Explanatory latent variable modeling of mathematical ability in primary school : crossing the border between psychometrics and psychology

Hickendorff, M.

CHAPTER 4

Individual differences in strategy use on division problems: Mental versus written computation

This chapter has been published as Hickendorff, M., Van Putten, C. M., Verhelst, N. D., & Heiser, W. J. (2010). Individual differences in strategy use on division problems: Mental versus written computation. *Journal of Educational Psychology, 102*, 438-452.

The research was supported by CITO, National Institute for Educational Measurement.

ABSTRACT

Individual differences in strategy use (choice and accuracy) were analyzed. A sample of 362 Grade 6 students solved complex division problems under two different conditions. In the choice condition they were allowed to use either a mental or a written strategy. In the subsequent no-choice condition, they were required to use a written strategy. Latent class analysis showed that there were 3 subgroups of students with respect to their pattern of strategy choices: primarily using a written strategy (more girls than boys); primarily using a mental strategy (more boys than girls); and using a written strategy on more difficult items, but a mental strategy on the easier ones (almost no weak mathematical achievers). Strategy accuracies were analyzed with explanatory IRT modeling. A between-subjects comparison in the choice condition showed that written strategies were usually more accurate than mental strategies, especially for the weak achievers. A within-subject comparison showed that the performance of students who used mental calculation on a particular item in the choice condition, improved by requiring the use of a written strategy in the no-choice condition.

4.1 INTRODUCTION

In arithmetic, children know and use multiple strategies (e.g., Lemaire & Siegler, 1995; Siegler, 1988a). These strategies and their characteristics have been extensively studied in elementary addition and subtraction (e.g., Carr & Jessup, 1997; Carr & Davis, 2001; Torbeyns et al., 2004b), in elementary multiplication (e.g., Lemaire & Siegler, 1995; Siegler & Lemaire, 1997) and in mental multidigit addition and subtraction (e.g., Beishuizen, 1993; Beishuizen, Van Putten, & Van Mulken, 1997; Blöte et al., 2001; Fuson et al., 1997; Torbeyns et al., 2006). In these domains of mathematics, near consensus is reached on what strategies children use (Torbeyns et al., 2006). In contrast, strategies for division have received considerably less attention (Robinson et al., 2006). In particular, few studies are devoted to more complex division problems in higher grades of primary school (Van Putten et al., 2005). However, complex arithmetic in general (i.e., operations with multidigit numbers for which one may use a written procedure) and complex division in particular are interesting domains in the light of the reform in mathematics education. In addition, the most recent national assessment in the Netherlands reported a large performance decline on these domains since 1987 (J. Janssen et al., 2005). Therefore, it is important to systematically study solution strategies for solving complex division problems, which was the purpose of the present study. Specifically, the aim was to analyze individual differences in strategy choice and strategy accuracy in solving complex division problems by students at the end of primary school (Grade 6).

4.1.1 Solution strategies

Strategy competence is a much-studied topic in cognitive and educational psychology. Lemaire and Siegler (1995) distinguished four dimensions of strategic competence on which individuals may differ: their strategy repertoire (which strategies are used), their strategy distribution (the frequency with which the strategies are used), their strategy efficiency (strategy speed and/or accuracy), and their strategy selection (how strategies are chosen, related to problem and individual strategy characteristics). These dimensions are central to the current study.

In cognitive models such as the Adaptive Strategy Choice Model (ASCM, Siegler & Shipley, 1995), the choice of a strategy on a particular problem is a function of individual strategy performance characteristics for that problem and strategy-choice criteria held by the individual. People tend to choose their strategy adaptively: they choose the fastest and most accurate strategy for a given problem out of their strategy repertoire. Strategy speed and accuracy may vary from individual to individual (Siegler & Lemaire, 1997). Furthermore, individuals differ in the stringency of the threshold for choosing a strategy, the confidence criterion (Siegler, 1988a, 1988b). For example, Siegler (1988a) has found three subgroups of first-graders with regard to their strategy choices, and labeled them as good students, not-so-good-students, and perfectionists. The latter group contrasted with the first two groups with respect to the required certainty that a particular strategy yields the correct answer for choosing that strategy: the perfectionists had a high confidence-criterion (Siegler, 1988a, 1988b). Hecht (2006)) found similar subgroups in adult's multiplication.

In addition to task and individual strategy performance characteristics, individual differences in strategy choices are also the result of the influence of other cognitive and socio-emotional or socio-cultural variables (Torbeyns, Ghesquière, & Verschaffel, 2009). In mathematics, gender differences and achievement level effects on strategy choices have been found. Previous studies reported that girls show a greater reliance on rules and procedures (i.e., they may set their confidence criterion higher), whereas boys seem to use more intuitive strategies (Carr & Davis, 2001; Carr & Jessup, 1997; Gallagher et al., 2000; Timmermans et al., 2007). These findings may also be related to more general gender-related differences in mathematics, such as that girls have lower levels of confidence with mathematics (Mullis et al., 2008; Timmermans et al., 2007; Vermeer et al., 2000). With respect to achievement level, Torbeyns et al. (2006)) found that above-average achievers were more adaptive in their strategy choices than below-average achievers on addition and subtraction up to 100. Similar differences were found by Foxman and Beishuizen (2003) in mental calculation strategy choices of top, middle, and bottom attainment band students. Finally, on complex (multidigit) division problems, weak mathematical achievers more often used mental strategies and less often used written strategies, compared to medium and strong mathematical achievers (Hickendorff et al., 2009b).

Individual differences in strategy use can only be found by analyzing individual profiles of strategy choice and strategy performance over items; aggregate measures such as means or correlations obscure the profiles (Gilmore & Bryant, 2006; Mabbott & Bisanz, 2003). Until recently, strategy use was usually studied by the so-called *Choice* method: letting students solve several problems, subsequently coding their overt and/or covert strategy use, and relate these strategies to the correctness of the answers and time needed to execute the strategy. Such a procedure was also followed in the few studies devoted to solution strategies for complex division (Anghileri et al., 2002; Van Putten et al., 2005). Siegler and Lemaire (1997) have argued that such studies are flawed by selection effects, because strategy performance was assessed in a condition in which students could choose which strategy they used. In that case, estimates of strategy characteristics may be biased by selection effects in two ways. Selection of students could play a role: for example, it could be that the accuracy of a particular strategy is overestimated because it is applied relatively more often by better than by weaker students. Second, selection of items could play a role: it could for example be that the accuracy of a particular strategy is underestimated because it is applied more often on difficult problems than on easy ones.

To estimate strategy characteristics in an unbiased manner, Siegler and Lemaire (1997) have proposed the *Choice/No-Choice* method. Each participant is tested under two different types of conditions. In the *Choice* condition, participants are free to choose their strategy. In the *No-Choice* conditions, participants have to use a specific strategy to solve all items, so that accuracy and speed of that strategy are assessed unbiasedly. Several studies in mathematics have applied the Choice/No-Choice methodology to study solution strategies (e.g., Lemaire & Lecacheur, 2002; Luwel, Verschaffel, Onghena, & De Corte, 2003; Torbeyns et al., 2004b, 2006). An important aspect of the Choice/No-Choice methodology is that the adaptiveness of strategy choices can be evaluated on an individual level, i.e., whether a subject chooses that strategy that *for him or her* is most efficient (Luwel et al., 2003).

4.1.2 Complex division

In the present paper, complex division is defined as division problems in which the quotient is a multidigit number (e.g., $872 \div 4 = 218$), and the divisor may be multidigit too (e.g., $736 \div 23 = 32$). This contrasts with simple division (division problems from the multiplication tables), in which the quotient is a single digit (e.g., $48 \div 8 = 6$; Robinson et al., 2006). Complex division is of special interest for three reasons. First, complex division is an understudied topic thus far. Second, with the mathematics education reform in the Netherlands, instruction in solving complex division problems has changed to the largest extent compared to other arithmetical domains. Third, a large decline in achievement has been observed on this domain in the Netherlands. These latter two points will be discussed below.

Mathematics education has experienced a reform process of international scope over the last couple of decades (Kilpatrick et al., 2001). A common international trend is that students should become active learners who construct their own mathematics (Blöte et al., 2001). In the Netherlands this reform movement is known by the name of Realistic Mathematics Education (RME) (Freudenthal, 1973; Gravemeijer, 1997b; Treffers, 1987). In RME, students' deep understanding of mathematics is pursued, instead of mastery of rules and procedures. Students should acquire insight and flexibility in their use of strategies (Kilpatrick et al., 2001). Instruction is based on the key principle of guided reinvention (Freudenthal, 1973), implying that instructors should give students the opportunity to reinvent the mathematics they have to learn for themselves, according to a mapped out learning route, starting at the informal or intuitive strategies students have. Another characteristic of the reform is that mental arithmetic plays a central role in mathematics education (Blöte et al., 2001). At present, nearly all Dutch primary schools use mathematics textbooks based on RME principles (J. Janssen et al., 2005).

The RME approach to solving complex division problems starts from the informal strategies that young children employ for division. These are direct counting, repeated addition, use of a multiplicative operation (reversed multiplication), and repeated subtraction (Ambrose et al., 2003; Mulligan & Mitchelmore, 1997; Neuman, 1999; Robinson et al., 2006). In the Netherlands, Treffers (1987) introduced progressive schematization, building on the informal strategy of repeated subtraction of multiples (chunks) of the divisor. The resulting solution strategy is increasingly schematized and abbreviated. Therefore, two aspects of the solution strategy can vary: the level of abbreviation and the level of schematization. Figure 4.1 shows examples of strategies varying in abbreviation of chunking (low-level versus high-level, discussed in more detail later) and in whether or not the schematic notation of repeated subtraction is used. In general, there are two fundamental differences between these types of chunking strategies and the traditional algorithm for long division (also in Figure 4.1). First, in the traditional algorithm it is necessary that each subtraction of a multiple of the divisor is optimal, while this is not necessary in the chunking approaches. Second, understanding the place values of the digits in the dividend is not important for applying the traditional algorithm in a correct way, while the place values of the numbers are left intact in the progressive schematization approach. In RME-based textbooks, the traditional long division algorithm is replaced by the alternative approach as introduced by RME. Therefore, complex division can be said to be a prototype of the RME approach (Van Putten et al., 2005), which makes it an interesting domain to study.

Another reason why research on strategies for complex arithmetic, especially division, is needed, is that national assessment results from the Netherlands showed that achievement on these domains has decreased considerably since 1987 (J. Janssen et al., 2005). This decrease occurred between four consecutive large scale national assessments of mathematics achievement at the end of primary school, carried out in 1987, 1992, 1997, and 2004 by the Dutch National Institute for Educational Measurement (CITO). Results showed that achievement has increased on numerical estimation and general number concepts, and to a lesser extent on calculations with percentages and mental addition and subtraction. However, results showed a decline of performance on complex arithmetic. Students who were in their final year of primary school in 2004 performed less well than students who were at their final year in 1987 on complex addition and subtraction, and especially on complex multiplication and division. Between 1987 and 2004, achievement in complex multiplication and division has dropped with more than one standard deviation on the ability scale, with an accelerating trend (J. Janssen et al., 2005).

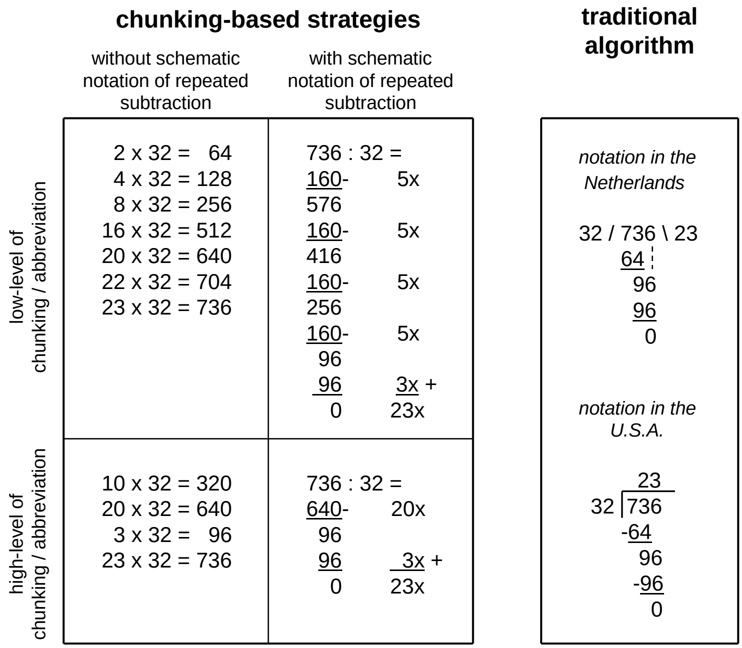

FIGURE 4.1 *Examples of solution strategies for the problem 736 ÷ 32.*

A recent study related this achievement decline on complex division on the two most recent national assessments of 1997 and 2004 to the solution strategies used (Hickendorff et al., 2009b). Results showed that two changes appeared to have contributed to the decline. First, strategy use had shifted: use of written procedures decreased (attributable to a decrease in the use of the traditional long division algorithm), while an increased percentage of items was answered without calculations written down on scrap paper. This shift could be attributed to boys much more than to girls. The strategy shift was unfortunate, since answering these problems with a nonwritten strategy yielded fewer correct answers than when students used a written strategy. Second, each of the solution strategies yielded less correct answers in 2004 than in 1997, with approximately the same amount of accuracy decrease per strategy. So, the performance decline on complex division seems to be related to a change in strategy use, in particular to a decrease in the use of written strategies, and to a change in strategy accuracy as well (Hickendorff et al., 2009b). However, this study was descriptive and therefore limited in several aspects, among which the aforementioned selection effects.

4.1.3 Current study

The purpose of the current study was to study strategy use for solving complex division problems in a more systematic way than has been done in the descriptive studies devoted to this subject. Therefore, a partial Choice/No-Choice design was used: in the Choice condition students could choose whether they used a written or mental strategy in solving a set of complex division problems, and in the subsequent No-Choice condition they had to use a written strategy on a set of parallel problems. Students were interviewed on their nonwritten strategies in the Choice condition.

The main focus in this study was on the distinction between written and mental strategies, because national assessment findings showed that this was a very relevant distinction with respect to the observed performance decrease: use of mental strategies increased over time, but their success rates stayed far behind those of written strategies (Hickendorff et al., 2009b). In the present study, mental computation was defined as carrying out arithmetical operations without the use of any recording devices such as pen and paper, similar to the definitions of for example Reys (1984), Timmermans et al. (2007), and Varol and Farran (2007). Written strategies included all forms of calculations in which some part of the solution process was written down on paper, ranging from only recording intermediate solution steps to written algorithmic procedures.

The design of the current study was a partial implementation of the Choice/No-Choice design, because there was only one No-Choice condition (forced written strategy use). So, there was no No-Choice condition in which students had to answer these problems using mental computation. We believed it would be too large a burden for a great number of students if they had to use a mental strategy on this type of problems on which they would usually need scrap paper. Many students may have become frustrated and unmotivated to continue such a task. Another difference with many of the other studies using the Choice/No-Choice methodology was that response times were not recorded, because we applied a classroom administration procedure, comparable to the procedure in the national assessments.

The present study's first aim was to describe the repertoire and distribution of the written strategy types, as well as of the mental calculation strategies. Two aspects were of particular interest. One aspect concerned the mental strategies. From previous studies, it was not entirely clear that students who did not write down their strategy or intermediate solution steps had used a mental calculation strategy to reach their answer. It could also be that they provided only an estimate of the precise answer, or that they just guessed because they did not know how to solve the problem. The other aspect concerned the repertoire and distribution of (forced) written strategies that were used by students who applied a mental strategy when they could choose. Specific points of focus were to investigate whether these students did have written procedures in their repertoire, and how their written strategy choices compared to those of students who already used a written procedure when they were free to choose.

The second aim was to analyze individual differences in the extent to which students chose mental or written procedures, and relate these individual differences to problem characteristics and to the student characteristics gender and general mathematics level. Regarding problem features, we distinguished problems with a large cognitive demand from problems requiring less cognitive effort. In the former category, the dividend, divisor, and outcome were either large or decimal numbers. On these problems, more written strategies were expected. Problems in the latter category were such that either the dividend could be easily split up, or that a compensation procedure (rounding off the dividend) could be used. We expected that solving these problems would require less cognitive effort, and therefore, less need to write down a solution strategy or intermediate steps. Regarding student characteristics, we first hypothesized that boys would be more inclined than girls to use a mental strategy, replicating the findings of the national assessments (Hickendorff et al., 2009b) and other studies on gender differences in strategy choice (e.g., Carr & Davis, 2001; Carr & Jessup, 1997; Timmermans et al., 2007). Second, we expected that students with lower levels of mathematics achievement would make less adaptive strategy choices than higher achieving students, such as has been found by Foxman and Beishuizen (2003) and Torbeyns et al. (2006).

The final aim of this study was to compare the relative accuracy of written and mental calculation strategies in a more systematic way than was possible in the national assessments (Hickendorff et al., 2009b) and previous studies into complex division (Anghileri et al., 2002; Van Putten et al., 2005). The national assessments showed that the accuracy of written procedures was higher than the accuracy of answering without writing anything down, suggesting that encouraging the use of of a written strategy would improve performance. However, these results were based on comparing the performance results *between* students and items, and thus biased by selection effects. In the present study, these comparisons were made *within* students and items. We hypothesized that the accuracy of students who used mental calculation on a particular item would increase on a parallel item, by forcing them to use a written strategy, since writing down of the solution procedure helps both in recording key information and in schematizing the solution process (Ruthven, 1998). For students who already chose to write down their solution steps in the Choice condition, we expected that forcing them to do so in the No-Choice condition would not be harmful (but also not beneficial) since it would be their usual way of solving the problem.

4.2 METHOD

4.2.1 Participants

The present sample consisted of 362 students from Grade 6, with a mean age of 12.0 years (SD = .51), ranging from 10.4 to 14.5 years. There were 193 boys, 161 girls, and 8 students with missing gender information. The students originated from 12 different schools, located in different regions in the Netherlands with varying levels of urbanization. Each of these schools used a mathematics textbook based on the RME principles, although they did not use the same textbook.

4.2.2 Materials

Experimental task

The experimental task consisted of a total of 13 items: 4 pairs of parallel items yielding 8 items, and 5 additional unpaired items. [1] The complete item set is presented in Appendix 4.A. Parallel item versions were constructed such that the two items within each pair were as similar as possible with respect to item context and number characteristics of the divisor, dividend, and outcome, but would not result in testing effects that would occur when 2 sets of identical items were presented to each student. All 13 items presented a complex division problem that was embedded in a realistic situation of sharing or dividing.

1 The complete task administered to the students consisted of 22 items. The 9 items that were not part of the present study were administered for other purposes of CITO.

As discussed in our hypothesis on the effect of problem features, we distinguished items with large cognitive demands – the four items pairs – from items requiring less cognitive effort – the five unpaired items. In the four item pairs, several item characteristics were varied, taking pair 1 as the base. In pair 2, the dividend and divisor were decimal numbers. In pair 3, the outcome (quotient) was much larger than in the other items, so that students had to subtract more chunks and/or larger chunks. Finally, in pair 4 the outcome was a decimal number, so students had to deal with a remainder. The five unpaired items (items 5 to 9 in Appendix 4.A) were constructed so that they could well be solved mentally, or by only recording some key numbers. Specifically, the dividend and divisor of items 6, 8, and 9 were chosen so that they could be mapped directly (e.g., $3240 \div 4$ could easily be split up in $32(00) \div 4$ and $40 \div 4$). Items 5 and 7 were constructed so that a compensation procedure (rounding the dividend) would be an efficient approach (e.g., $2475 \div 25 = 2500 \div 25 - 1$).

The task was divided into two parts containing separate instructions. Lay-out was designed in such a way that when students had finished the first part (items in the Choice-condition), they could not see the next part.

Standardized mathematics test

The general mathematical ability of each student was measured by the standardized mathematics subtest from the 2007-version of CITO's End of Primary School Test (CITO, 2007). This instrument is widely used in the Netherlands, and assesses the level of achievement of Grade 6 students on mathematics, language, study skills, and world orientation. The 60-item subtest on mathematics has high internal reliability (KR20 = .92; CITO, 2007). In addition to the sum score (the number of items answered correctly), the percentile rank for each student, based on a total of more than 150,000 participants of the End of Primary School Test 2007, could be calculated.

4.2.3 Conditions

In the experimental task, items were administered in 2 different conditions: 9 items were administered in the Choice condition, and 4 items in the No-Choice condition. The 9 items in the Choice condition consisted of a particular version of each of the 4 pairs of parallel items, and of the 5 unpaired items. In this condition, students were free to choose whether they used the scrap paper or not in solving each of the problems presented. If students wanted to write down notes or solution procedures, they had to use the space next to each item for doing this.

In the No-Choice condition, the parallel versions of the first 4 items in the Choice condition were administered. In this condition, students were instructed that they had to write down their solution procedure in a calculation box presented next to each of the items. To encourage the use of the calculation box even more, students were told that if they would not write anything down in the calculation box, the answer they would state would be scored as incorrect.

The Choice condition always preceded the No-Choice condition, to prevent carry-over effects of having to write down solution steps on the free strategy choice. The assignment of item version to condition was counterbalanced, yielding two different forms of the task. In Form A, item versions *a* of each of the 4 item pairs were presented in the Choice condition, and their counterparts (versions *b*) were presented in the No-Choice condition. In Form B, this assignment was reversed. Half of the sample completed Form A, the other half Form B.

4.2.4 Procedure

Classroom administration of the experimental task

The experimental task was administered in the classroom. One of the two specific task forms (A or B) was assigned to each class. The teacher instructed the students that this test consisted of two parts, and that they could start with the first part (the items in the Choice condition), but that they could not start the second part before all students in the classroom had finished the first part. In addition, students were instructed that they could use the space next to each item to write down notes or calculations, and that it was not allowed to use a separate piece of paper.

When the last student in the classroom had finished the first part of the task, students could turn the page to the next part of the task, and the teacher read the instructions about using the calculation box out loud. All students then started the second part at the same time. When all students had finished both parts of the task, they handed in their test booklet. They could take as much time as they needed, so there was no time pressure.

Individual interviews

In 8 of the 12 schools, students interviews took place. After all students had finished both parts of the paper-and-pencil task and had handed in their booklets, the experimenters selected those who failed to write out calculations on at least 1 of the first 4 items. Due to time limitations, only a sample of 89 students from this selection (stratified on a teacher-based judgment of their general mathematics level) were interviewed. They were asked about their solution procedure on problems in the Choice condition on which they had not written down anything but their answer.

Interviews took place individually in a room outside the class, approximately one to three hours after they had finished the experimental task. The experimenters emphasized that they were only interested in *how* the student had solved the problem, so that the students needed not to worry about making mistakes. The experimenters asked the students whether they could remember how they had calculated the answer they had given and whether they could demonstrate the solution steps thinking aloud, for each item solved without written working. These interviews were audiotaped, and the experimenter also made notes.

Standardized mathematics test

The students completed the 2007 End of Primary School Test (CITO, 2007), as part of their final year's standardized assessment. This assessment took place approximately one month after the students participated in the current study.

4.2.5 Solution strategies

The strategy use on the experimental task of all 362 students was categorized. First, strategies were crudely categorized into one of three categories, based on the notes or solution procedures that were written down in the booklets containing the items. These categories were *written* strategies, *mental* strategies (answers stated without written work), or *skipped* items. A more fine-grained classification specified the type of written strategies or the type of mental strategies, respectively.

Types of written strategies were classified based on the work students had written down. A first consideration was whether a traditional strategy or a chunking-based approach was used. The latter category was subdivided into 4 categories, based on the combination of two aspects (see also Figure 4.1). First, this was the level of abbreviation or chunking: divided into *low-level chunking* or *high-level chunking*. For a strategy to be categorized as high-level chunking, the first chunk needed to be at least the largest possible power of 10 times the divisor that fits into the dividend (for example, when solving $782 \div 34$ the first chunk should be at least 10×34, and when solving $4080 \div 20$ it should be at least 100×20). Furthermore, the remaining chunks needed also to be sufficiently efficient (so, no long tail after a first efficient chunk). If these criteria did not hold, the strategy was categorized as low-level chunking. This category included also: trying out several solutions, no chunking at all (for instance, when solving $782 \div 34$ repeatedly subtracting single 34s from 782), or splitting of the dividend (e.g., when solving $33 \div 12$ doing $30 \div 12 + 3 \div 12$). The second aspect on which written strategies were classified was whether or not a schematic notation of repeated subtraction was used. A final category was included containing *wrong* procedures (either the wrong operation or splitting of the divisor, e.g., when solving $782 \div 32$ doing $782 \div 30 + 782 \div 2$) and *unclear* strategies. Interrater reliability of categorization was assessed by computing Cohen's κ (Cohen, 1960) on 200 randomly selected observations (student-by-item combinations) that were coded by two independent experts. For the crude classification (written, mental or skipped) κ was .95, indicating almost perfect agreement. For the fine-grained classification κ was .76, indicating substantial and satisfactory agreement.

Based on audiotaped interviews, the type of strategies that the sample of students interviewed used in solving the items they had answered without written work in the experimental task was inferred. In the large majority of these instances (92%), students reported that they calculated the answer mentally ("in their head"). Within these mental calculation strategies, the following four categories were distinguished. Similar to the types of written strategies, these were first *low-level chunking* and second *high-level chunking* (of course, always without the schematic notation of repeated subtraction). In the third category, students reported that they did not try to calculate the exact answer and that the answer given was a *guess* or a *numerical estimation*. The final category comprised *wrong* and *unclear* mental calculation procedures.

4.3 Results

For 354 subjects, the CITO mathematics test score as well as gender information were available. On the mathematics test, an average score of 45.2 items correct ($SD = 9.3$) out of a total of 60 items was obtained. The present sample of students performed

TABLE 4.1 *Descriptive statistics of strategy use and strategy accuracy.*

	Choice									No-Choice			
strategy	it. 1	it. 2	it. 3	it. 4	it. 5	it. 6	it. 7	it. 8	it. 9	it. 1	it. 2	it. 3	it. 4
	strategy choice: number of students using each strategy												
mental strategy	44	106	42	92	141	155	157	235	196	3	8	2	13
written strategy (total)	313	253	317	264	218	205	202	125	161	354	350	355	343
high-level schema	263	122	234	176	124	114	107	77	93	304	187	281	222
high-level no schema	27	58	39	40	59	57	67	24	55	26	80	30	62
low-level	18	63	37	26	25	26	22	19	10	19	64	37	37
wrong/unclear	5	10	7	22	10	8	6	5	3	5	19	7	22
total (non-skipped)	357	359	359	356	359	360	359	360	357	357	358	357	356
	strategy accuracy: proportion correct per strategy												
mental strategy	.57	.57	.40	.53	.82	.81	.75	.95	.71	n.a.	n.a.	n.a.	n.a.
written strategy (total)	.79	.82	.75	.59	.76	.79	.85	.90	.83	.82	.77	.76	.62
high-level schema	.83	.87	.81	.66	.81	.89	.92	.99	.83	.88	.84	.81	.68
high-level no schema	.78	.90	.77	.60	.83	.77	.84	.83	.87	.73	.84	.73	.71
low-level	.39	.75	.49	.62	.52	.65	.73	.89	.90	.32	.73	.54	.46
total (non-skipped)	.76	.75	.72	.58	.79	.80	.81	.93	.77	.82	.76	.76	.60

slightly better than the national sample of test takers: the median (population) percentile rank was 57.0, with (population) quartiles of 38.3 and 78.0 (these values would have been 50, 25, and 75, respectively, if the distribution of mathematics achievement of this sample would have been completely representative of the total population of participants of the End of Primary School Test 2007). Furthermore, there were gender differences in standardized mathematics achievement: girls answered significantly fewer items correctly ($M = 42.8$, $\mathrm{SD} = 9.3$) than boys did ($M = 47.2$, $\mathrm{SD} = 8.9$), $t(352) = 4.43$, $p < .001$. In the total population, there were similar differences between boys and girls on this mathematics test, so the present sample is representative in that respect.

Table 4.1 shows descriptive statistics of strategy choice (upper part) and strategy accuracy (lower part) for the items in the Choice as well as in the No-Choice condition on the experimental task. The two low-level chunking categories (with and without schematic notation) were taken together because of small numbers of observations. Furthermore, in the present sample there turned out to be only 3 out of the 362 students who used the traditional algorithm. Therefore, the traditional algorithm was grouped with the strategies of high-level chunking with a schema of repeated subtraction. Several aspects of Table 4.1 will be discussed in the following sections in which each of the research objectives is addressed.

4.3.1 Repertoire and distribution of written and mental solution strategies (first aim)

Table 4.1 shows that in the Choice condition, items 1 to 4 were solved with a mental strategy by 12% to 29% of the students. Not surprisingly, these percentages were higher on items 5 to 9, that were devised such that they could be solved mentally more easily than the first 4 items, ranging from 39% to 65%. Items were skipped only by very small numbers of students (2 to 6 students). In the No-Choice condition, items 1 to 4 were answered without written working by 3, 8, 2, and 13 students, respectively. These observations will be left out of the analyses comparing the Choice and No-Choice conditions. Evidently, the experimental manipulation to force students to write their solution steps in the calculation box was successful for the large majority of observations.

Written strategies

Table 4.1 also shows that the vast majority of written procedures consisted of high-level chunking, ranging from 71% on item 2 to 93% on item 1. High-level chunking was applied usually with the schematic notation of repeated subtraction. However, on items 5 to 9 and also on item 2, this schematic notation was relatively less often used than on items 1, 3, and 4. Another interesting result (not presented in Table 4.1) is that on items 5 and 7, which were devised so that applying a compensation strategy would be an efficient approach, compensation was not used very often within the written solution strategies. On item 5, 13% of the 218 written solutions involved compensation of the dividend, and on item 7 this was only in 7% of the 202 written solutions.

Table 4.2 shows the relative distributions of written strategies on items 1 to 4 in the No-Choice condition, separately for students who used a written or a mental computation procedure on the parallel item in the Choice condition. Per item, students were excluded who skipped the item in Choice and/or in No-Choice. The Fisher's Exact test-statistic testing the association between using a mental or written procedure on an item in the Choice condition on the one hand, and the distribution of written strategies used on that item on the other hand, is reported.[2] For each item, the association was significant (see last row of Table 4.2).

2 Statistical testing of the Pearson χ^2-statistic for independence of rows and columns of a contingency table was not feasible. This was because the assumption of most of the cells having an expected cell count of at least 5 was violated in 2 of the contingency tables. Fisher's Exact test was used instead. This test uses exact distributions instead of large sample approximations, such as the Pearson χ^2 does. See for example Agresti (2002).

TABLE 4.2 *Distributions of written strategies in the No-Choice condition, separate for students who solved that item with a mental (m) or written (w) strategy in the Choice condition.*

	item 1		item 2		item 3		item 4	
strategy	m	w	m	w	m	w	m	w
low-level	.07	.05	.16	.19	.18	.10	.18	.08
high-level	.22	.05	.26	.22	.23	.06	.27	.16
high-level schema	.66	.89	.44	.57	.51	.83	.43	.72
wrong/unclear	.05	.01	.14	.02	.08	.01	.12	.05
N	41	310	98	251	39	314	82	257
Fisher's Exact Test	18.1*		20.0*		21.7*		23.3*	

*$p < .001$

On all 4 items, students who used mental computation in the Choice condition relatively more often applied wrong or unclear procedures on that item in the No-Choice condition than students who used a written procedure in the Choice condition. Schematic high-level chunking was relatively more often used by students who already used a written procedure in the Choice condition. In addition, low-level chunking and high-level chunking without schematic notation were relatively more often used by students applying mental calculation in the Choice condition, except on item 2.

Mental strategies

Table 4.3 displays the distribution of the types of mental computation strategies used on items 1 to 9 presented in the Choice condition, based on the sample of 89 students who were interviewed. A clear difference between items 1 to 4 on the one hand, and items 5 to 9 on the other hand appears. Although on all items high-level chunking carried out mentally was the dominant strategy, this dominance was much more pronounced on items 5 to 9 (92% to 100% high-level chunking) than on items 1 to 4 (56% to 65% high-level chunking). On items 1 to 4, the students interviewed sometimes made a guess or numerical estimate, applied low-level chunking mentally, or used a wrong or unclear strategy. These strategies were almost never observed on items 5 to 9.

Another important observation is that within these mental calculation procedures, compensation strategies were used very often on items 5 and 7: on item 5 in 95% of the 43 mental strategies, and on item 7 in 75% of the 36 mental strategies. These high

TABLE 4.3 *Distribution of mental computation strategies on items in the Choice condition.*

strategy	it. 1	it. 2	it. 3	it. 4	it. 5	it. 6	it. 7	it. 8	it. 9
guess/estimate	.05	.06	.00	.27	.00	.00	.00	.00	.00
low-level	.15	.19	.22	.07	.00	.05	.00	.00	.04
high-level	.65	.56	.56	.56	.98	.95	.97	1.00	.92
wrong/unclear	.15	.19	.22	.10	.02	.00	.03	.00	.04
N (interviewed)	20	48	18	41	43	42	36	58	52

percentages are in contrast with the 13% and 7% of the written strategies on items 5 and 7, respectively. These differences in proportions of use of compensation between written and mental strategies are statistically significant on both items (for item 5, $\chi^2(1, N = 261) = 119.0, p < .001$, and for item 7, $\chi^2(1, N = 238) = 91.4, p < .001$).

4.3.2 Individual differences in strategy choices (second aim)

Thus far, strategy distributions were analyzed per item. In this section, we focus on individual differences in the extent to which students chose a written or a mental solution strategy on the items administered in the Choice condition. In the Choice condition, 38% of the students chose at least once for a mental solution strategy on the first 4 items. Considering all 9 items in the Choice condition, this percentage was 79%.

Multivariate analysis: Latent class models

Since each student chooses a strategy on each of the 9 items administered in the Choice condition, resulting data are multivariate. There are within-subject dependencies between the 9 strategy choice variables, and we need analysis techniques that can take these dependencies or correlations into account. As Hickendorff et al. (2009b) argued, latent variable modeling is appropriate. The latent variable (either categorical or continuous) accounts for the correlations within individuals by mapping the multivariate responses on the latent variable. Individual differences between students are allowed for as well, because each individual is allocated a particular position on the latent variable.

To analyze individual differences in the extent to which students chose written or mental procedures (on the items presented in the Choice condition), we used latent class analyses (LCA) (e.g., Goodman, 1974; Lazarsfeld & Henry, 1968. In LCA, the latent variable is assumed to be categorical, representing latent (unobserved) classes or subgroups of subjects. The basic latent class model is $f(\mathbf{y}) = \sum_{k=1}^{K} P(k) \prod_i P(y_i|k)$. Classes run from $k = 1, \ldots, K$, and $\mathbf{y}$ is a vector containing the observed data: the strategy categorization (written/mental) on all items $i = 1, \ldots, 9$ presented to the student in the Choice condition. Resulting parameters are the class probabilities or sizes $P(k)$ and the conditional probabilities $P(y_i|k)$. The latter reflect for each latent class k the probabilities of solving item i with a written and a mental strategy, respectively. As latent class software, we used LEM (Vermunt, 1997), a general and versatile program for categorical data analysis. To decide on the number of latent classes underlying the data, we relied on the BIC-criterion. The BIC-value (Bayesian Information Criterion) is a model fit statistic that penalizes the fit of a model with the number of parameters estimated. Lower BIC-values represent a better trade-off between model fit and parsimony of the model.

In addition to searching for subgroups of students who could be characterized by a specific tendency to apply mental calculation on each of the items, we were interested in potential effects of background variables (gender and general mathematics level) on these class sizes. In other words: are the relative sizes of the latent subgroups different for boys and girls, and/or for students with a weak, medium or strong mathematics level? The existence of these kind of class size differences can be tested by inserting covariate(s) in latent class models, meaning that the covariate (such as gender) predicts class membership (Vermunt & Magidson, 2002). The LC-model with one observed covariate z can be expressed as $f(\mathbf{y}|z) = \sum_{k=1}^{K} P(k|z) \prod_i P(y_i|k)$. Now, class probabilities sum to 1, conditional on the level of the covariate, i.e. $\sum_{k=1}^{K} P(k|z) = 1$. The contribution of a covariate can be tested by a Likelihood Ratio (LR) Test statistic Λ, in which the improvement in fit of the model *with* the covariate relative to the fit of the model *without* the covariate is tested for statistical significance.

Model fitting steps

The 9 dependent variables of the LCA were, for each item in the Choice condition, whether it was solved by a written procedure or by mental calculation. Skipped items were coded missing. There were two explanatory variables: gender (2 categories) and general mathematics level (3 categories, based on the student's obtained percentile rank scores on the standardized mathematics test: weak students with a (population) percentile rank of 33 or below, medium students with a percentile rank between 34 and 66, and strong students with a percentile rank of 67 or higher). Eight students had missing values on either or both of these explanatory variables, and were excluded from these analyses, so $N = 354$.

First, we had to decide how many latent classes to interpret. The model with 3 latent classes had the lowest BIC-value. Taking this 3-class model as the base model, our next step was to assess the effect of gender and general mathematics level (GML) on the sizes of these classes, by including these variables as covariates in the latent class model. Both general mathematics level ($\Lambda = 29.7$, $df = 4$, $p < .001$) and gender ($\Lambda = 38.3$, $df = 2$, $p < .001$) had a significant effect on latent class sizes. Furthermore, gender had a significant effect on the latent class formation independent from GML ($\Lambda = 22.6$, $df = 4$, $p < .001$). Finally, gender and GML did not interact in their prediction of latent class probabilities ($\Lambda = 2.0$, $df = 4$, $p = .74$).

Interpreting the best fitting model

So, we found that a model with 3 latent classes showed the best fit, and gender as well as general mathematics level significantly affected class sizes. First, the latent classes are characterized. Figure 4.2 shows, for each class separately, the probability to use mental calculation on each item. The first class, containing 18% of the students, was characterized by a high probability to apply mental calculation on each item. However, items 1 and 3 were less likely to be solved mentally than the other items. The third class (43% of the students) was characterized by low probabilities to use mental calculation, so these were students who mostly used written procedures. Item 8 had the highest probability to be solved by mental calculation, but this probability was still only .33. The second class (39% of the students) was in between the first and the third class. Students in this class were influenced by item characteristics in their choice between written procedures and mental calculation. On the first 4 items, their probability to use mental calculation did not exceed .17. However, on items 5 to 9 their tendency to use mental calculation was between .55 (item 5) and .85 (item 8).

All three classes showed a similar pattern of strategy choice over the first 4 items, reflecting that students adapted their strategy choices to problem features to some extent. Items 1 and 3 were least likely to be solved by written procedures, followed by items 2 and 4. Items 1 and 3 were whole number division problems, of which item 3 was

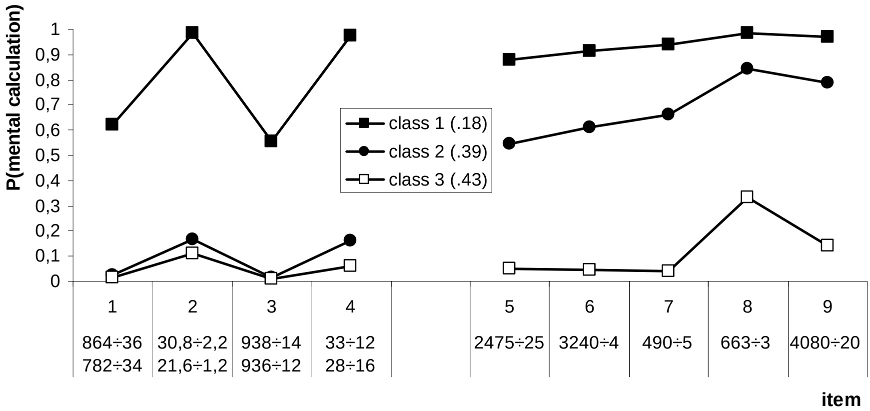

FIGURE 4.2 *Probability of applying mental calculation in 3 latent classes.*

designed to have the highest cognitive demands when doing mental calculation. Items 2 and 4 both dealt with decimal numbers, in the dividend and divisor or in the quotient, respectively. From the remaining 5 items, item 8 triggered mental calculation the most, also for students that were the least inclined to use mental calculation (class 3). This item may have had the least cognitive demands, because 663 can be divided by 3 on a single-digit basis. Although direct mapping of the divisor to the dividend was also possible on items 6 and 8, this may have had a little higher cognitive demands than on item 6, because it was not possible on a single-digit basis, but on a two-digit basis (i.e., 3240 could not be split up in single digits 3, 2, 4, and 0, but instead had to be split up in 32 and 40). Items 5 and 7 were special in the sense that if students did not use a compensation strategy, many steps would have to be taken, requiring high cognitive demands if they would do this mentally. However, when using a compensation strategy, fewer steps need to be taken, which can be well done mentally. So, choosing between mental and written strategies on these items may reflect whether a compensation strategy was used or not.

Table 4.4 shows the effects of gender and general mathematics level (GML), by presenting the class sizes conditional on the combinations of gender and GML[3]. In general, boys were much more likely to be classified in the first class of mainly mental

3 Although gender and GML did not interact in their prediction of class probabilities, class probabilities are still presented for combinations of gender and GML. This was done because boys and girls were not equally spread over the levels of mathematics achievement, which confounded with interpreting the effects of gender and GML separately.

TABLE 4.4 *Estimated class probabilities, conditional on gender and GML. Standard errors (SEs) between brackets.*

gender	GML	class 1 mental	class 2 ment/writ	class 3 written	*N*
	weak	.40 (.09)	.09 (.06)	.51 (.09)	27
boys	medium	.22 (.05)	.50 (.07)	.28 (.05)	66
	strong	.25 (.04)	.52 (.05)	.23 (.05)	100
	weak	.09 (.03)	.06 (.04)	.84 (.05)	49
girls	medium	.06 (.02)	.41 (.07)	.53 (.07)	64
	strong	.07 (.03)	.47 (.08)	.46 (.08)	48
total		.18	.39	.43	1.00

calculation than girls, while the majority of the girls could be characterized as consistently using written procedures. Boys were also slightly more often than girls classified in the class of students that combined written and mental calculation. With regard to general mathematics level, the main differences were found between weak students on the one hand, and medium or strong achieving students on the other hand. These differences predominantly arised in the second and third latent class. About half of the medium and strong students (boys and girls) combined written and mental calculation, while only 6 to 9% of the weak students did so. In contrast to switching between mental and written strategies, weak students were more inclined either to consistently use a written procedure (weak girls) or to consistenly use a mental procedure (weak boys). So, weak students adapted their strategy choices to problem features to a lesser extent than medium and strong students did.

Besides characteristics on the student level, school level characteristics might also have an effect on strategy choices and hence on latent class membership. Since there were 12 schools, entering school as another covariate in the latent class model would result in an unstable model. As an alternative, students were assigned to the class for which they had the highest posterior probability (modal assignment), yielding a classification of students. Relative class sizes differed between the 12 schools: the size of the first latent class (mainly mental) ranged from 0% to 28% between schools, the size of the second latent class (mental/written) ranged from 34% to 53%, and the size of the third latent class (mainly written) ranged from 26% to 55%. However, these school differences were not significant (Fisher's Exact Test = 29.2, $p = .12$), and neither were class size differences with respect to mathematics textbook used (Fisher's Exact Test = 2.8, $p = .59$) or to indicators of social-economical status (Fisher's Exact Test = 3.8, $p = .40$).

4.3.3 Strategy accuracies (third aim)

The lower part of Table 4.1 presents descriptive statistics of the accuracy of each strategy. On the first 4 items in the Choice condition, accuracy of written strategies seems higher than of mental strategies, but this difference will be statistically tested below. Furthermore, comparing the accuracies on a particular item presented in the Choice condition with the accuracies on its parallel version presented in the No-Choice condition does not yield clear differences. However, in the strategy accuracies in the No-Choice condition of Table 4.1, no distinction was made with respect to the type of strategy (mental/written) chosen on the parallel version of that item in the Choice condition. This type of trial (student-by-item combination) information is crucial in the current analyses, because we only expect an accuracy difference between the two conditions on an item for students who chose a mental strategy in the Choice condition *on that particular item*. Only on these student-by-item combinations, strategy use had changed from a mental strategy in the Choice condition to a written strategy in the No-Choice condition.

Multivariate analysis: IRT-models

Strategy accuracy data are also multivarate. Therefore, latent variable modeling is again suitable. To analyze accuracy differences between mental and written strategies, we applied explanatory item response theory (IRT) analyses (De Boeck & Wilson, 2004; Rijmen et al., 2003). In IRT modeling, a continuous latent variable θ is introduced, usually interpreted as the latent ability of each subject. In the most simple IRT measurement model, the Rasch model, the probability of a correct response ($y = 1$) of subject p on item i can be expressed as a logistic (S-shaped) function of the difference between the latent ability of that subject θ_p and the item difficulty β_i, i.e., $P(y_{pi} = 1) = \frac{\exp(\theta_p - \beta_i)}{1+\exp(\theta_p - \beta_i)}$.

Such descriptive or IRT measurement models can be extended by an explanatory part (Rijmen et al., 2003; Wilson & De Boeck, 2004), meaning that covariates or predictor variables are included. These explanatory variables can be subject predictors, item predictors, or subject-by-item predictors. In the present analyses, there were two subject predictors, gender and general mathematics level. There were three predictors on the item level: item number (1 to 9) dummies, parallel item version (a or b), and condition

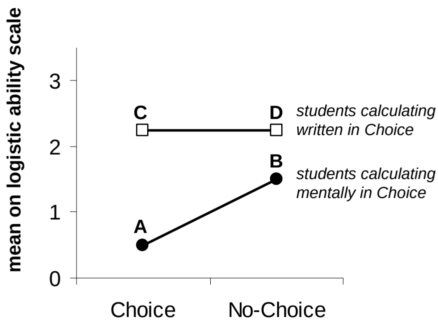

FIGURE 4.3 *Hypothesized group means on logistic latent ability scale for one item pair.*

of administration (Choice/No-Choice). Furthermore, there was one subject-by-item predictor: the strategy a subject used in solving an item (mental or written) in the Choice condition (see Hickendorff et al., 2009b, for further details on using strategy data in explanatory IRT-analyses).

We were interested in the effects of using a mental or a written strategy on the probability to solve an item correctly (i.e., the strategy accuracies), controlled for item difficulty level β_i and potential differences in difficulty level between the two versions of an item. Specifically, we had three hypotheses concerning performance on the item pairs, also presented graphically in Figure 4.3. First, we expected that in the Choice condition, the performance of students using a written strategy would be higher than that of students using a mental strategy, i.e. C > A in Figure 4.3 (a between-subjects comparison). Second, we expected that performance of students who calculated mentally on an item in the Choice condition would be higher on the parallel item presented in the No-Choice condition, i.e. B > A (a within-subjects comparison). Third, we hypothesized that performance of students who used a written strategy on an item in the Choice condition would be equal to their performance on the parallel item presented in the No-Choice condition, i.e. D = C (a within-subjects comparison). In addition, we hypothesized that also on the unpaired items 5 to 9, the performance of students using a written strategy would be higher than that of students using a mental strategy (between-subjects comparisons). All explanatory IRT analyses in the present study were carried out in SAS (SAS Institute, 2002, see also De Boeck & Wilson, 2004).

Model fitting steps

We fitted a series of IRT-models, starting with the general model as illustrated in Figure 4.3. As a first step, several restrictions were imposed to make the model more parsimonious. The effects of condition (Choice/No-Choice) on accuracy were restricted to be equal for items 1 to 4: i.e., the difference between B and A in Figure 4.3 had to be the same for each of the items, and also the D - C difference had to be the same. This restriction resulted in a non-significant loss in model fit ($\Lambda = 3.6$, $df = 6$, $p = .74$), so this was a legitimate simplification of the model. In contrast, restricting the accuracy differences between written and mental strategies in the Choice condition to be equal for all items (i.e., the difference between C and A in Figure 4.3 had to be the same for items 1 to 9) did result in a significant decrease in model fit ($\Lambda = 44.6$, $df = 8$, $p < .001$). So, the differences between chosen mental and written strategies accuracies were item-specific.

As a second step, student characteristics gender and the standardized score on the mathematics achievement test were inserted as predictors in the model. Both variables appeared to have a significant effect on latent ability (for gender, $\Lambda = 8.9$, $df = 1$, $p = .003$; for mathematics achievement, $\Lambda = 176.3$, $df = 1$, $p < .001$). When mathematics achievement was incorporated in the model, adding gender did not have a significant effect on performance anymore ($\Lambda = .1$, $df = 1$, $p = .81$), so gender differences in general mathematics level mediated gender differences in performance on complex division. Mathematics achievement did have a significant effect on performance, as could be expected, and was explored further in the third step. Specifically, two interaction effects of mathematics achievement were tested. It was tested whether the within-subject effect of condition (No-Choice compared to Choice for mental calculators: B - A difference) was dependent on the mathematics level of the student, but this interaction effect was not significant ($\Lambda = .0$, $df = 1$, $p = .90$). However, mathematics achievement and strategy used in the Choice condition (mental/written, between-subjects) did have a significant interaction effect with each other on accuracy ($\Lambda = 17.4$, $df = 1$, $p < .001$). This implied that the item-specific differences between the accuracies of written and mental strategies (the C - A differences) depended on the mathematics achievement level of the student.

Interpreting the best fitting model

Results of the best fitting model are presented graphically in Figure 4.4. We start our interpretation with results on the first hypothesis: the relative accuracies of written and

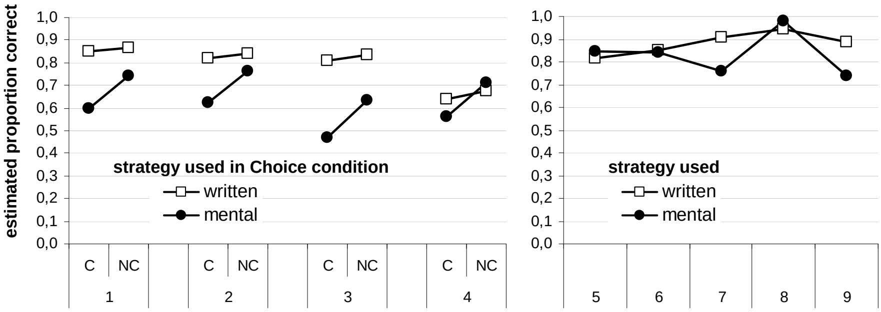

FIGURE 4.4 *Estimated probabilities to solve items 1 to 9 correctly for students at the mean level of mathematics achievement. Left plot: items administered in Choice as well as No-Choice condition, per item students who used mental calculation on that item in the Choice condition are separated from those who used a written procedure. Right plot: items only administered in Choice condition.*

mental strategies on the items in the Choice condition (between-subjects comparison). On more than half of the items, students choosing a written strategy were more accurate than those choosing a mental strategy, but these differences decreased with higher levels of mathematics achievement. For students at the mean level of mathematics achievement, written strategies were significantly more accurate than mental strategies on items 1, 2, 3, 7, and 9. On the remaining items (4, 5, 6, and 8), the accuracy difference between written and mental strategies was not significant for average achievers.

So, although selection effects on these strategy accuracy differences could not be ruled out completely (because they are based on between-subjects comparisons), we could correct for the mathematical achievement level of the students. We found that written strategies were at least as accurate as and usually more accurate than mental strategies. These differences, however, depended on the mathematics achievement level of students, being largest for weak students and smallest for strong students.

Within-subjects comparison of performance on parallel items showed results on the second hypothesis: for those who used a written procedure when they could choose, performance in Choice and No-Choice did not differ significantly from each other (in Figure 4.4, the lines with white squares on items 1 to 4 are almost horizontal). However, as formulated in the third hypothesis, for those who used mental calculation on an item when they were free to choose, performance was higher in the No-Choice condition on the parallel item (when they were forced to use a written strategy). This difference was present on all four items (for items 1 to 4, the lines with black circles in Figure 4.4 show an ascending trend). The average effect of forcing mentally calculating students to write down their solution strategy is .68 (on the logistic scale), SE = .22, $p = .002$, effect size = .87. This large effect is equivalent to raising the probability of a correct answer from 50% to 66%. Furthermore, the size of the effect did not depend on the student's mathematics achievement level.

4.4 Discussion

The main results of this study were that when students were free to choose how they solved the presented problems on complex division, there were individual differences in these strategy choices. Specifically, about 20% of the students predominantly used mental calculation procedures on all items (more boys than girls), 40% quite consistently used a written procedure (more girls than boys), and the remaining 40% used mental calculation on more easy items but written procedures on the more difficult ones (almost no weak students). There were also individual differences in strategy accuracies. Mental calculation was less accurate than applying a written procedure on several items, especially for students having a weak mathematical level. More importantly, when students who used mental calculation to solve a problem were forced to write down their solution steps on a parallel item, they were usually capable of applying a written procedure, and their performance improved. This effect was unaffected by the level of mathematical achievement of the student. So, the present study showed that mental calculation may be a less accurate strategy than writing down notes or calculations on complex arithmetic problems, not only between but also within students and items. Therefore, encouraging students to make use of scrap paper in solving this type of complex division problem would probably improve performance.

4.4.1 Individual differences in strategy use

There was an association between the type of strategy used in the Choice condition (mental/written) and the types of written strategies used in the No-Choice condition. In particular, mental calculators used less structured written strategies when forced to write down their solution steps than written calculators did. This difference may be the result of a more general student characteristic affecting strategy choice: the tendency to use algorithmic strategies (perhaps related to a larger emphasis on accuracy) versus the tendency to use more intuitive, less structured strategies. These different tendencies have been found in studies on gender differences in strategy use (Carr & Davis, 2001; Carr & Jessup, 1997; Timmermans et al., 2007). Similar gender differences in strategy choices were also found in the present study: it were mainly boys who relied on mental calculation, girls were much less inclined to do so.

Furthermore, there were differences between items in the extent that they triggered mental calculation, as came forward from the strategy frequencies (Table 4.1) as well as from the latent class analysis (Figure 4.2). Mental calculation was applied less often on the first 4 items than on the remaining 5 items (which had a smaller cognitive demand due to the number characteristics) to some extent. So, students showed that they adapted their strategy choices to problem characteristics, but individual differences were present in the general tendency to apply a mental or a written strategy. In addition, there were differences between students of different levels of mathematical performance in the extent to which they spontaneously adapted their strategy choices to problem features. There were almost no weak students combining mental and written procedures, while almost half of the students with a medium or strong mathematical level did switch to mental calculation on the easier items, adapting to the item characteristics. So, medium and strong students showed some flexibility in their strategy choices, while weak students did not, a result resembling that of Foxman and Beishuizen (2003) and Torbeyns et al. (2006) with mental calculation.

Another interesting finding with respect to the adaptivity of strategy choices concerns application of the compensation strategy on items 5 and 7. The number characteristics of these two items were such, that rounding the dividend would be a very efficient approach. We found only a small proportion of the written strategies making use of compensation, while, in contrast, the majority of the sample of mental strategies involved compensation. A possible explanation for this difference could be that students who were aware of the possibility of compensation given the number characteristics of these items, could solve these items by compensation in their head. In contrast, students unaware of this possibility of compensation required many more solution steps, for which scrap paper would be useful. Another explanation could be that a third variable, such as mathematical insight, influenced both the awareness of the possibility of compensation, and the skills needed to solve these problems mentally. In either way, the fact that those applying a compensation strategy usually did it mentally, while those who did not use a compensation strategy predominantly used a written strategy, indicates that to some extent an adaptive strategy choice was made.

Another aspect of adaptivity is whether students adapted their strategy choice to their individual strategy accuracies. We can only discuss this topic with respect to accuracy of (forced) written strategies. There is evidence that several students (mainly boys) did not choose adaptively between written and mental calculation, because their performance with mental calculation was lower than when they were forced to write down their solution steps. This performance difference did not depend on the mathematics achievement level of the student, so even high achievers can be said to be unadaptive in this respect. However, it should be noted that adaptiveness in the present study could only be related to accuracy, and not to speed of strategy execution. It could well be that mental calculation is faster than written calculations, and that this plays an important role in students' choices as well.

4.4.2 Methodological considerations

In designing the present study, several methodological choices were made. First, parallel items were used to assess the effect of administration condition. By doing this, it was implicitly assumed that how students solved one version of the item in the No-Choice condition represents how the other item version (presented in the Choice condition) would have been solved in the No-Choice condition, and vice versa. This assumption cannot be tested, because students were never administered the same item version twice. However, because we counterbalanced item version over the administration conditions, we could assess that the parallel item versions did not differ significantly in difficulty level from each other, on any of the four items that had parallel versions.

Second, it was decided not to implement a No-Choice condition in which students would have had to calculate the answer to all items with a mental procedure, because we expected many students to struggle with obligatory mental calculation on these problems with large numbers. As a consequence of having only one No-Choice condition, the difference in accuracy between solving a problem with a mental or a written strategy could not be assessed completely unbiasedly, and therefore, conclusions about adaptivity of strategy choices with respect to individual strategy accuracies could only be drawn on the basis of results for the written strategy. However, we could correct for the students' level of achievement in mathematics in general, ruling out this source of selection effects. In addition, we were able to assess whether the distribution and accuracy of the written strategies in the No-Choice condition were different for students using a mental or a written strategy when they were free to choose. Another disadvantage of having only one No-Choice condition, which was always preceded by the Choice condition, was that alternative explanations for the positive effect on performance of forcing the mental calculators to use a written strategy could not be ruled out. For example, accuracy may increase just from forcing students to use another strategy, irrespective of the particular strategy used, because they have to be more effortful and deliberate in executing their non-chosen strategy. Another explanation may be that by mental calculation, conceptual knowledge is activated that may have a beneficial effect on subsequent problem solving. Future research should include a No-Choice condition in which students would have to use a mental strategy, with a problem set requiring less cognitive effort.

A third methodological consideration was our procedure of classroom administration of the task, which had several consequences. An advantage was that the testing situation resembled that of the national assessments and classroom practice in general. However, a disadvantage was that it was not possible to gather data on strategy speed, because that would have required individual testing. Therefore, the speed of execution could not be incorporated as a strategy performance component. As a result, the adaptivity of strategy choices could only be assessed with respect to accuracy and not to speed, while speed probably is an important predictor of strategy choice as well. Moreover, students may weigh speed and accuracy differently (i.e., they may hold a different speed/accuracy trade-off), which may be an alternative explanation of found class size differences with respect to gender and mathematics level. Future research should take accuracy as well as speed into account to tap these issues.

Another consequence of classroom administration was that the experimental task could only be followed by the interviews one to several hours later. Robinson (2001) showed that retrospective reporting is a valid measure of cognitive processes in children's subtraction, supporting the validity of our procedure. However, it should be noted that in Robinson's (2001) study, children had to report on their solution strategy immediately after they had stated their answer to an item, on a trial-by-trial basis. In contrast, in the present study students completed all items before they were asked whether they could report how they solved the items. This time lag may have negatively affected the veridicality of the verbal reports, i.e., the reports may not have been accurate descriptions of the strategy used due to forgetting and fabrication (Ericsson & Simon, 1993; Russo, Johnson, & Stephens, 1989). On the other hand, having the students verbally report only after all items were completed safeguarded against potential reactivity bias, i.e., the strategy choice and accuracies being affected by verbalization requirements (Russo et al., 1989). However, results on the interview data should be interpreted with caution. Future research should test students individually, and let them report on their strategy use immediately after answering each problem.

4.4.3 Educational implications

The most important implication of this study for school practice probably lies in promoting the value of writing down solution steps on more difficult complex arithmetic problems. As noted before, students nowadays are less inclined than students were a decade ago to use a written strategy in solving these kind of problems on complex division (Hickendorff et al., 2009b). In the present study we showed that both in comparisons between as well as within students, mental calculation may be less accurate than written calculation. That raises the question what role school practice plays in the strategy choices students make. It might be that the large emphasis on mental calculation in RME has had the side-effect that some students overuse mental calculation. A recommendation is that teachers emphasize these possible benefits of writing down notes or calculations to their students.

Another interesting finding was the near absence of the traditional algorithm for long division. Apparently, instruction regarding division was completely based on RME principles, at least in the 12 schools that were part of our sample. Instruction in the traditional algorithm has been discredited because it was said to be a mechanistic trick, in which understanding and insight in the numbers and their interrelations are not fostered. However, we argue that it might be possible to build in the traditional algorithm as the optimal form of abbreviation at the end point of the learning trajectory.

APPENDIX 4.A ITEM SET

item	**version *a***	**version *b***
1	Fanny takes piano lessons. She has to pay 782 euros for 34 lessons. How much does one lesson cost?	Marleen takes piano lessons. She has to pay 864 euros for 36 lessons. How much does one lesson cost?
2	The machine packs up 1.2 kilos of chocolate per minute. How many minutes does it take to pack up 21.6 kilos of chocolate?	The machine packs up 2.2 kilos of candies per minute. How many minutes does it take to pack up 30.8 kilos of candies?
3	FEEDING BOX SUITED FOR 12 COWS The farmer had 936 cows. How many of these feeding boxes does he need?	FEEDING BOX SUITED FOR 14 PIGS The farmer had 938 pigs. How many of these feeding boxes does he need?
4	16 children go to the playground. Together, they have to pay 28 euros. How many euros is that per child?	12 children go to the museum. Together, they have to pay 33 euros. How many euros is that per child?
5	Saskia sells DVDs for € 25 a piece. She received € 2475. How many DVDs did she sell?	
6	A cycle path of 3240 meters will be covered with concrete plates. Each plate is 4 meters long. How many plates are needed for the entire cycle path?	
7	Grandma divides 490 euros among her 5 grandchildren. How many euros does each grandchild get?	
8	The DVD-player costs € 663,-. Jasper pays this amount in 3 times, each time the same amount. How much does he have to pay each time?	
9	The farmer has 4080 liters of milk in his cooling basin. The milk is distributed over milk cans: each milk can is filled with 20 liters of milk. In total, how many milk cans will the farmer fill?	

Note. Items are translated from Dutch. Illustrations of items 5, 6, 8, and 9 are not shown.