

CHILDREN'S STRATEGIES FOR NUMEROSITY JUDGEMENT IN SQUARE GRIDS OF DIFFERENT SIZES

Koen LUWEL, Lieven VERSCHAFFEL, Patrick ONGHENA,
and Erik DE CORTE
University of Leuven

This study investigates the development of subjects' strategies for the judgement of numerosity from the theoretical perspective of "strategic change" (Lemaire & Siegler, 1995) as initiated by Verschaffel, De Corte, Lamote and Dherdt (1998). In the present study we assessed second and sixth graders' strategy use in determining numerosities of colored blocks in square grids of three different sizes. Converging evidence from response times and error rates showed that three distinct major numerosity judgement strategies were used: (a) an addition strategy by means of which (groups of) blocks are counted (and added), (b) a subtraction strategy in which the number of empty squares is subtracted from the (estimated or computed) total number of squares in the grid, and (c) a rough estimation strategy, whereby the number of blocks is determined in a quick but imprecise way. In terms of Lemaire and Siegler's model of strategic change, second and sixth graders did not differ in the kind of strategies used but in the efficiency with which they applied them. From a methodological point of view, this study demonstrates the potential of Beem's (1993, 1999) two- and three-phase segmented curve model for detecting subjects' strategy use in cognitive tasks.

In most traditional research on strategy use in a cognitive task children are described as using strategy X to solve the task at a particular age, strategy Y when they are somewhat older, and yet another strategy Z at still another age. Recent research on children's strategy development (for an overview see Siegler, 1996), however, has revealed that children of a specific age use a variety of strategies to solve a particular problem. This multiple strategy use allows a flexible adaptation to inherent task characteristics such as problem difficulty as well as to changing situational demands such as the need to answer quickly or accurately.

Koen Luwel, Lieven Verschaffel, and Erik De Corte: Center for Instructional Psychology and Technology; Patrick Onghena: Center for Methodology of Empirical Educational Research.

This study is supported by the grant G. 0157.98 from the Fund of Scientific Research.

The authors would like to thank Dr. Leo Beem for developing the three-phase segmented curve model by means of which the data were analysed, Evie Maesen for collecting the data, and two anonymous reviewers for their valuable comments on an earlier version of the paper.

Correspondence concerning this article should be addressed to Koen Luwel, Center for Instructional Psychology and Technology, University of Leuven, Vesaliusstraat 2, B-3000 Leuven, Belgium. Electronic mail may be sent to Koen.Luwel@ped.kuleuven.ac.be.

Based on these studies, Lemaire and Siegler (1995) developed a model that distinguishes among four dimensions of strategic competence. Changes in any of those dimensions can yield overall improvements in speed and accuracy of performance: (a) acquisition of new strategies and abandonment of old ones, (b) shift towards greater use of more efficient available strategies, (c) improvement in the fluency and efficiency with which strategies are executed, and (d) increase of the adaptive nature of the choice among available strategies.

Although strategic variability and strategic change within children has been observed in a whole range of domains such as simple addition (Siegler & Robinson, 1982), subtraction (Siegler, 1987), scientific reasoning (Kuhn & Phelps, 1982), time telling (Siegler & McGilly, 1989), serial recall (McGilly & Siegler, 1990), and spelling (Marsh, Friedman, Welch, & Desberg, 1980), these phenomena have hardly been studied in one of the oldest domains of experimental psychology, namely the judgement of numerosities (see e.g., Binet, 1890; Cattell, 1886; Jevons, 1871; Wundt, 1896). As far as we know, only one study tried to apply Lemaire and Siegler's (1995) theoretical perspective of strategic change to the content domain of the judgement of numerosities (Verschaffel, De Corte, Lamote, & Dhert, 1998).

In Verschaffel et al.'s (1998) study, participants of three different age groups (i.e., university students, sixth graders, and second graders) had to determine the numerosity of 100 quantities of square blocks presented in a square 10 x 10 grid. According to these authors, this particular task allows for the use of at least two fundamentally different kinds of strategies. First, an addition strategy, in which (groups of) blocks are counted (and added), and, second, a subtraction strategy, in which the number of empty squares is subtracted from the total number of squares in the grid.

The first hypothesis of Verschaffel et al. (1998) was that the stereotyped use of the addition strategy would decrease with age, whereas the mixed and adaptive use of the addition and the subtraction strategy would increase with age. According to the second hypothesis, the combined and adaptive use of both strategies would - regardless of age - be accompanied by a higher performance on the numerosity judgement task.

Subjects' solution strategies were mainly identified by the pattern of response times on all trials by means of Beem's (1993, 1995) "program for fitting two-phase segmented curves models with an unknown change point". More specifically, the fit of each subject's response-time pattern with, respectively, a one phase linear model and a two-phase segmented curve model was computed.

With respect to the first hypothesis, converging evidence from response accuracies, response times, and retrospective reports provided support for

the hypothesized developmental trend towards the greater use of the more efficient subtraction strategy. With respect to the second hypothesis, a strong and positive relationship was found between the combined and adaptive use of the addition and the subtraction strategy, on the one hand, and accuracy of numerosity judgement, on the other. However, this latter finding was obtained in the two oldest age groups only. In terms of Lemaire and Siegler's (1995) model of strategic change, this study showed that the difference between the distinct age groups was *not* a result of the use of different strategies, but rather was due to the more frequent, more fluent and more adaptive use of the subtraction strategy.

This study raised two questions on which the present study is based upon. First, although the data of the second graders contained - as expected - much less evidence for the use of the subtraction strategy than those of the sixth graders and the university students, there were nevertheless already some undeniable traces of it in most subjects from this youngest age group. Whereas these traces remained typically restricted to a very small number of trials wherein the grid was almost completely filled with blocks (i.e., the trials with 95 blocks or more), they nevertheless indicated that the adaptive use of the subtraction strategy begins earlier than expected. However, this raises the question whether this unexpectedly early appearance of the subtraction strategy was caused by the rather peculiar nature of the grid size used in this experiment, namely the 10×10 grid, which makes the use of this subtraction strategy exceptionally attractive and easy. Indeed, 10×10 is one of the rare grids for which the total amount of squares in the grid (i.e., the so-called "anchor") consists of a "round" number, i.e., $(10 \times 10 =) 100$. This specific anchor has at least two advantages that make it easy to compute with, in comparison to other grid sizes. First of all, the multiplication fact one has to know to arrive quickly at the anchor, namely $10 \times 10 = 100$, is typically one of the first facts of the multiplication table children know by heart. Another advantage is that subtracting the number of empty squares from this anchor ($100 - a = ?$) requires fewer solution steps and less mental effort than subtracting the same number of empty squares from other total grid sizes, such as 64 (in case of a 8×8 grid) or 144 (in case of a 12×12 grid). Because of these computational advantages, the 10×10 grid can be expected to elicit the subtraction strategy more quickly and more strongly than other grids.

Second, a number of participants seemed to have used a third - unanticipated - strategy besides the two strategies that were at the heart of Verschaffel et al.'s (1998) rational and empirical analysis. Indeed, the graphs of the response-time patterns and of the accuracy rates suggested that in all three age groups a significant number of participants applied for the trials in the middle range (i.e., the trials with both a large number of blocks and empty squares) a third strategy. The basic characteristics of this strategy are

that it is relatively quick, that it leads to rather imprecise answers, and that its duration does not seem to be seriously affected by the numerosity to be determined. Verschaffel et al. describe this unexpected third strategy - which we will call the estimation strategy - as "a rather primitive procedure in which the number of blocks is determined instantaneously, as in the well-documented 'subitizing' procedure by which small quantities up to ± 5 are determined" (p. 365). This primitive procedure for quick and rough estimations of numerosities larger than ± 5 , has also been identified by other authors (Kaufman, Lord, Reese, & Volkmann, 1949; Klahr & Wallace, 1973; Newman, Friedman, & Gockley, 1987; Svenson & Sjöberg, 1978), however, without providing a full account of it. As an illustration, Figure 1 contains an example of a response-time pattern of a participant who solved the 100 trials only by means of the addition and the subtraction strategy (Figure 1a) contrasted with an example of a response-time pattern of a participant who seemed to have combined these two strategies with the estimation strategy (Figure 1b).

The occurrence of this estimation strategy has important theoretical and methodological consequences. From a theoretical point of view, it necessitates a refinement of the rational task analysis that lied at the basis of Verschaffel et al.'s (1998) study by incorporating this third strategy. From a methodological point of view, it jeopardizes the use of Beem's (1993, 1995) two-phase segmented curve model, which is only able to identify (the shift between) two strategies in a set of data points. Therefore, an elaboration of Beem's two-phase segmentation model towards a three-phase model is required to get a better picture of the use of the different strategies in the response-time patterns.

The major aims of the present study were twofold. First, starting from the theoretical framework of Lemaire and Siegler (1995), we expected to observe larger differences between second and sixth graders on each of the four above mentioned dimensions of this framework when working with grid sizes for which the advantages of using the sophisticated subtraction strategy are less evident. We were especially interested in whether young children would still show evidence of the efficient and adaptive use of this subtraction strategy with less facilitating grid sizes.

Second, we wanted to investigate the potential of Beem's (1999) three-phase segmented curve model with two unknown change points. The availability of this elaborated model made it possible to split the data set in three instead of only two segments. We expected that this possibility would allow us to capture the occurrence of and the shift to the subtraction strategy in a more accurate way than by means of Beem's (1993, 1995) two-phase segmented curve model with only one unknown change point, which was used in Verschaffel et al.'s (1998) original study.

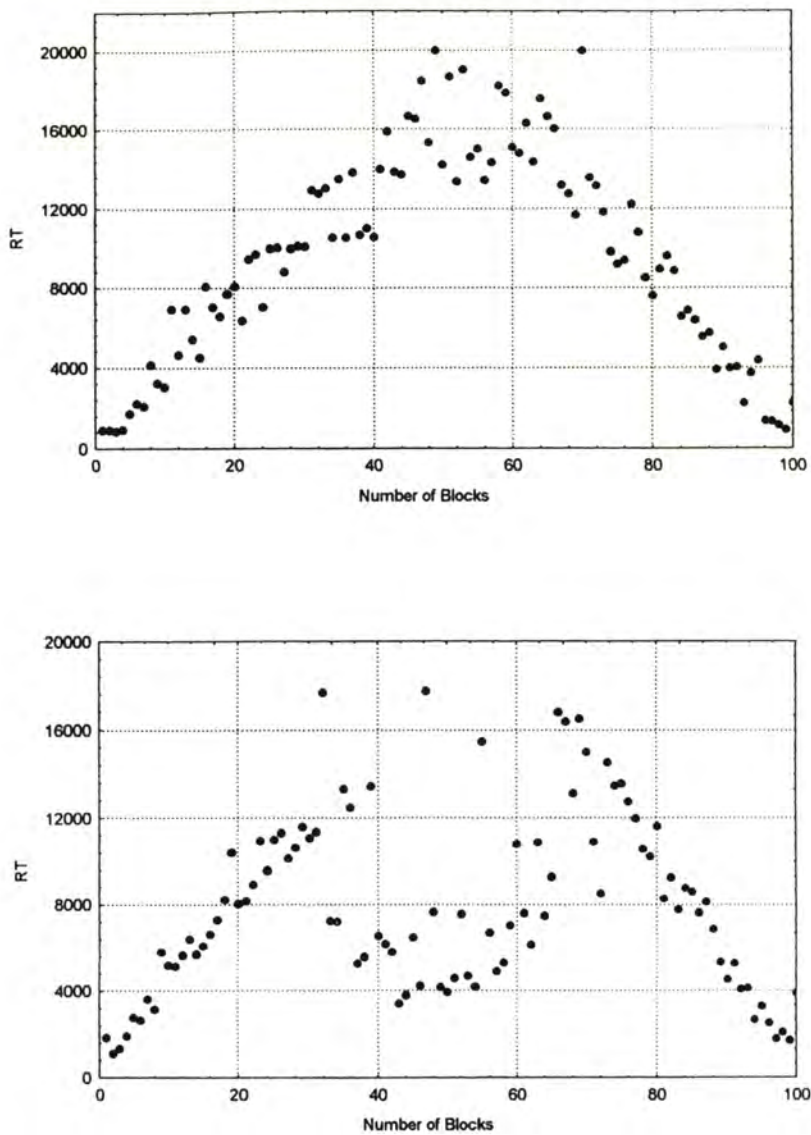


Figure 1. Examples of response-time patterns (in ms) of (a) subject 3 (university students) and (b) subject 1 (sixth graders) from the study of Verschaffel et al. (1998).

Rational Task Analysis and Hypotheses

In this section we present an adapted rational task analysis in which the estimation strategy is incorporated as a possible "coping strategy" for the cases wherein neither the addition nor the subtraction strategy seems feasible. Afterwards, we sum up the specific research hypotheses and questions that can be derived from it. However, we first give some pivotal information about the task and the way that it was administered, which is indispensable to follow the line of argumentation in this section.

The same task as used by Verschaffel et al. (1998) was administered to groups of second and sixth graders. But instead of using a single square grid with a side of 10 squares, all subjects were presented a 7 x 7, an 8 x 8, and a 9 x 9 grid. For each grid size all possible numerosities were administered to each participant. In the case of a 7 x 7 grid, for instance, the task consisted of 49 trials. In every trial, each of the 49 squares in the grid was either empty or filled with a colored block. Consequently, the grid was always filled with a number of blocks ranging from 1 to 49. Each trial was presented for a maximum duration of 20 s and the participant was asked to determine the total number of blocks. After each trial, the computer recorded the pupil's answer as well as the response time.

Taking into account the findings of Verschaffel et al. (1998) concerning the occurrence of the estimation strategy on the trials in the middle range, as well as Lemaire and Siegler's (1995) theory of strategic change, we adapted Verschaffel et al.'s rational task analysis in the following way: We hypothesized a development in the ability to make numerosity judgements in square grids from a "novice" level, wherein the subtraction strategy is completely absent, to an "expert level" involving the frequent, efficient and adaptive use of the subtraction strategy. More specifically, the following combinations of solution strategies were hypothesized for experts and novices.

Experts are expected to apply the addition strategy for trials with few blocks and many empty squares, and the subtraction strategy for trials with many blocks and few empty squares. This will lead to quick and accurate responses on trials with few blocks, but also on trials with large numbers of blocks. Because of the nature of the addition and the subtraction strategies, the response times for trials solved by addition will more or less linearly increase with increasing number of blocks, while response times for trials solved by means of subtraction, will decrease with increasing number size. On trials with too many blocks and too many empty squares to be determined within the available time by means of the addition or the subtraction strategy (i.e., the middle region), experts will resort to a coping strategy, the so-called estimation strategy. Due to its nature, application of the estimation strategy will elicit less accurate responses with relatively quick response

times that lie within in the same range. Whether experts will make use of the estimation strategy, and, if so, on which range of numerosities, will depend on subject variables such as the speed with which they can determine and add small numerosities and their motivation to do so, on the one hand, and task variables such as the actual size of the grid and its presentation time, on the other.

Novices are expected to apply the addition strategy for trials with small numerosities. However, once they are not able or motivated to make a precise determination of a given numerosity within the given time limit, the addition strategy will be replaced by the cognitively less demanding but more error-prone estimation strategy. Novices (who do not know the subtraction strategy yet) will continue to use this estimation strategy, even for trials wherein the grid is (almost) completely filled with blocks. This implies that for trials for which they are no longer able or willing to apply the addition strategy, their response times will be flat and their response accuracies will be low.

Figure 2¹ contains the response-time patterns that follow from this rational task analysis. Graphs (a) and (b) describe, respectively, the pattern of the response times for novices who use the addition strategy solely (Pattern 1) and for those who use it in combination with the estimation strategy (Pattern 2). Graphs (c) and (d) show, respectively, the response-time pattern for experts who use the addition and subtraction strategy (Pattern 3), and for experts who adaptively use the three numerosity judgement strategies involved in our rational task analysis (Pattern 4).

This rational task analysis led us to the following research hypotheses and questions. First, since the numerical, computational, and metacognitive knowledge and skills are assumed to develop with age, we hypothesized that the accuracy of the numerosity judgements would increase with age (Hypothesis 1a). More specifically, we predicted that, for each of the three grid sizes, the mean error rates (i.e., the mean overestimation or underestimation of the number of blocks actually presented) would be smaller for sixth graders than for second graders. Moreover, we hypothesized an effect of grid size. The greater the grid size, the bigger the enumerations, computations and/or estimations to be made, and - thus - the greater the chance of obtaining a response that deviates more significantly from the given numerosity (Hypothesis 1b). Finally, we predicted an interaction between the factors grade and grid size, in the sense that the difference in numerosity judgement accuracy between the second and the sixth graders would increase with growing grid size (Hypothesis 1c).

¹ A complete account of this figure will be provided further in the text.

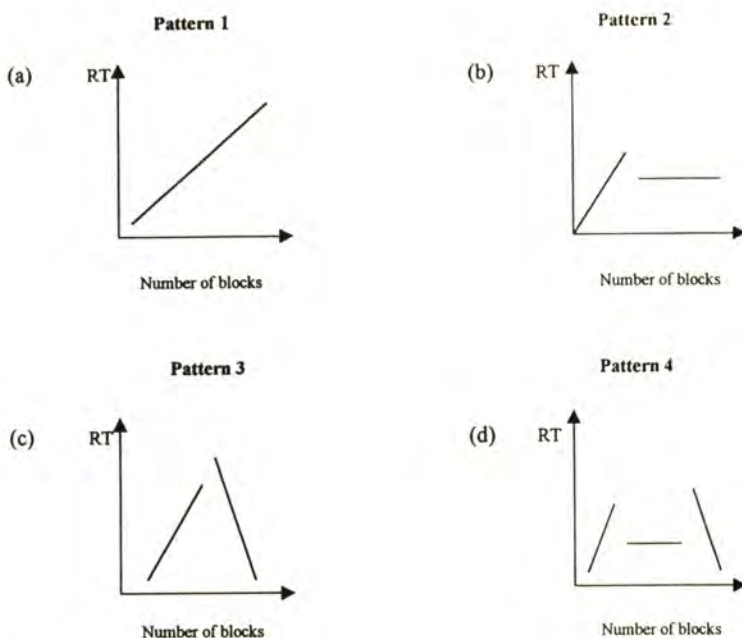


Figure 2. Hypothetical response-time patterns with (a) application of the addition strategy, (b) use of the addition and estimation strategy, (c) execution of the addition and subtraction strategy, and (d) application of the addition, estimation, and subtraction strategy.

Second, we hypothesized that the occurrence and the efficiency of the use of the subtraction strategy would increase with age (Hypothesis 2). Therefore, we anticipated that, for each of the three grid sizes, there would be more sixth graders than second graders showing evidence of the use of this sophisticated strategy, and also that sixth graders would use it more effectively than second graders.

Third, based on the rational task analysis presented above, we expected considerable differences in the response-time patterns of the two age groups. More specifically, we anticipated that for each of the three grid sizes there would be more second than sixth graders whose response-time patterns resemble Pattern 1 and 2 in Figure 2 and more sixth graders with response-time patterns resembling Pattern 3 and 4 in Figure 2 (Hypothesis 3a). We also anticipated an effect of grid size on the division of pupils' response-time patterns over these four hypothetical patterns. More specifically, we predicted that a growth in grid size would lead to an increase in the proportion of response-time patterns with a segment reflecting the use of the estimation

strategy (i.e., Pattern 2 and 4 in Figure 2), and, correspondingly, to a decrease in the proportion of response-time patterns without such a segment (i.e., Pattern 1 and 3 from Figure 2) (Hypothesis 3b). In addition, we also expected an interaction effect between grade and grid size. Since the skills required to efficiently execute the addition and subtraction strategy are more developed in sixth than in second graders, we anticipated that there would be a greater proportion of second graders compared to sixth graders who would show evidence for the estimation strategy already in the smallest grid, i.e., the 7 x 7 grid. We also expected that this difference between both age groups would disappear with increasing grid size, since the greater the grid size, the smaller the chance that even sixth graders can solve the whole range of trials by means of the addition and the subtraction strategy only (Hypothesis 3c).

Fourth, based on the characterization of the three strategies for judging numerosity involved in our rational task analysis, we predicted that - for each of the three grid sizes and for the two age groups - the mean error rates would be considerably larger for the trials solved by means of the estimation strategy than those for the trials solved by the addition strategy or subtraction strategy (at least for those participants who applied it properly). Moreover, we predicted that there would be no difference in the mean error rates between the trials solved by the addition strategy, on the one hand, and the trials solved by the subtraction strategy, on the other hand (Hypothesis 4).

Method

Participants

Sixty-nine pupils participated in the study: 39 second graders (7-8 years old) and 30 sixth graders (11-12 years old). (The reason for having a greater number of second than sixth graders will be explained below.) The second graders were chosen from four classes in two schools, whereas the sixth graders were chosen from three classes in two schools. On the basis of their results on their last mathematics exam, ten pupils were selected from each class to participate in the experiment: the five pupils who scored best and the five pupils with the weakest scores on the exam. Both sexes were equally represented in our sample.

In the current mathematics curriculum in Flanders (Belgium) little or no explicit and systematic attention is spent at counting and estimating large numerosities (Verschaffel & De Corte, 1996). Therefore, it can be assumed that the subtraction strategy has not been systematically taught to the participants. Like Verschaffel et al. (1998), we took second graders as the youngest age group because they were in the midst of the process of mas-

tering the conceptual and procedural underpinnings of the subtraction strategy. At the time they were tested, second graders in Flanders have already been familiarized with the basic meaning of the arithmetic operations of addition, subtraction and multiplication (which are all required when making numerosity judgements by means of the subtraction strategy). Moreover, they have already explored the number line up to 100, and they have been taught how to add and to subtract with numbers up to 100, although many of them still have great difficulties with subtractions like $81 - 15 = ?$ or $64 - 18 = ?$ Finally, they have already passed the first stages of learning how to multiply numbers from 1 to 10, but not all multiplication facts are yet known by heart. Since we wanted to compare the findings of the present experiment with those of Verschaffel et al.'s study, we took sixth graders as the older age group.

Materials

The task was presented to the participants using a computer with a resolution of 640 x 480. Stimuli were square grids consisting of either 7 x 7, 8 x 8 or 9 x 9 little square units. These square units could either be "on" (i.e., being filled with a green-colored block) or "off" (remaining empty, i.e., being filled with the same black color as the background of the whole screen). Each square had a size of 1 x 1 cm. A small black line separated the green squares, whereas the empty squares were not intersected. The outline of the square grid was visible and was colored red.

Procedure

All participants were tested individually during three separate sessions. By presenting each grid size (7 x 7, 8 x 8 and 9 x 9) in a separate session, fatigue-effects were minimized. The order of the sessions was counterbalanced. In each session, a one-minute pause was given after every 25 trials.

Dependent on the grid size, participants ran a different number of trials: respectively 49, 64 and 81 trials for the 7 x 7, 8 x 8 and 9 x 9 grid. The sequence of the stimuli within a given grid size and the placement of the blocks in the grid were completely randomized by the computer. After each trial, the computer recorded participant's answer and response time.

A small pilot study had revealed that some children - typically second graders - who showed evidence of the subtraction strategy, worked with a wrong anchor (i.e., the total number of squares in the grid). Some approached the task with the assumption that the size of the first grid (e.g.,

8 x 8) did not change throughout the whole experiment (and thus erroneously took the next grid - e.g., 9 x 9 - for a grid consisting of 64 instead of 81 squares); others conceived each of the three grids as being a 10 x 10 grid (and thus as consisting of 100 squares). As a consequence of these wrong anchors, these children's numerosity judgements of the items on which the subtraction procedure was applied, were always far removed from the actual numerosities, which made the comparison of the responses and response-time data between the two age groups highly problematic. In an attempt to minimize such misinterpretations of the task (especially by children of the youngest age group), four cards were presented at the beginning of each session, each card showing a completely filled grid of a different size (i.e., 7 x 7, 8 x 8, 9 x 9 and 10 x 10). Pupils were asked to determine the number of squares in each grid, and we corrected them in case they were wrong. This procedure was repeated at the beginning of each session. However, because we doubted that this attempt to sensitize children to the size and changeability of the anchor would be sufficient, especially for the children from the youngest age group, we decided to increase the number of second graders in the experiment from 30 to 40. We ended up with 39 second graders in our study since the data of one child were lost.

After the introduction and the presentation of the four cards, participants were introduced to the numerosity judgement task on the computer. As in the study of Verschaffel et al. (1998), they were told that a certain number of screen displays would be presented, each showing a grid with a number of green blocks, and that their task was to judge the number of blocks as accurately as possible within the given time constraint of 20 s. To help them get accustomed to the task and to understand what was meant by an accurate numerosity judgement, 10 example trials of a random number of blocks were given before the start of the actual experiment in each session. During these 10 example trials, participants received immediate feedback about the accuracy of their answer: a message appeared on the screen telling them whether their response was considered as an accurate or as an inaccurate judgement. A 10% deviation criterion was used to decide whether a numerosity judgement was considered accurate or not. This feedback was dropped as soon as the actual experiment started. The goal of these example trials was twofold: (a) to get participants familiarized with the criterion that would be used when analyzing whether an answer was correct or not, and (b) to allow participants to determine the total number of squares in the grid (i.e., the anchor) before the actual start of the experimental data collection.

As in the previous study (Verschaffel et al., 1998), each stimulus was presented for a maximum duration of 20 s. Participants were asked to verbally state their answer as soon as they knew it. The experimenter then immediately pressed a key, the clock stopped and the screen was blanked. A new

screen appeared in which the experimenter entered the response by means of the numerical pad of the keyboard. When no response was given within 20 s, the screen went blank and the participant was asked to make a numerosity judgement anyhow. After the experimenter had entered the response, a new stimulus was presented on the screen.

Results

Before the data were analyzed, some outliers were removed. As in Verschaffel et al's (1998) study, a distinction was made between two groups of outliers:

1. *Inverse errors*: errors due to the fact that - at the end of the numerosity judgement process - the pupil has forgotten whether (s)he was engaged in an addition or subtraction strategy, leading to numerosity judgement errors approximating the number of empty squares instead of the actual numerosity of the blocks (by definition, these types of errors could only be traced in the data of the subjects who had been identified as using the subtraction strategy on a correct anchor; see below)².

2. *Typing errors*: extreme errors that were the result of typing errors (e.g. 488 instead of 48). All numerosity judgement errors that had a minimal deviation of 90 from the number of presented blocks, were considered as typing errors. Furthermore, we removed numerosity judgements with a value of 0. This 0 value was used by the experimenter to indicate that the pupil had not made any numerosity judgement at all.

Based upon these criteria, we removed 91 outliers from a total of 13386 data points (i.e., 0.007%). We now turn to the presentation of the results related to the different hypotheses formulated above starting with the findings regarding the effect of age and grid size on numerosity judgement accuracy. Unless stated otherwise, an alpha level of .05 was used for all statistical tests. Exact *p*-values will be reported, but very small values are rounded to $p < .0001$.

² To be classified as an inverse error, the following three conditions had to be fulfilled (a) the response deviates at least two standard deviations from the subject's mean error rate for all trials in a specific grid, (b) the response fits the so-called complement rule, stating that the sum of the subject's response and the actual number of blocks in the trial must equal the total number of squares in the grid ± 5 (or ± 10 when the response deviates at least three standard deviations from the subject's mean error rate), (c) inverse errors are assumed to occur solely in the first and last third of the trials.

Influence of Age and Grid Size on the Accuracy of the Numerosity Judgements

To test the three predictions derived from our first hypothesis - namely that there would be an effect of age and grid size on the mean error rates as well as an interaction effect between both variables -, an analysis of variance was carried out with age as independent between-subjects variables and grid size as an independent within-subjects variable. The dependent measure was the mean error rate. Since previous analyses of variance had revealed that the control variables sex, class group, and grid size order did not have any effect on mean error rates, these variables were discarded from the present analysis.

The results of this analysis showed a main effect of age, $F(1, 67) = 33.13$, $p < .0001$. The mean error rates of sixth graders ($M = 3.28$) were significantly smaller than these of the second graders ($M = 9.04$). Somewhat surprisingly, the predicted main effect of grid size as well as the predicted interaction effect between age and grid size were not observed (See Figure 3). An account of the absence of both predicted effects will be given in the next paragraph.

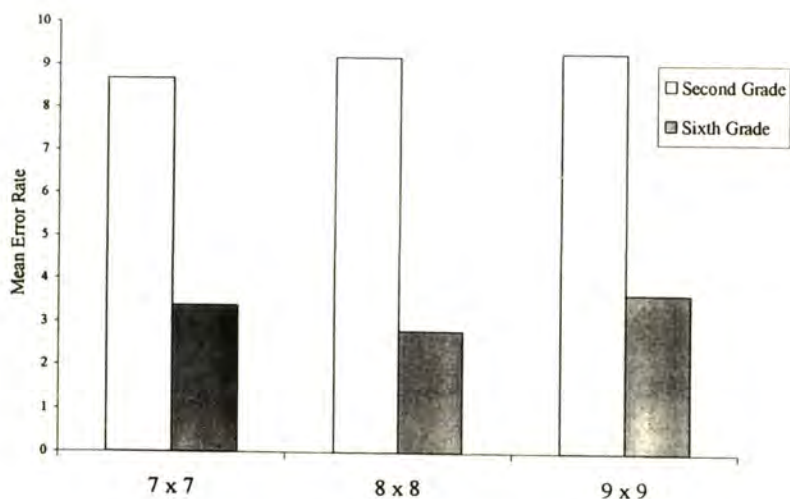


Figure 3. Mean error rates of second and sixth graders for each of the three grid sizes.

Use of the Subtraction Strategy

To test our second hypothesis, we computed for the different grid sizes and age groups the number of participants whose response patterns showed evidence of (a) correct use of the subtraction strategy, (b) incorrect use of the subtraction strategy, and (c) no use of the subtraction strategy. Before presenting the results of this analysis, we explain how we distinguished between these three categories.

To decide whether a pupil had applied the subtraction strategy for a given grid size, we looked at the deviation scores (i.e., the absolute difference between the judged and the actual numerosity of blocks) in the interval of trials with the five largest numbers of blocks. If this interval had elicited at least two times exactly the same deviation score (including 0), it was assumed that this pupil had subtracted at least these two times from the same anchor, and - thus - had applied the subtraction strategy at least occasionally. The rationale behind this criterion is the assumption that the addition strategy does not allow a sufficiently accurate determination of such a large number of blocks within the given presentation time (20 s) and that the probability of twice obtaining the same deviation score by means of the rough estimation strategy is extremely low. Whereas the *occurrence* of (at least two) identical deviation scores was used to decide whether a pupil had applied the subtraction strategy, the *magnitude* of the deviation from the correct numerosities was used to decide if that pupil had applied this subtraction strategy properly, and, more specifically, if (s)he had subtracted from the correct anchor. If the difference between the given answers and the actual number of blocks was 0, we decided that the pupil had subtracted from the *correct* anchor. For example, if a child answered with the exact numerosity on at least two of the five trials with the highest numerosities from the 8 x 8 grid (i.e., 60, 61, 62, 63 and 64), thus leading to a deviation score of 0 on these two trials, this child was considered as a correct subtraction strategy user for this grid size. However, if at least two of the five answers on these trials deviated with the same value (different from 0), we assumed that the pupil had subtracted at least two times from an *incorrect* anchor when using the subtraction strategy, and - therefore - categorised him or her as someone who applied the subtraction strategy incorrectly. For instance, if a child responded to the four highest stimuli from the 7 x 7 grid (i.e., 49, 48, 47 and 46) with, respectively, 100, 99, 98 and 97 - leading each time to a deviation score of 51 - we assumed that (s)he had erroneously worked with the anchor of 100 instead of 49 when applying the subtraction strategy. Pupils who did not produce at least two responses that were exactly the same as the given numerosity or who did not have exactly the same deviation from the correct answer within the interval of five largest numerosities, were assumed not to

have applied the subtraction strategy for that particular grid size.

Following this criterion, all pupils were categorised-for each grid size separately - into one of the three categories mentioned above. Table 1 shows the result of this classification.

Table 1. *Proportion of Second and Sixth Graders According to Their Application of the Subtraction Strategy for each of the Three Grid Sizes*

Grid size	Application of subtraction strategy		
	No	Correct	Incorrect
Second Graders			
7 X 7	0.08 ($n = 3$)	0.36 ($n = 14$)	0.56 ($n = 22$)
8 X 8	0.15 ($n = 6$)	0.21 ($n = 8$)	0.64 ($n = 25$)
9 X 9	0.15 ($n = 6$)	0.28 ($n = 11$)	0.56 ($n = 22$)
<i>M</i>	0.13	0.28	0.59
Sixth Graders			
7 X 7	0.00 ($n = 0$)	0.70 ($n = 21$)	0.30 ($n = 9$)
8 X 8	0.03 ($n = 1$)	0.80 ($n = 24$)	0.17 ($n = 5$)
9 X 9	0.03 ($n = 1$)	0.83 ($n = 25$)	0.14 ($n = 4$)
<i>M</i>	0.02	0.78	0.20

The way in which the anchor was determined can be considered as an outcome variable containing three categories (no anchor, correct anchor, and incorrect anchor). To distinguish between the outcome categories that show an effect of age and those who do not, we carried out three weighted least squares analyses for repeated categorical data (Koch, Landis, Freeman, Freeman, & Lehnen, 1977) in which we combined each time two of the three outcome categories and compared this combination with the remaining outcome category. Combining categories in such a way resulted each time into ($2^3 =$) 8 profiles.

In the first analysis we compared the outcome category "correct anchor" with the combined categories "incorrect anchor" and "no anchor". This analysis revealed a main effect of age, $\chi^2(1, N = 69) = 40.94, p < .0001$. Inspection of Table 1 shows that, for each of the grid sizes, there was a greater proportion of sixth graders than of second graders with a correctly determined anchor. Furthermore, we observed a significant interaction effect between age and grid size, $\chi^2(2, N = 69) = 6.42, p = .04$. Additional contrast testing revealed that - besides the significant main effect of age in each of the three grid sizes -, within the youngest age group, there was only a significant difference in the proportion of participants with a correct anchor

between the 7 x 7 and the 8 x 8 grid, $\chi^2(1, N = 69) = 5.09, p = .02$. For the sixth graders, we only observed a significant difference in the proportion of subjects with a correctly determined anchor between the 7 x 7 and 9 x 9 grid.

A second weighted least squares analysis compared the outcome category "no anchor" with the combination of the categories "correct anchor" and "incorrect anchor". This analysis did not show any significant difference. In other words, there was no difference in the proportion of second and sixth graders who did not determine an anchor at all.

The third weighted least squares analysis compared the "incorrect anchor" category with the combination of the "correct anchor" and the "no anchor" categories. This analysis yielded a significant main effect of age, $\chi^2(1, N = 69) = 26.94, p < .0001$. A closer look at Table 1 revealed that, for each of the three grid sizes, the proportion of second graders with an incorrect anchor was significantly larger than the proportion of sixth graders in the same outcome category.

As in the previous study with a much easier grid size (Verschaffel et al., 1998), the majority of the second graders applied the subtraction strategy, although this time many of them did it in a defective manner using an inappropriate anchor. Surprisingly, even in the oldest age group a significant number of children (19%) worked with an incorrect anchor. Apparently, our efforts aimed at sensitising pupils to the correct size of the anchor (as described in the Methods section) were quite unsuccessful. The majority of the incorrect anchors in the group of the second graders had a magnitude of 100 (respectively 68% for 7 x 7, 52% for 8 x 8, and 64% for 9 x 9 compared to, respectively, 25%, 20% and 0% for the sixth graders). Probably, this remarkable finding was due to the disturbing influence of the so-called "hundred-field", a didactical model that is used in most second grade classes in Flanders to visually support a good understanding of the structure of the number range up to 100 as well as of the operations of addition and subtraction within this number domain (Beishuizen, 1993). When they saw the grid, some of the second graders may have immediately associated it with this "hundred-field" and therefore may have assumed that this grid had a size of 100 blocks too.

The fact that a lot of second graders had an anchor of 100 explains the absence of the predicted main effect of grid size and the interaction effect between age and grid size in the previous paragraph of the Results section. Indeed, the larger the grid size, the smaller the size of the systematic error the pupils produced, since the difference between the judged anchor (100) and the actual anchor (resp. 49, 64, 81) decreased with increasing grid size. As a consequence, the mean error rates of the second graders remained rather stable with an increasing grid size (See Figure 3).

Summarising the results for Hypothesis 2, we conclude that the better

overall performance of the sixth graders was not due to the fact that there were *more* pupils of this age group who used the subtraction strategy, but rather to the fact that this sophisticated strategy was used *more properly* by the sixth graders. The second graders seem to have experienced great difficulties with the determination of the anchor, probably as a side effect of their recent instructional histories.

Response-Time Patterns

To test the predictions related to the effects of the subject and task variables on pupils' response-time patterns (Hypothesis 3), we compared the individual response-time data with the four hypothetical patterns in Figure 2 by means of three types of statistical models: a linear model, a two-phase and a three-phase segmented curve model.

In the well-known linear models, the relationship between the independent (x) and the dependent (y) variable is described by one linear regression equation of the form $y = a + bx + e$ (1). The parameters a and b , which are to be estimated, denote the intercept and the slope of the regression line, whereas e is the error term.

The two- and three-phase segmented curve models are less well-known and used, although they are, according to Beem (1993, 1995, 1999; Ippel & Beem, 1987), ideally suited for the study of strategy shifts like those involved in the present study. As the name "two-phase segmented curve" indicates, the relationship between the independent and the dependent variable is not described by one but by two regression lines, which hold for different ranges of the independent variable. The two-phase segmented model can be formally described as:

$$y = a_1 + b_1x + e \quad (x \leq s) \quad (2)$$

$$y = a_2 + b_2x + e \quad (x > s) \quad (3)$$

For values of the independent variable up to s the first regression equation is used, while for values higher than s the second equation - with a different intercept and slope - holds. The parameter s is called the "change point" or "break point".

The three-phase segmented curve model is an extension of the two-phase segmented curve model in which the relationship between the independent and the dependent variable is described by three regression lines instead of two. The three-phase segmented curve model has the following formal description:

$$y = a_1 + b_1x + e \quad (x \leq s_1) \quad (4)$$

$$y = a_2 + b_2x + e \quad (s_1 < x \leq s_2) \quad (5)$$

$$y = a_3 + b_3x + e \quad (x > s_2) \quad (6)$$

Completely analogous to the two-phase segmented model, the three-phase segmented model computes three different regression equations and estimates two change points.

Defining the four hypothetical response-time patterns from Figure 2 in terms of the different parameters of the statistical models presented above, leads to the following characterisation of each pattern:

1. Pattern 1 (always addition): no change point, and the only b -parameter is positive.
2. Pattern 2 (first addition, then estimation): one change point, a positive b -parameter for the first regression line, and a b_2 -parameter with a value close to zero.
3. Pattern 3 (first addition, then subtraction): one change point, a positive b_1 -parameter and a negative b_2 -parameter.
4. Pattern 4 (first addition, then estimation, and finally subtraction): two change points, a positive b_1 -parameter, a b_2 -parameter with a value close to zero, and a negative b_3 -parameter.

We followed a stepwise procedure to identify the possible presence of one of the different hypothetical patterns from Figure 2 in pupils' response-time data. First, participants with an incorrectly determined anchor were discarded from the analysis, since our rational task analysis suggests that subtracting from a different anchor results in response times that are incomparable with the response times that are obtained by subtracting from a correct anchor. All other individual response-time patterns were then tested for the occurrence of two change points. If we did not find any evidence for the presence of two change points, the same pattern was tested again for the occurrence of one change point. When we did not detect a change point at all, we assumed that the particular participant had only used the addition strategy, and, therefore, this response-time pattern was classified as being of type 1. Response-time patterns that showed one or two change points, were further tested for a possible fit with the hypothetical patterns 2, 3, or 4 by entering the b -parameters in the different regression equations into chi-square tests (Beem, 1993, 1999). A more detailed account of this procedure can be found in Luwel, Verschaffel, Onghena and De Corte (1999).

Averaged over the different grid sizes, 88% of second graders and 81% of sixth graders involved in the analysis fitted one of the four hypothetical patterns. Table 2 shows the distribution in both age groups over the four hypothetical patterns.

Table 2. *Proportion of Second and Sixth Graders with Response-Time Patterns Corresponding to the Four Hypothetical Patterns for each of the Three Grid Sizes*

Grid size	Pattern			
	1	2	3	4
Second graders				
7 X 7	0.12 ($n = 2$)	0.70 ($n = 12$)	0.12 ($n = 2$)	0.06 ($n = 1$)
8 X 8	0.00 ($n = 0$)	0.64 ($n = 7$)	0.18 ($n = 2$)	0.18 ($n = 2$)
9 X 9	0.21 ($n = 3$)	0.36 ($n = 5$)	0.00 ($n = 0$)	0.43 ($n = 6$)
<i>M</i>	0.11	0.57	0.10	0.22
Sixth graders				
7 X 7	0.05 ($n = 1$)	0.17 ($n = 3$)	0.50 ($n = 9$)	0.28 ($n = 5$)
8 X 8	0.00 ($n = 0$)	0.12 ($n = 2$)	0.44 ($n = 8$)	0.44 ($n = 8$)
9 X 9	0.05 ($n = 1$)	0.14 ($n = 3$)	0.19 ($n = 4$)	0.62 ($n = 13$)
<i>M</i>	0.03	0.14	0.38	0.45

The above mentioned analysis led to an outcome consisting of five categories: a fit with respectively Pattern 1, 2, 3 or 4, or no fit at all. For each participant, we determined his specific outcome profile for the three grid sizes and entered the different profiles in a weighted least squares analysis for repeated categorical data.

To test Hypothesis 3a, stating that there would be more second than sixth graders who would show a fit with Pattern 1 or 2 while the reverse would be true for Pattern 3 and 4, the following two weighted least squares analyses were carried out. In the first analysis we combined the outcome categories "fit with Pattern 1" and "fit with Pattern 2" and compared it with the combination of the outcome categories containing a fit with Pattern 3 or 4, or no fit. Results of this analysis showed a significant main effect of age, $\chi^2(1, N = 69) = 4.82, p = .03$. For each of the grid sizes, the proportion of second graders who fitted Pattern 1 or 2 was significantly larger than the proportion of sixth graders who fitted these patterns. Furthermore, there was a significant main effect of grid size, $\chi^2(2, N = 69) = 7.76, p = .02$. Additional contrast analyses revealed only one significant difference: the proportion of participants showing a fit with Pattern 1 or 2 was significantly larger in the 7 x 7 grid than in the 8 x 8 grid, $\chi^2(1, N = 69) = 7.54, p = .006$. The second analysis was similar to the first one, but this time we combined the outcome categories consisting of a fit with Pattern 3 or 4 and compared this combination with the combined outcome category "fit with Pattern 1 or 2, or no fit at all". This analysis revealed a significant main effect of age, $\chi^2(1, N = 69) = 37.98, p < .0001$. The proportion of sixth graders fitting Pattern 3 or 4 was signifi-

cantly larger than the proportion of second graders who fitted these two patterns.

Hypothesis 3b, stating that a growth in grid size would result in a larger proportion of participants fitting Pattern 2 and Pattern 4, was tested by a weighted least squares analysis in which we combined the outcome categories of a fit with Pattern 2 and 4 and compared them with the combination of a fit with Pattern 1, or 3, or no fit at all. Contrary to our predictions, this analysis did not reveal a significant effect of grid size. Actually, we even found a decreasing number of second graders fitting Pattern 2 with an increasing grid size.

Hypothesis 3c, in which we predicted a decrease in the difference in the fitting patterns between the two age groups with growing grid size, was tested via a weighted least squares analysis in which we compared three outcome categories: "fit with Pattern 2", "fit with Pattern 4", and a combination of a fit with Pattern 1 or 3, or no fit at all. A priori contrast analyses confirmed our prediction: although the proportion of second graders fitting Pattern 2 was significant larger than the proportion of sixth graders fitting Pattern 4 in the 7 x 7 grid, $\chi^2(1, N = 69) = 20.96, p < .0001$, this significant difference disappeared with an increasing grid size.

To summarize, the weighted least squares analyses confirmed Hypothesis 3a: there were more second than sixth graders fitting a novice pattern (i.e., Pattern 1 or 2) and more sixth than second graders fitting an expert pattern (i.e., Pattern 3 or 4). Hypothesis 3b, however, was not confirmed by the data, as we found no significant increase of subjects using the estimation strategy (i.e., Pattern 2 or 4) with a growing grid size. The last part of our third hypothesis was again confirmed by the data. Second graders started to apply the estimation strategy intensively from a smaller grid size on than sixth graders.

The large number of second graders fitting Pattern 2 (i.e., subjects that are assumed to apply only the addition and the estimation strategy) seems to be inconsistent with the analysis of the responses and also with the visual inspection of response-times patterns, which both revealed that a large number of second graders showed traces of using the subtraction strategy. How can this inconsistency be explained? In our opinion, this has to do with the nature of Beem's (1993, 1995, 1999) models. The analysis of responses as well as the visual inspection of response-time patterns showed that second graders typically applied the subtraction strategy only on these trials where almost all squares are filled with blocks. As a consequence, the change point in the graph of response times (See Figure 2) is located very close to the end point of the abscissa, and this reduces the chance that Beem's models will identify a strategy switch at this point in the data set.

Magnitude of Error Rates as a Function of Strategy.

Hypothesis 4 stated that the mean error rates produced by the estimation strategy would be significantly larger than the mean error rates elicited by the addition or subtraction strategy, whereas there would be no significant difference in mean error rates resulting from the last two strategies. To test this hypothesis we selected all participants of whom the response-time patterns fitted Pattern 2, 3 or 4. Based upon the number of change points detected in their response-time patterns, we accordingly divided the data set of their error rates into two or three segments and we determined the mean error rate of each segment. These mean error rates were entered in an analysis of variance with grid size (and age) as independent between-subjects variables and segment as an independent within-subjects variable. Grid size was considered here as a between-subjects variable since there were very few participants who fitted the same model in more than one grid size. The data of these rare participants who actually fitted the same model in more than one grid size, were excluded from the analysis in such a way that only the mean error rates for one particular grid size of these pupils remained in the analysis³.

For Pattern 2, this analysis revealed a significant main effect of segment, $F(1, 19) = 13.00$, $p = .002$. The mean error rates in Segment 1 ($M = 1.90$) were significantly smaller than those in Segment 2 ($M = 6.86$).

With respect to Pattern 3, we were only able to conduct the analysis of variance for the sixth graders since there were too few second graders who fitted this pattern. As expected, this analysis showed no significant difference in mean error rates between Segment 1 and Segment 2.

For Pattern 4, a significant main effect of segment was observed, $F(2, 44) = 35.35$, $p < .0001$. An a posteriori Tukey test revealed that the mean error rate of Segment 2 ($M = 4.73$) was significantly larger than the mean error rate of Segment 1 ($M = 0.53$), $p = .0001$, and of Segment 3 ($M = 1.49$), $p = .0001$. As for Pattern 3, there was no significant difference in mean error rate between Segment 1 and Segment 3.

In sum, as expected, the mean error rates of the numerosity judgements for the trials in the middle region produced by means of the estimation strategy were much larger than those produced by the addition or subtraction strategy for the trials with, respectively, small and large numerosities of blocks. Furthermore, there was no difference between the mean error rates produced on the trials with small numerosities by means of the addition

³ These data were deleted in such a way, that the remaining number of participants was more or less the same for each grid size.

strategy, on the one hand, and those produced on the items with large numerosities by means of the subtraction strategy, on the other.

Discussion

The present study examined the strategic aspects of second and sixth graders' numerosity judgements of colored blocks in different square grids. This research was based on a previous study by Verschaffel et al. (1998) that investigated the age differences in numerosity judgement strategies in a 10 x 10 grid. Interpreting their results by means of Lemaire and Siegler's (1995) theoretical framework regarding strategic change, Verschaffel et al. concluded that the distinct age groups did not differ in the set of strategies they used, but rather that the older participants applied the sophisticated subtraction strategy more frequently, more fluently and more adaptively. The unexpected early appearance of the subtraction strategy among second graders raised the question whether this phenomenon could be due to the rather peculiar nature of the grid size that Verschaffel et al. used in their experiment, a 10 x 10 grid. This specific grid size has some computational advantages compared to other grid sizes, whereby it can be expected to elicit the subtraction strategy more quickly and strongly than other grids. Therefore, the present study investigated second and sixth graders' numerosity judgement strategies in three other grids sized 7 x 7, 8 x 8 and 9 x 9. A difference among second and sixth graders was expected with respect to the first dimension of the theoretical framework of Lemaire and Siegler, namely the specific repertoire of strategies that pupils from both age groups used for the different grid sizes.

Furthermore, we wanted to consider the theoretical and methodological consequences of Verschaffel et al.'s (1998) finding about the existence of a third strategy for determining numerosities besides the addition and subtraction strategy, namely the estimation strategy, which they could not yet incorporate in their rational task analysis nor in their statistical model for strategy identification. To identify the occurrence of each of the three above-mentioned strategies in children's response-time patterns, Beem's (1993, 1995, 1999) two-phase segmented curve model was complemented with a three-phase segmented curve model in the present study.

A first important outcome of the present study was that not only the sixth graders, but also most second graders showed evidence of using the sophisticated subtraction strategy on the less provocative grid sizes. Actually, there was almost no difference in the proportion of subtraction strategy users in the two age groups (as measured by analysing their responses on the trials with the highest numerosities). However, whereas the majority of the sixth

graders applied the subtraction strategy correctly (i.e., they subtracted the determined number of empty squares from a *correct* anchor) according to this analysis, most second graders applied it *incorrectly* by working with an incorrect anchor. What caused this incorrect use of the sophisticated subtraction strategy? Was it because the second graders were not yet familiar with "rectangular area" as a situational model of multiplication (Verschaffel & De Corte, 1996) or because they did not know their multiplication tables by heart? In our opinion, it would be unwarranted to jump to this conclusion. We suspect that second graders' difficulties in using a correct anchor were mainly due to the superficial resemblance of the outlook of the experimental task with the didactic model of the "hundred-field" (Beishuizen, 1993), which played a pivotal role in the mathematics lessons of the children from the youngest age group.

Beem's (1993, 1995, 1999) two-phase and three-phase segmented curve model allowed a scrutinised analysis of pupils' response-time patterns in accordance with the distinct hypothetical response-time patterns derived from the adapted rational task analysis. The majority of the response-time patterns fitted one of the hypothesized data patterns. Most of the predictions based on the adapted rational task analysis were confirmed. In every grid size, the majority of sixth graders showed a fit with one of the two expert patterns (i.e., Pattern 3 or 4), whereas the majority of second graders fitted one of the two novice patterns (i.e., Pattern 1 or 2). There was also evidence for second graders relying already on the estimation strategy in a smaller grid size compared to sixth graders. However, we did not find any evidence for a significant increase of subjects using the estimation strategy (i.e., Pattern 2 or 4) with a growing grid size. Finally, in line with what was found by Verschaffel et al. (1998), few response-time patterns fitted Pattern 1.

The magnitude of the mean error rates of the numerosity judgements for the different segments of the response-time patterns demarcated by the segmented curve models, provided supplementary evidence for the assumed strategies. As expected, numerosity judgements produced by means of the estimation strategy were significantly less accurate than those produced by the addition and subtraction strategy. Moreover, there was no significant difference in accuracy between answers based on the addition and on the subtraction strategy.

At a more general level, the results of the present experiment can be interpreted in terms of Lemaire and Siegler's (1995) theoretical model of strategic change. With respect to the first dimension of their model (i.e., the kind of strategies used), second and sixth graders used the same set of strategies. However, it can be assumed that there will be a difference on this first dimension if one would work with younger children who do not master yet the declarative and procedural knowledge about subtraction and multiplica-

tion which underlies the sophisticated subtraction strategy. Unfortunately, we are not able to draw any conclusions with respect to the second dimension, namely the relative frequency of each strategy. The plan was to investigate this dimension of the model by comparing second and sixth graders' change-point data. The location of the change points would allow us to determine the number of trials on which a particular strategy was applied and, hence, would give us a measure of the relative frequency of each underlying strategy. However, due to the unequal division of both age groups over the hypothetical data patterns and, especially, to the large number of second grader's showing an incorrect anchor, this analysis could not be done. Concerning the third dimension (i.e., the efficiency with which strategies are executed), we also observed a clear difference in efficiency between second and sixth graders. More sixth graders applied the subtraction strategy correctly, resulting in more accurate numerosity judgements as compared to second graders. Finally, the results of the present study do not allow to make strong statements about the development of the fourth dimension, namely the adaptiveness of children's strategy choices. Of course, the finding that the distribution of pupils' strategies over the distinct trials was generally in line with the predictions derived from the rational task analysis, can be considered as an indication of the flexible and adaptive nature of their strategy use. However, the present study yields no answer to the question whether the sixth graders were actually *more* adaptive in their strategy choices than the second graders. To answer this question, one would have to set up an experiment that uses the "choice/no-choice"-paradigm (Siegler & Lemaire, 1997). In this paradigm every subject is administered the same task in two conditions. In one condition ("no-choice") the subject is forced to solve all trials in the task by using one specific strategy; this procedure is repeated for each strategy distinguished in the rational task analysis. The accuracy and/or response time data gathered in this "no-choice"-condition allow the experimenter to determine for each subject and for each trial the most adaptive strategy choice in terms of speed and accuracy. The subject's optimal strategy profile that is derived from his data in the "no-choice"-condition can then be compared with the strategy profile that he exhibited in the "choice"-condition, wherein he could choose among the different available strategies.

After having summarized the results of the present study and having interpreted them in terms of Lemaire and Siegler's (1995) theory of strategic change, we close with some methodological comments and considerations.

First, the present study is the first one in which Beem's (1999) three-phase segmented curve model has been applied. Whereas the study has shown the possibilities of Beem's two- and three-phase segmented curve models for detecting strategy use and strategy change based on response-

time patterns, these models have also some weaknesses. First, when the number of observations within a particular segment is small, the null-hypothesis concerning the b -parameter in the chi-square test cannot be rejected, even for very large values of this b -parameter. Second, the models seem quite insensitive in detecting strategy shifts early in the beginning or near the end of the data pattern. In this respect, we point to the observation that whereas the analysis of participants' answers as well as the visual inspection of the response-time data patterns sometimes both yielded strong indications that second graders had used the subtraction strategy at least on those trials where the grid was almost completely filled with blocks, the statistical analysis with the segmented curve models did not identify these children as users of the subtraction strategy.

Second, although the estimation strategy was incorporated in our rational task analysis and although the availability of Beem's (1999) three-phase segmented curve model allowed us to demarcate this strategy from the addition and subtraction strategies, the present study still does not provide a clear picture about the exact nature of this strategy. The results of the statistical analyses and the visual inspections of the graphs of the response-time patterns suggest that the so-called estimation strategy is actually an amalgam of various unknown processes for determining numerosity other than by means of addition and subtraction strategies which both lead to systematic and predictable linear patterns of response-time and accuracy data. Indeed, in many cases the trials of the middle range did not display a clear (linear) pattern but rather a "cloud" of very different response times (including both extremely short response times and response times reaching the ceiling of 20 s). Further research, involving on-line data-gathering techniques like eye-movement registration and thinking aloud seem necessary to further unravel the nature of the strategies for the judgement of numerosities in the middle region.

Third, in the present study we were not able to analyse in a systematic and fine-grained way the development in children's strategy use and strategy choice between different sessions or even within a single session, since too few participants, especially among the second graders, had determined the anchor correctly for all three grid sizes, - which is necessary for a systematic investigation of this developmental issue. Nevertheless, we are aware that some participants may have behaved in line with novice pattern 1 or 2 for the first session, but may have gradually or suddenly changed their overall approach to the numerosity judgement task towards expert pattern 3 or 4 during the second or the third session. Such kinds of strategic change could be due to a general learning effect, the discovery of a more efficient strategy or of a more efficient way to choose among the available strategies. Moreover, some children may have shifted from one model to another - for

instance, from pattern 2 to pattern 4 - even during a single session. We are planning to study this aspect more deeply in future research.

Finally, we point out that at the end of the experiment every pupil was also interviewed about the way they had handled the task during the experiment. This interview included several open and closed questions about the "what", the "when" and the "why" of their strategy choices. However, due to place restrictions we could not insert this part of the results in the present article. They will be presented and discussed in a separate report that will focus on the relationship between the response-time and error-rate data reported in the present article, on the one hand, and pupils' verbal protocols, on the other.

References

- Beem, A. L., (1993). *Segcurve: A program for fitting two-phase segmented curve models with an unknown change point* [Program manual]. Leiden, The Netherlands: Leiden University, Educational Computing Research Unit.
- Beem, A. L., (1995). A program for fitting two-phase segmented curve models with an unknown change point, with an application to the analysis of strategy shifts in a cognitive task. *Behavior Research Methods, Instruments, & Computers*, 27, 392-399.
- Beem, A. L., (1999). *Seg3curve: A program for fitting three-phase segmented curve models with two unknown change points* [Program manual]. Amsterdam, The Netherlands: University of Amsterdam, Department of Biological Psychology.
- Beishuizen, M. (1993). Mental strategies and materials or models for addition and subtraction up to 100 in Dutch second grades. *Journal for Research in Mathematics Education*, 24, 294-323.
- Binet, A. (1890). La perception des longueurs et des nombres chez quelques petits enfants. *Revue Philosophique de la France et de l'Étranger*, 30, 68-81.
- Cattell, J. M. (1886). Über die Trägheit der Netzhaut und des Sehcentrums. *Philosophische Studien*, 3, 94-127.
- Ippel, M. J., & Beem, A. L. (1987). A theory of antagonistic strategies. In E. De Corte, H. Lodewijks, R. Parmentier, & P. Span (Eds.), *Learning and instruction: European research in an international context* (Vol. 1, pp. 111-121). Leuven: University Press.
- Jevons, W. S. (1871). The power of numerical discrimination. *Nature*, 3, 281-282.
- Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkman, J. (1949). The discrimination of visual number. *American Journal of Psychology*, 62, 498-525.
- Klahr, D., & Wallace, J. G. (1973). The role of quantification operators in the development of conservation of quantity. *Cognitive Psychology*, 4, 301-327.
- Koch, G. G., Landis, J. R., Freeman, J. L., Freeman, D. H., & Lehnen, R. G. (1977). A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics*, 33, 133-158.
- Kuhn, D., & Phelps, E. (1982). The development of problem-solving strategies. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior*

- (Vol. 17, pp. 1-44). New York: Academic Press.
- 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.
- Luwel, K., Verschaffel, L., Onghena, P., & De Corte, E. (1999). *Detecting strategies in response-time data by means of segmented curve models* (Tech. Rep.). Leuven, Belgium: University of Leuven, Center for Instructional Psychology and Technology.
- Marsh, G., Friedman, M., Welch, V., & Desberg, P. (1980). The development of strategies in spelling. In U. Frith (Ed.), *Cognitive processes in spelling* (pp. 339-354). London: Academic Press.
- McGilly, K., & Siegler, R. S. (1990). The influence of encoding and strategic knowledge on children's choices among serial recall strategies. *Developmental Psychology*, 26, 931-941.
- Newman, S. R., Friedman, C. A., & Gockley, D. R. (1987). Children's use of multiple-counting skills: Adaptation to task factors. *Journal of Experimental Child Psychology*, 44, 268-282.
- Siegler, R. S. (1987). Strategy choices in subtraction. In J. A. Sloboda & D. Rogers (Eds.), *Cognitive processes in mathematics* (pp. 81-106). Oxford: Clarendon Press.
- Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.
- Siegler, R. S., & Lemaire, P. (1997). Older and younger adult's strategy choices in multiplication: Testing predictions of ASCM using the choice/no choice method. *Journal of Experimental Psychology: General*, 126, 71-92.
- Siegler, R. S., & McGilly, K. (1989). Strategy choices in children's time-telling. In I. Levin & D. Zakay (Eds.), *Time and human cognition: A life-span perspective* (pp. 185-218). Amsterdam: North-Holland.
- Siegler, R. S., & Robinson, M. (1982). The development of numerical understandings. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior* (Vol. 16, pp. 241-312). New York: Academic Press.
- Stokes, M. E., Davis, C. S., & Koch, G. C. (1996). *Categorical data analysis using the SAS system*. Cary, NC: SAS Institute Inc.
- Svenson, O., & Sjöberg, K. (1978). Subitizing and counting processes in young children. *Scandinavian Journal of Psychology*, 19, 247-250.
- Verschaffel, L., & De Corte, E. (1996). Number and arithmetic. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (Vol. 1, pp. 99-138). Dordrecht: Kluwer.
- Verschaffel, L., De Corte, E., Lamote, C., & Dhert, N. (1998). The acquisition and use of an adaptive strategy for estimating numerosity. *European Journal of Psychology of Education*, 13, 347-370.
- Wundt, W. (1896). *Grundriss der Psychologie*. Leipzig: Engelmann.

Received March, 2000

Revision received September, 2000

Accepted November, 2000