Solving multiplication and division problems: latent variable modeling of students' solution strategies and performance

Fagginger Auer, M.F.

Author: Fagginger Auer, M.F.
Title: Solving multiplication and division problems: latent variable modeling of students' solution strategies and performance
Issue Date: 2016-06-15

— 5 —

Affecting students' choices between mental and written solution strategies for division problems

Abstract

Making adaptive choices between strategies is a central element of current day mathematics, but not all students may be able to do so. Suboptimal choices between mental and written division strategies are indicated for lower mathematical ability students. Strategy choices in this domain were related to student and teacher factors for 323 sixth graders, and for 224 lower ability students an intervention promoting choices for relatively accurate written strategies was evaluated using a pretest-posttest design. Written strategy choices and performance increased considerably for students receiving intervention or control training, but not for students who did not receive any training. Results suggest that students' strategy choices may also be affected by targeting their motivation and the sociocultural context for strategy use.

5.1 Introduction

Tasks are executed using a variety of strategies during all phases of development (Siegler, 2007). For example, infants vary in their use of walking strategies (Snapp-Childs & Corbetta, 2009), first graders in their use of spelling strategies (Rittle-Johnson & Siegler, 1999), and older children in their use of transitive reasoning strategies (Sijtsma & Verweij, 1999). This large variance in strategies goes together with widely differing performance rates of the different strategies, thereby having

This chapter is currently submitted for publication as: Fagginger Auer, M. F., Van Putten, C. M., & Hickendorff, M. (submitted). *Affecting students' choices between mental and written solution strategies for division problems.*

We would like to thank the schools and students for their participation in the experiment, and the Dutch National Institute for Educational Measurement Cito for allowing use of the assessment items.

profound effects on performance levels. As such, strategies have been a topic of continued investigation.

Children's and adults' solution strategy use has been investigated for many cognitive tasks, such as mental rotation (A. B. Janssen & Geiser, 2010), class inclusion (Siegler & Svetina, 2006), and analogical reasoning (Stevenson, Touw, & Resing, 2011). A cognitive domain that has featured prominently in strategy research is arithmetic. Many studies have been conducted on elementary addition (e.g., Barrouillet & Lépine, 2005; Geary et al., 2004), subtraction (e.g., Barrouillet et al., 2008), multiplication (e.g., Van der Ven et al., 2012) and division (e.g., Mulligan & Mitchelmore, 1997), which concern operations in the number domain up to 100 that are taught in the lower grades of primary school. Some studies have also addressed strategy use on the more complex multidigit (involving larger numbers and decimal numbers) arithmetical tasks in the higher grades (e.g., Van Putten et al., 2005; Selter, 2001; Torbeyns, Ghesquière, & Verschaffel, 2009).

5.1.1 Determinants of strategy choices

Different aspects of strategy use for both elementary and multidigit arithmetical problems can be discerned (Lemaire & Siegler, 1995): individuals' strategy repertoire (which strategies are used); frequency (how often each strategy is used); efficiency (the accuracy and speed of each strategy); and adaptivity (whether the most suitable strategy for a given problem is used). These four aspects together shape arithmetical performance. With reforms that have taken place in various countries over the past decades (Kilpatrick et al., 2001), the aspect of adaptivity has become particularly important. Building on students' own strategic explorations and developing adaptive expertise in flexibly using an array of strategies now take a central place, instead of perfecting the execution of a single algorithm per problem type (Gravemeijer, 1997; Verschaffel et al., 2009). This makes choosing the most suitable strategy for a given problem (i.e., making an adaptive strategy choice) crucial.

There are several ways in which the adaptivity of a strategy choice can be defined, as described by Verschaffel et al. (2009). One way is to define adaptivity purely based on task variables: the characteristics of a problem determine which strategy is adaptive (e.g., the adaptive strategy choice for a problem like $1089 \div 11$ is compensation: $1100 \div 11 - 1$). However, individuals differ in their mastery of different strategies, and the strategy that is most effective for one person does not have to be that for another person. Therefore, a second way to define adaptivity also takes subject variables into account: the strategy that is the adaptive choice is the one that is most effective for a given problem for a particular person. A third way looks even further and includes context variables in the definition. These can be variables both in the direct context of the test (e.g., time restrictions and characteristics of preceding items) and in the broader sociocultural context. In their discussion of adaptive expertise in elementary mathematics education, Verschaffel et al. (2009) stress the importance of more educational research attention to these sociocultural context variables.

Ellis (1997) reviewed research on this topic and argues that the sociocultural context is very important in shaping individuals' strategy repertoire and choices. Students have an implicit understanding of which ways of problem solving are valued by their community: whether speed or accuracy is more important; whether mental strategies are valued over using external aids; whether using conventional procedures or original approaches is preferred; and whether asking for help in problem solving is desirable. Ellis (1997) describes examples of existing differences in strategy use between different cultures (e.g., Western, Asian, aborigine and Navajo cultures). What is also interesting, and moreover, highly practically relevant, is to investigate in what way the context may be manipulated to favorably influence strategy choices.

5.1.2 Influencing students' choices between mental and written division strategies

A case in which influencing students' strategy choices could have large beneficial effects for performance, is that of mental and written strategies for multidigit division problems. As previously described, the attention to traditional algorithms decreased during the reforms of mathematics education. In the Netherlands, this was most extreme for the operation of division, for which the traditional algorithm was abandoned in favor of a new standardized approach (Buijs, 2008; J. Janssen et al., 2005). The traditional and newer approach (see Table 5.1 for examples) differ in that the traditional algorithm is digit-based in the sense that it breaks the dividend up into digits (e.g., in Table 5.1, the 54 part of 544 is considered separately in subtracting 34, and the rest of the dividend is only considered in a later step), whereas the newer approach is whole-number-based and considers the dividend as a whole (e.g., in Table 5.1, 340 is subtracted from 544; Van den Heuvel-Panhuizen et al., 2009). Dutch national assessments in 1997 and 2004 showed the expected decrease in sixth graders' use of the digit-based algorithm, but use of the whole-number-based approach did not increase accordingly; instead, students made more

Table 5.1: Examples of the digit-based algorithm, whole-number-based algorithm, and non-algorithmic strategies applied to the division problem 544 ÷ 34.

digit-based algorithm	whole-number-based algorithm	non-algorithmic written strategies
34/544\16	544 : 34 =	10 × 34 = 340
34	340 - 10×	13 × 34 = 442
204	204	16 × 34 = 544
204	102 - 3×	
0	102	
	102 - 3×+	
	0 16×	

use of strategies without any written work (Hickendorff et al., 2009).

These mental strategies turned out to be very inaccurate compared to written strategies (digit-based or otherwise), suggesting a lack of adaptivity of strategy choices with regard to accuracy, and a large performance decline for multidigit division was observed on the assessments (Hickendorff et al., 2009). In follow-up studies, Fagginger Auer, Hickendorff, and Van Putten (2016) and Hickendorff et al. (2010) showed that requiring (lower mathematical ability) students who answer without any written work to write down calculations improved their performance. This shows that requiring the use of more efficient strategies can affect performance favorably in the short term, providing a concrete suggestion for educational practice. A valuable extension of this finding would be an investigation of instructional contexts that increase students' *choices* for efficient strategies in the longer term, thereby instilling more sustainable improvements in performance.

5.1.3 Present study

The present study is intended as a first step of such an investigation of the determinants of sixth grade students' choices between mental and written division strategies. In the first part of the study, existing differences in these strategy choices are related to students' motivations and attitudes in mathematics and to the sociocultural context for mathematics provided by the students' teachers. In the second part of the study, an intervention designed to increase students' free choices for written rather than mental strategies (and thereby, their performance) is evaluated. Since mental strategies appear especially inaccurate for lower ability students (Fagginger Auer et al., 2016; Hickendorff et al., 2010), our intervention focuses on this group. Using a pretest-posttest design, an intervention training condition consisting of training sessions designed to promote writing down calculations is compared to a control training condition where strategy use is not targeted, and to a no training condition.

A meta-analysis by Kroesbergen and Van Luit (2003) on mathematics interventions for low ability students showed that effect sizes were larger for interventions that featured direct instruction and self-instruction compared to interventions with mediated instruction, and smaller effect sizes for interventions with computer-assisted instruction and peer tutoring compared to interventions without those elements. More specifically, in another meta-analysis on this topic, Gersten et al. (2009) identified explicit instruction as an important component of effective interventions. This explicit instruction involves a step-by-step problem solving plan for a specific type of problems, that is demonstrated by an instructor and that students are asked to use. In order to maximize the potential efficacy of the intervention training in the present study, this training therefore involves direct instruction by a human, adult instructor using a step-by-step plan.

Hypotheses

The investigation of determinants of existing differences in mental versus written division strategy choices is exploratory in nature, and involves of a number of potentially relevant factors. Several of the aspects of the sociocultural context (as seen by the teacher) described by Ellis (1997) as influential with regard to strategy choices are considered: importance of speed versus accuracy, preference for mental strategies versus use of external aids, and preference for conventional versus original approaches. In addition, students' self-rated functioning in mathematics and motivation, teachers' characteristics, and the mathematics textbook and division algorithm instruction are considered.

As for the effects of the intervention: written strategy choices are expected to increase more from pretest to posttest in the intervention than in the control training group, given that they are is only promoted in the former group. Given the higher accuracy of written compared to mental strategies, performance is therefore expected to increase more in the intervention than in the control training group (though the control group should also improve because of the additional practice and attention that students receive). In the no training group, no large changes in strategy choices or performance are expected because of the lack of training and the limited amount of time that passes between the pretest and posttest.

The effect of the intervention training may depend on students' characteristics. As boys appear to use more mental strategies for division than girls (Fagginger Auer et al., 2013; Hickendorff et al., 2009, 2010), there is more room for improvement through training in boys than in girls. Mathematical ability level may also be relevant, as mental strategies are especially inaccurate for lower ability students (Fagginger Auer et al., 2016; Hickendorff et al., 2010), and therefore increases in the use of written strategies may affect performance more when ability is lower. Finally, training may have a larger effect on performance when students' working memory capacity is lower, because then the working memory resources freed up by writing down calculations make more of a difference (in line with cognitive load theory; Paas, Renkl, & Sweller, 2003). This is especially relevant in our sample, given that students with a lower mathematical ability tend to have a lower working memory capacity than higher ability students (Friso-van den Bos, Van der Ven, Kroesbergen, & Van Luit, 2013).

5.2 Method

5.2.1 Participants

A total of 323 sixth graders (53 percent girls) with a mean age of 11 years and 8 months (SD = 5 months) from 19 different classes at 15 different schools participated in the study. For all students, a general mathematical ability score from a widely used standardized national student monitoring system (J. Janssen et al., 2010) was available. All students participated in the pretest and posttest, but training was only given to the 147 students with mathematical ability percentile scores between 10 and 50. Students scoring in the lowest performing decile (7 percent in our sample) were excluded, because atypical problems such as dyscalculia could occur in this group. Of the selected students, 74 received intervention training and 73 control training. They were assigned to a training condition using random assignment with gender, ability quartile and school as blocking variables.

For an indication of development independent of training, performance and strategy choices were also investigated for students who did not receive any training. However, no students with the same ability level as the students who received training were available, so data from the 77 students in the adjoining ability groups (the quartile just above the median and the lowest decile) was used, as in a regression discontinuity design (Hahn, Todd, & Van der Klaauw, 2001). The ability scale scores in the untrained group were on average somewhat higher and they were more varied ($M = 101.9$, $SD = 13.4$) than in the control training ($M = 97.9$, $SD = 5.3$) and intervention training group ($M = 97.3$, $SD = 5.5$).

5.2.2 Materials

Pretest and posttest

The pretest used to asses students' division strategy choices and performance contained the twelve multidigit division problems given in Table 5.2 (for the problems not yet released for publication as they will be in future assessments, parallel versions are given in italics). These problems were taken from the two most recent national assessments of mathematical ability at the end of primary school (J. Janssen et al., 2005; Scheltens et al., 2013), so that our results could be interpreted relative to the national results that called for this line of mathematical strategy research. All problems were situated in realistic problem solving context (e.g., determining how many bundles of 40 tulips can be made from 2500 tulips), except for the problem $31.2 \div 1.2$. The test also contained twelve problems involving other mathematical operations (all from the most recent national assessment), so that it more closely resembled a regular mathematics test to students. The posttest was identical to the pretest to allow for a direct comparison of results, and with the tests being a month apart and students solving similar problems on a daily basis in mathematics lessons during that period, it was very unlikely that students remembered any of the (complex) solutions.

Accuracy (correct or incorrect) and use of written work (yes or no) were scored for every problem. For solutions with written work, a further distinction was made between three strategy categories: the digit-based algorithm; the whole-number-based algorithm; and non-algorithmic written strategies (see Table 5.1 for examples).

Training problems

The problems used in the three training sessions in between the pretest and posttest were three sets of parallel versions of the twelve problems in those tests.

Student and teacher questionnaires

The students filled out a questionnaire on their attitude towards mathematics and mental mathematical strategies consisting of seven questions. The teachers filled out a questionnaire of fifteen questions on their attitude towards and instruction

Table 5.2: The division problems that students had to solve at the pretest and posttest.

problems			
$1536 \div 16 = 96$	$872 \div 4 = 218$	*31.2 ÷ 1.2 = 26*	*6496 ÷ 14 = 464*
544 ÷ 34 = 16	$11585 \div 14 = 827.5$	*47.25 ÷ 7 = 6.75*	$157.50 \div 7.50 = 21$
$2500 \div 40 = 62$	$1470 \div 12 = 122.50$	$736 \div 32 = 23$	$16300 \div 420 = 39$

Note: Parallel versions of problems not yet released for publication are in italics.

of division algorithms, writing down calculations, and various aspects of flexible strategy use. Both questionnaires can be found in the Appendix.

Working memory tests

The verbal working memory capacity of students who received training was assessed using a computerized version (Stevenson, Saarloos, Wijers, & De Bot, in preparation) of the digit span test from the WISC-III (Wechsler, 1991), and their spatial working memory using a computerized version (Stevenson et al., in preparation) of the Corsi block test (Corsi, 1972).

5.2.3 Procedure

The experiment was conducted over a period of five weeks in the fall of 2014. In the first week, the students first completed the pretest in a maximum of 45 minutes in their classroom. They then did the two working memory tasks on the computer and filled out the student questionnaire. The teacher also filled out the teacher questionnaire in this first week. In the following three weeks, the students participated in three individual training sessions of fifteen minutes each (one per week) with the experimenter. The experiment was concluded in the fifth week, in which students did the posttest in again a maximum of 45 minutes in their classrooms.

The training sessions consisted of the students working on the set of training problems for that week. The experimenter evaluated each solution when it was written down and told the student whether it was correct or incorrect. When correct, the students proceeded to the next problem. When incorrect, the student tried again. Accuracy feedback was provided again, and regardless of whether the solution was correct this time, the student proceeded to the next problem. The session was terminated when fifteen minutes had passed.

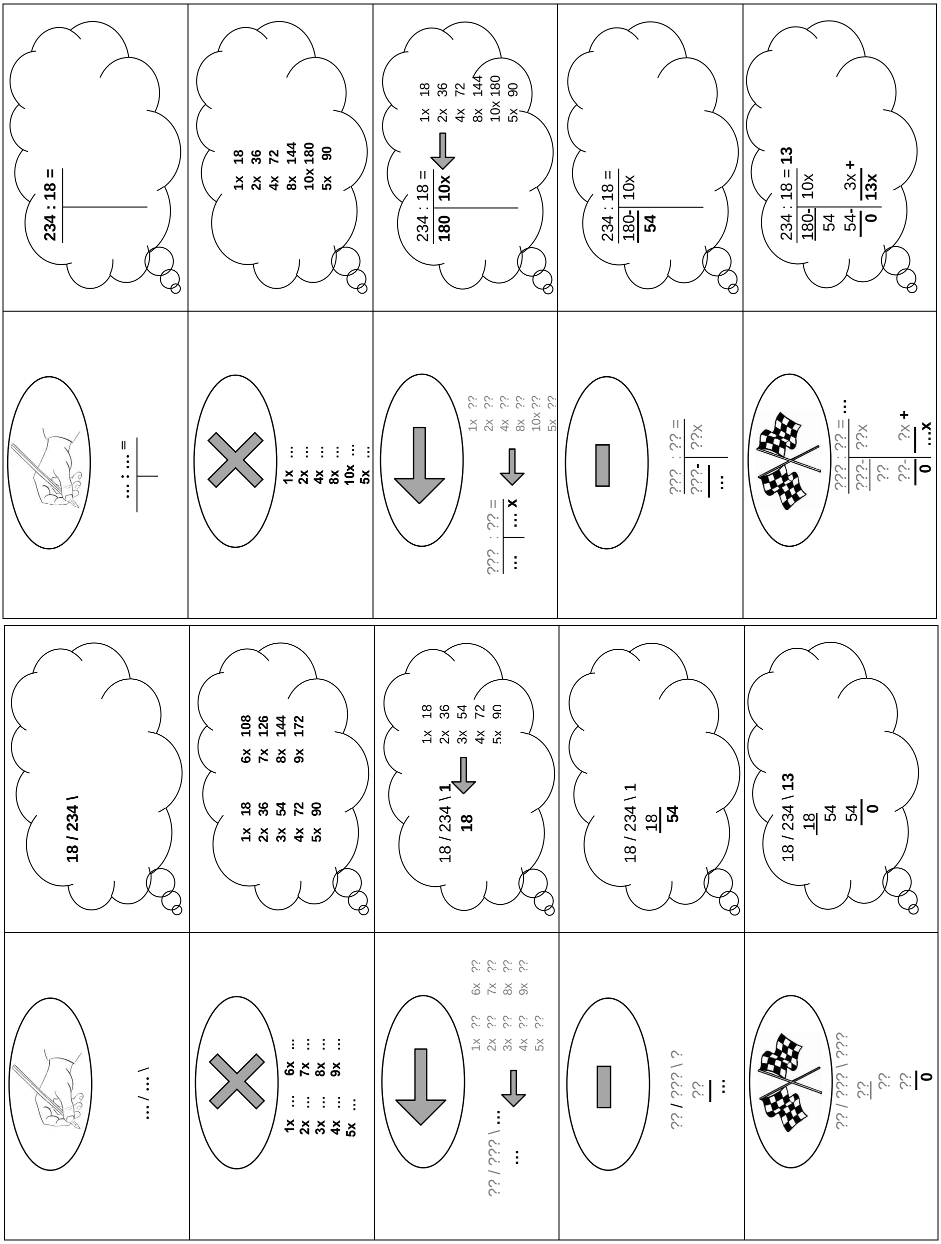

Figure 5.1: The step-by-step plans (the lower one for students using the digit-based algorithm, and the upper one for students using the whole-number-based algorithm).

Though these elements of the training were the same for the control and training conditions, two important aspects differed. The first is that students in the control condition were free in how they solved the problems (just as in the pretest), whereas the students in the intervention condition had to write down their calculations in a way that would allow another child to see how they had solved the problem (but otherwise, strategy choice was free). In addition, while students in the intervention condition made their first attempt at solving the problem independently (using a written strategy of their own choice), if they failed, they were provided with systematic feedback on writing down calculations in a standardized way at the second attempt. The students in the control condition received no such feedback and made both their first and second attempt independently.

A step-by-step plan was used for providing the feedback on writing down calculations in the intervention condition, while there was no such plan in the control training condition. The step-by-step plan was always on the table for the intervention training students so they could use it whenever they wanted, and when intervention students were stuck in their problem solving, the experimenter used the plan and standardized instructions to help the students with writing down calculations. No feedback was given on the accuracy of what students wrote down (e.g., mistakes in the multiplication table), except for the final solution.

There was a version of the plan for students taught the digit-based algorithm and one for students taught the whole-number-based algorithm (see Figure 5.1). Both versions consist of five highly similar steps (with step 3 and 4 repeated as often as necessary): (1) writing down the problem; (2) writing down a multiplication table (optional step); (3) writing down a number (possibly from that table) to subtract; (4) writing down the subtraction of that number; and (5) finishing when zero is reached, which in the case of the whole-number-based algorithm requires a final addition of the repeated subtractions. Each step is represented by a symbol to make the step easy to identify and remember (the symbols in the ellipses on the left side of the scheme). Below this symbol, a general representation of the step is given, with question marks for problem-specific numbers already present at that step and dots for the numbers to be written down in that step. On the right-hand side of the plan, an example of the execution of each step for the particular problem $234 \div 18$ is given in a thinking cloud. On both sides, the elements to be written down in the current step are in bold font.

5.2.4 Statistical analysis

Correlation analyses

To explore possible relations between the questions on the student and teacher questionnaires and students' written strategy choices on the pretest, correlations rather than formal models were used because of the high number of questions involved. Point-biserial correlations were used for dichotomous questionnaire responses and Spearman's rank correlations for scales.

Explanatory IRT models

More formal tests were conducted using explanatory item response theory (IRT) models. As argued by Stevenson, Hickendorff, Resing, Heiser, and de Boeck (2013), measuring learning and change has inherent problems that can be addressed using explanatory IRT. These are problems such as the dependence of the meaning of scale units for change on pretest score, because of the non-interval measurement level of non-IRT scores (e.g., an increase of one in the number correct does not necessarily mean the same for a person who already had a nearly perfect score as for someone who had a lower score).

IRT models place persons and items on a common latent scale (Embretson & Reise, 2000). The distance between the persons and items on that scale determines the probability of a correct response: if person ability and item difficulty are close together that probability is around fifty percent, whereas it is lower if ability is lower than difficulty, and higher if ability is higher than difficulty. In its most basic form, the (Rasch) model for the probability of a correct response of person p with ability θ_p on item i with difficulty β_i is $P(y_{pi} = 1|\theta_p) = \frac{exp(\theta_p - \beta_i)}{1+exp(\theta_p - \beta_i)}$. The estimated ability parameters for persons are more likely to have an interval measurement level than simple sum scores.

This model becomes explanatory when explanatory factors for items' difficulty or persons' ability are included, which can be item covariates (not used in the present study), person covariates (condition and student gender, ability score and working memory in the present study), and person-by-item covariates (solution strategy choice in the present study). This type of models can be estimated as multilevel logistic regression models using general purpose generalized linear mixed model (GLMM) software, by fitting a binomial model with solution accuracy (correct or incorrect) as the dependent variable, a random intercept for students as the ability parameter, and the covariates of interest as fixed effects (De Boeck &

Wilson, 2004).

In the present study, different explanatory IRT models were fitted using the `lme4` package in `R` (Bates, Maechler, Bolker, & Walker, 2014; De Boeck et al., 2011). All models were random person-random item Rasch models (RPRI; De Boeck, 2008), with a random intercept for students, and also a random intercept for the item effects (as they were considered a draw from the larger domain of multidigit division). The different covariates were added in stepwise fashion (as in Stevenson et al., 2013), so that the added value of each addition could be evaluated by comparing the models based on the Aikaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and likelihood ratio tests. The AIC and BIC balance model fit and parsimony and lower values of these criteria are better, and a significant likelihood ratio test indicates that of the two models that are compared, the more complex model fits significantly better. Of the final best fitting model according to these various criteria, the regression parameters were interpreted. Since our research question did not only concern accuracy (correct vs. incorrect) but also strategy choice (written vs. not written), and IRT models accommodate dichotomous variables regardless of content, strategy use was modeled in the same way. The person parameter θ_p then reflects individual differences in the tendency to use written strategies.

For an indication of the size of significant effects, the probability P of a correct response or of using a written strategy is given for different levels of the covariate, with all other covariates in the model set at the mean in the sample. For example, for the effect of testing occasion (pretest or posttest), the probability of a correct solution for an average student on an average problem on the pretest and on the posttest is given. For numeric covariates (e.g., ability score) the effects of a difference of one standard deviation around the mean ($M - 0.5SD$ to $M + 0.5SD$) are given.

5.3 Results

5.3.1 Relation between student and teacher factors and written strategy choices

First, an exploration of pre-existing differences in choices for written strategies based on students' attitudes with regard to mathematics and teachers' strategy instruction was made using the pretest data. Students used written strategies in 62 percent of their pretest solutions, which varied between 51 percent for the problem $31.2 \div 1.2$ and 87 percent for the problem $11585 \div 14$.

Student questionnaire

The Appendix shows what all students ($N = 323$) reported on the student questionnaire on their mathematical attitudes. The proportion of students choosing each alternative is given in brackets after the respective alternative. After each question, the correlation between the question response and the overall proportion of pretest division problems solved with written strategies is also given.

On average, the students had a slightly positive attitude towards mathematics ($M = 3.2$ on a 5-point scale) and were slightly positive about their mathematical ability ($M = 3.3$), and the more positive their attitude and the higher their judgment of ability, the higher their frequency of choices for written strategies ($r(322) = .17$ and $r(322) = .21$ respectively). Students reported putting quite some effort into math ($M = 4.3$) and almost all (98 percent) reported valuing accuracy over speed, but these factors were unrelated to written strategy choices. A majority of students (72 percent) found it more important to be able to solve mathematical problems with than without paper, and this was positively related to using written strategies ($r(318) = .19$). Students reported sometimes answering without writing down a calculation ($M = 2.8$), and indeed, reporting more frequent mental calculation was negatively related to using written strategies ($r(322) = -.17$).

Students also reported on reasons they had for not writing down calculations, on the occasions that they used this approach (which were less frequent for some students and more frequent for others). The most popular reason (chosen by 60 percent of students) was because they did not feel it was necessary, followed by doing it because it was faster (37 percent), because of not feeling like it (19 percent), and because of guessing the solution instead of calculating it (19 percent). Some students also reported better accuracy with mental strategies (13 percent) and finding it smarter to be able to solve a problem mentally (11 percent). Virtually no students (1 percent) perceived mental calculation as cooler. Indicating not finding writing down calculations necessary as a reason for not doing it was positively related to written strategy choices ($r(322) = .20$), whereas indicating not feeling like writing anything down and considering mental calculation more accurate as reasons were negatively related to written strategy choices ($r(322) = -.12$ and $r(322) = -.23$).

Teacher questionnaire

The Appendix also shows what the teachers of the students ($N = 19$) reported on the teacher questionnaire on their strategy instruction. As for the student questionnaire, the proportion of teachers choosing each alternative is given, and the mean is given for the 5-point scales. Correlations were also calculated, but none of them were significant, possibly due to low power because of the small N.

A small majority of the teachers was male (58 percent) and the teachers were on average 38 years old. Almost half (47 percent) used the textbook 'Wereld in Getallen' and their students solved 54 percent of the problems using written strategies, while the students of teachers using other textbooks ('Pluspunt', 'Alles telt' and 'Rekenrijk') used written strategies on 66 to 69 percent of the problems.

Most teachers taught their students the whole-number-based algorithm exclusively (58 percent) or in combination with the digit-based algorithm (26 percent), and 16 percent taught their students the digit-based algorithm exclusively. On average, teachers did not prefer one algorithm over the other ($M = 3.0$), but did prefer use of an algorithm to non-algorithmic approaches ($M = 2.2$). During their own training, the whole-number-based algorithm (53 percent) or digit-based algorithm (42 percent) was emphasized, and for one teacher both algorithms. On average, teachers found performing calculations well on paper and mentally equally important for their students ($M = 3.0$). They reported instructing their students in writing down calculations frequently (on average almost daily, $M = 4.2$).

Concerning multidigit division problems specifically, teachers on average found writing down calculations somewhat more important for their students than trying to do it mentally ($M = 2.4$) and valued accuracy somewhat over speed ($M = 2.5$). Making a good estimation of the solution was more important than being able to determine the exact solution ($M = 3.5$), as was knowing more solution procedures than just one ($M = 3.4$). Teachers considered using an algorithm versus choosing a custom solution strategy on average equally important ($M = 3.0$), and valued convenient shortcut strategies somewhat more than using a method that can always be applied ($M = 3.3$).

5.3.2 Content of the training

After the pretest, students with a mathematical ability percentile rank between 10 and 50 ($N = 147$) received intervention or control training. During the three training sessions, the students in the intervention condition completed on average

5.1 division problems per session and the students in the control condition 6.1 problems. The number of problems that students attempted a second time (when the solution was incorrect the first time) was 1.6 for the intervention and 1.8 for the control condition. During all the second attempts of a session combined, intervention students received feedback 3.3 times on average. This feedback most often concerned writing down a multiplication table (0.8 times) and selecting a number from that table (1.1 times), and less often the writing down of the problem (0.5 times), subtracting the selected number (0.5 times) and finishing the procedure (0.5 times).

As instructed, the students in the intervention condition virtually always wrote down a calculation (for 98, 99 and 99 percent of the problems in the first, second and third session respectively). Though not instructed to do so, the students in the control condition also often wrote down a calculation and this appeared to increase over sessions, with 81 percent in the first session and 87 and 93 percent in the second and third session. The use of written calculations that were algorithmic (digit-based or whole-number-based) increased over sessions in both groups and appeared higher overall in the intervention condition (84, 93 and 96 percent in the three sessions in the intervention condition and 63, 71 and 76 percent in the control condition).

5.3.3 Effects of the intervention and control training

The effects of the training were evaluated using a series of explanatory IRT models on the pretest and posttest data with successively more predictors (see Table 5.3).

Written strategy choices

First a baseline model for the probability of a written strategy choice was fitted with only random intercepts for students and problems and no covariates (model M_0). In model M_1, main effects were added for the student characteristics gender, ability and working memory capacity, which improved fit according to all criteria (see Table 5.3). Fit was further improved by adding a main effect for testing occasion (pretest or posttest; model M_2). However, the change in written strategy choices from pretest to posttest did not significantly differ for the control and intervention training groups (model M_3). Adding interactions between condition, testing occasion and student characteristics also did not improve the model (these models are not included in Table 5.3 for brevity).

Table 5.3: Explanatory IRT models for training effects on written strategy choices and accuracy (all comparisons are to M_{n-1}).

		intervention vs. control training (**$N = 147$**)				
strat.	added fixed effects	LL	# pars	AIC	BIC	LRT
M_0		-1337.6	3	2681.1	2699.4	
M_1	gender, ability and WM	-1315.7	6	2643.3	2679.8	$\chi^2(3) = 43.8$, $p < .001$
M_2	testing occasion	-1216.5	7	2447.0	2489.5	$\chi^2(1) = 198.3$, $p < .001$
M_3	condition×occasion	-1215.6	9	2449.2	2503.9	$\chi^2(2) = 1.7$, $p = .42$
acc.	added fixed effects	LL	# pars	AIC	BIC	LRT
M_0		-1801.0	3	3607.9	3626.1	
M_1	gender, ability and WM	-1785.3	6	3582.5	3619.0	$\chi^2(3) = 31.4$, $p < .001$
M_2	testing occasion	-1711.1	7	3436.3	3478.8	$\chi^2(1) = 148.3$, $p < .001$
M_3	condition×occasion	-1710.8	9	3439.6	3494.2	$\chi^2(2) = 0.7$, $p = .70$
		training vs. no training (**$N = 224$**)				
strat.	added fixed effects	LL	# pars	AIC	BIC	LRT
M_0		-2107.4	3	4220.8	4240.3	
M_1	gender and ability	-2069.2	5	4148.3	4180.8	$\chi^2(2) = 76.5$, $p < .001$
M_2	testing occasion	-2016.9	6	4045.8	4084.8	$\chi^2(1) = 104.6$, $p < .001$
M_3	condition×occasion	-1962.8	8	3941.7	3993.7	$\chi^2(2) = 108.1$, $p < .001$
M_{4a}	gender×condition×occasion	-1962.2	11	3946.5	4018.0	$\chi^2(3) = 1.2$, $p = .76$
M_{4b}	ability×condition×occasion	-1961.5	11	3945.0	4016.4	$\chi^2(3) = 2.7$, $p = .43$
acc.	added fixed effects	LL	# pars	AIC	BIC	LRT
M_0		-2724.1	3	5454.2	5473.7	
M_1	gender and ability	-2668.5	5	5347.0	5379.4	$\chi^2(2) = 111.2$, $p < .001$
M_2	testing occasion	-2610.9	6	5233.8	5272.8	$\chi^2(1) = 115.2$, $p < .001$
M_3	condition×occasion	-2593.2	8	5202.4	5254.4	$\chi^2(2) = 35.4$, $p < .001$
M_{4a}	gender×condition×occasion	-2592.7	11	5207.4	5278.9	$\chi^2(3) = 1.0$, $p = .81$
M_{4b}	ability×condition×occasion	-2591.6	11	5205.2	5276.7	$\chi^2(3) = 3.2$, $p = .36$

Table 5.4: Strategy use proportions on the pretest and posttest in the intervention, control and no training conditions.

	pretest			posttest		
training	interv.	control	none	interv.	control	none
digit algorithm	.09	.09	.19	.13	.13	.20
number algorithm	.37	.40	.32	.61	.62	.32
non-alg. written	.19	.19	.15	.13	.08	.12
no written work	.35	.30	.34	.13	.17	.37
other	.01	.02	.01	.00	.00	.01

The best fitting model, M_2, shows that girls used more written strategies ($P = .94$) than boys ($P = .74$), $z = -6.0$, $p < .001$, and that general mathematics ability score was positively associated with using written strategies ($P = .80$ vs. $P = .92$ for one standard deviation difference), $z = 4.3$, $p < .001$. Working memory (sum score of the verbal and spatial working memory scores) had no significant effect, $z = -0.6$, $p = .55$. Students used more written strategies at the posttest ($P = .94$) than at the pretest ($P = .76$), $z = 13.5$, $p < .001$.

Table 5.4 gives a more detailed categorization of strategies than just written or non-written, as intervention and control training may differ in the type of written strategies they elicit. It shows that the frequency of use of the digit-based and whole-number-based algorithms, non-algorithmic written strategies, non-written strategies and other strategies is almost identical (differences of no more than 5 percentage points) in the two training groups - both at the pretest and at the posttest. In both groups, similar increases in the use of both algorithms and decreases in the use of non-written strategies and non-algorithmic strategies occurred.

Accuracy

As for written strategy choices, first a baseline model for the probability of a correct response was fitted (M_0), and again, this model was improved by adding student gender, ability and working memory (M_1) and by adding testing occasion ($M2$), but not by adding condition effects (M_3). The best fitting model, M_2, shows that girls ($P = .43$) performed better than boys ($P = .28$), $z = -3.8$, $p < .001$, and that general mathematics ability score was positively associated with performance ($P = .28$ vs. $P = .43$ for one SD difference), $z = 4.5$, $p < .001$. Working memory had no significant effect, $z = 0.04$, $p = .97$. Students performed better at the posttest ($P = .48$) than at the pretest ($P = .24$), $z = 11.9$, $p < .001$.

The difference in accuracy between written and non-written strategies was investigated by fitting a model for accuracy with main effects for all previous predictors (student characteristics, testing occasion, and condition) and strategy choice (written or not), and all first-order interactions between strategy choice and the other predictors. This showed that written strategies were much more accurate ($P = .40$) than non-written strategies ($P = .19$), $z = 4.1$, $p < .001$, and that this did not depend significantly on testing occasion, $z = 1.1$, $p = .27$, gender, $z = 0.0$, $p = .99$, ability, $z = 1.0$, $p = .32$, working memory, $z = 0.3$, $p = .75$, or condition, $z = -1.0$, $p = .33$.

5.3.4 Differences with no training group

Given the similar changes in strategy choices and accuracy in both training groups, it was investigated whether these changes also occurred in students who did not receive any training. The previous analyses were repeated, this time comparing trained students ($N = 147$) to untrained students from adjoining ability groups ($N = 77$). Working memory was omitted from these models, as this was only assessed for the children who received training.

Written strategy choices

This time, the fit of the models for written strategy choices was best for model M_3 (which also included an effect of condition; see Table 5.3). The effect of the intervention did not differ significantly by gender or ability level (models M_{4a} and M_{4b}). Model M_3 once more showed more written strategy choices for girls ($P = .90$) than boys ($P = .63$), $z = -6.9$, $p < .001$, and a positive association with ability ($P = .72$ vs. $P = .86$ for a difference of one SD), $z = 6.9$, $p < .001$. There was no significant effect of testing occasion, $z = -1.4$, $p = .15$, and no overall difference between the trained and untrained students, $z = 0.5$, $p = .64$. However, the change in use of written strategies from pretest to posttest was different for trained ($P = .75$ to $P = .93$) than for untrained students ($P = .73$ to $P = .69$), $z = 9.8$, $p < .001$.

Comparisons of more specific strategies in Table 5.4 show that at pretest, the untrained students appear to have used the digit-based algorithm somewhat more often and the whole-number-based algorithm somewhat less often than the trained students. Most notably, however, strategy choices on the pretest and posttest are almost identical for the untrained children, whereas the trained children increased their use of algorithms and decreased their use of non-written strategies and non-algorithmic strategies.

Accuracy

The fit of the models for accuracy was also best for model M_3 with the condition effect (see Table 5.3). This model again showed higher accuracy for girls ($P = .41$) than boys ($P = .28$), $z = -4.3$, $p < .001$, and a positive association with ability ($P = .26$ vs. $P = .44$ for one SD difference), $z = 10.1$, $p < .001$. There was no significant effect of testing occasion, $z = -1.4$, $p = .15$, and no overall difference between the trained and untrained students, $z = -1.8$, $p = .07$. However, the increase in accuracy from pretest to posttest was higher for trained ($P = .25$ to $P = .49$) than for untrained students ($P = .31$ to $P = .35$), $z = 5.9$, $p < .001$.

Written strategies were again found to be much more accurate ($P = .41$) than non-written strategies ($P = .21$), $z = 3.0$, $p = .002$, and this did not depend significantly on testing occasion, $z = 1.6$, $p = .12$, gender, $z = 0.2$, $p = .88$, ability, $z = 0.8$, $p = .44$, or condition, $z = 1.1$, $p = .28$.

5.4 Discussion

The determinants of students' choices between mental and written division strategies were investigated. First, an exploration was carried out of the relation between existing differences in these choices and students' motivations and attitudes in mathematics and the sociocultural context for mathematics provided by the students' teachers. For an important part, students' choices for mental strategies appear to be related to their motivation: mental strategies are used more by students who report liking mathematics less and being less good at it, and who report not writing down calculations because they do not feel like it. Mental strategies are also used more by students reporting higher accuracy with these strategies. Though this higher accuracy could be true for high ability students (Fagginger Auer et al., 2016), it mostly appears to be a misjudgment as the reporting of it is negatively correlated with ability level, $r(322) = -.24$, $p < .001$.

No statistically significant relations between teacher reports and students' strategy choices were found, even though several aspects of the sociocultural context described as influential on mathematical strategies by Ellis (1997) were investigated, but this could very well be due to a lack of power (there were only 19 teachers in our sample). Overall, teachers reported frequent instruction in writing down calculations, preferred use of an algorithm to non-algorithmic approaches, and valued written strategies somewhat over mental strategies and accuracy somewhat over speed. These reports suggest a sociocultural context in which there is room for written strategies, but where it is not the highest priority.

In the second part of the study, an intervention training designed to promote lower mathematical ability students' choices for written rather than mental strategies (and thereby, their performance) was evaluated. As intended, written strategy choices and accuracy were considerably higher after training than before training. However, similar changes occurred in the control training condition. This means that the extra elements of the intervention training specifically targeted at strategy use did not add to the effect of the training. The common elements of the control and intervention training do appear to be responsible for the observed changes in strategy choices and accuracy, as no such changes occurred in the students who received no training (though these students were of a different ability level, limiting the comparison). An important question is therefore which of the training elements not specifically targeted at strategy use nonetheless affected it.

5.4.1 Elements of the intervention and control training

Practicing written strategies

While writing down calculations was not required during control training (it was a specific part of the intervention training), it did occur frequently in this condition. During the first control training session, calculations were written down for 81 percent of the problems - considerably more than the 70 percent during the pretest. This increased up to 93 percent in the third training session. As such, students practiced written calculations almost as much in the control training as in the intervention training condition, reducing the contrast between the two conditions.

The generally higher level of written strategy choices in the control training compared to the pretest may be due to the different settings in which the pretest and training occurred: in a classroom versus one-on-one with an experimenter. An individual setting is likely to increase students' motivation to do well, and since the student questionnaire suggested that an important reason for using mental strategies is a lack of motivation, this increased motivation may cause the students to use less mental strategies. Another possibility is that students use written strategies because they think the experimenter may expect or prefer that (i.e., demand characteristics; Orne, 1962), in line with the students' teachers' light inclination towards written rather than mental strategies. Supporting the explanation of the higher level of written strategy choices by setting (individual versus classroom), the increase in written strategy choices from pretest to first training session was followed by a decrease from final training session to posttest (93 to 87 percent).

A possible cause of the further increase in the use of written strategies over sessions in the control training group is the direct accuracy feedback after each solution, and the requirement to do a problem again when the first solution was incorrect. Direct accuracy feedback allows for an immediate evaluation of the success of the strategy that was applied, and this evaluation should often be in favor of written rather than mental strategies given the considerably higher accuracy of the former. Combined with the extra effort associated with an incorrect solution (redoing the problem), this is likely to be an important incentive for written strategy choices. The possibility of accuracy feedback promoting mathematical strategy change was also demonstrated by Ellis, Klahr, and Siegler (1993).

Step-by-step plan

The only training element that was truly unique to the intervention condition was the step-by-step plan for writing down calculations. Though the meta-analysis on mathematics interventions for low ability students by Gersten et al. (2009) identified such plans as an important component of effective interventions, the lack of differences between the training conditions shows that the plan did not make a significant contribution in our study. Indeed, students turned out to require little feedback based on the plan, and the feedback that was given mostly concerned an optional element of written division algorithms (the multiplication table). This suggests that by sixth grade, even lower ability students do not require further instruction in the notation of the division algorithm (even though the algorithm was introduced only one or two years earlier).

Given that the only real difference between the control and intervention training turned out to be mostly redundant, there was no chance for student characteristics to interact with type of training in the effect on changes from pretest to posttest. Our hypotheses regarding the effects of gender, ability and working memory were therefore not confirmed. An interaction with having training or not could have been detected if present given the differences found between these two conditions, but was also not found. Working memory was not included in these analyses, as it was only measured in the children who received training, and ability scores were different in the training and no training conditions. Gender, however, could very well have interacted with condition: as expected from the literature (Fagginger Auer et al., 2013; Hickendorff et al., 2009, 2010), boys used written strategies far less frequently than girls, and therefore had more room to improve with training than girls. However, training may not eliminate boys' general preference for more intuitive, less formal strategies (Carr & Jessup, 1997; Davis & Carr, 2002), which may therefore continue to limit their choices for (formal) written strategies to some extent.

5.4.2 Future directions

The results of the present study provide several suggestions for future research on strategy training programs. Firstly, they underline the necessity of very careful consideration of the content of the control condition(s). With regard to control groups, U. Fischer, Moeller, Cress, and Nuerk (2013) stress the importance of these groups being performance-matched to the intervention group, as learning trajectories are highly dependent on ability level, and equal in motivational appeal and training time, as these two non-specific factors also contribute to performance. The untrained group in the present study does not meet these demands, which may have inflated the effects we found (U. Fischer et al., 2013), but the control training group certainly does. In fact, the control training even matched the intervention training too closely, which shows that attention should also be devoted to which control training elements may be (unintentionally) effective.

Some of the elements of the present study are promising for future training investigations. The results suggest that direct accuracy feedback (possibly with some cost involved in incorrect solutions) may be conducive to beneficial changes in strategy choices. They also show that considerable changes in strategy choices and improvements in performance may be achieved with as few as three training sessions of fifteen minutes (in line with the finding of Kroesbergen & Van Luit, 2003, that longer mathematics interventions are not necessarily more effective). A follow-up test after a longer period of time (e.g., several months) should be used to establish whether the changes are lasting.

The results also provide two suggestions for other possible ways to influence students' choices between mental and written strategies. A first possibility is to target students' motivation: since strategy choices appear to be related to motivation, increasing students' motivation may also increase their choices for written strategies. In a review, Middleton and Spanias (1999) concluded that students' motivation in mathematics depends for an important part on their perception of success in this area, but also that it can be positively affected by instruction. This may be achieved with teacher practices such as asking students to make daily recordings of what they learned or excelled at, and prompting them to attribute failures to lack of effort and encouraging them to try harder (Siegle & McCoach, 2007). However, the relation found in the present study was purely correlational, so it should be established experimentally whether changes in motivation actually lead to changes in strategy choices.

A second possibility for increasing students' choices for written strategies lies in the sociocultural context for mathematical strategy use provided by the teacher. The results from the teacher questionnaire show that while teachers generally give instruction on writing down calculations frequently, they only have a slight preference for written over mental strategies and for accuracy over speed. Since cultural values regarding the use of external aids (e.g., paper and pencil) in constructing solutions and regarding accuracy versus speed can have large effects on students' strategy choices (Ellis, 1997), targeting these aspects of the sociocultural context could affect written strategy choices beneficially. This might be done by having teachers express more appreciation of the use of external aids in problem solving, and of accuracy compared to speed, since written strategies offer more accuracy and mental strategies more speed (Fagginger Auer et al., 2016).

5.A Student questionnaire

The proportion of students choosing each alternative is given in between brackets, and for five-point scales, the mean is also given. The correlations are between the question response and the frequency of written strategy choices on the pretest.

1. How much do you like math? ($M = 3.24$) ($r(322) = .17$, $p = .002$)
   not at all (.06) / not so much (.13) / it's okay (.40) / quite a bit (.32) / a lot (.08)
2. How much effort do you put into doing math? ($M = 4.29$) ($r(323) = .08$, $p = .17$)
   none (.00) / not so much (.02) / a bit (.06) / quite a lot (.54) / a lot (.39)
3. How good do you think you are at math? ($M = 3.27$) ($r(322) = .21$, $p < .001$)
   not good at all (.04) / not so good (.17) / okay (.31) / quite good (.44) / very good (.04)
4. What is more important to you when you solve a mathematics problem?
   ($r(320) = .06$, $p = .28$)
   solving the problem quickly (.02) / finding the correct solution (.98)

5. What is more important to you when you solve a mathematics problem? ($r(318) = .19$, $p = .001$)
   being able to do it mentally (.28) / being able do it using paper (.72)
6. How often do you solve problems without writing down a calculation? ($M = 2.80$) ($r(322) = -.17$, $p = .002$)
   almost never (.11) / not often (.24) / sometimes (.43) / often (.19) / very often (.03)
7. When you do not write down a calculation, why is that? *(tick boxes that apply)*
   - because it is faster (.37) ($r(322) = -.04$, $p = .52$)
   - because then you get a correct solution more often (.13) ($r(322) = -.23$, $p < .001$)
   - because doing mental calculation shows you are smart (.11) ($r(322) = -.02$, $p = .71$)
   - because it is cooler to do mental calculation (.01) ($r(322) = -.18$, $p = .001$)
   - because you do not feel like writing anything down (.19) ($r(322) = -.12$, $p = .03$)
   - because you guessed the solution (.19) ($r(322) = -.05$, $p = .37$)
   - because it is not necessary to write down a calculation (.60) ($r(322) = .20$, $p < .001$)

5.B Teacher questionnaire

The proportion of teachers choosing each alternative is given in between brackets, and for five-point scales, the mean is also given. The correlations are between the question response and the frequency of the teachers' students' written strategy choices on the pretest.

1. What is your gender? *male (.58) / female (.42)* ($r(19) = .03$, $p = .91$)
2. What is your birth year? ... ($M = 1976$) ($r(19) = -.23$, $p = .35$)
3. Which mathematics textbook do you use in sixth grade? *Alles Telt (.21)* ($M = .66$)/ *Wereld in Getallen (.47)* ($M = .54$) / *Pluspunt (.26)* ($M = .69$) / *Rekenrijk (.05)* ($M = .69$)
4. Do you teach your students the whole-number-based algorithm, digit-based algorithm or non-algorithmic approaches for solving multidigit problems (such as $544 \div 34$ or $12.6 \div 1.4$)? When multiple approaches apply, tick multiple boxes.
   whole-number-based algorithm (.58) / both whole-number-based and digit-based algorithm (.26) / digit-based algorithm (.16) ($r(19) = .07$, $p = .77$)

5. To what extent do you as a teacher prefer a division algorithm?
strong preference whole-number-based - strong preference digit-based (5-point scale) *(* $M = 3.0$ *)* *(* $r(19) = .28$, $p = .24$ *)*

6. To what extent do you as a teacher prefer an algorithmic over a non-algorithmic approach?
strong preference algorithmic - strong preference non-algorithmic (5-point scale) *(* $M = 2.2$ *)* *(* $r(19) = -.15$, $p = .55$ *)*

7. Which division approach was emphasized most during your own training?
whole-number-based algorithm (.53) / both whole-number-based and digit-based algorithm (.05) / digit-based algorithm (.42) *(* $r(19) = .25$, $p = .29$ *)*

8. Which ability do you find more important in general for your students?
performing calculations well on paper - performing calculations well mentally (5-point scale) *(* $M = 3.0$ *)* *(* $r(19) = .02$, $p = .92$ *)*

9. How often do you instruct your students in writing down intermediate steps or calculations? *almost never - daily (5-point-scale)* *(* $M = 4.2$ *)* *(* $r(19) = .07$, $p = .77$ *)*

10. What is more important to you when your students solve multidigit division problems? *(six 5-point scales)*

    - *that they write down all calculations - that they try to do it mentally* *(* $M = 2.4$ *)* *(* $r(19) = .06$, $p = .82$ *)*
    - *that they keep trying until they get the correct solution, even if that takes a lot of time - that they can do it quickly, even if they sometimes make mistake* *(* $M = 2.5$ *)* *(* $r(19) = -.08$, $p = .78$ *)*
    - *that they can determine the exact answer - that they can make a good estimation of the answer* *(* $M = 3.5$ *)* *(* $r(19) = .35$, $p = .15$ *)*
    - *that they know one solution procedure - that they know multiple solution procedures* *(* $M = 3.4$ *)* *(* $r(19) = .35$, $p = .15$ *)*
    - *that they use an algorithm - that they choose their own solution strategy* *(* $M = 3.0$ *)* *(* $r(19) = .24$, $p = .33$ *)*
    - *that use a method that can always be applied - that they use convenient shortcut strategies (such as* $1089 \div 11 = 1100 \div 11 - 1$ *)* *(* $M = 3.3$ *)* *(* $r(19) = .19$, $p = .44$ *)*