1991; Siegler, 1996). Children with poor working memory are more likely to have forgotten the operands once the problem is solved, preventing simultaneous activation of the operands and the answer. They are also likely to make more procedural mistakes, leading to incorrect associations between operands and answer (Geary, 1993; Geary et al., 1991; Noël, Seron, & Trovarelli, 2004), and are therefore expected to lag behind in the formation of strong connections between operands and answer, which prevents them from using retrieval-based strategies (Siegler, 1988), even though these more mature strategies pose a much lower load on working memory than counting does (Hecht, 2002).

Empirical evidence also supports the hypothesis that children with poor working memory use less mature strategies. In second to fourth grade, children with high working memory used the retrieval strategy more often than children with lower working memory when solving addition problems (Barouillet & Lépine, 2005; Wu et al., 2008). In first and third grade, high working memory skills were also associated with less finger counting, and in third and fifth grade with more frequent use of decomposition strategies (Geary, Hoard, Byrd-Craven, & DeSoto, 2004; Noël et al., 2004). No relationship between working memory and retrieval was found in the higher grades (Geary et al., 2004), confirming that working memory is most important during the early stages of the math learning process, as the counting strategies that are used in the early stages of the learning process pose a high load on working memory resources. Success in these early stages likely enables transitions to more mature strategies that require fewer steps and are therefore faster and less error-prone and that eventually enable relying on retrieval, which barely requires working memory resources. This was confirmed in other studies (Imbo & Vandierendonck, 2007; but see Wu et al., 2008).

Some questions still remain. Most studies described above included only one measurement and often the problems that were used were relatively easy for the age range studied. This makes it difficult to investigate the initial steps in learning mathematical skills, the period in which, as described above, working memory is presumably most important (Geary et al., 2004; Imbo & Vandierendonck, 2007). What we therefore need is an integrative study in which the dynamics of development in strategy selection and accuracy are modeled together and related to working memory. With the present study we attempted to provide such a view of one mathematical skill, namely single-digit multiplication in second grade (7-8 years). With a microgenetic study, in which data are collected intensively during a relatively short period of development, we aimed to unravel the dynamics of development in strategy selection and accuracy of problem solving, and the relation of working memory with both.

Our first aim was to confirm the validity of research question related to the Overlapping Waves model by Siegler (1996), which states that the frequencies of mathematical solution strategies increase and decrease as the child becomes more proficient. Thus far, this model has been used primarily metaphorically. In this study we tried to validate this model statistically by creating a categorical growth model that describes the shapes of these overlapping waves as a function of increasing mathematical ability of the child. This approach allowed a more detailed, yet quantitative approach of the analysis of the development of strategy use than has hitherto been used, something greatly needed in microgenetic research (Cheshire, Muldoon, Francis, Lewis, & Ball, 2007).

The second question considered the role of working memory. We investigated whether individual differences in working memory were related to the maturity of the selected strategies and to development in this maturity, and whether working memory was also related to development in accuracy. We expected that children with high working memory scores would be capable of a fast formation of problem-answer associations and would therefore show a steep learning curve, resulting in faster use of more mature strategies such as retrieval. In addition, we expected that regardless of the strategy that was chosen, children with high working memory would make fewer procedural mistakes and therefore obtain higher accuracy scores. In other words, we expected working memory to be related to the development of both strategy selection and strategy execution.

Method

Participants

A total of 98 Dutch second graders (52 boys, 46 girls), from seven primary schools in diverse neighborhoods (predominantly lower and middle class) participated. Mean age was 7;9 years (*SD* = 4 months). Parental consent was obtained for all children. A total of 79 children participated in all eight sessions, 15 missed one session and 4 were absent twice, mainly due to illness.

Procedure

Once a week, during a period of eight weeks, always on the same weekday, the children were given a math test that was administered individually in a separate room by a trained research assistant (mostly graduate students). When a problem was presented in a story, the research assistant read the story out loud while the child could also read the text. The child was encouraged to use paper and pencil to solve the problem. However, the majority of children did not do this. Therefore directly after each problem the research assistant asked the child how (s)he had solved the problem. Children of this age are capable of reporting this (Wu et al., 2008). The procedure lasted approximately 20 minutes per child per session.

Measurements

Multiplication

We administered fifteen math problems with a large difficulty range (from 3 x 2 to 6 x 8), which enabled us to select the math problems that were most sensitive to the children's mathematical development during the study, i.e., the problems that did not show floor or ceiling effects, for the final analysis. The easiest items also served as positive fillers, to keep the children, especially those with low mathematical abilities, motivated.

The items varied per session: they were drawn from a pool of 31 items that were grouped in fifteen subsets of increasing difficulty level based, on previous research (Klinkenberg, Straatemeier, & Van der Maas, 2010). Each week, one item from each subset was selected randomly in order to avoid test-specific learning effects but at the same time maintaining approximately equal difficulty of the test throughout the eight weeks. This yielded a total of fifteen problems per week. The problems that were selected are presented in Table 5.1.

Eight problems were always presented as bare problems; the remaining seven problems were presented in a context consisting of a short story with an illustration. The story was always presented in three sentences, the first always containing the first factor, the second sentence containing the second factor and the third sentence consisting of the multiplication question. In the accompanying illustration both factors

Table 5.1 The math problems used in each session

Math problem	Context	W1	W2	W3	W4	W5	W6	W7	W8
1	No	7 x 2	**5 x 3**	**5 x 3**	7 x 2	**5 x 3**	7 x 2	7 x 2	7 x 2
2	No	2 x 4	2 x 4	2 x 4	2 x 4	2 x 4	2 x 4	2 x 4	2 x 4
3	Yes	4 x 2	**3 x 2**	**3 x 2**	**3 x 2**	4 x 2	4 x 2	**3 x 2**	4 x 2
4	Yes	8 x 2	8 x 2	8 x 2	8 x 2	8 x 2	8 x 2	**9 x 2**	8 x 2
5	No	2 x 8	2 x 8	2 x 8	**2 x 7**	2 x 8	**2 x 7**	*9 x 2*	2 x 7
6	Yes	4 x 4	**5 x 4**	4 x 4	**5 x 4**	3 x 7	4 x 4	4 x 4	4 x 4
7	No	7 x 3	7 x 3	7 x 3	7 x 3	7 x 3	7 x 3	7 x 3	7 x 3
8	Yes	3 x 4	3 x 4	**6 x 5**	**6 x 5**	**6 x 5**	**6 x 5**	3 x 4	3 x 4
9	No	9 x 5	**6 x 6**	9 x 5	9 x 5	9 x 5	*8 x 5*	*8 x 5*	**6 x 6**
10	Yes	9 x 3	9 x 3	9 x 3	9 x 3	9 x 3	9 x 3	9 x 3	9 x 3
11	No	3 x 8	3 x 8	3 x 8	3 x 8	**5 x 7**	**5 x 7**	**5 x 7**	**5 x 7**
12	Yes	8 x 3	**4 x 6**	**4 x 6**	**4 x 6**	**4 x 6**	**4 x 6**	8 x 3	8 x 3
13	No	4 x 9	**9 x 4**	4 x 9	**9 x 4**	4 x 9	**9 x 4**	**9 x 4**	4 x 9
14	Yes	8 x 6	8 x 6	**6 x 9**	**6 x 9**	**6 x 9**	8 x 6	8 x 6	8 x 6
15	No	4 x 8	4 x 8	4 x 8	**6 x 8**	**6 x 8**	**6 x 8**	4 x 8	**6 x 8**

Note. W1-W8 = week 1 – week 8. To aid visibility, the first-appearing unique problem of each subset is presented in default font, the second in boldface and the third in italics.

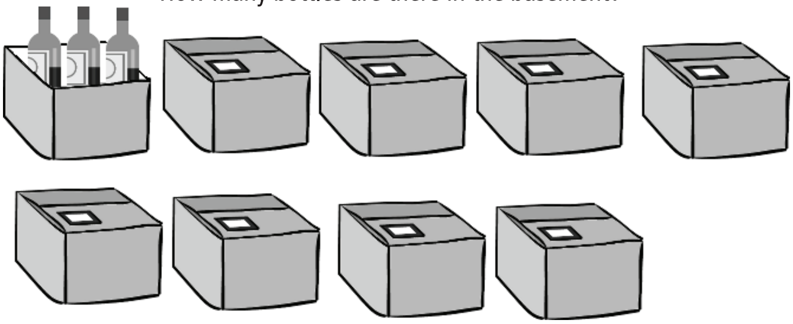


Figure 5.1 Example math item.

could also be obtained by counting. However, part of the problem was always occluded, so the answer could not be obtained by mere one-to-one counting. An example item can be found in Figure 5.1. The context was different in each week. For each week the order of the problems was randomized once: all problems were presented to each child in the same order. The first two problems were always selected from the easiest five of the session. Problems were presented in a booklet. Each page contained two problems; the problem that the child was not working on was occluded by the research assistant. The two problems were not allowed to share one of the factors.

Working memory

Digit Span Backwards. An adaptation of the Dutch version of Digit Span Backwards from the Automated Working Memory Assessment test battery (Alloway, 2007) was administered. The child heard a digit sequence and had to repeat it backwards, starting with a two-digit sequence. After three correct answers, the sequence increased by one digit. When two mistakes were made in trials of the same length, the task was discontinued. The number of correct sequences was used as a final score. The observed range was 2 – 10 correct trials.

Odd One Out. This task was also adapted from the Automated Working Memory Assessment (Alloway, 2007). A series of stimuli was shown consecutively. Each stimulus consisted of three boxes with shapes presented next to each other. One of the shapes differed from the other two. The child was asked to point at the deviant shape. Then the next stimulus was presented. At the end of each trial three empty boxes appeared; the child had to point at the locations of the previously shown deviant shapes in the same order in which they had appeared. An answer was correct if each location was recalled correctly in the right order. The task started with only one item; after three correct answers of the same length, the sequence increased by one. When two mistakes were made in trials of the same length, the task was discontinued. The number of correct sequences was used as a final score. The observed range was 3 – 16 correct trials.

Keep Track. A computerized version of the Keep Track task was created. This task was adapted from Van der Sluis et al. (2007). The child was shown pictures, each of which belonged to one of the following five categories: sky (sun, moon, stars, cloud), fruit (strawberry, pear, cherry, banana), shapes (square, triangle, circle, heart), animals (dog, cat, fish, bird), and toys (teddy bear, scooter, 'LEGO®', car). The pictures were shown in series of ten; each picture was displayed for 3.5 seconds. The child was asked beforehand to pay special attention to one or more designated categories. During the series, the child had to name each picture. At the end, the child had to recall the last item of the designated categories. In each series, one, two, or three items of each designated category were presented. The remainder consisted of filler items from other categories. During the series, small pictures symbolizing the to-be-remembered categories were shown in the bottom of the screen, serving as a reminder. The number of to-be-remembered categories increased from one to four. There were two series of each difficulty level, yielding a total of eight series. Each correct answer was noted. Depending on the length of the sequence, there could thus be a maximum of one, two, three, or four correct answers in a trial, yielding a maximum possible score of twenty correct responses. The total number of correct responses was recorded. Prior to testing, the child was familiarized with the pictures and the categories. The test was preceded by a practice item that was repeated when necessary. The observed range was 9 – 20 correct items.

Coding and reliability

The solution strategies that were reported by the children were written down by the research assistants. These verbatim reports were coded afterwards by four trained raters: a total of 35 different types of strategies and hybrid strategies were noted. These were further reduced to five strategy categories that were highly similar to those used by Mabbot and Bisanz (2003), in order of maturity: (1) incorrect strategies, (2) counting strategies, (3) repeated addition, (4) derived facts, and (5) retrieval. Incorrect strategies are conceptually wrong; examples are repeating a factor (7 x 4 = 7), addition (7 x 4 = 11), guessing, and stating 'I don't know'. Counting strategies included finger counting, making a drawing and counting the items and counting out loud. Repeated addition involved changing the multiplication problem into an addition problem (5 x 7 = 7 + 7 + 7 + 7 + 7), sometimes in smaller steps (3 x 7 = 5 + 5 + 5 + 2 + 2 + 2), and the derived facts category consisted of all strategies in which the child used a 'shortcut', requiring fewer steps than repeated addition; examples are doubling (for 6 x 4: 4 + 4 + 4 = 12; 12 + 12 = 24), or using the easier 5 or 10 times tables (9 x 3 = 10 x 3 – 3). Retrieval means that the child retrieved the answer from longterm memory without reporting or overtly showing a computational procedure. Whenever a child used a strategy containing elements from more than one category (e.g., solving 6 x 7 by stating 5 x 7 = 35, then counting fingers from 36 to 42), the item was allocated to the most mature category. This happened to 3.3% of all data. When a child could not report which strategy had been used, or clearly reported something else (e.g., 4 x 8 = 10 + 10 + 10 + 2 = 32), the strategy was coded as missing. This happened to 0.7% of the data. A subset of 556 problems (4.9%) was coded by two independent raters: Cohen's κ was .95 for the five strategy categories.

Data analysis

The data are complex and also required complex statistical analyses. In order to test the first hypothesis that development can be modeled statistically as overlapping waves, we used an integration of IRT modeling for categorical data and growth modeling, using Mplus version 6 (Muthén & Muthén, 1998-2010, see example 6.4).

In this model it is assumed that the five strategies can be ordered in terms of maturity, such that the strategies form an ordinal scale representing increasing maturity. The model also assumes the existence of an underlying continuous latent ability predicting the probability of use for each strategy. The higher the latent strategy ability, the higher the probability that relatively mature strategies would be selected over immature ones; while the lower the latent strategy ability, the higher the probability that immature strategies were selected over mature ones. The result is a latent strategy maturity scale. The use of the various strategies is nonlinearly and probabilistically related to the position along this strategy scale. The scale (see Figure 5.4 as an illustration) can be used to represent inter-individual differences, but also to represent intra-individual development. The scale was constrained to be the same on all eight measurement occasions (measurement invariance) but the position of participants along the scale (x-axis) is hypothesized to shift towards higher ability (to the right) over the eight weeks. This development or movement along strategy ability dimension over the eight weeks was modeled by a linear latent growth model. A linear growth curve can be described by an intercept and a slope. The intercept reflects the overall or the general ability throughout the study. The mean of the intercept is the estimated average ability of the group, while the variance captures individual differences in this ability. The slope reflects the growth in ability during the course of the eight weeks of the study and also has a mean and variance. In our particular model the intercept reflects average and individual differences in latent strategy ability, i.e., the position along the x-axis in Figure 5.4. Since in this model time was centered around zero (i.e., t = 0 halfway the study, between week 4 and 5), the intercept can be interpreted as the ability halfway the study. The slope reflects (individual differences in) the degree to which participants move to the right along the x-axis in Figure 5.4, and this model was scaled such that the slope can be interpreted as the improvement throughout the entire 8 weeks.

Differences in difficulty level between the problems were accounted for in the model by allowing the strategy probabilities to vary between problems, thus accommodating our hypothesis that regardless of the latent strategy ability, more mature strategies would be applied to the easier problems (Mabbott & Bisanz, 2003). The problems within a subset (e.g., 7 x 2 and 5 x 3 for problem 1) were allowed to differ.

In order to address the second research question, the investigation of relations between working memory and strategy selection and problem solving accuracy, first a growth model was made for accuracy, similar to the model for strategy selection. In this model, the latent accuracy ability scale indicates the probability that a child produced the correct answer to a certain problem, given a certain ability level: the higher the accuracy ability, the higher the likelihood that the child produced the correct answer. This model also had an intercept and a slope, and again, the resulting curves were allowed to differ between problems, in order to accommodate the hypothesis that easier problems were solved correctly more often than more difficult problems. The intercepts and slopes of the strategy and accuracy growth models were then related to each other, and working memory and gender were added as time invariant predictors of these intercepts and slopes.

The fit indices of a model indicate the validity of the model: in our case that also means that the fit of the model indicates how well development can be modeled as overlapping waves. However, given eight measurement occasions, many different problems, five strategy categories and two accuracy categories, the model complexity was enormous and the usual fit indices that can be used to judge the validity of a model were therefore not always available. Categorical growth models can be estimated using the Weighted Least Squares (WLS) or the Maximum Likelihood (ML) estimator in Mplus (Muthén, 1998-2004, Appendix 1). WLS provides absolute fit indices but could only be used for the separate growth models (accuracy and strategy use), as it could not handle the complexity of the model in which these were related to each other. ML could handle the complete model but provides only relative fit indices (e.g. AIC) for models as complex as at issue here. This final analysis was therefore performed using ML. Note, however, that the separate growth models were also estimated with ML, and results were comparable to the WLS results.

When WLS estimation was possible, the fit indices are reported as recommended by Blunch (2008): χ^2 (chi square) with its p-value, CFI, RMSEA and AIC. χ^2 should be as low as possible and its p-value should be non-significant. However, for complex models p is almost always significant; another guideline is the NC, or the ratio of χ^2/df, which should not exceed 2 (Kline, 2005). CFI > .95 can be considered a good fit, and > .90 an acceptable fit. RMSEA is good if < .05 and acceptable if < .08. AIC, finally, cannot be used as an absolute fit index but is used to compare models: the lowest AIC value is preferred.

In categorical growth modeling, a link must be used that describes the shapes of the curves. Mplus supports the use of probit and logit links for ML estimation but only probit for WLS estimation (Agresti, 2002; Muthén & Muthén, 1998-2010). Therefore the probit link, which bases the shapes of the curves on the Gaussian curve, was used throughout, but note that when possible, analyses were also performed with a logit link, resulting in similar Item Characteristic Curves (given a scaling difference) and virtually identical relations between the different variables.

Results

Since each math problem was allowed to have different parameters (necessary to accommodate differences in difficulty) the model complexity increased exponentially with each math problem that was added to the set. The computational limits arising from this high model complexity allowed a maximum of eight problems to be analyzed together. The first five problems and problem eight appeared to be too easy, with too little variance in terms of strategy use and too few errors (ceiling effects). Moreover, for technical reasons the categorical growth model has difficulties with a strategy that is not used at all in a particular week. This affected two more multiplication problems (6 and 12), but by imputing three missing data points (0.05% of the data) these could be retained; in all three cases an 'incorrect strategy' was imputed for a child with missing data that had applied the incorrect strategy to this problem on at least one other occasion. Finally, problem 14 was also removed, as there appeared to be peculiarities with this problem that were likely due to some context stories that some children failed to understand. Problems 6, 7, 9-13, and 15 were therefore used for further analyses, which yielded sufficient power and variation in difficulty of the problems. The group means of the frequencies of each strategy type are shown in Figure 5.2; both averaged over all eight selected problems and, as an illustration, for two individual problems. Figure 5.2 shows that overall the use of retrieval increased while the use of the other strategies decreased, but there are differences between the problems that can be attributed to the difficulty of each problem: e.g., retrieval was applied less to the more difficult problem. A similar result was obtained for accuracy, expressed as the proportion of correct answers to these problems, as is shown in Figure 5.3: in general there was an increase in accuracy, and the easier problem was answered correctly more often than the more difficult problem. As expected, the more mature the strategy, the higher the likelihood of a correct answer: the proportion of correct answers for the incorrect strategy was .02, for counting it was .55, for repeated addition .62, derived facts .68, and for retrieval .85.

The graphs represent group data and therefore do not contain sufficient information considering individual differences in development. A horizontal line for a certain strategy does not imply stability in the use of this strategy at the individual level: its frequency might have decreased in some children while increasing in others. In order to account for these individual differences, growth models were created, as explained in the previous section.

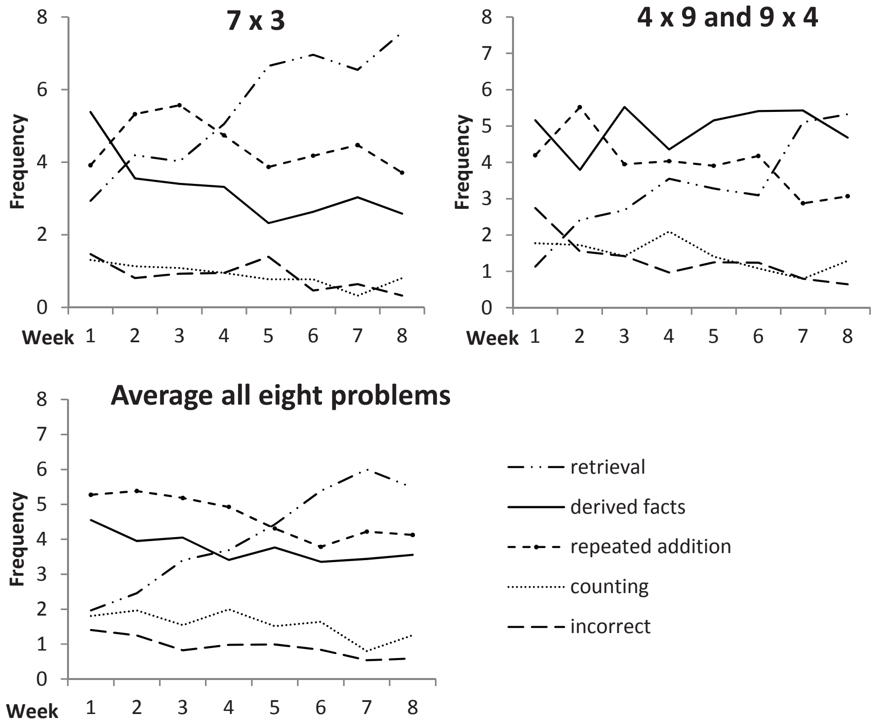


Figure 5.2 Descriptive strategy data: observed average frequencies of selected strategies, averaged for all eight problems and, as an illustration, for two example problems.

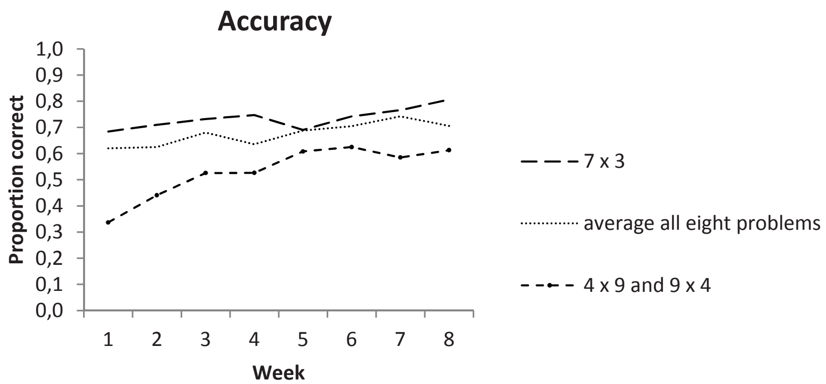


Figure 5.3 Descriptive accuracy data: proportion of problems answered correctly over time, averaged for all eight problems and, as an illustration, for two example problems.

To answer the first research question, whether development can be modeled like overlapping waves, a growth model for strategy selection was built and estimated using WLS. The fit of this model was good, considering its complexity, $\chi^2(2151) = 2937.42$, $p < .001$, NC = 1.37, CFI = .90, RMSEA = .06. Figure 5.4 shows the predicted probabilities of each strategy as a function of the latent strategy ability for two of the math problems.

The Figure clearly shows a pattern of overlapping waves. Since the model was based on all eight problems, both figures represent the same ability (i.e., a child's score on the x-axis is the same in both figures), but, as can be seen in Figure 5.4, the resulting curves are problem-specific. For example, a child with an ability of 0.5 has a probability of around .5 to use retrieval to solve 9 x 3, but this probability is only around .25 for solving 4 x 6; for this problem the child with ability 0.5 is much more likely to use the derived facts

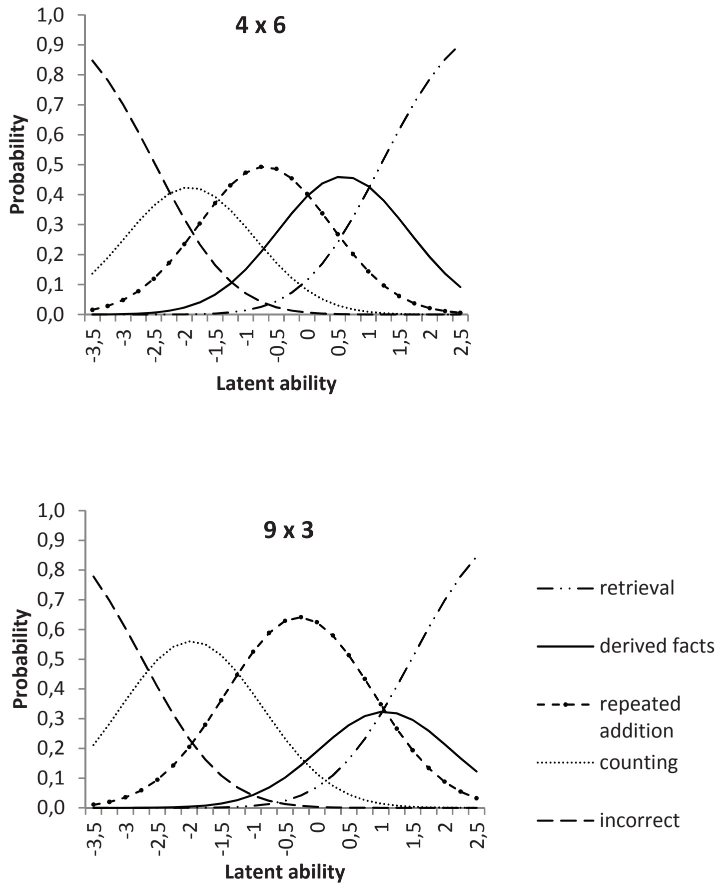


Figure 5.4 Modeled strategy probability curves for two math problems: 4 x 6 and, slightly more difficult, 9 x 3.

strategy, with a probability of approximately .5. The order of the curves is the same for each problem, with incorrect strategies being the leftmost and retrieval the rightmost curves. The curves do differ in location: for easier problems curves are shifted to the left, indicating that for each ability level more mature strategies were applied to these problems. Curves also differ in height, reflecting that regardless of the latent strategy ability, certain strategies were applied more to some problems than to others: e.g., the curve for derived facts is much lower for 9 x 3 than for 4 x 6, showing that this strategy was more likely to be applied to 4 x 6.

The mean intercept of latent strategy ability was set to 0 (*SD* = 1.02): this means that, given that we centered time around 0, halfway the study the average child had a strategy ability score of 0. The mean slope was 0.97 (*SD* = 0.90), so during the course of the study the average child progressed 0.97 units to the right on the latent strategy ability scale.

In order to answer the second research question, regarding the role of working memory in both strategy selection and accuracy, a similar growth model was created to fit the accuracy data. The fit of this model was also good, $\chi^2(1998) = 2071.75$, $p = .12$, NC = 1.04, CFI = .96, RMSEA = .02. The latent accuracy ability scale predicts the probabilities of obtaining correct answers to problems of varying difficulties. Three example curves are shown in Figure 5.5. The curves show that the higher the latent accuracy ability, the larger the probability is that a correct answer was obtained. The Figure illustrates that 4 x 6 was an easier problem than 6 x 8, as the former curve is shifted further towards the left: 4 x 6 was more likely to be solved correctly for each possible ability level. The mean ability was again set to 0 (*SD* = .44) : halfway the study the average child had an accuracy ability score of 0. The slope was .28 (*SD* = .45): the average child shifted .28 units to the right during the course of the study.

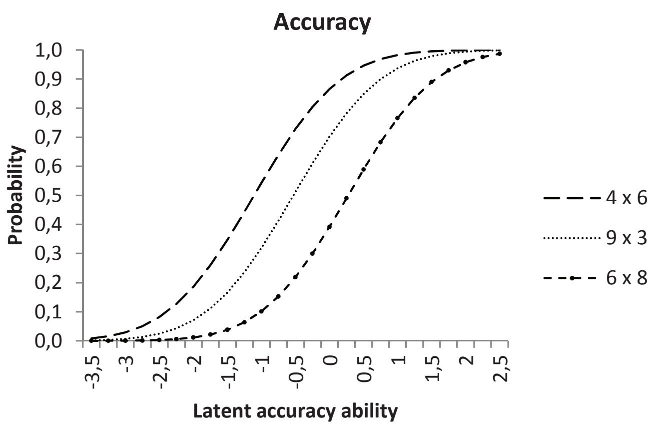


Figure 5.5 Modeled accuracy probability curves for three different math problems. The more difficult the problem, the more the curve is shifted towards the right.

In the final step, the strategy and accuracy growth models were combined in one model in which the intercepts and slopes of both models were allowed to covary. In addition, a latent working memory factor, based on the three working memory tasks, was added as a predictor. Gender was also entered as a control variable, as it improved the fit of the model as reflected by the AIC. This model proved too complex for WLS and was therefore estimated using ML, so absolute fit indices were not provided. However, as described before, the fit indices of the two separate growth models were reasonable to good. Other indicators also suggest that the final model is valid: the model estimation process converged and significant relationships were obtained as described in more detail below, indicating that confidence intervals of these parameters could be estimated with sufficient certainty. The results of the final analysis are presented in Figure 5.6.

The Figure shows that the intercepts of both models were strongly related to each other: children who chose more mature strategies also had high accuracy scores. The two slopes were also significantly related to each other, meaning that children who improved more in strategy selection also improved more in accuracy. There was also a significant positive relationship between the slope of strategy selection and the intercept of accuracy: children with a high ability in accuracy improved more in strategy selection. Working memory was significantly related to both the strategy selection intercept and the accuracy intercept: children with good working memory skills had higher general

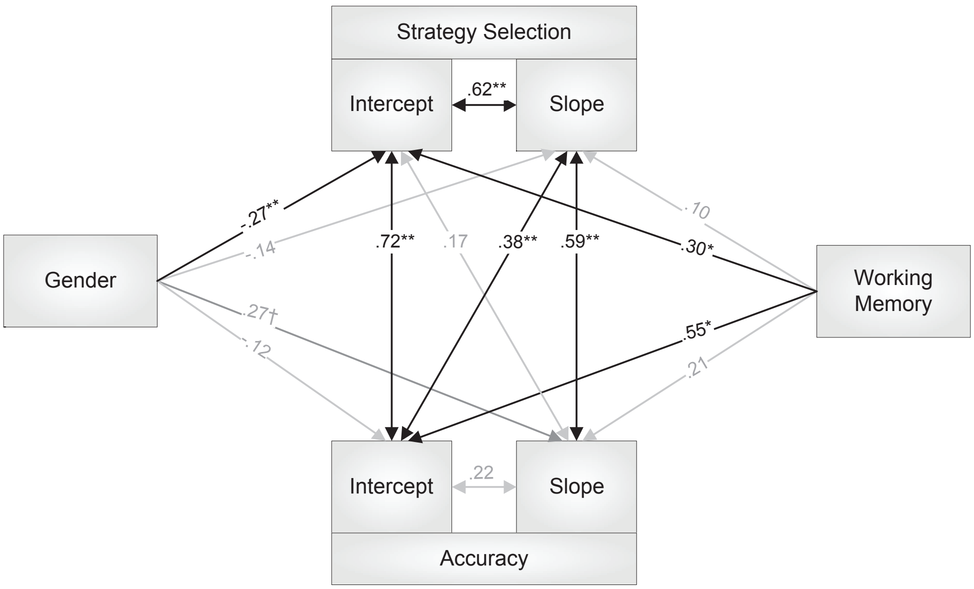


Figure 5.6 Final model: two connected growth models with two predictors. Black arrows represent significant relations, marginally significant relations are dark gray and non-significant relations are light gray. † $p = .06$, * $p < .05$, ** $p < .01$.

abilities in both domains. The size of the estimated relation between working memory and the slope of accuracy was also notable (β = .24) but because of the large confidence interval for this estimate this was not significant. Gender, finally, showed a somewhat contradictory pattern: boys had a higher strategy selection intercept than girls, but girls had a marginally significant higher accuracy slope, indicating that boys in general chose more mature strategies than girls, but girls improved more in accuracy. Note that the correlation between intercept and slope, especially of the strategy selection model, is also rather large: children with a high ability in strategy selection also improved more. That also means that although there is no significant direct effect between working memory and the strategy selection slope, there is still an indirect effect through the intercept (and also the intercept of accuracy), making this total effect substantial.

Discussion

In the present study, two main questions were addressed. First, it was investigated whether strategy development in children learning single digit multiplication could be modeled statistically according to the Overlapping Waves model by Siegler (1996). Secondly, we addressed the relation between working memory and mathematical strategy selection and problem solving accuracy.

In order to answer the first question, we created a statistical model that describes the development of strategy use in children learning single-digit multiplication. This model confirmed the validity of the Overlapping Waves model. As children became more proficient in single-digit multiplication, they moved from incorrect strategies through counting strategies, repeated addition, and derived facts, towards the retrieval strategy. This did not occur in an all-or-none fashion; rather, at each time point children had a repertoire of strategies at their disposal, as predicted by Siegler (1996, 2007). As the children's mathematical ability increased, this repertoire became dominated by increasingly more mature strategies.

The model that was created required two rather strong assumptions: (1) that the strategies can be ordered in terms of maturity, while this order and the relative distance between the curves does not change during development, and (2) that there is a single latent ability underlying strategy selection. The resulting model had reasonably good fit indices, indicating that these assumptions are plausible. The results of this model can be promising in the future, as a manner of quantifying strategy choices. This enables the use of more advanced statistical techniques than often used in previous studies, which often treated different strategies as separate variables and analyzed them separately. Strategies, however, are not independent, as they are multiple mutually exclusive possible outcomes of a single variable: when one strategy is chosen, another strategy is not, so an increase in the use of one strategy necessarily implies a decrease in at least one other strategy.

Our analyses acknowledged this fact, by treating strategies as having probabilities that rise from a latent ability. The analyses also allowed the creation of latent growth models, thus enabling the analysis of development.

Moreover, the analyses allowed an investigation of our second research question: the relationship of working memory with both the development of strategy selection and accuracy. Although relationships have been found between working memory and strategy selection (e.g., Barouillet & Lépine, 2005; Geary et al., 2004; Wu et al., 2008), and between working memory and accuracy (e.g., Agostino, Johnson, & Pascual-Leone, 2010; Lee et al., 2004; Swanson & Kim, 2007; Van der Ven et al., 2011b), we incorporated both analyses simultaneously and longitudinally by means of a microgenetic design. We found indications for a twofold role of working memory, which was significantly related to both strategy choice and accuracy. This shows that children with high working memory skills are more likely to choose mature strategies that more often result in a correct answer, but even when strategy choice is taken into account, their accuracy is still higher. Working memory was also related to improvement in both accuracy, and, albeit indirectly, strategy selection. These results give rise to the paradoxical situation that children with poor working memory predominantly use strategies that pose a high load on working memory. They suffer from a double deficit, using immature strategies that require many steps and working memory resources and are therefore error-prone, and also making more mistakes executing these strategies than children with higher working memory. It also means that children with high working memory develop mathematical skills at a faster pace than their peers with poorer working memory, who will likely need help in both selecting appropriate strategies and carrying these out without making procedural errors. On the other hand, a successful working memory intervention may yield improvements in both strategy selection and accuracy.

The small effect of gender that was found is consistent with previous findings that at a young age there are no gender differences in accuracy in mathematics (Carr & Jessup, 1997; Klein, Adi-Japha, & Hakak-Benizri, 2010), but girls rely more on counting strategies while boys prefer to use retrieval (Carr & Jessup, 1997). The marginally significant finding that girls improved more in accuracy than boys may suggest that boys try to use mature strategies too soon, before having made the necessary connections between problem and answer, but this hypothesis awaits further confirmation.

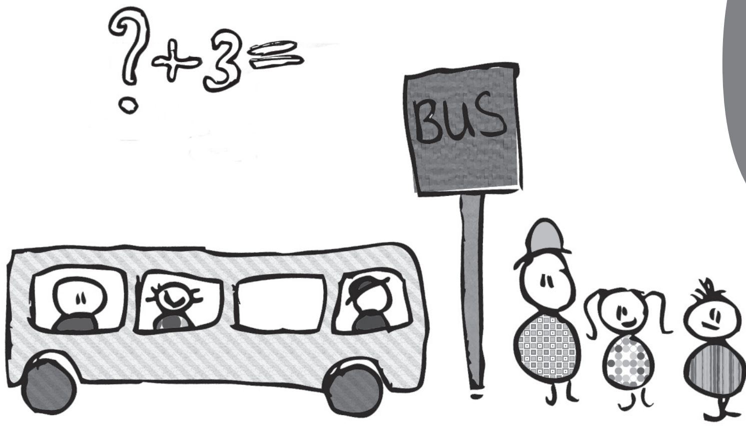
Another question that is still open is whether children with low working memory obtained lower accuracy scores than necessary because their strategy choice was not adaptive: possibly they would have performed better, had they chosen different, more mature strategies. Indeed, adults have been shown to choose less adaptive strategies when they have fewer working memory resources available because of a secondary task (Imbo, Duverne, & Lemaire, 2007; but see Imbo & Vandierendonck, 2007). Nevertheless, it is unlikely that this accounts for all our findings. Even when correcting for strategy choice, there was still a strong direct connection between working memory and accuracy, which indicates that regardless of the strategy they chose, children with low working memory made more mistakes. In addition, there is evidence that even young children are capable of choosing their strategies adaptively: they have a strong tendency towards choosing the most mature strategy they can handle. This also holds for children with mathematical difficulties, albeit less so than their normally developing peers (Torbeyns, Verschaffel, & Ghesquière, 2002, 2004). It is therefore likely that choosing more mature strategies would have led to even more mistakes in these children. Nevertheless, intervention studies are desirable to establish the causality of the relationships between working memory, strategy choice and accuracy.

Future studies may also show how exactly the use of one strategy influences future strategy choices. It will also be interesting to investigate whether all children progress through the ability scale in a similar way, only differing in speed with which they make this progress. The fit of the model was high enough to suggest that for most children this assumption holds, but there might be a minority of children that develop differently. Studies targeting different populations, such as children with mathematical disabilities, are needed to answer this question. In addition, the Overlapping Waves model needs confirmation in other mathematical skills, such as visuospatial skills or other arithmetical skills such as addition.

A limitation of the study was that the analyses were constrained by technological limits. Even with a relatively small number of different categories for strategy use (though larger than in similar studies), the complexity of the model met the boundaries of computational possibilities, as each problem that was added to the model increased the complexity exponentially. Therefore, absolute fit indices could sometimes not be provided. Nevertheless, the number of problems that were analyzed together, eight problems for each week, was already large, especially given the fact that the data were not aggregated over these problems, but instead, item-characteristic curves were obtained. In addition, by administering even more problems, we were able to select eight developmentally sensitive problems, so all children showed clear development during the study: this prevented the incidence of ceiling and/or floor effects in children in both extremes of mathematical abilities.

Moreover, we made a contribution to the existing body of analytical techniques in microgenetic research: by applying latent growth modeling to categorical data. This is an existing technique that relies on proven statistical tools, latent growth modeling and IRT, but to our knowledge this has not been done before in microgenetic studies, even though it is a powerful tool for analyzing microgenetic development, as it allows an integrative quantitative analysis of development in strategy use. With this method we were able to model strategy selection over time and, in addition, we showed that working memory in children was related to both strategy selection and accuracy.

6



Summary and general discussion

Introduction

The ability to plan ahead and adjust behavior flexibly, referred to as executive functioning, is an important ability to function well in everyday life. It has therefore received increasing attention in the past decades. This dissertation was aimed to contribute to this ever-growing body of knowledge.

One important aim of the dissertation was to scrutinize the structure of executive functioning. It has been claimed that there are three main executive functions (EF): inhibition, or the ability to suppress a dominant response, shifting, which is the ability to switch flexibly between rules, and updating, the ability to store information in working memory and revise and update this when relevant new information enters. The latter concept is sometimes referred to as working memory. Serious doubts, however, have arisen concerning the validity of this distinction. It is difficult to measure executive functions, and both the selection of tasks and the ways in which scores are derived from these tasks may strongly influence the results of a study. In this dissertation, the distinction in three executive functions has been investigated with carefully selected tasks and scoring methods, and it was found that the distinction in inhibition, shifting, and updating may no longer be tenable, at least in children aged 6-8 years. An alternative approach is therefore proposed in this discussion.

Because EF are responsible for flexible and adaptive behavior, they are often considered as predictive of and responsible for other skills for which this flexibility is important. Learning academic skills, such as learning mathematics, are one example: since a young child has not acquired many routines for solving problems yet, it is not sufficient to rely on routine behavior. Instead, new problem solving strategies must be learned and applied when appropriate. Indeed, relations between EF and mathematics have often been found. Nevertheless, although EF are presumably most important during the learning process, longitudinal studies, necessary to capture this learning process are scarce. In this dissertation growth of both executive functions and mathematics was investigated and it was found that the development of especially updating was strongly related to the development of mathematical skills.

In the final study, the nature of this relationship between updating and mathematics was investigated further. It was assumed that strong updating abilities enable the formation of a richer network of mathematical knowledge, which enables the use of more sophisticated mathematical strategies that build on these networks. Moreover, strong updating abilities were expected to lead to fewer procedural errors. A microgenetic study, in which the development of a skill is investigated with dense, in-depth observations, was therefore carried out. Children's single-digit multiplication skills were assessed eight times on a weekly basis. Both the strategy children used and the accuracy of their answer were analyzed, and as expected, it was found that updating was related to both aspects of mathematics.

These main findings are addressed in more detail in this discussion. Suggestions for future research and practical implications are also discussed.

Structure of executive functions

In this dissertation the structure of executive functions was analyzed. It was investigated whether a proposed division in three separate abilities, namely inhibition, shifting, and updating, could be confirmed statistically. This distinction was based on a suggestion by Baddeley (1996) and, using factor analysis, has already been confirmed in some studies (Espy et al., 2004; Fisk & Sharp, 2004; Hughes, 1998; Miyake et al., 2000). Other studies, however, did not find evidence in favor of this distinction, especially studies investigating children (Huizinga, Dolan, & Van der Molen, 2006; St Clair-Thompson & Gathercole, 2006; Van der Sluis, De Jong, & Van der Leij, 2007; Wiebe, Espy, & Charak, 2008). These results might be a reflection of the strong development of executive functions during childhood (Huizinga et al., 2006), but another reason may be the impurity problem: the fact that executive functions influence other cognitive skills and therefore cannot be measured in isolation. This problem can be addressed with factor analysis, extracting only the shared variance of multiple tasks that have been designed to measure a common construct (Miyake et al., 2000). This procedure yields a purer measure of executive functioning. But even when this is done, the possibility remains that the factor structure that emerges is 'contaminated' with measures of other, lower-order cognitive processes that are also shared by the different tasks. In this dissertation it was therefore investigated whether the three-factor structure could be replicated in children aged 6-8 years while the chances of obtaining an impure factor solution were minimized. Care was therefore taken that tasks varied as much as possible in non-executive aspects. A test battery with multiple tests per hypothesized executive function was administered and, as in previous studies, it was investigated whether a three-factor solution was tenable.

In a pilot study, reported in *Chapter 2*, an exploratory factor analysis confirmed the three-factor structure in a small sample. This promising result gave rise to the larger longitudinal study, presented in *Chapter 3* and *Chapter 4*. In this study, a series of confirmatory factor analyses was performed, in which the hypothesized three-factor model was compared to alternative models. One alternative was a one-factor model, in which executive functions are represented as an undifferentiated whole. Three other alternative models were two-factor models, each containing one combined factor of two executive functions. All models were comprehensive models, containing measures of both accuracy and speed and controlling for baseline speed. The results from the longitudinal study showed that a two-factor solution fitted the data better than the three-factor model: inhibition and shifting could not be distinguished as separate factors. This finding was highly robust across different models in which test scores were obtained in different ways, but the results contradict the findings of the pilot study. This illustrates the current lack of consensus in the literature as described before: the three-factor structure of inhibition, shifting, and updating is found in some studies but not in others.

Since the longitudinal study presented in this dissertation contained a large sample, carefully selected tests, precise and comprehensive measurements of both speed and accuracy, four repeated measures during a period of two years, and robust results when scoring methods were varied, the results of this study, namely a two-factor structure with an updating factor and a factor on which both inhibition and shifting tasks loaded, seem most convincing. These results are in line with the increasing number of studies that also failed to replicate the three-factor structure. However, it must be noted that these studies did not converge on an alternative factor solution. Almost every study that deviated from the original inhibition, shifting, and updating distinction did so in a different way. This suggests that a new framework is needed, that is grounded in strong theoretical foundations. The studies described in this dissertation were designed to examine the validity of the distinction in inhibition, shifting, and updating, and not to design or test an alternative theoretical framework. Nevertheless, although suggestions regarding a different structure will await future corroboration, a critical review of the results obtained in this dissertation may yield elements that might be important in a new theoretical framework.

One alternative to the three-factor structure is a hierarchy of the three functions, as has also been proposed by Garon, Bryson and Smith (2008), who stated that inhibition is the lowest-order skill, for updating tasks both inhibition and updating skills are needed, and shifting requires all three skills. This approach does justice to the complexity of executive functions, as it acknowledges the reliance on multiple executive skills in executive function tasks that seemingly measure only one EF skill. Nevertheless, a possible hierarchy does not seem to be this straightforward, as for example a complex updating task may require more resources than a simple shifting task.

Instead, it may be more fruitful to abandon the approach of higher order executive function 'units' such as shifting, and focus instead on an analysis of the underlying processes that are important when performing an executive function task. Shifting, for example, is likely to be not one skill, but rather, a shifting task such as the Animal Shifting task requires many skills, such as maintenance of the task requirements throughout the task (yellow background = name the fruit, purple background = name the animal), encoding of the stimulus (e.g., a yellow background, a picture of a bird and a picture of an apple), engagement of the accompanying task set (yellow background = name the fruit), and formation and articulation of the appropriate response ("apple"). In a line of research in experimental psychology, this approach is more common: executive functions or executive control in a broader sense are regarded as the result of dynamic interactions between lower-order processes (Vandierendonck, Szmalec, Deschuyteneer,

& Depoorter, 2007). Vandierendonck et al. describe how the coordination of resolution of conflict between task goal representations in working memory and processes that are triggered automatically can lead to executive control. This dynamics of lower-order processes may eventually even yield one statistical factor, possibly leading to the false conclusions that this factor represents one coherent and static skill (Van der Maas et al., 2006). The approach of finding lower-order processes that are involved in higher-order cognitive functioning may shed more light on the present results.

In *Chapter 3* it has been argued that both inhibition and shifting require the resolution of conflicting information, as a selection must be made between two possible responses: inhibition tasks require the suppression of a dominant response and shifting tasks require the selection of the correct response between two possible responses. Both tasks therefore require activation of the correct stimulus-response map, based on the requirements of a particular task. This activation must be sufficiently strong to overcome activation of disturbing alternative and possibly dominant responses. Both types of tasks therefore also require a suppression mechanism that prevents these alternative responses from gaining too much strength.

This suppression mechanism might be further subdivided into two possible processes: filtering and deactivation. Filtering refers to a process taking place during the stimulus encoding stage, when the stimulus features are being processed. At this stage it is possible to direct attention to the relevant parts of the stimulus, so that activation of a dominant response is prevented. This way, the dominant response is not activated, because information that leads to this dominant response is not processed. When doing the traditional Color-Word Stroop task (Stroop, 1935), in which a color word is presented in a different color and the color of the ink rather than the spelled word must be named, a participant may not focus on the words but instead focus on infinity, so the words appear blurred but the color is still visible. In the Animal Stroop, a task that was used in this study, participants may develop similar strategies, such as focusing on a curly tail to recognize a pig. This way, the dominant response that must be suppressed, namely the tendency to recognize and name the head of the animal, is hardly present: interfering information is filtered.

When the dominant response is not filtered, deactivation is needed. This refers to the ability to counter the activation of the dominant response, possibly by allocating attention away from the irrelevant activation. In the case of the Color-Word Stroop Task, deactivation means that the word is read and processed, but the strength of this response is diminished by focusing attention away from the words and towards the ink color, allowing the ink color to be named instead.

Inhibition tasks are said to measure suppression of a dominant response: deactivation should be the important process. However, accurate task performance can be reached when filtering, deactivation, or both are used, and on most traditional inhibition tasks it is not possible to deduce which one was predominantly used by the participant. Moreover, the importance of each process is likely to be subject to change during development, and even during a task. While doing the task, the participant may discover strategies to diminish the activation of irrelevant representations by using filtering techniques instead. This decreases the executive task load substantially.

The activation of irrelevant representations may also be diminished automatically with increased task experience. When multiple stimuli are administered successively, as is necessary to obtain reliable speed and accuracy scores, the strength of the activation related to the required response may increase. This way, the relative strength of the dominant representation decreases, possibly even up to the point where this response is no longer dominant. During the Simon task the association left hand–mouse and right hand–dragon may become increasingly stronger, until it is stronger than the assumed location preference association. This leads to a paradoxical situation: on the one hand, a large number of stimuli is needed to obtain reliable scores. On the other hand, by administering a large number of stimuli, the processes that are responsible for correct performance on the task may change.

The three inhibition tasks that were administered in the longitudinal study, presented in *Chapter 3* and *Chapter 4*, have been inspected for possible speed changes during the task. The evidence was mixed: it was found that especially in the first, control block responses slowed down over time. This effect was smaller or not present during later blocks (Haster, 2009). It is, however, impossible to infer which processes were active, based on reaction time data only: the possible effects mentioned above may lead to faster responses, while effects of fatigue and boredom may slow down responses.

The tasks that were used in the pilot study, presented in *Chapter 2*, may have been less prone to the effects mentioned above. All tasks contained far fewer stimuli than in the longitudinal study, and the Tower of London (Shallice, 1982), a task that was used to measure inhibition in this study, is devised such that filtering the input was not possible. In this task, colored balls must be moved from one peg to another in order to copy a goal configuration in a minimum number of steps. In order to do this, the urge to start immediately must always be suppressed, as starting without a plan is likely to lead to incorrect solutions. The fact that filtering was not possible for this task might explain why in the pilot study a separate inhibition factor could be constructed.

Another notable finding concerning inhibition and shifting was reported in *Chapter 3*. In the model presented in this chapter, both speed and accuracy measures from the inhibition and shifting tasks were included. In addition, baseline speed was accounted for in this model by including separate baseline speed factors. While the accuracy measures loaded significantly on the inhibition/shifting factor, the speed loadings on this factor were not significant. Speed measures did load significantly and strongly on the two baseline speed factors and internal consistency of the tasks was high, so measurement unreliability as a possible cause seems unlikely. Two alternative explanations for this unexpected finding remain, both of which state that there is no fundamental difference between baseline speed and EF. The first possibility is that, contrary to what is often assumed, baseline speed tasks require the same executive resources as EF tasks. EF involvement in baseline speed tasks has already been shown (Szmalec, Vandierendonck, & Kemps, 2005). If this is true, then controlling for baseline speed measures possibly means controlling for too much. This is not unlikely, since baseline speed has been shown to be a predictor of academic performance (Berg, 2008; D'Amico & Passolunghi, 2009; Swanson & Kim, 2007).

A second possibility is that the speed measures of shifting and especially inhibition tasks do not reflect executive processing, but merely reflect non-executive processes. If the features from the incoming stimulus are filtered such by the participant that there is no dominant response to suppress, reaction time of these stimuli may merely measure baseline speed. Responses may be slowed down, not because different processes are recruited, but because the stimuli are more complex (e.g., multiple basic shapes rather than one) and therefore require more (visuospatial) processing time. Failure to correct for baseline speed properly in a factor analysis may lead to a factor that seems to reflect inhibition or shifting, but instead represents the overadditivity effect: the effect that people with a slow baseline speed slow down disproportionally on any other task that is more difficult (Faust, Balota, Spieler, & Ferraro, 1999).

If one or both of these explanations are valid, the question arises why the accuracy loadings were significant. It is possible that the speed measures largely reflect presumably non-executive processes such as motor and articulation speed, while accuracy is a better reflection of the ability to choose between competing responses. It must, however, also be noted that baseline accuracy was not controlled for, as there was not sufficient variation in baseline accuracy measures. In addition, most studies investigating inhibition, shifting, and updating thus far did not derive or compare accuracy and speed measures from the same task, so this finding needs confirmation in future studies.

In order to perform well on EF tasks, it is not only important to suppress incorrect or irrelevant activation. It is also essential to be able to activate important information that must be remembered, be it the requirements of the task or information that must later somehow be reproduced. Both the ability to keep multiple items activated and the ability to sustain this activation for a longer period of time and in the face of other, disturbing information are important.

These two elements of activation and suppression might be combined in a new model of a structure of executive functions. A possible model is presented in Figure 6.1. This Figure shows the idea that inhibition, shifting, and updating do not reflect primary executive skills. Rather, tasks measuring one of these hypothesized constructs measure a variety of the same underlying cognitive skills, but, depending on the executive function they were designed to measure, they measure these skills to a different degree.

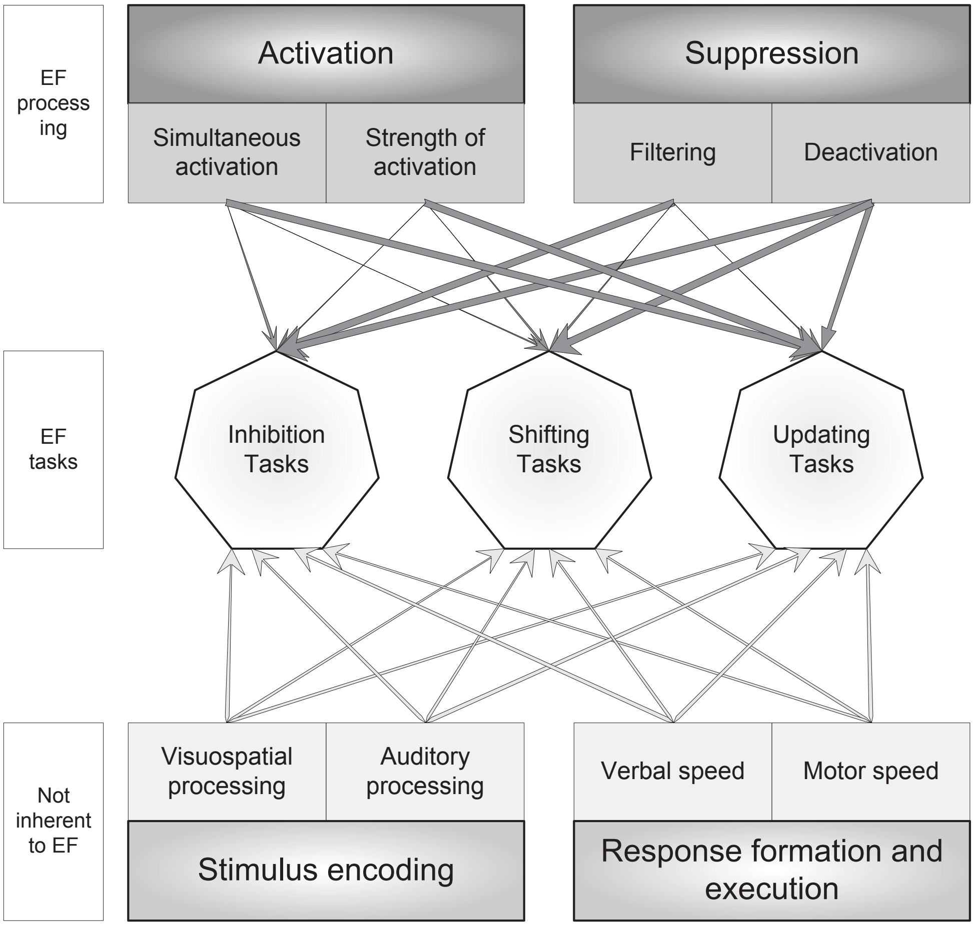


Figure 6.1 A possible alternative model of executive functions. The model poses that performance on executive function task is influenced by several processes. The relative importance of each process in the upper part is reflected in the size of the arrow. In the lower part this cannot be done, as the relative importance is task specific.

Figure 6.1 is divided into three parts. The center displays the three previously presumed executive functions: inhibition, shifting, and updating. However, according to this model, performance on these tasks is not achieved by processes uniquely associated with any of these constructs. The word 'task' has therefore been added to each executive function, in order to stress that these three concepts are not primary cognitive abilities. Instead, performance on these tasks is achieved by coordinating several underlying processes, which are displayed above and below the three factors.

In previous research it has already been acknowledged that so-called impurity of task scores arises because of lower-order processes that are not executive; in other words, processes that are not related to flexible decision making (Miyake et al., 2000). A by no means exhaustive list of these processes is displayed in the lower part of Figure 6.1. These processes may be subdivided further into processes important during the stimulus encoding stage (e.g., visuospatial and auditory processing) and processes important during the response formation and response execution stage (e.g., motor and verbal speed). If tasks are chosen that vary systematically in the load they pose on these skills, factor analysis will eliminate these factors to a large degree.

The upper half of the Figure presents a possible new distinction in executive skills: skills that are key to flexible decision making. The postulation made here moves away from the model by Baddeley and Hitch (1974), towards models that conceptualize working memory as the activated part of long-term memory (Agostino, Johnson, & Pascual-Leone, 2010; Cowan, 1988; Roncadin, Pascual-Leone, Rich, & Dennis, 2007). In these models there is a balance between the ability to activate representations, displayed here on the upper left side of the Figure, and the ability to suppress this activation when necessary, displayed on the upper right side. The model presented here can be regarded as an extension of these models, as an alternative explanation to the previously made distinction of executive functions in inhibition, shifting, and updating.

The activation part in the upper left can be divided further into the ability to activate a large number of items simultaneously and the strength of this activation. The latter process refers to the ability to prevent spontaneous decay, and to sustain activation when interfering information enters. Empirical evidence for spontaneous decay, merely as a function of time, is mixed (Cowan, 2010; Lewandowsky, Oberauer, & Brown, 2009; Zhang & Luck, 2009), but especially resistance against interference seems important.

Limiting the occurrence of interference is a second important skill, presented in the right part of the figure. This suppression of items may be further subdivided into (1) the ability to filter or prevent activation during the stimulus encoding stage and (2) deactivation, the ability to suppress this activation after it occurs by directing attention away from this activation. The previously used concepts of cognitive inhibition (Nigg, 2000) and resistance to proactive interference (Friedman & Miyake, 2004), both defined as the ability to prevent unwanted contents of working memory, seem to refer implicitly to this same ability of deactivation.

Tasks supposedly measuring inhibition, shifting, and updating are all expected to require these skills to a different degree. The strength of these hypothesized relations between the skills in old distinction (in the middle of the Figure) and the skills in the new distinction (in the upper part of the Figure) is represented graphically by the thickness of the arrows: thick lines represent higher predicted importance. (Note that the executive function tasks also vary in the demands posed on the skills mentioned in the lower part of the figure, but this variation is not systematically related to the three different executive functions; the variation is task-specific instead, so the thickness of the line will vary per task). If this model is true, it is theoretically impossible to extract factors related to the skills in the upper half of the Figure if tasks are used that are supposed to measure inhibition, shifting, and updating: all tasks measure a mixture of the same underlying executive processes. The importance of the different processes in the upper half of the Figure is also task-specific, so the result of factor analysis is likely to vary between studies, depending on the specific tasks that are used.

It is expected that for inhibition tasks, activation is necessary to some degree, as the task requirements must be kept in an activated state. As long as the task is simple, the activation part does not require many resources, but the activation of the task rules must be strong enough to overcome activation of the dominant, interfering representation. Suppression is important, both by filtering, or selectively focusing attention away from stimulus features that automatically activate the dominant and to-be-suppressed response, and by deactivation, or redirecting attention, away from the dominant response set, and towards stimulus features that are relevant to the task. For shifting tasks, a similar pattern is expected, with the exception that filtering is not as important for shifting, because there is no systematic stimulus feature that has to be ignored.

Updating tasks are expected to require a different pattern of resources: in order to perform well on these tasks, it is necessary to maintain activation of a large number of items and to sustain this activation during a prolonged period of time. Filtering is of minor importance, but deactivation is important when a representation that was previously useful becomes obsolete in the light of new information. When a child has to remember the last animal, *cat* must remain active as long as it is the last animal that was presented, but it must be deactivated when a dog stimulus appears.

A complicating factor is that the processes mentioned in the upper half of Figure 6.1 may partially not be executive, in the sense that they are not entirely under voluntary control. Instead a combination of conscious, top-down and subconscious, bottom-up processes may be involved. Representations in long-term memory may be activated voluntarily, but also automatically upon the presentation of a stimulus. Sustaining this activation may be achieved by conscious strategies, such as subvocal rehearsal: it is known that turning attention towards the to-be-remembered items every now and then, thus refreshing the activation, enhances performance (Barouillet, Gavens, Vergauwe, Gaillard, & Camos, 2009). It may also be an automatic process to some extent, i.e., spontaneous decay. Similarly, suppression of activation may also be achieved voluntarily or automatically. It has been suggested that outside the focus of attention, activated items are prone to decay and interference from similar, incoming stimuli, while items inside the focus of attention are protected against these influences (Cowan, 1988, 2010).

Another issue to be resolved is the distinction between updating and working memory, and the conceptual confusion arising from the lack of understanding of the similarities and differences between these two concepts. This can also be better understood in the light of Figure 6.1. As described before, updating is defined as the ability to activate information in working memory and to change this information when necessary, and is typically measured with *n*-back tasks or keeping track tasks. In *n*-back tasks, the required response must not be based on the last stimulus that was presented, but on the stimulus that was presented *n* stimuli ago, e.g., the location of a dot must be indicated after two more dots have been presented. In keeping track tasks the final item of a certain category must be remembered, and when an item is presented the participant does not know if it is the final one. Both *n*-back and keeping track tasks require the continuous monitoring of incoming information: new incoming information may require the information that is active to be 'overwritten'. Working memory, on the other hand, is defined as the ability to store and process information simultaneously, and is typically measured with complex span tasks. These tasks require activation of items that is robust to interference. This means that an important factor determining performance on both types of tasks is the temporary activation of multiple representations simultaneously. Keeping track of the temporal order in which these items were activated is also important for both types of tasks. But there are also differences between the two: while updating also relies heavily on the ability to *deactivate* representations when they become obsolete in the light of new information, for working memory this deactivation is only important when a new trial is presented and items from the old trial must be deactivated. Time intervals between trials are much larger than in updating tasks, and this delay will therefore cause a natural decay of this activation. More important for working memory tasks is maintenance of activation over a longer period of time and interference, since the processing tasks that are an essential part of working memory tasks also take up time. Also important for working memory tasks is maintenance of sufficient activation of the relevant items when attentional resources must also be spent on the processing task. In short, both types of tasks require similar processes, since multiple representations must be activated simultaneously, but for updating tasks deactivation is relatively more important, while for working memory tasks strength of activation is key. The overlap between the tasks requirements is large, and the strategy that is used during the task may cause the overlap to be even larger: updating tasks require a continuous change in activation of items, but a task like 'Keep Track' could also be carried out in a different way. If a child does not update the activation of 'last-presented items' continuously, but rather tries to remember as much as possible during the sequence, and thinks back of the last items after the series has finished, then the requirements needed for successful performance closely resemble those of working memory. It might be possible to ensure that participants use the continuous updating strategy they are supposed to use by demanding continuous rehearsal of the last items. For children below seven this demand may be impossible, as they do not tend to use subvocal rehearsal spontaneously as much as older children and adults (Gathercole, Adams, & Hitch, 1994; Tam, Jarrold, Baddeley, & Sabatos-DeVito, 2010). Together, this explains the confusion that has arisen between the two concepts of updating and working memory, as well as the fact that tasks measuring both concepts load on the same factor in a factor analysis (St Clair-Thompson & Gathercole, 2006).

The model presented in Figure 6.1 might also provide an explanation of the difference between short-term memory and working memory: a distinction that has not received much attention in this dissertation (it is only referred to in *Chapter 2*), but that is important in the literature. Short-term memory, defined in the Baddeley and Hitch (1974) model as the phonological loop and the visuospatial sketchpad, reflects the ability to store information for a brief period of time, whereas working memory refers to the ability to store and process information. Short-term memory is usually assessed with simple span tasks, in which a series of stimuli must be repeated. Examples are word and digit span tasks (phonological loop) and Corsi blocks (visuospatial sketchpad). Working memory, on the other hand, is measured with complex span tasks that also include a processing component, such as listening span and counting span. Working memory has been found to be more strongly related to fluid intelligence (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002) and to be a better predictor of scholastic achievements (Gathercole & Pickering, 2001; Geary, Hoard, & Hamson, 1999; Lee, Ng, Ng, & Lim, 2004; Noël, 2009) than short-term memory. In terms of the activation part of the present model, short-term memory tasks only requires the simultaneous activation of multiple representations, whereas working memory tasks also require this activation to last longer, as the processing task takes more time. That means that short-term memory and working memory are not distinct entities, but working memory is an extension of short-term memory (Conway et al., 2002; Cowan, 2010).

It must be stressed that the model as presented here is still hypothetical and serves as a first attempt to explain the contradictory results that arose in executive function studies that tried to replicate the factor structure of inhibition, shifting, and updating. Since the present data do not allow a statistical validation of this model, it awaits confirmation in future research. This will be a challenge, as these processes are likely to be dynamically interrelated and therefore impossible to measure in isolation of the other processes. In addition, this dynamics may differ between individuals, as a function of the strategies that are applied, and the processes that are involved may even change during the course of a single task. It must also be stressed again that executive functions are strongly in development during childhood, and that a model therefore needs to be tested in a wide age range. The results obtained in this dissertation showed a stable factor structure during first and second grade, but this cannot be generalized to other ages.

Learning mathematics

In *Chapter 4*, the development of mathematical performance during the first two years of primary education was investigated. Cito math tests, which are norm-referenced national tests that measure a variety of mathematical skills (Janssen, Scheltens, & Kraemer, 2005) were administered four times and development of the children's scores was modeled. In these models, a linear growth trajectory provided a good fit to the data. There was no significant relation between the mathematical level of the children and their improvement during the study. This means that on average, high-scoring children did not learn mathematics faster or slower than their low-scoring peers.

In *Chapter 5*, a microgenetic design was applied to gain more insight in the mathematics learning process of children. It was already known that strategies that children apply in their attempts to solve math problems vary over time and are indicative of their success in solving these problems (Imbo & Vandierendonck, 2007; Lemaire, 2010; Lemaire & Siegler, 1995). Therefore strategy use of children was analyzed further in one domain: single-digit multiplication. Siegler (1996) proposed that when learning a new skill, strategies come and go in a manner resembling overlapping waves: after a certain strategy is discovered, the frequency with which it is used initially increases but decreases later, when the child discovers more advanced strategies that slowly replace the older, less mature strategy. An important feature of the model is that it states that the strategies that are used do not replace each other in an all-or-none fashion: rather, at each time point children have a repertoire of strategies, of which the relative use slowly changes.

The Overlapping Waves model was applied to the strategies that children applied to single-digit multiplication problems during a period of eight weeks. The strategies that children used were classified into five types of strategies: in order of increasing maturity these were wrong strategies (e.g., addition: 7 x 4 = 11), counting strategies (e.g., finger counting), repeated addition (5 x 4 = 4 + 4 + 4 + 4 + 4), derived facts (6 x 3 = 5 x 3 + 3 = 15 + 3) and retrieval (answer is directly retrieved from memory). Thus far, the Overlapping Waves model had only been used metaphorically. With the use of categorical growth modeling, it was possible to validate the model statistically: it was possible to construct one underlying scale representing 'strategy maturity'. This scale predicted the probability of use for each of the five strategies. The higher the latent strategy ability of a child, the larger the probability that relatively mature strategies would be selected over immature strategies. During the study, the estimates of children's strategy maturity gradually increased, thus reflecting development in mathematics proficiency. This model fitted the data well. In *Chapter 5*, Figure 5.4 illustrates how the changes in probabilities of the use of each strategy, predicted as a function of increasing ability, resemble the shape of overlapping waves, as was predicted by Siegler (1996). Furthermore, the strategies children used were strongly related to the accuracy of their answers, and growth in strategy maturity during the eight weeks of the study was strongly related to growth in accuracy. This shows that strategy use is an important factor in children's success or failure in mathematics.

The strategies that were used did not only differ between children, but they were also problem-specific: for easier problems such as 3 x 2, more mature strategies were applied than for more difficult problems such as 8 x 6. This confirms previous studies in which it was found that children are adaptive in the strategies they apply (Torbeyns, Verschaffel, & Ghesquière, 2002, 2004). In addition, the study showed that categorical growth modeling is a powerful tool in modeling development of categorical data. This method enabled a more detailed and quantitative analysis of children's development in strategy use in mathematics, illustrating the importance of the inclusion of children's strategy use in research.

Executive functions and mathematics

The main finding of this dissertation is the importance of working memory or updating in math learning. While it was already known that there is a relation between working memory performance and math achievements (Passolunghi, Mammarella, & Altoè, 2008; St Clair-Thompson & Gathercole, 2006; Swanson & Kim, 2007; Van der Sluis et al., 2007), in this dissertation more details were revealed about the nature of this relationship. Already in the pilot study in *Chapter 2*, the relations between mathematics and updating were more pronounced than with inhibition and shifting. In *Chapter 4* we only found a unique relation between mathematics and updating; relations with inhibition and shifting were significant, but this significance disappeared in the presence of the updating factor. This does not necessarily mean that suppression of a dominant response or switching between tasks is of no or only minor importance for mathematical performance. The results may be understood better in the light of the alternative structure of executive functions as discussed in the previous section. This model proposes that each executive function task from the inhibition/shifting/updating paradigm assesses a variety of the same skills, but each to a different degree. That means that a mixture of the same skills is measured multiple times. For updating and working memory tasks, the overall load is highest, as is reflected in the thicker arrows pointing towards updating. This suggests that the additional value of inhibition and shifting tasks is not large enough to add significant explained variance. Moreover, for updating and working memory tasks especially the ability to activate multiple representations simultaneously and for a prolonged period of time is important. It is likely that this ability is key to mathematics learning. Consolidation of math facts is important in math learning, and this can only be done properly when the operands and the answer are activated together, so a neural connection between them can be formed and strengthened (Geary, Brown, & Samaranayake, 1991). This means that the information from the operands must still be activated when the answer to the problem is found, since only then the operands can be associated with the answer. In addition, in complex math problems all information from the problem must be activated together in order to form a meaningful representation of the problem. When a child is unable to do this, vital parts of the problem will be missing in the problem representation.

A second explanation may be that, as mentioned before, children may have used a filtering strategy rather than a deactivation strategy for especially inhibition tasks. Deactivation of a dominant response is likely to be important for mathematics, since the contents of working memory are limited and scarce resources should not be spent on irrelevant information. In addition, deactivation of activated immature strategies may also be important. Filtering may not be such an important skill: it is impossible to know which parts of a math problem are important before they are studied, so filtering the incoming information beforehand will not be a useful skill. It may, however, be argued that it is important to filter other stimuli from the environment and attend selectively to the math problems, but it is questionable whether specifically this ability was tapped by inhibition and shifting tasks as used in this study: resistance to interference from the environment is important for performance on any task, also the working memory tasks.

Thirdly, it must also be noted that the age range in this study was limited to early primary education. Different executive function processes may play a role in later stages of mathematics learning. when math problems are more complicated and the child has acquired more math strategies that possibly need to be suppressed.

Another finding is that the relations between working memory and mathematics as found in this study, especially in *Chapter 4*, were stronger than often reported. This might be due to the fact that we used factor analysis, and thus obtained a purer measure of working memory than in many other studies. Since the math test that was administered in the longitudinal study covers a variety of mathematical skills, this strong relationship also suggests that working memory is needed for many aspects of mathematics.

In *Chapter 4*, indications were also found that development of updating is related to development in math achievements. This stresses that the learning process is key to explaining the relationship between working memory and mathematics, and single measurements are therefore not sufficient to capture the complexities that are involved.

In the microgenetic study, presented in *Chapter 5*, the relationship between working memory and mathematics was investigated in more depth than in most studies. In this chapter the term working memory was used rather than updating, as by the time this study was carried out, it was clear that inhibition and shifting did not contribute substantially to mathematical performance. It was therefore decided to include only the updating factor. The facts that (1) the term updating is hardly used in isolation but the term working memory is, and (2) the tasks that comprised the factor were both updating and working memory tasks, led to the choice of a change in terminology. Nevertheless, the term refers to the same concepts: the working memory scores that were used in the microgenetic study were obtained in the longitudinal study.

By investigating the strategies children use, a first step was made to unravel the mechanisms that underlie the often-reported relationship between working memory and mathematics. Most studies thus far only included the number of items answered correctly on a math test, which makes it impossible to know how children achieved their answers. In addition, most studies employed single measurements, which does not do justice to the dynamic nature of the relationship: working memory is supposedly most important during the learning process. The microgenetic study was designed to include these two important elements. Relations were found between working memory and both strategy and accuracy. This means that children with good working memory abilities are capable of using more mature strategies, which are less error-prone by themselves, but these children make even fewer mistakes than can be explained by strategy selection alone. The results suggest that both the strategies children apply and their success at executing these strategies are partially determined by their working memory abilities. Good working memory abilities may prevent the occurrence of procedural errors, explaining the relation between working memory and accuracy. The relation with working memory and strategy selection may be that strong working memory abilities aid in solving problems fast. Finding correct answers quickly may stimulate the formation of strong connections between representations of the problem and the answer (Geary et al., 1991). These connections are needed to progress to more mature strategies. The use of more mature strategies may then lead to faster and less error-prone problem solving, thus promoting further growth in mathematical proficiency. First evidence in favor of this theory was found on a small scale in this microgenetic study. Small-scale development is likely to reflect learning on a macrolevel (Siegler & Svetina, 2002): math development during years. These results illustrate again how the relationship between executive functions and mathematics is dynamic.

Directions for future research

The results of this dissertation give rise to a number of potentially interesting areas of new research. These are again divided into two parts: the structure of executive functions and relations between executive functions and mathematics.

The structure of executive functions

Considering the structure of executive functions, the increasing evidence that the distinction in inhibition, shifting, and updating is no longer tenable indicates that what is needed now is a new theoretical framework. An attempt has been made in this chapter by suggesting an activation/suppression model. This could be tested empirically by systematically varying the load of each proposed subprocess: simultaneous activation of multiple items and strength of this activation, filtering, and deactivation by reallocation of attention. This is not an easy challenge, because participants may vary in the strategy they use and therefore in the load of each process involved. These strategies may well be employed subconsciously, especially in young children, so they cannot be measured by means of interviews.

It must also be noted that assessing the efficiency of the different subprocesses in isolation may not do justice to everyday life situations, in which many different skills are required together. Successful performance is more than the sum of all subprocesses: it probably heavily depends on the coordination of these different processes (Roncadin et al., 2007). Using a combination of simple and complex tasks may shed more light on this issue.

Brain imaging techniques may also be a valuable tool to identify and localize the processes that are required by different tasks. Some studies have already been carried out: one such study was able to demonstrate the role of attention in verbal short-term memory (Majerus et al., 2006). A challenge for brain imaging studies will be that activation patterns should not be aggregated over groups of individuals, because interindividual differences in activation patterns should not be considered noise but may yield valuable insight in individual differences in strategies.

EF and mathematics

Throughout the study, there was one robust finding: updating, or working memory, was more strongly related to mathematics than inhibition and shifting. At first sight this may suggest that working memory is key to mathematical performance, while studying inhibition and shifting in relation to scholastic achievement may eventually not be a fruitful approach. Nevertheless, as suggested before, it may also be the case that working memory tasks cover a variety of cognitive processes to such a degree that there is no additional value of other tasks. Since the nature of the relation between executive functions and mathematics has thus far been left implicit in most studies, it is as yet impossible to answer this question. In the future we therefore need a theoretically sound framework, such as the framework suggested in the previous section, and we need to find a solid theoretical basis to explain why exactly each proposed process should be involved in mathematics. Only when this is done, specific hypotheses can be formulated and tested. This also means that assessment of mathematical skills must go beyond the number of items answered correctly on a math test: it must also be analyzed further what mathematics entails exactly. Just like executive functions, mathematics is an umbrella term for many different skills, even in the beginning of primary education when basic concepts are acquired. Skills that children acquire in these early years include, but are not limited to:

- Understanding the base-10 structure of numbers
- Simple arithmetic: addition, subtraction, and multiplication, both mentally and with pencil-and-paper aids
- Visuospatial skills: e.g., measuring lengths and counting blocks in 3D structures.
- Clock reading
- Money (understanding and using coins and bills)

In previous studies, the math tests that are used differed widely. In some studies a single mathematical skill was assessed, such as speeded single-digit addition and subtraction in older children who should have already acquired these skills (Van der Sluis et al., 2007). Other studies used a broader measure, such as the arithmetic subtest of the Wide Range Achievement Test (Berg, 2008; Passolunghi & Siegel, 2004; Swanson & Beebe-Frankenberger, 2004). A closer analysis of the underlying abilities that are necessary to carry out these mathematical operations would be helpful: e.g., does the test require mathematical reasoning or knowledge of mathematical facts? Does it require visuospatial or verbal reasoning? Once this is done, much more precise connections between executive functions and mathematics can be established than has been achieved thus far.

In this dissertation, two approaches were used to make a start in addressing the diversity of mathematics. In the longitudinal study, the math scores that were used represented a diversity of skills in one single score. In the microgenetic study, one particular skill, single-digit multiplication, was singled out and studied in detail. The results that were obtained were similar, suggesting that working memory is needed for many mathematical skills. Nevertheless, it will be worthwhile in future research to distinguish more different aspects of mathematics and study relations with executive functioning separately. A further unreported attempt has been made to distinguish different mathematical skills in the data from the longitudinal study, but this proved an impossible exercise with the Cito math tests that were used: psychometric characteristics did not allow analyses of growth curves of subsets of the tests.

Another yet unanswered question is how to design appropriate interventions for mathematically disabled children with a working memory deficit. All present studies were correlational and can therefore not establish a causal relation. The results from the microgenetic study suggest that the strategies that children use are key for mathematical performance, confirming previous studies (Imbo & Vandierendonck, 2007; Lemaire, 2010). The microgenetic study showed that over time children used increasingly more mature strategies. Their strategy choice was strongly related to accuracy of the answer. Causality cannot be inferred from the study, which means that there are at least two possible (not mutually exclusive) mechanisms that account for the relation. It is possible that the use of more mature strategies by itself leads to more correct answers, as these strategies require fewer steps and therefore yield fewer opportunities for procedural errors to be made. If this is true, interventions should mainly target strategy selection: children should learn to apply mature strategies. If this means that these children reach incorrect answers in the beginning, it is not such a problem. It is, however, also possible that solving problems correctly strengthens the neural connections that are needed to use more mature strategies. In this case, encouraging the use of more mature strategies before children are capable of using them will be useless at best, but may lead to the formation of incorrect connections at worst. Finally, it is possible that other, non-specified mathematical skills increase during learning that influence the development of both maturity of strategy choice and accuracy.

It may be worthwhile to unravel these mechanisms in future research. In order to establish causal relations, it is necessary to employ an experimental design, i.e. an intervention study. A possible design would be a comparison of different interventions that target only strategy maturity or only accuracy: one intervention that encourages the use of mature strategies, one that aids children in reaching accurate solutions to whichever strategy they choose and, optionally, an intervention that combines both elements. Care must be taken that the intervention that targets accuracy does not resort to mere drilling of the tables, as this would also influence strategy choice indirectly. Rather, the child should be assisted in correctly carrying out their own desired procedure. A comparison of these interventions may show which element is key to mathematics learning: the strategies that are chosen, or the likelihood that the obtained answer is correct and possibly also the speed with which this correct answer was found. Working memory may be a mediating factor predicting success of these interventions: for children with low working memory abilities, it may be especially important to obtain the correct answer quickly. These children may also have more difficulties in carrying out new and more mature strategies, as they may be delayed in forming the neural basis that it necessary to execute these strategies. It may therefore be possible that for children with average or above average working memory abilities an intervention targeting strategy use is most efficient, while for children with poorer working memory an intervention targeting the accuracy of the answer is more beneficial.

Some studies have already applied designs that resemble this structure. One such study (Milo & Ruijssenaars, 2002) compared three different interventions in teaching children with mild learning difficulties addition and subtraction up to 100. Two interventions aimed at teaching these children one particular strategy: in one group this was a mature strategy, in the other less mature strategy. The third group was taught to use the strategy they preferred. The first group, which was taught the mature strategy, outperformed both other groups. This may suggest that in this particular group, namely children with mild learning difficulties, teaching a mature strategy is more fruitful. However, the intervention in the flexible group was not specifically targeted at quickly reaching the correct answer, an element that is presumably crucial to enable the transition to more mature strategies.

Practical implications

The focus of this dissertation was theoretical and claims about practical implications can therefore only made with great caution. Perhaps the dissertation even revealed that the gaps in our knowledge about mathematics learning and the underlying processes are so wide that these need to be closed before practical implications can be made with confidence.

Some preliminary and cautious attempts can be made to make a start in this process. First, the failure to replicate the three-factor structure of inhibition, shifting, and updating shows that it is difficult to distinguish different executive functions. This means that practitioners assessing executive functions in their clinical practice must be cautious in the conclusions they draw when children perform poorly on these tasks. There are probably many processes that contribute to performance on these tests, which makes it difficult to establish the nature of the problems that cause these children to perform poorly, especially when the number of tests that is administered is small.

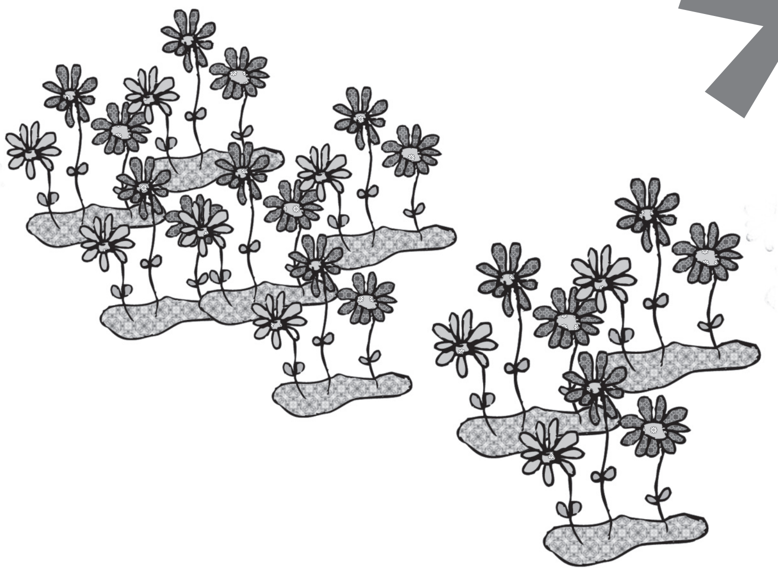
Second, a robust finding is that working memory is strongly related to mathematical performance. This indicates that children with poor working memory are at risk of developing later mathematical difficulties (c.f. Toll, Van der Ven, Kroesbergen, & Van Luit, 2010). This yields opportunities to identify these at-risk children in an early stage and it yields insight in the reasons why these children lag behind their normally developing peers.

How these at-risk children may be assisted best, is still unclear. The fact that improvement in working memory was also related to improvement in mathematics, suggests that working memory is flexible. Early working memory interventions may therefore aid this group in improving on their working memory skills, which may enhance their mathematical learning potential and therefore also their mathematical skills. Initial evidence supports the effectiveness of training working memory in children with poor working memory; this training effect also generalized to mathematical skills (Holmes, Gathercole, & Dunning, 2009; Kroesbergen, Van ‘t Noordende, & Kolkman, 2011), but these initial promising results need validation in larger samples. In addition, these training studies targeted multiple areas, such as attention and verbal and visual working memory simultaneously. The key elements and optimal duration and number of sessions must therefore still be established. It may also be possible to train teachers to incorporate elements of these trainings in their normal teaching routines.

Despite these promising initial results, it might still be possible that the poor working memory abilities of at least some low performing children cannot be remediated sufficiently. In this case, an alternative approach would be to be encourage these children to compensate for their memory weaknesses by using external aids, such as abacuses. This might enable them to obtain answers fast enough to strengthen the necessary neural connections. When presented with complex problems, an external aid could be the use of paper and pencil to build an external meaningful problem representation. However, using external aids also reduces the opportunities for these children to practice their working memory skills, which may potentially lead to an even larger delay in working memory abilities. It is beyond the scope of this dissertation to resolve this dilemma, but it is well possible that the best solution is a well-balanced combination of these two strategies, tailored to the child's abilities.

Conclusion

The dissertation allows two main conclusions to be drawn. The first conclusion is that the fractionation of executive functions in inhibition, shifting, and updating, is no longer tenable. The empirical evidence in this dissertation contributes to the accumulating body of studies that could not validate such a distinction. In this chapter, an alternative model has been proposed, but this model needs empirical validation that cannot be provided yet. The second conclusion is that working memory is strongly related to the acquisition of mathematical skills in young children and that strategy selection is an important factor mediating this relationship.



References

Agostino, A., Johnson, J., & Pascual-Leone, J. (2010). Executive functions underlying multiplicative reasoning: Problem type matters. *Journal of Experimental Child Psychology, 105*, 286-305. doi: 10.1016/j.jecp.2009.09.006

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley.

Alloway, T. P. (2007). *Automated Working Memory Assessment*. London: Pearson Assessment. Translated and reproduced by permission of Pearson Assessment.

Andres, P. (2003). Frontal cortex as the central executive of working memory: Time to revise our view. *Cortex, 39*, 871-895. doi: 10.1016/S0010-9452(08)70868-2

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556-559.

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology, 49A*, 5-28. doi: 10.1080/027249896392784

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417-423. doi: 10.1016/S1364-6613(00)01538-2

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York: Academic Press.

Barouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: A time-based resource-sharing model account. *Developmental Psychology, 45*, 477-490. doi: 10.1037/a0014615

Barouillet, P., & Lépine, R. (2005). Working memory and children's use of retrieval to solve addition problems. *Journal of Experimental Child Psychology, 91*, 183-204. doi: 10.1016/j.jecp.2005.03.002

Berends, I. E., & Van Lieshout, E. C. D. M. (2009). The effect of illustrations in arithmetic problem-solving: Effects of increased cognitive load. *Learning and Instruction, 19*, 345-353. doi: 10.1016/j.learninstruc.2008.06.012

Berg, D. H. (2008). Working memory and arithmetic calculation in children: The contributory roles of processing speed, short-term memory, and reading. *Journal of Experimental Child Psychology, 99*, 288-308. doi: 10.1016/j.jecp.2007.12.002

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*, 647-663. doi: 10.1111/j.1467-8624.2007.01019.x

Blunch, N. J. (2008). *Introduction to structural equation modelling using SPSS and AMOS*. London: SAGE Publications Ltd.

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology, 26*, 571-593. doi: 10.1207/s15326942dn2602_3

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*, 205-228. doi: 10.1080/87565640801982312

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*, 421-442. doi: 10.1080/87565649909540759

Bull, R., & Scerif, G. (2001). Executive functions as a predictor of children's mathematical ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19*, 273-293. doi: 10.1207/S15326942DN1903_3

Bunge, S. A., & Wright, S. B. (2007). Neurodevelopmental changes in working memory and cognitive control. *Current Opinion in Neurobiology, 17*, 243-250. doi: 10.1016/j.conb.2007.02.005

Burgess, P. W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society, 4*, 547–558.

Canobi, K. H. (2004). Individual differences in children's addition and subtraction knowledge. *Cognitive Development, 19*, 81-93. doi: 10.1016/j.cogdev.2003.10.001

Carr, M., & Jessup, D. L. (1997). Gender differences in first-grade mathematics strategy use: Social and metacognitive influences. *Journal of Educational Psychology, 89*, 318-328.

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

Censabella, S., & Noël, M. P. (2008). The inhibition capacities of children with mathematical disabilities. *Child Neuropsychology, 14*, 1-20. doi: 10.1080/09297040601052318

Cheshire, A., Muldoon, K. P., Francis, B., Lewis, C. N., & Ball, L. J. (2007). Modelling change: New opportunities in the analysis of microgenetic data. *Infant and Child Development, 16*, 119-134. doi: 10.1002/icd498

Clark, C. A. C., Pritchard, V. E., & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology, 46*, 1176-1191. doi: 10.1037/a0019672

Collette, F., Van der Linden, M., Laureys, S., Delfiore, G., Degueldre, C., Luxen, A., & Salmon, E. (2005). Exploring the unity and diversity of the neural substrates of executive functioning. *Human Brain Mapping, 25*, 409-423. doi: 10.1002/hbm.20118

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, 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. doi: 10.1016/S0160-2896(01)00096-4

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constrains within the human processing system. *Psychological Bulletin, 104*, 163-191.

Cowan, N. (2010). Multiple concurrent thoughts: The meaning and developmental neuropsychology of working memory. *Developmental Neuropsychology, 35*, 447-474. doi: 10.1080/87565641201049 4985

Cragg, L., & Gilmore, C. (2011). *Dual task performance demonstrates that the role of working memory in children's arithmetic performance differs by strategy*. Paper presented at the SRCD biennial meeting, Montreal.

Crone, E. A., Donohue, S. E., Honomichl, R., Wendelken, C., & Bunge, S. A. (2006). Brain regions mediating flexible rule use during development. *The Journal of Neuroscience, 26*, 11239-11247. doi: 10.1523/JNEUROSCI.2165-06.2006

D'Amico, A., & Guarnera, M. (2005). Exploring working memory in children with low arithmetical achievement. *Learning and Individual Differences, 15*, 189-202. doi: 10.1016/j.lindif.2005.01.002

D'Amico, A., & Passolunghi, M. C. (2009). Naming speed and effortful and automatic inhibition in children with arithmetic learning disabilities. *Learning and Individual Differences, 19*, 170-180. doi: 10.1016/j.lindif.2009.01.001

De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology, 103*, 186-201. doi: 10.1016/j.jecp.2009.01.004

Diamond, A., & Goldman-Rakic, P. S. (1989). Comparison of human infants and rhesus monkeys on Piaget's AB task: Evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research, 74*, 24-40.

Dibbets, P., & Jolles, J. (2006). The switch task for children: Measuring mental flexibility in young children. *Cognitive Development, 21*, 60-71. doi: 10.1016/j.cogdev.2005.09.004

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Espy, K. A. (2004). Using developmental, cognitive, and neuroscience approaches to understand executive control in young children. *Developmental Neuropsychology, 26*, 379-384. doi: 10.1207/s15326942dn2601_1

Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology, 26*, 465-486. doi: 10.1207/s15326942dn2601_6

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125*, 777-799.

Fisk, J. E., & Sharp, C. A. (2004). Age-related impairment in executive functioning: Updating, inhibition, shifting, and access. *Journal of Clinical and Experimental Neuropsychology, 26*, 874-890. doi: 10.1080/13803390490510680

Flynn, E., & Siegler, R. (2007). Measuring change: Current trends and future directions in microgenetic research. *Infant and Child Development, 16*, 135-149. doi: 10.1002/icd.502

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

Freudenthal, H. (1991). *Revisiting mathematics education*. Dordrecht, the Netherlands: Kluwer.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General, 133*, 101-135. doi: 10.1037/0096-3445.133.1.101

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

Gathercole, S. E., Adams, A. M., & Hitch, G. J. (1994). Do young children rehearse? An individual-differences analysis. *Memory & Cognition, 22*, 201-207.

Gathercole, S. E., & Pickering, S. J. (2000a). Assessment of working memory in six- and seven-year-old children. *Journal of Educational Psychology, 92*, 377-390. doi: 10.1037//0022-0663.92.2.377

Gathercole, S. E., & Pickering, S. J. (2000b). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology, 70*, 177-194.

Gathercole, S. E., & Pickering, S. J. (2001). Working memory deficits in children with special educational needs. *British Journal of Special Education, 28*, 89-97. doi: 10.1111/1467-8527.00225

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*, 177-190. doi: 10.1037/0012-1649.40.2.177

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

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37*, 4-15. doi: 10.1177/00222194040370010201

Geary, D. C., Brown, S. C., & Samaranayake, V. A. (1991). Cognitive addition: A short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Developmental Psychology, 27*, 787-797.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology, 88*, 121-151. doi: 10.1016/j.jecp.2004.03.002

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*, 1343-1359. doi: 10.1111/j.1467-8624.2007.01069.x

Geary, D. C., Hoard, M. K., & Hamson, C. O. (1999). Numerical and arithmetical cognition: Patterns of functions and deficits in children at risk for a mathematical disability. *Journal of Experimental Child Psychology, 74*, 213-239. doi: 10.1006/jecp.1999.2515

Godefroy, O. (2003). Frontal syndrome and disorders of executive functions. *Journal of Neurology, 250*, 1-6. doi: 10.1007/s00415-003-0918-2

Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology, 48*, 404-411.

Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review, 79*, 329-343. doi: 10.1037/h0032950

Haster, M. (2009). *The ins and outs of three different inhibition tasks: A study on the details of the Animal Stroop Task, the Local Global Task and the Simon Task.* Bachelor Thesis, Utrecht University, Utrecht.

Hecht, S. A. (2002). Counting on working memory in simple arithmetic when counting is used for problem solving. *Memory & Cognition, 30*, 447-455.

Hitch, G. J., & McAuley, E. (1991). Working memory in children with specific arithmetical learning difficulties. *British Journal of Psychology, 82*, 375-386.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science, 12*, F9-F15. doi: 10.1111/j.1467-7687.2009.00848.x

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*, 233-253.

Huizinga, M., Dolan, C. V., & Van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*, 2017-2036. doi: 10.1016/j.neuropsychologia.2006.01.010

Imbo, I., Duverne, S., & Lemaire, P. (2007). Working memory, strategy execution, and strategy selection in mental arithmetic. *The Quarterly Journal of Experimental Psychology, 60*, 1246-1264. doi: 10.1080/17470210600943419

Imbo, I., & Vandierendonck, A. (2007). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology, 96*, 284-309. doi: 10.1016/j.jecp.2006.09.001

Janssen, J., Scheltens, F., & Kraemer, J.-M. (2005). *Leerling- en onderwijsvolgsysteem rekenen-wiskunde [Student and education tracking system arithmetic-mathematics]*. Arnhem, the Netherlands: Cito.

Jenks, K. M., De Moor, J., & Van Lieshout, E. C. D. M. (2009). Arithmetic difficulties in children with cerebral palsy are related to executive function and working memory. *Journal of Child Psychology and Psychiatry, 50*, 824-833. doi: 10.1111/j.1469-7610.2008.02031.x

Jones, G., Gobet, F., & Pine, J. M. (2007). Linking working memory and long-term memory: A computational model of the learning of new words. *Developmental Science, 10*, 853-873. doi: 10.1111/j.1467-7687.2007.00638.x

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychology Review, 17*, 213-233. doi: 10.1007/s11065-007-9040-z

Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: Rethinking the study of developmental disorders. *Developmental Psychology, 45*, 56-63. doi: 10.1037/a0014506

Klein, P. S., Adi-Japha, E., & Hakak-Benizri, S. (2010). Mathematical thinking of kindergarten boys and girls: Similar achievement, different contributing processes. *Educational Studies in Mathematics, 73*, 233-246. doi: 10.1007/s10649-009-9216-y

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: The Guilford Press.

Klinkenberg, S., Straatemeier, M., & Van der Maas, H. L. J. (2011). Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education, Advance online publication*. doi: 10.1016/j.compedu.2011.02.003

Korkman, M., Kirk, U., & Kemp, S. (1998). *NEPSY, A developmental neuropsychological assessment*. San Antonio, TX: The Psychological Corporation.

Krajewski, K., & Schneider, W. (2009a). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: Findings from a four-year longitudinal study. *Learning and Instruction, 19*, 513-526. doi: 10.1016/j.learninstruc.2008.10.002

Krajewski, K., & Schneider, W. (2009b). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: Findings from a 3-year longitudinal study. *Journal of Experimental Child Psychology, 103*, 516-531. doi: 10.1016/j.jecp.2009.03.009

Kray, J., & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology and Aging, 15*, 126-147. doi: 10.1037//0882-7974.15.1.126

Kroesbergen, E. H., Van 't Noordende, J. E., & Kolkman, M. E. (2011). Number sense in low-performing kindergarten children: Effects of a working memory and a number sense training. In D. Molfese (Ed.), *Reading, Writing, Mathematics and the Developing Brain: Listening to Many Voices*. New York: Springer Publications.

Kroesbergen, E. H., Van 't Noordende, J. E., Kolkman, M. E., & Huiting, R. (2010). Math and working memory in kindergarten children: Domain-specific versus domain-general working memory training. *Manuscript submitted for publication*.

Kroesbergen, E. H., Van de Rijt, B. A. M., & Van Luit, J. E. H. (2007). Working memory and early mathematics: Possibilities for early identification of mathematics learning disabilities. *Advances in Learning and Behavioral Disabilities, 20*, 1-19. doi: 10.1016/S0735-004X(07)20001-1

Lee, K., Ng, S. F., Ng, E. L., & Lim, Z. Y. (2004). Working memory and literacy as predictors of performance on algebraic word problems. *Journal of Experimental Child Psychology, 89*, 140-158. doi: 10.1016/j.jecp.2004.07.001

Lee, K., Ng, S. F., Pe, M. L., Ang, S. Y., Hasshim, M. N. A. M., & Bull, R. (2011). The cognitive underpinnings of emerging mathematical skills: Executive functioning, patterns, numeracy, and arithmetic. *British Journal of Educational Psychology, Advance online publication*.

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

Lemaire, P. (2010). Executive functions and strategic aspects of arithmetic performance: The case of adults' and children's arithmetic. *Psychologica Belgica, 50*, 335-352.

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General, 124*, 83-97.

Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences, 13*, 120-126. doi: 10.1016/j.tics.2008.12.003

Linn, R. L., & Slinde, J. A. (1977). The determination of the significance of change between pre- and posttesting periods. *Review of Educational Research, 47*, 121-150. doi: 10.3102/00346543047001121

Llorente, A. M., Williams, J., Satz, P., & D'Elia, L. F. (2003). *Children's Color Trail Test*. Lutz: Psychological Assessment Resources, Inc.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin, 68*, 304-305.

Luciana, M., & Nelson, C. A. (1998). The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologia, 36*, 273-293. doi: 10.1016/S0028-3932(97)00109-7

Mabbott, D. J., & Bisanz, J. (2003). Developmental change and individual differences in children's multiplication. *Child Development, 74*, 1091-1107. doi: 10.1111/1467-8624.00594

Majerus, S., Poncelet, M., Van der Linden, M., Albouy, G., Salmon, E., Sterpenich, V., ... Maquet, P. (2006). The left intraparietal sulcus and verbal short-term memory: Focus of attention or serial order? *NeuroImage, 32*, 880-891. doi: 10.1016/j.neuroimage.2006.03.048

McLean, J. F., & Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology, 74*, 240-260. doi: 10.1006/jecp.1999.2516

Messer, M. H., Leseman, P. P. M., Boom, J., & Mayo, A. Y. (2010). Phonotactic probability effects in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology, 105*, 306-323. doi: 10.1016/j.jecp.2009.12.006

Milo, B., & Ruijssenaars, W. (2002). Strategiegebruik van leerlingen in het speciaal basisonderwijs bij optellen en aftrekken tot 100: Begeleiden of sturen? [Strategy use in addition and subtraction below 100 by pupils in special education: Guiding or structuring?]. *Pedagogische Studiën, 79*, 117-129.

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. doi: 10.1006/cogp.1999.0734

Muthén, B. O. (1998-2004). *Mplus Technical Appendices*. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide. Sixth edition*. Los Angeles, CA: Muthén & Muthén.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System*. IL: Riverside: Itasca.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*, 353-383. doi: 10.1016/0010-0285(77)90012-3

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

Noël, M. P. (2009). Counting on working memory when learning to count and to add: A preschool study. *Developmental Psychology, 45*, 1630-1643. doi: 10.1037/a0016224

Noël, M. P., Seron, X., & Trovarelli, F. (2004). Working memory as a predictor of addition skills and addition strategies in children. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 22*, 3-25.

Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. Davidson, G. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. 4, pp. 1-18). New York: Plenum Press.

Ottem, E. J., Lian, A., & Karlsen, P. J. (2007). Reasons for the growth of traditional memory span across age. *European Journal of Cognitive Psychology, 19*, 233-270. doi: 10.1080/09541440600684653

Panaoura, A., & Philippoua, G. (2007). The developmental change of young pupils' metacognitive ability in mathematics in relation to their cognitive abilities. *Cognitive Development, 22*, 149-164. doi: 10.1016/j.cogdev.2006.08.004

Passolunghi, M. C., Mammarella, I. C., & Altoè, G. (2008). Cognitive abilities as precursors of the early acquisition of mathematical skills during first through second grades. *Developmental Neuropsychology, 33*, 229-250. doi: 10.1080/87565640801982320

Passolunghi, M. C., & Siegel, L. S. (2001). Short-term memory, working memory, and inhibitory control in children with difficulties in arithmetic problem solving. *Journal of Experimental Child Psychology, 80*, 44-57. doi: 10.1006/jecp.2000.2626

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology, 88*, 348-367. doi: 10.1016/j.jecp.2004.04.002

Passolunghi, M. C., Vercelloni, B., & Schadee, H. (2007). The precursors of mathematics learning: Working memory, phonological ability and numerical competence. *Cognitive Development, 22*, 165-184. doi: 10.1016/j.cogdev.2006.09.001

Posner, M. I., & Rothbart, M. K. (2005). Influencing brain networks: Implications for education. *Trends in Cognitive Sciences, 9*, 99-103. doi: 10.1016/j.tics.2005.01.007

Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology, 91*, 137-157. doi: 10.1016/j.jecp.2005.01.004

Raven, J. (1962). *Coloured Progressive Matrices*. New York: The Psychological Corporation.

Raven, J. C. (1958). *Standard Progressive Matrices*. London: H.K. Lewis & Co, Ltd.

Resing, W. C. M., Ruijssenaars, A. J. J. M., & Bosma, T. (2002). Dynamic assessment: Using measures for learning potential in the diagnostic process. In G. M. Van der Aalsvoort, W. C. M. Resing & A. J. J. M. Ruijssenaars (Eds.), *Advances in cognition and educational practice* (Vol. 7, pp. 29-64). New York: Elsevier Science.

Roncadin, C., Pascual-Leone, J., Rich, J. B., & Dennis, M. (2007). Developmental relations between working memory and inhibitory control. *Journal of the International Neuropsychological Society, 15*, 59-67. doi: 10.1017/S1355617707070099

Rose, S. A., Feldman, J. F., Jankowski, J. J., & Van Rossem, R. (2005). Pathways from prematurity and infant abilities to later cognition. *Child Development, 76*, 1172-1184.

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

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London (Biology), 298*, 199-209.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skills. *Journal of Experimental Psychology: General, 117*, 258-275.

Siegler, R. S. (1996). *Emerging minds, the process of change in children's thinking*. New York: Oxford University Press.

Siegler, R. S. (2007). Cognitive variability. *Developmental Science, 10*, 104-109. doi: 10.1111/j.1467-7687.2007.00571.x

Siegler, R. S., & Svetina, M. (2002). A microgenetic/cross-sectional study of matrix completion: Comparing short-term and long-term change. *Child Development, 73*, 793-809.

Simon, J. R., & Berbaum, K. (1990). Effect of conflicting cues on information processing: The 'Stroop effect' vs. the 'Simon effect'. *Acta Psychologica, 73*, 159-170.

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology, 51*, 300-304.

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology, 59*, 745-759. doi: 10.1080/17470210500162854

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

Swanson, H. L. (2006a). Cognitive processes that underlie mathematical precociousness in young children. *Journal of Experimental Child Psychology, 93*, 239-264. doi: 10.1016/j.jecp.2005.09.006

Swanson, H. L. (2006b). Cross-sectional and incremental changes in working memory and mathematical problem solving. *Journal of Educational Psychology, 98*, 265-281. doi: 10.1037/0022-0663.98.2.265

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 3*, 471-491. doi: 10.1037/0022-0663.96.3.471

Swanson, H. L., & Jerman, O. (2006). Math disabilities: A selective meta-analysis of the literature. *Review of Educational Research, 76*, 249-274.

Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology, 79*, 294-321. doi: 10.1006/jecp.2000.2587

Swanson, L., & Kim, K. (2007). Working memory, short-term memory, and naming speed as predictors of children's mathematical performance. *Intelligence, 35*, 151-168. doi: 10.1016/j.intell.2006.07.001

Szmalec, A., Vandierendonck, A., & Kemps, E. (2005). Response selection involves executive control: Evidence from the selective interference paradigm. *Memory & Cognition, 33*, 531-541.

Tam, H., Jarrold, C., Baddeley, A. D., & Sabatos-DeVito, M. (2010). The development of memory maintenance: Children's use of phonological rehearsal and attentional refreshment in working memory tasks. *Journal of Experimental Child Psychology, 107*, 306-324. doi: 10.1016/j.jecp.2010.05.006

Tamnes, C. K., Østby, Y., Walhovd, K. B., Westlye, L. T., Due-Tønnessen, P., & Fjell, A. M. (2010). Neuroanatomical correlates of executive functions in children and adolescents: A magnetic resonance imaging (MRI) study of cortical thickness. *Neuropsychologia, 48*, 2496-2508. doi: 10.1016/j.neuropsychologia.2010.04.024

Toll, S. W. M., Van der Ven, S. H. G., Kroesbergen, E. H., & Van Luit, J. E. H. (2010). Executive functions as predictors of math learning disabilities. *Journal of Learning Disabilities, advance online publication*. doi: 10.1177/0022219410387302

Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2002). Efficiëntie en adaptiviteit van strategiegebruik bij elementaire rekensommen bestudeerd via de 'choice/no-choice'-methode [Efficiency and adaptivity of strategy use for elementary math problems investigated with the choice/no choice method]. *Pedagogische Studiën, 79*, 89-102.

Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2004). Strategy development in children with mathematical disabilities: Insights from the choice/no-choice method and the chronological-age/ability-level–match design. *Journal of Learning Disabilities, 37*, 119-131. doi: 10.1177/00222194040370020301

UNESCO Institute for Statistics. (1999). *Operational Manual for ISCED-1997*.

Van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review, 113*, 842-861. doi: 10.1037/0033-295X.113.4.842

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. doi: 10.1016/j.jecp.2003.12.002

Van der Sluis, S., De Jong, P. F., & Van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence*, 427-449. doi: 10.1016/j.intell.2006.09.001

Van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2011a). The development of executive functions and early mathematics: A dynamic relationship. *British Journal of Educational Psychology, Advance online publication*.

Van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2011b). The structure of executive functions in children: A closer examination of inhibition, shifting, and updating. *Manuscript submitted for publication*.

Vandierendonck, A., Szmalec, A., Deschuyteneer, M., & Depoorter, A. (2007). Towards a multicomponent view of executive control: The case of response selection. In N. Osaka, R. H. Logie & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory* (pp. 247-259). Oxford: Oxford University Press.

Wagenmakers, E. J., Van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14*, 3-22.

Wainer, H. (1991). Adjusting for differential base rates: Lord's paradox again. *Psychological Bulletin, 109*, 147-151.

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology, 102*, 43-53. doi: 10.1037/a0016738

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

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58*, 475-482.

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent Structure. *Developmental Psychology, 44*, 575-587. doi: 10.1037/0012-1649.44.2.575

Wilson, K. M., & Swanson, H. L. (2001). Are mathematics disabilities due to a domain-general or domain-specific working memory deficit? *Journal of Learning Disabilities, 34*, 237-248. doi: 10.1177/002221940103400304

Wright, I., Waterman, M., Prescott, H., & Murdoch-Eaton, D. (2003). A new Stroop-like measure of inhibitory function development: Typical developmental trends. *Journal of Child Psychology and Psychiatry, 44*, 561-575. doi: 10.1111/1469-7610.00145

Wu, S. S., Meyer, M. L., Maeda, U., Salimpoor, V., Tomiyama, S., Geary, D. C., & Menon, V. (2008). Standardized assessment of strategy use and working memory in early mental arithmetic performance. *Developmental Neuropsychology, 33*, 365-393. doi: 10.1080/87565640801982445

Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68*.

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science, 20*, 423-428. doi: 10.1111/j.1467-9280.2009.02322.x

Zook, N. A., Davalos, D. B., DeLosh, E. L., & Davis, H. P. (2004). Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks. *Brain and Cognition, 56*, 286-292. doi: 10.1016/j.bandc.2004.07.003

Samenvatting
(Summary in Dutch)

Een van de belangrijkste doelen van het basisonderwijs is om kinderen goed en vlot te leren rekenen. Alhoewel dit de meeste kinderen goed lukt, is er toch een minderheid die grote moeite blijft houden met het rekenen. Ook tussen kinderen met goede rekenvaardigheden zijn er grote verschillen. Deze kinderen krijgen allemaal ongeveer hetzelfde onderwijs, dus deze verschillen tussen kinderen moeten ergens anders vandaan komen. Vroeger werd er vaak naar intelligentie als onderliggende factor gekeken, maar een intelligentiescore omvat vele facetten van het cognitief functioneren en biedt dus een onvoldoende precieze verklaring. Bovendien verklaren prestaties op intelligentietests lang niet alle verschillen tussen kinderen. Er moet dus verder gezocht worden naar andere verklaringen.

In de laatste jaren is er steeds meer onderzoek gekomen dat zich richt op het executief functioneren als mogelijke onderliggende factor. Executieve functies zijn een verzamelterm voor alle cognitieve processen die ons in staat stellen tot doelbewust gedrag. Hierbij kan bijvoorbeeld gedacht worden aan vooruit plannen, controleren of alles volgens plan verloopt en gedrag bijsturen wanneer dit nodig is. Ook het vermogen flexibel te reageren en het vermogen je niet te laten afleiden door niet-relevante dingen die er in de omgeving gebeuren vallen onder het executief functioneren. Executieve functies zijn extra belangrijk wanneer iemand geen routines heeft ontwikkeld om met een situatie om te gaan, dus in nieuwe situaties.

Op school komen kinderen voortdurend in nieuwe leersituaties terecht: ze krijgen telkens nieuwe leerstof aangeboden. In het rekenenonderwijs leren kinderen steeds nieuwe typen sommen waarvoor ze zich goede oplossingsstrategieën eigen moeten maken. In het huidige onderwijs waarin het realistisch rekenen centraal staat, moeten kinderen bovendien uit een context vol plaatjes en verhaaltjes een som weten te halen, waarbij ze de juiste informatie moeten selecteren en zich niet moeten laten afleiden door informatie die niet relevant is. Het ligt dan ook voor de hand om te vermoeden dat executieve functies een belangrijke rol zouden kunnen spelen in het leren rekenen. Hiernaar wordt dan ook steeds meer onderzoek verricht. Deze dissertatie is een verdieping van dit onderzoek, waarin de structuur van executieve functies is onderzocht en er is gekeken naar hoe executieve functies en rekenvaardigheden zich ontwikkelen bij kinderen in groep 3 en 4.

Als eerste stond de structuur van executieve functies centraal. Voordat verbanden met rekenen onderzocht kunnen worden, moet immers duidelijk zijn wat executieve functies nu precies zijn. Alhoewel duidelijk is dat executieve functies belangrijk zijn voor planmatig, flexibel en doelgericht gedrag, is er minder consensus over hoe dit precies in zijn werk gaat. Zijn er meerdere, min of meer onafhankelijke processen die samenwerken en elkaar beïnvloeden, en die los van elkaar bij iemand sterk of minder sterk ontwikkeld kunnen zijn? Of is het executief functioneren één groot geheel waarin geen onderdelen te onderscheiden zijn? En als er meerdere executieve functies zijn, welke zijn dat dan? In het verleden zijn al pogingen gedaan om dat te onderzoeken. Een onderscheid dat vaak is gebruikt is een onderscheid in inhibitie, shifting, en updating als drie onafhankelijke executieve functies. Inhibitie is het vermogen om een dominante neiging te onderdrukken. Een taak die vaak wordt gebruikt om inhibitie te meten is de zogenaamde Stroop-taak, waarbij een kleurwoord is gedrukt in een andere kleur inkt, bijvoorbeeld **blauw**. De dominante neiging om het woord te lezen ('blauw') moet worden onderdrukt en in plaats daarvan moet de kleur van de inkt worden benoemd ('zwart'). Shifting houdt het flexibel kunnen wisselen tussen verschillende taken of regels in. Dit wordt vaak gemeten met een taak waarin er twee regels worden gegeven en de participant steeds moet kiezen welke regel er toegepast moet worden. Gekleurde vormen moeten bijvoorbeeld afwisselend op vorm en op kleur gesorteerd worden, afhankelijk van een bepaalde cue. Een voorbeeld uit het huidige onderzoek: als er een hond in beeld is moet er op kleur gesorteerd worden, maar als er een kikker in beeld is op vorm. Updating, tenslotte, is het vermogen om informatie kort vast te houden en die informatie steeds te vervangen door nieuwe, relevantere informatie. Dit is in dit onderzoek gemeten met een taak waarin kinderen allerlei plaatjes te zien kregen maar alleen het laatste plaatje van bepaalde categorieën moesten onthouden, bijvoorbeeld het laatste dier. Elk dier moest dus onthouden worden omdat het mogelijk de laatste kon zijn, totdat er een nieuw dier verscheen.

Inhibitie, shifting en updating zouden alledrie van belang kunnen zijn voor het rekenen: inhibitie zou belangrijk kunnen zijn omdat er irrelevante informatie uit een som moet kunnen worden onderdrukt. Bovendien moet een kind het gebruik van oude, ingesleten strategieën zoals tellen op de vingers kunnen onderdrukken wanneer er nieuwe, handigere strategieën worden aangeboden. Shifting zou een rol kunnen spelen bij het wisselen tussen strategieën: voor de ene som is de ene strategie handig, maar voor een andere som kan een andere aanpak juist efficiënt zijn. Updating, tenslotte, kan een rol spelen zijn omdat er belangrijke informatie uit een (verhaaltjes-)som onthouden moet worden en omdat tijdens het uitrekenen van een som de tussenstappen onthouden moeten worden.

Of inhibitie, shifting en updating echt bestaan als afgescheiden cognitieve processen is echter nog steeds discutabel; in sommige studies is daar empirische ondersteuning voor gevonden, maar in andere niet. Dit heeft te maken met het zogenaamde impurity problem. Dit houdt in dat het onmogelijk is om executieve functies afzonderlijk te meten. Er spelen ook altijd andere processen een rol bij het maken van een taak. Wanneer er bijvoorbeeld gewisseld moet worden tussen plaatjes sorteren op vorm en op kleur, moet eerst het plaatje in de hersenen verwerkt worden. Visuele verwerking speelt dus ook een rol. Bovendien moet het sorteren zo snel mogelijk gebeuren, dus het is ook van belang om een snelle motoriek te hebben. Een lage score op een dergelijke taak betekent dus niet per definitie dat een participant zwak is in shifting, maar kan ook door deze andere processen worden veroorzaakt. Een oplossing hiervoor is het aanbieden van meerdere taken per executieve functie; taken die zo veel mogelijk van elkaar verschillen in alle andere processen. Vervolgens kan er met een factoranalyse worden vastgesteld of deze taken bij elkaar clusteren en er drie aparte factoren te extraheren zijn: een factor voor inhibitie, een andere voor shifting en een derde factor voor updating. In het verleden is dit echter niet altijd even zorgvuldig gedaan, wat mogelijk een verklaring zou kunnen zijn voor de tegenstrijdige uitkomsten van eerder onderzoek. Een andere mogelijke bron van verschillen in uitkomsten tussen studies is dat de structuur van executieve functies in de kindertijd anders zou kunnen zijn dan op volwassen leeftijd. De hersenen zijn in de kindertijd immers nog volop in ontwikkeling. In deze dissertatie is daarom gekeken of de zogenaamde driefactorstructuur van inhibitie, shifting, en updating terug te vinden was bij kinderen in groep 3 en 4 en of de gevonden factoren inderdaad samenhingen met rekenprestaties en de ontwikkeling daarvan.

Eerst is er in een pilotstudie, gepresenteerd in *Hoofdstuk 2*, bij een groep van 26 kinderen in groep 3 gekeken of de structuur van inhibitie, shifting, en updating terug te vinden was. De kinderen hebben voor elke executieve functie twee taken gemaakt en met behulp van een exploratieve factoranalyse is er gekeken of de scores van de kinderen op deze taken clusterden tot drie onafhankelijke factoren. Dit bleek inderdaad het geval. Vervolgens zijn deze kinderen vier weken lang gevolgd in het leren optellen over het tiental (bijvoorbeeld 8 + 8). Er werd gekeken of er samenhang was tussen de drie executieve functies en de leerprestaties van de kinderen gedurende deze vier weken. Dit bleek voor updating het geval: kinderen die goed waren in updating, waren ook goed in de optelsommen en gingen bovendien ook meer vooruit. Voor shifting waren er ook verbanden aanwezig, maar zwakker. Verbanden met inhibitie bleken verwaarloosbaar. Er werd echter vermoed dat de rol van inhibitie op latere leeftijd groter zou kunnen worden, als de sommen moeilijker worden en kinderen beschikken over een uitgebreider rekenrepertoire. Er is dan immers ook meer dat potentieel onderdrukt moet worden. De veelbelovende resultaten van de pilot gaven daarom aanleiding tot het opzetten van een uitgebreidere, longitudinale studie waarin 211 kinderen gevolgd werden in groep 3 en groep 4. Vier keer is er bij deze kinderen een uitgebreide testbatterij van executieve functies afgenomen. Er werden steeds drie taken per executieve functie afgenomen en er werd niet alleen gekeken naar hoeveel fouten de kinderen maakten maar ook naar de snelheid van werken tijdens de taak. Bovendien werden de taken zo veel mogelijk gevarieerd om zo het impurity problem te vermijden. De Citotoetsen Rekenen-Wiskunde die de scholen zelf twee keer per jaar afnamen dienden als rekenscore en ongeveer de helft van de kinderen nam bovendien deel aan een diepteonderzoek naar het leren vermenigvuldigen.

In *Hoofdstuk 3* werd de structuur van de executieve functies grondig onder de loep genomen. Dit keer werd er gebruik gemaakt van Structural Equation Modeling: met een confirmatieve factoranalyse werd er weer gekeken of er bewijs te vinden was voor een driefactorstructuur van inhibitie, shifting, en updating. Dit bleek niet het geval: updating was duidelijk te onderscheiden, maar inhibitie en shifting laadden steeds samen op één factor. Bovendien bleek dat scores op inhibitie- en shiftingtaken ook heel goed voorspeld konden worden door simpele reactietijdmetingen, waarbij het kind bijvoorbeeld zo snel mogelijk een simpel plaatje moest benoemen of zo snel mogelijk op een bepaalde toets moest drukken als er een plaatje op het computerscherm verscheen. Kortom, er is geen bewijs gevonden voor de onderscheidbaarheid in drie afzonderlijke executieve functies in deze leeftijdsgroep: vooral het bestaan van inhibitie en shifting is discutabel.

In *Hoofdstuk 4* werd er gekeken naar de ontwikkeling van executieve functies gedurende de vier metingen die werden afgenomen in de loop van de twee jaar van het onderzoek. Bovendien werd in dit hoofdstuk de relatie onderzocht met prestaties van de kinderen op de citotoetsen Rekenen-Wiskunde die zij ook vier keer maakten gedurende deze periode. Aangezien in het vorige hoofdstuk bleek dat inhibitie en shifting niet van elkaar te onderscheiden waren, zijn zij ook in dit hoofdstuk samengenomen als één factor: inhibitie/shifting. Allereerst bleek dat op er elke meting een sterke samenhang was tussen updating en rekenen. Hoe beter een kind was in updating, des te hoger waren gemiddeld genomen ook de rekenprestaties van dit kind. Het verband tussen inhibitie/shifting en rekenen was veel zwakker en dit verband verdween bovendien wanneer inhibitie/shifting en updating samen werden geanalyseerd. Het meten van inhibitie en shifting bleek dus geen meerwaarde te bieden om verschillen in rekenprestatie tussen kinderen te kunnen verklaren.

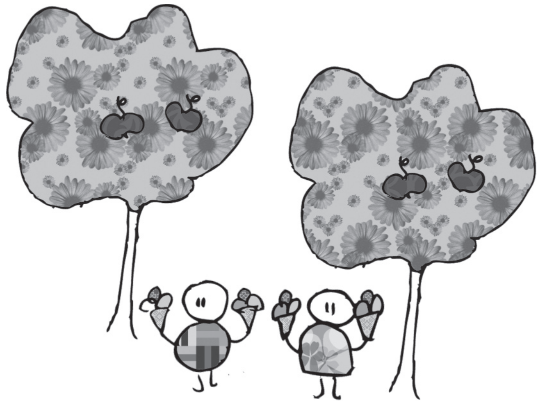
Vervolgens werden er latente groeimodellen (Latent Variable Growth Curve Models) gemaakt om de ontwikkeling te analyseren. Uit deze analyses bleek allereerst dat prestaties op zowel de updating- als de rekentaken lineair toenamen gedurende de twee jaar. Voor zowel updating als rekenvaardigheid gold bovendien dat er geen verband was tussen het algemene niveau (de intercept) van de kinderen en de vooruitgang gedurende de twee jaar (de slope). Dat betekent dat zwakke kinderen gemiddeld hun achterstand hielden: zij liepen deze niet in, maar hun achterstand werd ook niet groter. Op de inhibitie/shifting factor was er geen sprake van een duidelijke lineaire toename.

Updating en rekenvaardigheid bleken ook gedurende de ontwikkeling met elkaar samen te hangen. Niet alleen bleek dat het algemene niveau van updating samenhing met het algemene niveau van rekenen (een intercept-intercept verband), maar bovendien bleek dat kinderen die zich sterk hadden ontwikkeld op het gebied van updating, ook veel vooruitgegaan waren in rekenvaardigheid (een slope-slope verband). Dit ondersteunt de hypothese dat updating een belangrijke vaardigheid is die nodig is om te kunnen leren rekenen.

In *Hoofdstuk 5* is de relatie tussen updating en rekenvaardigheid verder onder de loep genomen. Er werd gekeken hoe het verband tussen updating en rekentoetsen tot stand komt: kiezen kinderen met goede updatingvaardigheden slimmere strategieën, maken ze ongeacht hun strategiekeuze minder fouten bij het uitvoeren van strategieën, of misschien allebei? Dit is onderzocht in een microgenetische studie, waarin heel gedetailleerd is gekeken hoe kinderen leren vermenigvuldigen. Gedurende acht weken in groep 4 hebben 98 kinderen uit de grote steekproef elke week een aantal vermenigvuldigsommen gemaakt, variërend van het vrij simpele 7 x 2 tot 6 x 8, wat voor deze kinderen een behoorlijk pittige som was. Er is daarbij niet alleen gekeken of de uitkomst goed was, maar ook naar welke strategieën de kinderen gebruikten. Er werd een onderscheid gemaakt tussen vijf typen strategieën: (1) foute strategieën, die niet tot een goed antwoord kunnen leiden, zoals optellen in plaats van vermenigvuldigen: 7 x 2 = 9, (2) telstrategieën, zoals tellen op de vingers, (3) herhaald optellen: 5 x 6 = 6 + 6 + 6 + 6 + 6, (4) handige strategieën die sneller zijn dan herhaald optellen omdat er gebruik wordt gemaakt van aanwezige kennis, bijvoorbeeld 6 x 3 = 5 x 3 + 3 = 15 + 3 en (5) geautomatiseerde kennis, waarbij het antwoord direct wordt opgehaald zonder dat er een berekening is uitgevoerd.

Het zogenaamde Overlapping Waves model stelt dat in een leerproces kinderen overgaan op het gebruik van steeds geavanceerdere strategieën, maar dat dit op een geleidelijke manier gebeurt: als een kind in het huidige onderzoek bijvoorbeeld net het herhaald optellen heeft ontdekt, zal het in het begin hier nog niet veel gebruik van maken en nog vaak terugvallen op de minder geavanceerde telstrategieën. Geleidelijk zal het gebruik van herhaald optellen echter toenemen, totdat er weer een nieuwe strategie wordt ontdekt en het patroon zich herhaalt. Dit model is echter nog nooit statistisch getoetst. In dit hoofdstuk is dit wel gebeurd en het model bleek inderdaad goed te passen. Bovendien bleek dat kinderen die slimmere oplossingsstrategieën gebruikten ook vaker het antwoord goed hadden. Tenslotte bleek dat kinderen die goed waren in updating slimmere strategieën gebruikten en bovendien minder rekenfouten maakten.

In *Hoofdstuk 6* worden de bevindingen van dit proefschrift besproken in een kritische discussie. Er is geconstateerd dat het onderscheid in inhibitie, shifting, en updating niet houdbaar is. Alhoewel er niet gegeneraliseerd kan worden naar andere leeftijdscategorieën dan in dit onderzoek, dus zo'n 6 tot 8 jaar, zijn er toch aanwijzingen dat er een nieuw theoretisch kader nodig is. In dit hoofdstuk is een nieuw model geponeerd dat uitgaat van twee basisvermogens die in de plaats komen van inhibitie, shifting, en updating: (1) het vermogen om informatie in het werkgeheugen te activeren en deze activatie in stand te houden en (2) het vermogen om irrelevante informatie in het werkgeheugen juist niet te activeren, door deze activatie te voorkomen of de aandacht op iets anders te richten. Er wordt besproken hoe dit model eerdere onderzoeksresultaten kan verklaren. Bovendien zijn er aanbevelingen gedaan voor toekomstig onderzoek: om de validiteit van het hierboven beschreven model te testen zal het empirisch getoetst moeten worden. Bovendien is het belangrijk om in vervolgonderzoek naar verbanden tussen executieve functies en rekenen een duidelijk onderscheid te maken tussen de verschillende executieve functies, aangezien ze niet allemaal even belangrijk bleken. Ook rekenen zou op een vergelijkbare manier in verschillende vaardigheden opgesplitst kunnen worden. Tenslotte bleek dat strategiekeuze een belangrijke factor is in het voorspellen van succes bij het rekenen, en updating bleek hierbij belangrijk. Dit geeft aan dat strategiegebruik een belangrijke factor is die in vervolgonderzoek meegenomen zou moeten worden en mogelijk ook een aanknopingspunt biedt voor interventiestudies voor zwakke rekenaars.



Dankwoord (Acknowledgements)

Zoals zo veel proefschriften had ook dit proefschrift zeker niet tot stand kunnen komen zonder de hulp van vele anderen. Morele en inhoudelijke steun maar ook zeker praktische hulp was onontbeerlijk. Zo waren daar de vele testassistenten die heel wat testuren met de kinderen hebben gemaakt. Willemijn Bolier, Gerdientje Danner-Vlaardingerbroek, Hanneke van den Geest, Merel Beek, Mathijs Gijsbers, Denise Meeuwissen, Nienke Streng, Noor Miedema, Karin Vander Heyden, Reggie Bax, Renée Dijks, Nicole Ermes, Nicolien Tip, Anne-Marthe Sigmond, Anniek Verheijen, Esther van Veelen, Mirjam Garritsen, Liesbeth Kreuk, Marieke Jansen, Barbara Müller, Sylke Toll, Baukje van Veen, Petra Wassenaar, Elise van Apeldoorn, Karolien van Ark, Annelieke van Heeswijk, Laurien Passtoors, Sanne Koops, Sharon Haenen, Wieke Dalenberg en Marleen Akkerman, bedankt! Het is leuk om te zien dat sommigen van jullie inmiddels zelf een promotietraject ingestapt zijn. Myrthe Haster en Constança van Ojen, jullie hebben allebei als getalenteerde studenten meegedacht over wat er nu toch allemaal een rol kon spelen bij die vreemde inhibitietaken. Ook daarvoor bedankt!

Ook de scholen die hebben deelgenomen aan het project waren uiteraard onmisbaar bij de totstandkoming van dit proefschrift. Mijn hartelijke dank gaat dan ook uit naar de kinderen, leerkrachten en directie van basisschool Het Zonnewiel in Almere, OBS De Bundel West in Apeldoorn, OBS De Swetten in Drachten, basisschool Panningen-Zuid in Panningen, de Josephschool in Pijnacker, Dorpsschool De Bron in Reeuwijk, basisschool Dr. H. Bavinck in Spakenburg, basisschool St.-Christoffel in Waalre, basisschool De Wilderen in Waalre en de Franciscusschool in Woerden. Fijn dat jullie ondanks jullie eigen volle schema's en beperkte vrije lokalen steeds weer tijd en ruimte vrij wisten maken voor dit tijdrovende project!

Kitty van der Veer en Renate Siebes, bedankt voor de mooie opmaak van het proefschrift. Zonder jullie was het lang zo mooi niet geworden!

Natuurlijk zijn er ook veel collega's die ik graag wil bedanken. Warrie Schuurman, Sylvia Girrel en alle andere collega's van La Grange: jullie zorgden voor het nodige tegenwicht. Inmiddels geef ik alweer bijna tien jaar bijles bij La Grange en al die tijd met veel plezier. Als mijn hoofd weer eens vol zat met alle analyses die weer eens niet lukten dan kon ik dat in die middag per week bij La Grange altijd van me af zetten. Ik heb er veel leerlingen zien komen en gaan: Annieck van Nimwegen, jou heb ik het langst begeleid van allemaal: bijna de gehele duur van deze promotie. Ook voor jou is het nu na je eindexamen tijd om je vleugels uit te slaan!

Ook aan de universiteit waren er uiteraard veel collega's die onmisbaar waren bij de voltooiing van dit proefschrift. Ellie van Eijk en Ruth Cramer, jullie stonden altijd klaar om bijvoorbeeld studentassistenten weer eens een contract te geven, ook al was dat vaak op het laatste moment. Jos Jaspers en Erica van Doorn, jullie waren er voor alle computerproblemen die met regelmaat de kop op staken en altijd hadden jullie een oplossing. Ook het secretariaat wil ik graag bedanken, in het bijzonder Jacqueline Tenkink-de Jong, voor alle praktische ondersteuning in de afgelopen jaren.

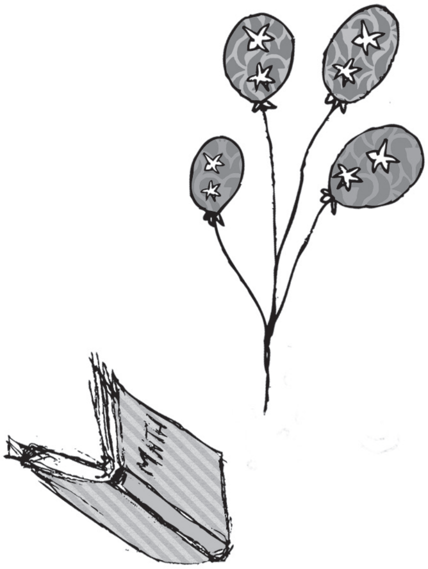
Alle mede-aio's en postdocs, bedankt voor alle gezellige lunchwandelingen, congresbezoeken en af en toe even samen stoom afblazen! Ilona van den Bos, Hanna Mulder en Josje Verhagen: bedankt voor de kritische blik op de conceptversie van dit proefschrift. Heleen Torringa, ons pad naar de promotie liep het laatste half jaar precies synchroon: jij een dag eerder dan ik. Leuk om de weg erheen met jou te kunnen delen! Ook alle overige collega's wil ik graag bedanken, in het bijzonder Lex Wijnroks: jouw immer kalme uitstraling en luisterend oor gecombineerd met je inhoudelijke interesse waren een weldaad in stressvolle periodes!

I had a wonderful stay in Singapore, thanks to Saskia Krebs, Jie Qi Lee, Jeremy Ng, Sharon Koh, Nabil Hasshim, Su Yin Ang, Monica Lim, Madeline Pe, and of course Kerry Lee, who made it all possible and even organized a real jungle tour for all of us in the middle of Singapore. Wading through the creek was brilliant, but I also learned a lot in terms of data analysis. Thank you all for the great experience!

Ook wil ik natuurlijk graag mijn begeleiders bedanken. Jullie hielden elkaar – en daardoor mij – in evenwicht. Jan Boom, jij had echt altijd mijn stukken grondig gelezen en zorgde daarbij voor de kritische noot. Dat laatste soms tot mijn frustratie, maar het eindresultaat werd er wel beter van! Ik heb bovendien veel nieuwe manieren van analyseren van je geleerd en ook de mogelijkheden van Excel verder ontdekt. Paul, jij wist steeds het onderzoek in een breder kader te plaatsen en de zin en betekenis ervan aan te geven als ik dat zelf even kwijt was door alle details. Ook wist je altijd wel weer ergens geld vandaan te halen voor congresbezoek en assistentie. En uiteraard Evelyn Kroesbergen, werken met jou was heel efficiënt: jij begreep altijd met een paar woorden precies wat ik bedoelde. Bovendien had je altijd een opbeurend woord klaar als dat nodig was!

Vicky en Jojanneke, mijn paranimfen. Niet alleen zijn jullie zelf inhoudelijk betrokken geweest bij het project, door data te verzamelen, te bewerken en te analyseren, maar ook om andere redenen waren jullie onmisbaar. Jojanneke, samen sloegen we onze werkfrustraties (niet altijd even elegant) eruit op de squashbaan. Jammer dat mijn schouder het afgelopen jaar roet in het eten gooide. Vicky, jij was onze trouwe planten- en dierenoppas die het nooit liet afweten als wij weer eens een verre of minder verre reis wilden maken. Bovendien wist je tijd vrij te maken om mee te denken over de promotie terwijl je zelf je bruiloft aan het voorbereiden was. En uiteraard waren jullie er allebei om af en toe even een kletspraatje te maken!

En dan uiteraard nog de familie. Jeanne-Marie, bij jou in de klas kon ik vanaf mijn studietijd al altijd terecht als ik weer eens wat kinderen nodig had voor een of ander onderzoek. Ook nu heeft jouw school – een van mijn oude basisscholen – meegedaan aan het onderzoek. Bruno en Nanny, jullie waren altijd geïnteresseerd in hoe het met de voortgang van het proefschrift ging. Pap, tot mijn grote verdriet heb jij de voltooiing van dit proefschrift niet meer mee mogen maken. Maar als iemand er een rotsvast vertrouwen in had dat het allemaal wel zou lukken, dan was jij het. Mam, op jou kon ik de afgelopen jaren altijd terugvallen zodat het inderdaad ook allemaal gelukt is. Wolter, jij bent zelf ook gepromoveerd terwijl ik met dit project bezig was. Je kende de frustraties en wist me steeds weer op te vrolijken. Welk continent het volgend jaar gaat worden weten we nu nog niet, maar samen kunnen we de wereld aan!



About the author

Curriculum vitae

Sanne van der Ven was born on March 16, 1982 in Waalre, the Netherlands. She obtained her high school degree (gymnasium) in 2000 from the Van Maerlantlyceum in Eindhoven, after which she attended University College Utrecht, the international honours college of Utrecht University. In 2003 she obtained her bachelor degree with honours , with a major in linguistics, neuroscience and psychology. She then studied Neuroscience & Cognition at Utrecht University. In 2006 she obtained her master's degree with honours. During these years she also obtained a bachelor's degree in psychology, also from Utrecht University. During the master programme she completed two research internships, concerning early precursors of dyslexia and the development of executive functions and mathematics. The latter project eventually resulted in this dissertation; from 2007 until 2011 Sanne worked as a PhD student at the faculty of social and behavioural sciences at Utrecht University. During her PhD she spent a month in Singapore at the National Institute of Education of Nanyang Technological University. Sanne also gained experience in teaching, as she supervised bachelor and master theses, gave various guest lectures and was involved in teaching a bachelor course.

Publications

Van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2011). The development of executive functions and early mathematics: A dynamic relationship. *British Journal of Educational Psychology, Advance online publication*. doi: 10.1111/j.2044-8279.2011.02035.x

Toll, S. W. M., Van der Ven, S. H. G., Kroesbergen, E. H., & Van Luit, J. E. H. (2010). Executive functions as predictors of math learning disabilities. *Journal of Learning Disabilities, advance online publication*. doi: 10.1177/0022219410387302

Kroesbergen, E. H., Van der Ven, S. H. G., Kolkman, M. E., Van Luit, J. E. H., & Leseman, P. P. M. (2009). Executieve functies en de ontwikkeling van (voorbereidende) rekenvaardigheid [Executive functions and the development of (preparatory) mathematical abilities]. *Pedagogische Studiën, 86*, 334-349.

Kroesbergen, E. H., Sontag, L., Van Steensel, R. C. M., Leseman, P. P. M., & Van der Ven, S. H. G. (2010). Plaatsing van leerlingen met een beperking in het reguliere onderwijs: De ontwikkeling van schoolvaardigheden, competentiebeleving en probleemgedrag [Children with special needs in mainstream education: The development of academic abilities, self-perception of competence and problem behavior]. *Pedagogische Studiën, 87*, 350-366.

Sontag, L., Kroesbergen, E. H., Leseman, P. P. M., Van Steensel, R., Van der Ven, S. H. G. & Van Wolput, B. (2007). *De werking van de leerlinggebonden financiering in het basisonderwijs en voortgezet onderwijs [the effectiveness of pupil-bound financing in primary and secondary education].* Tilburg: IVA.