UvA-DARE (Digital Academic Repository)

Microgenetic patterns of children's multiplication learning: Confirming the overlapping waves model by latent growth modeling

van der Ven, S.H.G.; Boom, J.; Kroesbergen, E.H.; Leseman, P.P.M.

DOI

[10.1016/j.jecp.2012.02.001](https://doi.org/10.1016/j.jecp.2012.02.001)

Publication date

2012

Document Version

Author accepted manuscript

Published in

Journal of Experimental Child Psychology

[Link to publication](#)

Citation for published version (APA):

van der Ven, S. H. G., Boom, J., Kroesbergen, E. H., & Leseman, P. P. M. (2012). Microgenetic patterns of children's multiplication learning: Confirming the overlapping waves model by latent growth modeling. *Journal of Experimental Child Psychology*, 113(1), 1-19. <https://doi.org/10.1016/j.jecp.2012.02.001>

General rights

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

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Microgenetic patterns of children's multiplication learning: Confirming the Overlapping Waves model by latent growth modeling

Sanne H.G. Van der Ven¹, Jan Boom², Evelyn H. Kroesbergen², & Paul P.M. Leseman²

¹University of Amsterdam

²Utrecht University

This is a post-print version (final draft, post-refereeing). For the fully typeset, publisher's version, please visit [dx.doi.org/10.1016/j.jecp.2012.02.001](https://doi.org/10.1016/j.jecp.2012.02.001) or send an email to the author: s.h.g.vandervan@uva.nl

Van der Ven, S. H. G., Boom, J., Kroesbergen, E. H., & Leseman, P. P. M. (2012).

Microgenetic patterns of children's multiplication learning: Confirming the overlapping waves model by latent growth modeling. *Journal of Experimental Child Psychology*, 113, 1-19. doi: 10.1016/j.jecp.2012.02.001

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 is related to development in accuracy and to what degree working memory is related to both. We expected that children with poor working memory are limited in their possibilities to make the associations that are necessary to progress to more mature strategies. This limitation would explain the often-found relationship between working memory and mathematical abilities.

To 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 of Overlapping Waves. Moreover, both the intercepts and the slopes of strategy selection and accuracy were strongly interrelated. Finally, working memory predicted both strategy selection and accuracy, confirming that working memory is related to mathematical problem solving in two ways, as it influences both the maturity of strategy choice and the probability of making procedural mistakes.

Microgenetic patterns of children's multiplication learning: Confirming the Overlapping Waves model by latent growth modeling

Mathematical proficiency is important in everyday life, and mathematics is one of the most important subjects children learn in school. Nevertheless, relatively little is known about how children learn mathematics, what mathematical strategies they use and how this repertoire changes over time, and how individual differences in mathematical proficiencies arise. Therefore dynamic in-depth studies are needed in which the progress that children make is followed closely, and in which measures for relevant underlying cognitive abilities are incorporated. In order to do this, we carried out a microgenetic study in which we investigated which strategies children employ when solving math problems and 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. Moreover, we expected that working memory partially determines children's success in mathematics, because working memory boundaries may influence which strategies children are able to execute, and how many procedural errors they make during the computational process.

A number of studies have already investigated strategy use in children learning mathematics: it was found that strategy selection is important in determining mathematical performance and that the strategy choice changes 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), posits that solution strategies resemble waves that rise and fall during development: after a certain strategy is discovered, the frequency with which it is used increases initially 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, when being faced with the single digit multiplication of 4×9 for the first time, a child may put up 9 fingers and count them 4 times. But once a child is sufficiently proficient in addition, repeated addition can be applied to solve the problem ($4 \times 9 = 9 + 9 + 9 + 9 = 36$). Since this strategy requires fewer steps than counting, it can be considered a more mature strategy. In the end, children rely most on memory retrieval for single digit multiplication problems (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 $10 \times 6 = 60$ can use this knowledge to solve 9×6 , as $9 \times 6 = 10 \times 6 - 6 = 60 - 6$.

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 3×2 by means of retrieval does not automatically imply that this child will solve 8×7 by means of retrieval too. Retrieval is only an efficient strategy if the child knows the math fact, 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 repeated addition decreased while the frequency of retrieval increased, 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 individual 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 (for a review, see Geary, 2004). The question arises why children with mathematical disabilities use less mature strategies than typically developing children.

We hypothesize that one important factor may be the ability to sustain and update simultaneous representations of the math problem. Solving math problems is complex: it requires keeping track of representations of (1) the math problem, (2) the steps that already have been carried out and the actions that still have to be performed and (3) the intermediate outcomes, yielded by the steps that have been carried out. If all three representations are maintained and updated fast and efficiently, the final solution is reached quickly. This will then strengthen the connections between a problem and the answer, or partial answers (Geary, Brown, & Samaranayake, 1991; Siegler, 1996), which enables the use of more mature strategies using shortcuts in the future. In our example problem of 4×9 , a child that is able to (1) maintain the goal: solving 4×9 , (2) update how many 9s have been added and (3) update what the total count is so far, will be able to solve the problem quickly. This child may therefore soon learn the association of the intermediate step of $9 + 9 = 18$ and therefore no longer need to count fingers when the same problem is shown again: this child is ready to progress from counting to repeated addition. A child that is less able to maintain and update these three representations will be slower and more likely to make procedural errors (counting one too many or adding too many or too few 9s because of losing track of the items). This leads to incorrect associations between operands and answer (Geary, 1993; Geary et al., 1991; Noël, Seron, & Trovarelli, 2004). This child is also more likely to have forgotten what the math problem was, once the answer is reached. For example, the child may not have learned that $9 + 9 = 18$ and has to rely on counting for a longer time.

Temporarily maintaining and updating information in problem solving reflects the operation of what is commonly referred to as the working memory system. There are currently several theoretical accounts of working memory, which have in common the distinction between storage and the operation of executive processes involving attention (for a review, see Ricker, AuBuchon, & Cowan, 2010). It is beyond the scope of this article to discuss the differences and similarities of these approaches extensively. In the present study we will use the term working memory for the executive processes involved in temporarily maintaining and

updating information, i.e., we do not refer to mere storage. In the method section we will explain how working memory was operationally defined and measured.

A relationship between working memory and mathematical abilities in children has been found in many studies, both in the normally developing population and in children with mathematical disabilities, who have been shown to have lower working memory abilities than their peers without learning disabilities (for a review, see Raghubar, Barnes, & Hecht, 2010).

Most of these studies focused on mathematical problem solving accuracy, leaving aside the question which strategies were used. However, strategy use may be an important factor explaining the relationship between working memory and mathematical performance. As in the example mentioned before, children with poor working memory are expected to have problems carrying out counting strategies, particularly when math problems contain large operands. Indeed, young children with low working memory have been shown to make more counting errors in simple addition problems (Geary et al., 1991). As these errors prevent the child from moving towards a more mature strategy, weak working memory can therefore impede children from progressing to more mature retrieval-based strategies (Siegler, 1988).

Because the less mature strategies require far more steps to keep track of than the more mature strategies, children with low working memory abilities are expected to face an extra difficulty: they cannot progress to more mature strategies, but these less mature strategies require far more working memory resources (Imbo & Vandierendonck, 2007). Especially the immature counting strategy poses a high load on working memory, because of the large number of steps (Cragg & Gilmore, 2011; Hecht, 2002).

Empirical evidence 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 more advanced 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 when strategies are chosen that require multiple steps to keep track of. Success in these early stages likely enables transitions to more mature strategies, which require fewer steps and are therefore faster and less error-prone, and eventually enable relying on retrieval. Retrieval requires few working memory resources (Imbo & Vandierendonck, 2007; but see Wu et al., 2008).

As described before, strategy use is likely to change over time and working memory is likely to be involved in the acquisition of new strategies as well as in the execution of well-learned strategies, while strategies may differ in the load they place on working memory. Nevertheless, so far, most studies included only one measurement and often the problems that were used were relatively easy for the age range studied. Lacking is an integrative study in which the dynamics of development in strategy selection and accuracy are modeled together and related to working memory.

The present study was designed to provide such an integrated analysis of the acquisition of one particular 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 aspects of math learning.

Our first aim was to confirm the validity of 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 the present 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 been applied thus far, something greatly needed in microgenetic research (Cheshire, Muldoon, Francis, Lewis, & Ball, 2007).

The second aim of the present research was to investigate the role of working memory in learning single-digit multiplication. We examined whether individual differences in working memory were related to the maturity of the selected

strategies and to development in using more mature strategies, 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 earlier use of more mature strategies such as retrieval. In addition, we expected that regardless of the strategy that was chosen, children with strong working memory abilities 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. The children also took part in a longitudinal study, described in more detail in Van der Ven, Kroesbergen, Boom, & Leseman (2011).

Procedure

Once a week, during a period of eight weeks, always on the same weekday, the children were given a math test, consisting of a booklet with 15 simple multiplication problems. The test was administered individually in a separate room by a trained research assistant (mostly graduate students). The problems were presented in a booklet. Each page contained two problems; the problem that the child was not working on was occluded.

Some of the problems were embedded in a story context. The research assistant read these stories out loud while the child could also read the text. Children were encouraged to solve the problem the way they wanted to and either explain or write down their strategy afterwards. They were allowed to use paper and pencil.

When they did not use this, as was the case for the majority of the children, 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 their strategies (Wu et al., 2008). The procedure lasted approximately 20 minutes per child per session.

Measurements

Math problems. Each week, the children solved 15 math problems. Variation in these problems was necessary to reduce the possible influence of unwanted memory effects on practiced problems (e.g. a child remembering that ‘the answer to that very big math problem is 48’). The 15 math problems were therefore drawn from a pool of 28 single-digit multiplication problems. We used a large range of difficulty (from 3×2 to 6×8) to ensure the presence of math problems that could capture the children’s mathematical development during the study. The easiest problems also served as positive fillers, to keep the children, especially those with low mathematical abilities, motivated.

Each week, 15 items were drawn from the pool of 28 items. To ensure a roughly constant difficulty of the test throughout the eight weeks, the 28 math problems were first grouped into 15 increasingly difficult subsets. Each subset contained 1-3 problems that were approximately equally difficult, based on previous research (Klinkenberg, Straatemeier, & Van der Maas, 2011). For each session, one item from each subset was selected randomly, yielding a total of 15 problems per session. The problem sets and the problem that was selected for each session are presented in Table 1.

The problems in eight of the problem sets (see Table 1, second column) were presented as plain numerical problems. The problems from the seven remaining sets were embedded in a context: they were presented as story problems and were accompanied by 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. The story contained sufficient information to solve the problem, but in the accompanying illustration both factors could also be obtained by counting. An example item can be found in Figure 1.

Table 1
Math problems.

Problem set	Con-text	Used in final analyses	W1	W2	W3	W4	W5	W6	W7	W8
1	No	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	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	No	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	No	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	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	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	Yes	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	No	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	Yes	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	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	Yes	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	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	Yes	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	No	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	Yes	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. For the purpose of presentation, the first-appearing problem of each problem set is presented in regular font, the second in boldface and the third in italics.

The specific story and illustration of each context problem was different in each session. The order of the problems was randomized on each session, with the restrictions that (1) the first two problems were always selected from the easiest five of the session, and (2) problems that were presented on the same page did not share one of the factors. Each child received the problems in the same order.

Working memory. It has been shown that working memory tasks, like other executive function tasks, suffer from the impurity problem (Huizinga, Dolan, & Van der Molen, 2006; Miyake et al., 2000; Van der Ven et al., 2011). This means that tasks that are supposed to be pure measures of executive functioning, also measure other, lower-level abilities.

“There are 9 boxes in the basement. Each box contains 3 bottles.
How many bottles are there in the basement?”

Figure 1. Example math item.

as possible on non-essential characteristics and subsequently carrying out a confirmatory factor analysis. The resulting factor reflects a more pure, higher-order measure of working memory, as has been successfully shown in many studies (Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Miyake et al., 2000; Van der Sluis, De Jong, & Van der Leij, 2007).

Therefore three tasks were administered, as different from each other as possible. The data from these three tasks have also been used in our longitudinal study (Van der Ven et al., 2011). We administered a verbal complex span task (Digit Span Backwards), a visuospatial complex span task (Odd One Out), and a memory updating task (Keep Track). In complex span tasks, a number of items must be remembered while at the same time another task is carried out. In updating tasks, the to-be-remembered information changes during the task: e.g., the last presented animal must be remembered.

Combining these two types of tasks minimized commonalities in factor structure due to the task administration, while still measuring the same concept. Complex span tasks and updating tasks have been shown to be strongly interconnected: they loaded on the same factor in a factor analysis (St Clair-Thompson & Gathercole, 2006) and evoked similar brain activation (Nulsen, Fox, & Hammond, 2010; Segalowitz, Wintink, & Cudmore, 2001).

As expected, in a factor analysis the three tasks loaded on one factor, which explained 51% of the total variance. The loadings on this factor were .59 for Digit Span Backwards, .75 for Odd One Out and .78 for Keep Track.

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. Test-retest reliability of this measure in AWMA is .64 (Alloway, Gathercole, & Pickering, 2006).

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 rectangles next to each other, each containing a shape. One of the shapes differed from the other two. The child was asked to point at the deviant shape. Then the next stimulus, consisting of three new rectangles with shapes, was presented. At the end of each trial three empty rectangles 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. Test-retest reliability of this measure in AWMA is .81 (Alloway et al., 2006)

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 designated categories were shown in the bottom of the screen, serving as a reminder.

The number of to-be-remembered categories within a series increased from one to four; always the final item of each category had to be recalled. There were two series of each difficulty level, yielding a total of eight series. Each correct answer was noted. Depending on the number of the to-be-remembered categories, 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. Published reliability measures are not available, but its internal consistency in this study, as measured by Cronbach's alpha, was .70.

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 ($7 \times 4 = 7$), addition ($7 \times 4 = 11$), 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 ($5 \times 7 = 7 + 7 + 7 + 7 + 7$), sometimes in smaller steps ($3 \times 7 = 5 + 5 + 5 + 2 + 2 + 2$), 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 6×4 : $4 + 4 + 4 = 12$; $12 + 12 = 24$), or using the easier 5 or 10 times tables ($9 \times 3 = 10 \times 3 - 3$). 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 6×7 by stating $5 \times 7 = 35$, 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., $4 \times 8 = 10 + 10 + 10 + 2 = 32$), 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.

Results

The data are complex and therefore required complex statistical analyses. For reasons explained in the next sections, eight of the fifteen math problem sets (together containing fifteen different problems) were used for further analyses: problem sets 6, 7, 9-13 and 15. The descriptive statistics of these eight problem sets are presented in Figure 2 and Figure 3. Figure 2 shows the raw group means of the frequencies of each strategy type; both averaged over all eight selected problem sets and, as an illustration, for two individual problems that were administered every week.

Figure 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 difficulty: e.g., retrieval was applied more often to 7×3 , which was the easier problem. A similar result was obtained for accuracy, expressed as the proportion of correct answers to these problems, as is illustrated in Figure 3: in general there was an increase in accuracy over time, and the easier 7×3 was solved correctly more often than the more difficult 9×4 . Furthermore, the more mature the strategy, the higher the likelihood of a correct answer, as expected: 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.

Figure 2. Descriptive strategy data: observed average frequencies of selected strategies, averaged for all eight problems and, as an illustration, for two example problems.

Figure 3. Descriptive accuracy data: proportion of problems answered correctly over time, averaged for all eight problems and, as an illustration, for two example problems.

Although informative at the group level, Figures 2 and 3 display results that are averaged over individuals and therefore do not contain information about individual differences in development. For example, a horizontal line for a certain strategy in group data does not imply stability in the use of this strategy at the individual level: its frequency might have decreased in some children while having increased in others. Moreover, there may be variability in strategy choice at the individual level. This was indeed the case: on average, each individual child used 2.68 of the 5 strategy types in each week to solve the eight problems. In 7.1 % of the cases only 1 strategy type was used, 2 strategy types were used in 33.8 %, 3 strategy types in 44.1 %, 3 strategy types in 14.0 % and all 5 types in 1.0% of all cases. When considering all eight weeks together, children used 4.04 different strategy types on average (1.0% used only 1 strategy type, 2.1% used 2 strategy types, 22.9% used 3 strategy types, 38.9% used 4 strategy types and 35.1% used all 5 strategy types).

This variability was not merely due to differences in difficulty level: even when children had answered an item correctly by means of retrieval, the most mature strategy that is possible, they did not always do so again when the same item was presented in a later session. Of the 949 instances when children used retrieval correctly to a problem that was presented again later, in 215 instances (22.7%) children regressed to a less mature strategy.

Overlapping Waves Model

We modeled variability in strategy selection and individual development over time with a growth model. Since strategy selection is a categorical variable, varying from 1 for a wrong strategy to 5 for retrieval, a growth model for categorical data, capable of showing an overlapping waves pattern of strategy selection as a function of increasing mathematical ability, was used. In order to be able to do this, it had to be assumed that (1) the five strategies can be ordered according to maturity, such that the strategies form an ordinal scale representing increasing maturity, and (2) there is an underlying continuous latent ability that predicts the likelihood that each strategy is used. The higher this latent strategy ability, the larger 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. Based on these assumptions the Graded Response Model

(Samejima, 1969) for multinomial categorical data could be applied to the indicator part of the model. If the resulting Overlapping Waves model would fits the observed data well enough it can subsequently be used investigated to what degree children progressed through these waves, i.e., if they chose increasingly more mature strategies over time. The model fits if (1) a clear curve arises for each strategy, if (2) these curves show partial overlap, and if (3) the model fit is sufficient.

Commonly, growth models have one indicator variable for each measurement. The models in this study, however, had far more indicators: one for every math problem in every week. The math problems varied in difficulty, and it is reasonable to assume that the curves are also different for each problem, as Figures 2 and 3 also illustrate. For example, we assumed that children would apply retrieval sooner to easier problems than to more difficult problems. These differences in difficulty necessitated the creation of curves for each problem separately, rather than merely averaging the results for different problems. An indicator variable was therefore included for every separate math problem on every measurement. For each child a score between 1 (incorrect) and 5 (retrieval) on each indicator was observed. The configuration of this growth model is illustrated in Figure 4.

Figure 4. Configuration of the strategy growth model.

The figure is an illustration of the strategy growth model, but the accuracy model is essentially the same. Note that the figure is simplified, as it displays only a small part of the model: the real number of indicators in the model equals the number of problems multiplied by the number of measurements. As every week 15 problems were administered, the number of indicators in each growth model theoretically equals $15 * 8 = 120$. However, the model complexity increases exponentially with each problem added to the set and, with this large number of indicators, the sample size would also have to increase exponentially for each extra problem (quickly increasing to tens of thousands or even more). The maximum number of problems to be analyzed with our sample of 98 children was 8 problems, as also mentioned before, meaning that in the final model there were $8 \text{ problems} * 8 \text{ measurements} = 64$ indicators for strategy use, and 64 more indicators for accuracy.

The 8 most informative problem sets were selected for these analyses. These appeared to be problem sets 6, 7, and 9-13 and 15: problem sets 1-5 and 8 were removed as they showed a ceiling effect, with many children solving these problems correctly using retrieval. Problem set 14 was removed because in week 7 some children did not understand the story context of the problem. The remaining problem sets yielded sufficient information and variation in difficulty of the problems. Because each problem set contained one to three different problems, of which one was selected randomly on each measurement, the total number of unique problems that were analyzed was fifteen¹; see also Table 1. Problem sets 6 and 12 needed a small imputation, because for technical reasons the model does not allow a strategy not to be used at all on a particular measurement. Therefore three missing data points (0.05% of the data) were imputed; in all three cases an ‘incorrect strategy’ was imputed for a child that had applied the incorrect strategy to the same problem on at least one other measurement.

The fit of growth models is usually evaluated with fit indices, such as the χ^2 (chi square) with its *p*-value, NC (Normed Chi square), CFI (Comparative Fit Index), and RMSEA (Root Mean Square Error of Approximation). The χ^2 should be as low as possible and its *p*-value non-significant. As for complex models *p* is almost always significant, another guideline is the NC, or the ratio of χ^2/df , which should not

¹ Actually, the number was sixteen, but problems 4 x 9 and 9 x 4 turned out to behave virtually the same in our analyses, so they were treated as one problem.

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. We used these indices whenever possible, but note that they are biased (i.e., too low) when the number of variables is high (Kenny & McCoach, 2003), such as in this study.

We used Mplus version 6 (Muthén & Muthén, 1998-2010, see example 6.4) to estimate all models. Mplus offers different estimators: Marginal Maximum Likelihood (MML) estimation takes the full covariance matrix into account and is therefore to be preferred. But if the size of the frequency table for the latent class indicator model becomes too large, as in our case, the chi-square test cannot be computed, so no meaningful fit indices can be obtained. WLS estimation is slightly simpler and therefore, up to a limit, still provides fit measures when MML does not, but parameter estimates may be less precise than with MML.

Strategy Model

The strategy choice model was successfully estimated with MML. The overall fit of this model, however, could not be evaluated because we had far too many indicator variables. Therefore, we used WLS estimation to assess fit. The fit seemed adequate, especially considering the large number of variables, $\chi^2(2214) = 3117.02$, $p < .001$, NC = 1.41, CFI = .89, RMSEA = .065 while the general trend in the ensuing figures was the same as estimated with MML. Moreover, the parameter estimates of the MML model, such as the intercept and slope means and variances, were all significant, and for all observed variables the proportion of explained variance was significant and reasonably high: between .39 and .65. This suggests that, as hypothesized, to a sufficient degree, (1) strategies can be ordered according to maturity and (2) individual differences (and improvement during the study) can be mapped onto an underlying continuous latent ability. This is in line with the Overlapping Waves hypothesis.

In order to confirm this hypothesis further, the shape of the curves was inspected. It is theoretically possible to obtain a good fit with a pattern of curves that does not resemble overlapping waves, e.g., one or two dominating strategies with all the other curves being low and flat. A set of five curves, one for each strategy, was obtained for each of the fifteen math problems and each set was visually inspected: the majority of the problems showed five clear curves. As an illustration, the curves

for two problems are presented in Figure 5: one easy problem (4 x 4) and one more difficult problem (9 x 3). The curves for the other problems are shown in Appendix A.

Figure 5. Modeled strategy probability curves for two example problems: 4 x 4 and, more difficult, 9 x 3. Curves for all problems are presented in Appendix A.

The Figure clearly shows a pattern of overlapping waves: apart from the extremes, for every ability level at least two strategies have a reasonable chance of being selected. The probabilities that each strategy is selected change as a function of increasing latent ability (displayed on the horizontal axis). Since the model was based on all problems simultaneously, both graphs in Figure 5 represent the same latent ability on the horizontal axis (i.e., a child's score on the horizontal axis is the same in both graphs), but the resulting curves are problem-specific. Note that the horizontal axis displays ability rather than time. This accommodates that at each measurement children differed from each other in their ability, and they also changed in ability at different rates. Development is therefore not a mere function of time. Nevertheless, time plays an implicit role: as time goes by, children move towards the right of the graph, each at their own speed. This is explained below in more detail.

The Figure illustrates that with increasing latent ability children were more likely to use more mature strategies. For example, a child with a latent strategy maturity of 1 used retrieval with a probability of around 60% to solve 4 x 4, which is far more likely than for a child with a latent strategy ability of 0 (the average level), for whom the chance of using retrieval was below 30%. The Figure also shows that children were more likely to select more mature strategies for the relatively easy problem of 4 x 4 than for the more difficult problem of 9 x 3: the child with ability 1,

who had a 60% chance of applying retrieval to 4 x 4, had a far lower probability, only around 35%, to apply retrieval to the more difficult problem of 9 x 3.

Development was also incorporated in the model. The model was created such that the shapes of the curves were constrained to remain the same over time (measurement invariance), but the position of participants along the horizontal axis shifted towards higher ability (to the right) over the eight weeks. In order to model this development, values for intercept and slope of this latent ability were estimated.

The intercept value reflects the children's overall latent ability: children with a high intercept chose more mature strategies. Improvement during the study was reflected in the slope, which can be interpreted as a gradual shift towards the right on the horizontal axis in Figure 5. Children with a high slope value developed faster and thus moved further towards the right during the eight weeks of the study. The mean intercept of latent strategy ability was set to zero ($SD = 1.02$) halfway through the study: between weeks 4 and 5. This means that halfway through the study the average child had a strategy ability score of 0. The slope was scaled such that its value can be interpreted as the total improvement during the study. The mean slope was 0.97 ($SD = 0.90$), so during the course of the entire study the average child progressed 0.97 units, so approximately a standard deviation, to the right on the latent strategy ability scale.

Accuracy Model

A similar growth model was created for accuracy. 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 fit of this model with WLS estimation was reasonable given its complexity, $\chi^2(2061) = 2538.88$, $p < .01$, $NC = 1.23$, $CFI = .85$, $RMSEA = .04$.

The curves of the problems are shown in Figure 6. That the higher the latent accuracy ability, the larger the probability that a correct answer was obtained. Easier problems are shown on the left; more difficult problems are shifted towards the right. The mean latent accuracy ability was also set to zero ($SD = 0.44$): halfway

through the study the average child had a latent accuracy ability score of 0. The slope was 0.28 ($SD = 0.45$): the average child shifted 0.28 units, approximately two thirds of a standard deviation, to the right during the course of the study.

Figure 6. Modeled accuracy probability curves for the different math problems. The more difficult the problem, the more the curve is shifted towards the right.

Combined model

In the final step, the two growth models for strategy selection and accuracy were combined in one comprehensive model. The intercepts and slopes of the strategy and accuracy growth models were allowed to covary. In addition, the latent working memory factor was added as a predictor to both intercepts and slopes. This model could be estimated with MML, but proved to be too complex for our approximation with WLS estimation, so there are no fit indices. However, the parameter estimates resembled the two partial models closely, suggesting that this model was also valid. The results of the final analysis are presented in Figure 7.

The Figure shows that the intercepts of both models were strongly interrelated: children who used more mature strategies also obtained higher accuracy scores.

** $p < .01$

Figure 7. Final model: two connected growth models with two predictors. To enhance visibility, the number of indicator variables has been reduced: the real number is 64 indicators for strategy selection and another 64 for accuracy. Only the sizes of the relations between latent factors are displayed. Black arrows represent significant relations ($p < .01$), non-significant relations ($p > .05$) are light gray.

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 intercepts: children with good working memory skills had higher general abilities in both strategy selection and accuracy. Note that the correlation between intercept and slope of the strategy selection model is also rather large: children with a high ability in strategy selection also improved more. Contrary to the expectations, there was no direct effect from working memory on the two slopes.

Discussion

In the present study, two main topics 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). Second, we investigated 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 progress 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 children's ability increased, more mature strategies tended to dominate their strategy selection.

The model that was created required two rather strong assumptions: (1) that the strategies can be ordered according to maturity, while this order and the relative distance between the curves do not change during development, and (2) that a single latent ability scale can be used to describe individual differences and development in the selection of different strategies. The resulting model had reasonably good fit indices, indicating that these assumptions are plausible. These results show that this approach is promising as a manner of quantifying strategy choices. It enables the use of more advanced statistical techniques.

Previous studies often treated different strategies as separate variables and analyzed them separately. Strategies, however, are not independent, given that they are multiple mutually exclusive possible outcomes of a single trial: 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 took this interdependence into account by treating strategies as different outcomes of a single variable. For every strategy, the probability that it is selected can be predicted from a single latent ability. The analyses also allowed the creation of latent growth models, thus enabling the analysis of development. Moreover, the analyses allowed curves for every individual problem that was administered. These curves show that easier problems are solved with more mature strategies. The curves also suggest a problem format effect: in the lower ability range, shown on the left side of each graph, the curve for counting has a higher peak for the contextual problems, while the 'wrong' strategy dominates the plain problems in this ability range. This may reflect that children of low ability did not understand the plain problems, but they could figure out how to use simple counting strategies to solve similar problems presented in a context format. This is an interesting finding for future research, which cannot be tested statistically in our study, as we did not administer the same problem in different formats.

Finally, 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 often 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 (for a review, see Raghubar et al., 2010) 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.

Children with high working memory abilities were more likely to choose mature strategies that more often resulted in a correct answer, but even when strategy choice was taken into account, their accuracy was still higher. These results give rise to the seemingly paradoxical situation that children with poor working memory predominantly use strategies that pose a high load on working memory. They face two difficulties: they use immature strategies that require many steps and working memory resources and are therefore error-prone, and they also make even more mistakes executing these strategies than children with higher working memory.

Contrary to our expectations, however, was the lack of a relation between working memory and the slopes of strategy selection and accuracy. That means that during the study, children with low working memory did not progress more slowly than children with high working memory: they were only lagging behind in both strategy selection and accuracy.

There are some possible explanations for this unexpected finding. The first is that in reality there is a relation between working memory and development, but our methodology did not capture it. We modeled linear development (on the latent scale), for reasons of parsimony, while in reality each child may show a unique development with phases of acceleration followed by deceleration. This may have obscured a possible effect with the slope. It might also be the case that the relation with the slope is most pronounced somewhat earlier in the learning process, when the majority of the children relied on counting. In the present study most children were approximately halfway through the learning process and only the weaker children used counting strategies frequently. Another possibility is that the children who were lagging behind received more teacher instruction or practiced more, which helped them to compensate for their slow progress.

It must be noted that working memory was far from perfectly related to strategy selection. This shows that other factors must also be involved. While working memory may provide the resources for children to solve a problem correctly and to learn math facts by strengthening connections between problem and answer, it may be seen as a necessary but not sufficient condition: children must also have insight in the structure of a math problem and in the corresponding problem solving procedure.

A 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: i.e., possibly they did not choose the strategies with the largest chance of success, given the difficulty of the problem and their mathematical abilities. They might have performed better if they had chosen 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 a lack of adaptive strategy selection accounts for all our findings, as there was a strong direct connection between working memory and accuracy (see Figure 7). This connection 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 tend to choose the most mature strategy they can handle. Even children with mathematical difficulties tend to make their choices adaptively, albeit less so than their normally developing peers (Torbeyns, Verschaffel, & Ghesquière, 2002 [ENREF 34](#), 2004). It is therefore possible 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, mainly differing in speed with which they make this progress. The fit of the model was good 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, and other arithmetical skills such as addition.

A limitation of the study was that the analyses were constrained by computational 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, fifteen different unique math problems in total, 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 problems in a large difficulty range, we were able to select the most developmentally sensitive problems, so all children showed clear development during the study. This selection procedure 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 methods in microgenetic research by applying latent growth modeling to categorical data. This is a method that relies on well-known statistical techniques and theories. Yet, to our knowledge this approach has not been used 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.

References

- Alloway, T. P. (2007). *Automated Working Memory Assessment*. London: Pearson Assessment. Translated and reproduced by permission of Pearson Assessment.
- Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77*, 1698-1716. doi: 10.1111/j.1467-8624.2006.00968.x
- 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
- 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
- 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.
- 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
- 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

- Hecht, S. A. (2002). Counting on working memory in simple arithmetic when counting is used for problem solving. *Memory & Cognition*, 30, 447-455.
- 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
- Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, 10, 333-351. doi: 10.1207/s15328007sem1003_1
- 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
- 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. doi: 10.1348/026151003321164627
- 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.

- 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
- 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, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide. Sixth edition*. Los Angeles, CA: Muthén & Muthén.
- 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.
- Nulsen, C. E., Fox, A. M., & Hammond, G. R. (2010). Processes contributing to forward and backward span: An ERP investigation. *NeuroReport*, 21, 298-302. doi: 10.1097/WNR.0b013e32833730f0
- Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, 20, 110-122. doi: doi:10.1016/j.lindif.2009.10.005
- Ricker, T. J., AuBuchon, A. M., & Cowan, N. (2010). Working Memory. *WIREs Cognitive Science*, 1, 573-585. doi: 10.1002/wcs.50
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 34, 100-114.
- Segalowitz, S. J., Wintink, A. J., & Cudmore, L. J. (2001). P3 topographical change with task familiarization and task complexity. *Cognitive Brain Research*, 12, 451-457. doi: 10.1016/S0926-6410(01)00082-9
- 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

- 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
- 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
- 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. (2011). The development of executive functions and early mathematics: A dynamic relationship. *British Journal of Educational Psychology, Advance online publication*, 1-20. doi: 10.1111/j.2044-8279.2011.02035.x
- 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

Appendix A.

Strategy curves as a function of latent ability for all math problems. Problems are ordered according to relative difficulty, based on the children's accuracy in the present study: 4 x 4 was the easiest problem and 6 x 8 was most difficult.