Universiteit Leiden
The Netherlands

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

Hickendorff, M.

Citation

Hickendorff, M. (2011, October 25). *Explanatory latent variable modeling of mathematical ability in primary school : crossing the border between psychometrics and psychology*. Retrieved from https://hdl.handle.net/1887/17979

Version:	Not Applicable (or Unknown)
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/17979

Note: To cite this publication please use the final published version (if applicable).

CHAPTER 3

Complex multiplication and division in Dutch educational assessments: What can solution strategies tell us?

This chapter has been submitted for publication as Hickendorff, M. & Van Putten, C. M. *Complex multiplication and division in Dutch educational assessments: What can solution strategies tell us?*

The research was supported by CITO, National Institute for Educational Measurement. We would like to thank all Psychology undergraduate students who participated in the coding of strategy use.

Abstract

The aim of the current study was to get more insight in sixth graders' performance level in multidigit multiplication and division that was found to be decreasing over time and lagging behind educational standards in large-scale national assessments in the Netherlands, where primary school mathematics education is characterized by reform-based learning/teaching trajectories. In secondary analyses of these assessment data, we extended the focus from achievement to aspects of strategic competence, by taking solution strategies that students used into account. In the first part of this paper, the negative performance trend between the 1997 and 2004 assessment cycles in multiplication problem solving was examined, by analyzing changes in strategy choice, overall differences in accuracy between strategies, and changes in these strategy accuracies. Findings showed that two changes contributed to the performance decline: a shift in students' typical strategy choice from a more accurate strategy (the traditional algorithm) to less accurate ones (non-traditional partitioning strategies and answering without written work, the increase in the latter strategy mainly observed in boys), as well as a general decline of accuracy rate within each strategy. In the second part, the influence of instruction on students' strategy choice in multiplication and division problems was analyzed. Findings showed that the teacher's instructional approach affected students' strategy choice, most profoundly in division problem solving.

3.1 Introduction

National and international large-scale educational assessments aim to report on the outcomes of the educational system in various content domains such as reading, writing, science, and mathematics. To evaluate the learning outcomes, a reference framework is needed. This can be either a comparison between countries as is done in the international comparative assessments (e.g., TIMSS, PIRLS, PISA), a comparison to the educational standards or attainment targets that are set within a country, or a comparison to performance levels from previous assessment cycles to find a trend over time.

The reports of educational assessments are usually limited to descriptive and correlational data on students' achievement, and therefore explanations for found differences or trends require further study. In such further studies, insights from educational psychology are essential to give direction to the exploration of potential explanatory mechanisms. In the current study, the focus is on one candidate mechanism: solution strategy use. The main research question is to what extent (change in) students' strategy choice explains (change in) their achievement, and in turn, to what extent instructional approach influences students' strategy choice, in the domain of complex or multidigit multiplication and division. We tried to answer this question by carrying out secondary analyses on data of the two most recent Dutch national assessments at the end of primary school (1997 and 2004 cycles), extending the focus on achievement to aspects of strategic competence (e.g. Lemaire & Siegler, 1995) by studying solution strategies that students used. The aim of the current study was to get more insight in the performance level of Dutch sixth graders in complex multiplication and division, that was found to be decreasing over time and lagging behind educational standards.

3.1.1 Dutch results of educational assessments of mathematics achievement

Recent national and international assessments showed a varying pattern of results regarding mathematics performance in primary schools in the Netherlands. On the positive side, national results of the most recent cycle of PPON (Dutch assessment of mathematics education at the end of primary school, i.e., 12-year-olds) in 2004 showed improvements over time on some mathematics competencies, in particular on numerical estimation and on number sense (J. Janssen et al., 2005; Van der Schoot, 2008; see also Figure 3.1). Moreover, TIMSS-2007 (Meelissen & Drent, 2008; Mullis et al., 2008) results showed that Dutch fourth graders performed at the top level internationally, and also in PISA-2009 (OECD, 2010) Dutch 15-year-olds' mathematics performance took in an international top position. On the downside, however, there are also some results that are less positive. Both TIMSS and PISA reported a negative ability trend over time in the Netherlands. In addition, national assessments showed that on some mathematics domains performance decreased substantially since the first assessment in 1987 (see Figure 3.1). Furthermore, in many mathematics domains the educational standards were not reached (Van der Schoot, 2008).

Particularly, performance in *complex operations* – i.e., addition, subtraction, multiplication, division, and combined operations with multidigit numbers on which paper and pencil may be used – is worrisome. Not only did performance decrease most severely on these domains, with an accelerating trend (Figure 3.1), but also the percentage of students who reached the educational standards was lowest. A group of experts operationalized the educational standards and defined a 'sufficient' level of performance per domain that had to be reached by 70-75% of all students. In PPON 2004, this level was reached by

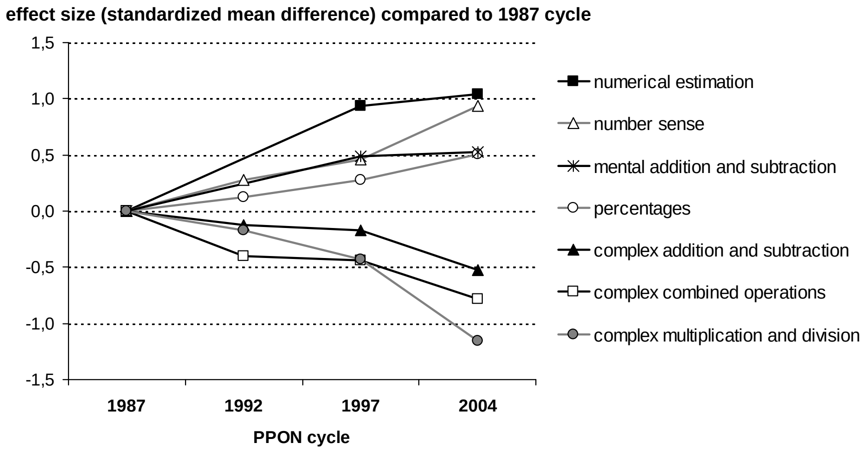

Figure 3.1 *Largest trends over time from Dutch national assessments (PPONs) of mathematics education at the end of primary school (Van der Schoot, 2008, p. 22), in effect sizes (standardized mean difference) with 1987 as baseline level. Effects statistically corrected for students' gender, number of school years, and socio-economical background, socio-economical composition of school, and mathematics textbook used.*

27% of the students in addition and subtraction, by 12% in multiplication and division, and by 16% in problems involving combining operations. On these domains, learning outcomes thus lagged far behind the goals.

The aim of the current study was therefore to gain more insight in students' lagging and decreasing performance level in the domain of complex multiplication and division. Our main approach was to extend the focus on achievement by including aspects of *strategic competence* (Lemaire & Siegler, 1995). We focused on complex multiplication and division for several reasons. First, as discussed above, performance decreased most severely on these operations, and it stayed furthest away from the educational standards. Second, compared to addition and subtraction, multiplication and division have received far less research attention, and especially multidigit multiplication and division are understudied research topics. Finally, instruction in how to solve multidigit operations has changed under influence of mathematics education reform, in particular on complex division, where the traditional algorithm for long division has completely disappeared from mathematics textbooks and the learning/teaching trajectory (Van den Heuvel-Panhuizen, 2008).

3.1.2 Solution strategies

It has been well-established that children know and use multiple strategies in mathematics, and these strategies have different characteristics such as accuracy and speed (e.g., Beishuizen, 1993; Blöte et al., 2001; Lemaire & Siegler, 1995; Torbeyns, Verschaffel, & Ghesquière, 2004b, 2006; Van Putten et al., 2005). Therefore, solution strategy use may be an important predictor of achievement, and thereby also a potential mediator between (change in) instruction and (change in) achievement.

Mathematics education and instruction in primary school have undergone a large reform of international scope (e.g., Kilpatrick et al., 2001). In the Netherlands, the reform movement goes by the name of realistic mathematics education (RME), and it has become the dominant didactical theory in mathematics education practice. In the 1997 assessment, over 90% of the schools used a mathematics textbook that was based on the RME principles (J. Janssen et al., 1999); in the 2004 assessment this increased to nearly 100% (J. Janssen et al., 2005).

Solution strategies play an important role in this reform in at least two ways. First, the learning/teaching trajectory for solving complex arithmetic problems has changed, from top-down instruction of standard written algorithms to building on children's informal or naive strategies that are progressively formalized (Freudenthal, 1973; Treffers, 1987, 1993; Van den Heuvel-Panhuizen, 2008), a process in which mental arithmetic has become very important (Blöte et al., 2001). Second, the reform aims at attaining *adaptive expertise* instead of *routine expertise*: instruction should foster the ability to solve mathematics problems efficiently, creatively, and flexibly, with a diversity of strategies (Baroody & Dowker, 2003; Torbeyns, De Smedt, Ghesquière, & Verschaffel, 2009b). The question is to what extent the instructional changes in complex arithmetic affected strategy use, and consequently, achievement.

Hickendorff, Heiser, Van Putten, and Verhelst (2009b) investigated the role of solution strategies in explaining the performance decrease in complex division problems observed in the Dutch national assessments. They carried out secondary analyses on the assessment material of 1997 and 2004 by coding the solution strategies that students used to solve the division problems (based on their written work). Findings showed shifts between the two assessment cycles in strategy choice as well as in strategy accuracy, both contributing to the explanation of the performance decrease. The use of the accurate traditional long division algorithm decreased at the cost of an increase in problems that were answered without any written work (most likely mental calculation), a strategy that was much less accurate. Moreover, each of the main strategies led to fewer correct answers (i.e., was less accurate) in 2004 than it was in 1997.

In a follow-up study, Hickendorff, Van Putten, Verhelst, and Heiser (2010) analyzed the most relevant strategy split – mental versus written computation – more rigorously by collecting new data according to a partial *choice/no-choice* design (Siegler & Lemaire, 1997). Findings showed that for students who spontaneously chose a mental computation strategy to solve a complex division problem, the probability of a correct answer increased on average by 16 percent points on a parallel problem on which they were forced to write down their working. This suggested that the choice for a mental strategy on these problems was not optimal or adaptive with respect to accuracy, contrasting with the prediction in cognitive models of strategy choice that individuals choose their solution strategy adaptively (e.g., Shrager & Siegler, 1998; Siegler & Shipley, 1995). Moreover, the findings had clear implications for educational practice: encouraging students to write down their solution steps in solving complex division problems would probably improve performance.

These two studies illustrate the mutual value of bringing together the field of large-scale educational assessments and the field of educational and cognitive psychology. In the current study, Hickendorff et al.'s (2009b) analyses of strategy use on the complex division problems in the Dutch assessments are extended in two important ways. First, the domain of study is broadened to complex multiplication. Second, information on the instructional approach the teachers applied (that was, unfortunately, only available in the 2004 assessment) was used as a predictor of strategy use. Below, we discuss these two topics in more detail.

3.1.3 Complex or multidigit multiplication strategies and instruction

The majority of studies that focus on multiplication strategies in children and adults considered simple multiplication under 100, i.e., multiplying two single-digit numbers (Anghileri, 1989; Imbo & Vandierendonck, 2007; Lemaire & Siegler, 1995; Mabbott & Bisanz, 2003; Mulligan & Mitchelmore, 1997; Sherin & Fuson, 2005; Siegler, 1988b). The following solution strategies were identified for solving simple multiplication problems like 3×4: counting procedures (unitary counting, $1,2,3,4,\ldots,5,6,7,8,\ldots,9,10,11,12$, as well as using a counting string, $4,8,12$), repeated addition (adding an operand the appropriate number of times, $4+4+4=12$), transforming the problem (referring to related operations or related facts, $2\times4=8$, $8+4=12$), and retrieval (knowing the answer by heart). With increasing age and experience, retrieval becomes the dominant strategy for simple multiplication.

In contrast, in multidigit or complex multiplication problems retrieval is not a feasible strategy, and computational strategies are required to derive the answer. Ambrose et al. (2003) analyzed the development of multidigit multiplication strategies and described three classes of strategies: concrete modeling strategies (which the authors note to be of limited use when two multidigit numbers have to be multiplied), adding and doubling strategies (including repeated addition), and partitioning strategies using tenfolds of one or both of the operands (see also Figure 3.2). Note that combinations of these classes of strategies are also possible (as was also described by Sherin and Fuson (2005), who called this hybrid strategies). For example, in Figure 3.2, the strategy in which one of the operands is decimally split also includes the additive strategy of doubling.

The RME learning-teaching trajectory in multidigit multiplication has its roots in the aforementioned developmental pattern, and can be characterized by progressive schematization and abbreviation of the informal solution strategies (Treffers, 1987; Van den Heuvel-Panhuizen, 2008). Buijs (2008) analyzed the recent RME-based textbooks, and found a common learning trajectory that starts from the repeated addition strategy, that is abbreviated by grouping, eventually with groups of ten times one of the operands. This leads to splitting or partitioning strategies in which one of the operands is decimally split. Partitioning strategies in which the solution steps are written down systematically in a more or less fixed order (which Buijs labeled 'stylized mental strategies', also called 'column multiplication' in the RME literature Van den Heuvel-Panhuizen, 2008) are suitable as transition phase toward the standard written algorithm for multiplication: it works with whole numbers instead of single-digits (like informal strategies), but it proceeds in a more or less standard way (like the traditional algorithm).

For multiplication, the end point of the RME-based learning trajectory still is the traditional algorithm in which calculation proceeds by multiplying single digits in a fixed order, from small to large (see Figure 3.2), although it does not have to be attained by all students; 'column multiplication' is considered a full alternative. In contrast, in the RME-

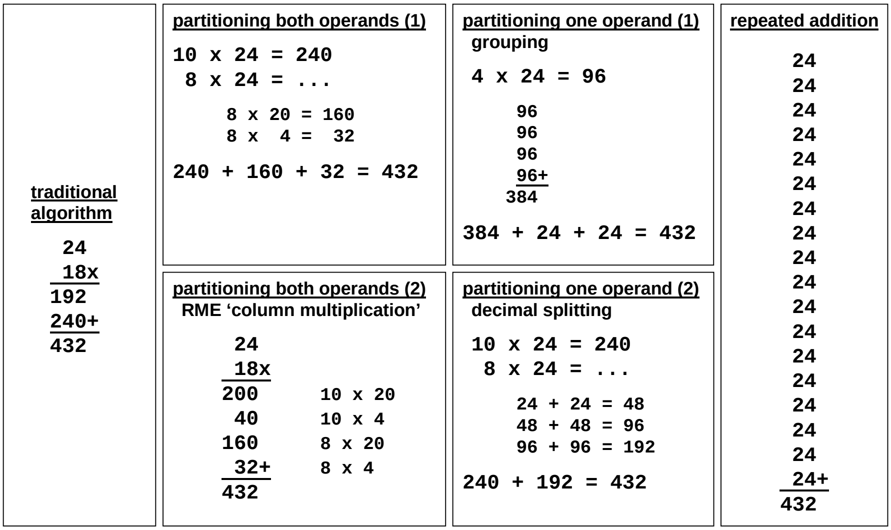

FIGURE 3.2 *Example strategies for multidigit multiplication for the problem 18 $\times$ 24.*

based learning trajectory for multidigit division, the traditional long division algorithm has completely disappeared (Van den Heuvel-Panhuizen, 2008). These instructional differences call into question to what extent they affect students' strategy choices in these operations. Therefore, the influence of teacher's instructional approach on students' strategy choice in complex multiplication and complex division problem solving is compared. The results may yield insights into the extent that teachers can influence students' strategic behavior, and by that, their achievement too. Furthermore, in contrast to the relation between simple multiplication and division (e.g., Campbell, Fuchs-Lacelle, & Phenix, 2006; De Brauwer & Fias, 2009; Mauro, LeFevre, & Morris, 2003), the relation between complex multiplication and division problem solving has not been studied before to our knowledge, so the current study extends the existing research body by studying these operations simultaneously.

3.1.4 Differences between students

Student level variables have been found to influence strategy choice and performance in mathematics. We focus in particular on the student characteristics gender, mathematics achievement level, and socio-economical background. Arguably, other variables such as students' motivation and attitudes (Vermeer, Boekaerts, & Seegers, 2000) and other home background and resources variables (Mullis et al., 2008; Vermeer et al., 2000) are found to be important determinants of mathematics achievement as well, but unfortunately, we have no information on that in the data.

Regarding mathematics achievement level, it has been frequently (but not uniformly, see Torbeyns, Verschaffel, & Ghesquière, 2005) reported that students of higher mathematical ability choose more adaptively or flexibly between strategies than students of low mathematical ability (Foxman & Beishuizen, 2003; Hickendorff et al., 2010; Torbeyns, De Smedt, et al., 2009b; Torbeyns, Verschaffel, & Ghesquière, 2002, 2004a; Torbeyns et al., 2006). In complex division, Hickendorff et al. (2009b, 2010) reported that sixth graders with a higher mathematics achievement level more often used written strategies (the traditional long division algorithm as well as repeated addition/subtraction strategies, see also Figure 1 in Hickendorff et al., 2010) than students with a lower mathematics level. Moreover, differences in accuracy between the strategies decreased with higher mathematics level. In other words, for high achievers it made less difference regarding accuracy which strategy they chose than it did for low achievers.

Gender differences in mathematics performance have often been reported. Large-scale international assessments TIMSS-2007 (Mullis et al., 2008) and PISA-2009 (OECD, 2010) showed that boys tend to outperform girls in most of the participating countries, including the Netherlands. This pattern is supported by Dutch national assessments findings: on most mathematical domains boys outperformed girls in third grade (Kraemer et al., 2005) and in sixth grade (J. Janssen et al., 2005). Furthermore, boys and girls have been found to differ in the strategy choices they make on mathematics problems: girls have a higher tendency to (quite consistently) rely on rules and procedures, whereas boys are more inclined to use more intuitive strategies (Carr & Davis, 2001; Carr & Jessup, 1997; Gallagher et al., 2000; Hickendorff et al., 2010; Timmermans et al., 2007). Furthermore, Hickendorff et al. (2010) found that the shift in strategy use towards mental computation in solving complex division problems was mainly attributable to boys.

Finally, students socio-economical background has an effect on mathematics performance. TIMSS-2007 reported effects of parents' highest level of education (positively related to mathematics performance), the language spoken at home (lower performance if different than the test language) and whether parents were born in a different country (lower performance) (Mullis et al., 2008). Results from the Dutch national assessments on the effects of parents' origin and education were similar (J. Janssen et al., 2005). Moreover, in complex division, students with lower socio-economical background more often answered without written work and less often with one of the two written strategies (traditional algorithm and non-traditional strategies) (Hickendorff et al., 2009b).

3.1.5 The current study

The central aim of the current study was to get more insight in Dutch sixth graders' performance level in complex multiplication and division that was found to be decreasing over time and lagging behind educational standards, by using insights from educational psychology. In secondary analyses of national assessment data, we studied the role of (change in) solution strategy use in explaining (change in) achievement, and in turn, the influence of instructional approach on students' strategy choice. Because information on the instructional approach was only available in the 2004 cycle and not in the 1997 cycle, we set up this study in two separate parts. In the first part, we focused on the effect of solution strategy use on achievement in complex multiplication, thereby extending previous work of Hickendorff et al. (2009b) in complex division. Specifically, we aimed to get more insight in the performance decrease between 1997 and 2004 in multiplication, by analyzing changes in students' typical strategy choice, overall differences in accuracy between strategies, and changes in these strategy accuracies. Moreover, we also addressed the effects of the student characteristics gender, mathematics achievement level, and socio-economical background. Findings may yield educational implications and recommendations on how to turn the negative trend around.

In the second part, we focused on the – possibly different – influence of teacher's strategy instruction on students' strategy choice in multiplication and division. To that end, solution strategy data on multiplication and division problems from the 2004 assessment data were combined. Because instruction in how to solve multiplication problems (end point is traditional algorithm) differs from instruction in how to solve division problems (traditional algorithm disappeared from the Dutch mathematics textbooks), the question is to what extent that reflects in students' strategy choice, potentially yielding implications for educational practice on the influence of the teacher's instruction on students behavior (strategy choice and performance).

3.2 Part I: Changes in strategy choice and strategy accuracy in multiplication

3.2.1 Method

Sample

In the present study, parts of the material of the two most recent national assessments carried out by CITO (Dutch National Institute for Educational Measurement) are analyzed in depth. These studies were carried out in May/June 1997 (J. Janssen et al., 1999) and in May/June 2004 (J. Janssen et al., 2005). For each assessment cycle, schools were sampled from the national population of primary schools, stratified with respect to three socio-economical status categories. In 1997, 253 primary schools with in total 5314 sixth graders (12-year-olds) participated. In the 2004 sample, there were 122 primary schools with in total 3078 students. Schools used various mathematics textbooks, although the large majority (over 90% of the schools in 1997, and almost 100% of the schools in 2004) used textbooks based on RME principles.

Subsets of the total samples of 1997 and 2004 were used in the present analysis: we included only students to whom items on complex multiplication were administered. In 1997, that subset consisted of 551 students with mean age 12 years 4 months (SD = 5 months; range = 11;2 - 14;0) from 218 different primary schools. It consisted of 995 students with mean age 12 years 4 months (SD = 4 months, range = 11;1 - 14;0) from 123 schools in 2004. So, the analyses in part I of this study are based on observations of 1,546 students in total.

In the 1997 sample, there were 45.9% boys and 49.9% girls (remaining 4.2% missing data); in the 2004 sample there were 49.6% boys and 48.8% girls (1.5% missing data). Information on the socio-economical background of the students was available too, based on the background and education of the parents: students with foreign parents with low level of education/occupation (SES-2) and all other students (SES-1)[1]. In 1997, the distribution of students was 87.0% SES-1 and 9.1% SES-2 (4.0% missing data). In 2004, these percentages were 84.0% SES-1 and 14.5% SES-2 (1.5% missing data).

Material and Procedure

In the two assessment cycles together, there were 16 different complex multiplication problems administered, of which five problems were administered in both 1997 and 2004. These five problems were the anchor items, serving as a common basis for comparisons over time. Table 3.1 shows several characteristics of the multiplication problems: the actual multiplicative operation required, whether or not the problem was presented in a realistic context, and the proportion correct in 1997 and 2004 (if observed). On the five common problems (items 7-11), the proportion correct was lower in 2004 than in was in 1997 with differences ranging from .05 (item 10) to .16 (item 11), illustrating the achievement decrease between the two consecutive assessments.

The design of the assessment tests was different in 1997 than it was in 2004. In the 1997 assessment, there were in total 24 different mathematics content domains, and for each domain a subtest was assembled. Students were administered three to four of these subtests. One content domain was *complex multiplication and division*, and its subtest contained 12 problems on multiplication (of which one was eliminated from the scale in the test calibration phase) and 12 on division (also one item was eliminated). Therefore, from the 1997 cycle there were responses of 551 students to 11 different multiplication problems, see also Figure 3.3. In the 2004 cycle, each subtest contained items from different content domains instead of from only one domain as in 1997. Specifically, items were systematically distributed over test booklets in an incomplete test design. In total, there were 18 different test booklets, of which 8 booklets contained items on complex multiplication (and division). There were 10 different multiplication problems used in 2004. Figure 3.3 shows the distribution of these problems (7-16) over the test booklets.

1 In the Dutch educational system, funding of schools is based on an index of parental background and education of the students. There are three major categories: at least one foreign (non Dutch) parent with a low level of education and/or occupation, Dutch parents with a low level of education and/or occupation, and all other students. The definition of the second category has become more stringent between the 1997 and 2004 cycles: in 1997, students were in this category if only one of the parents had a low level of education/occupation, while in 2004 both parents had to have a low level of education/occupation (J. Janssen et al., 2005). As a consequence, the first two categories are incomparable between the two cycles. Therefore, these two categories were combined in the current study, in the category SES-1 (as was also done by J. Janssen et al., 2005).

TABLE 3.1 *Specifications of the multiplication problems*.*

			% correct	
item	problem	context	1997	2004
1	25 × 22	yes	.86	-
2	704 × 25	yes	.62	-
3	178 × 12	yes	.73	-
4	1.800 × 1.75	yes	.31	-
5	86 × 60	no	.77	-
6	109 × 87	no	.70	-
7	*24 × 57*	yes	.76	.62
8	*9.6 × 6.4*	no	.43	.30
9	0.18 × 750	no	.51	.41
10	*16 × 13.2*	yes	.48	.43
11	38 × 56	yes	.75	.59
12	*1.500× 1.60*	yes	-	.53
13	*28 × 27.50*	yes	-	.48
14	*4380 × 3.50*	yes	-	.31
15	99 × 99	no	-	.43
16	42 × 52	no	-	.61

*Italicized problems concern problems that are not released for publication by CITO, and therefore a parallel version (with respect to number characteristics of the operands and outcome) is presented here.

995 students in the 2004 cycle completed one of these eight booklets, and thus responded to three to five different multiplication problems per student.

The testing procedure was very similar in both assessment cycles. Test booklets were administered in classroom setting and each student worked through the problems individually, without time pressure. On each page of the test booklet, several items were printed. Next to each item there was blank space that students could use to write down calculations. In 2004, test instruction was as follows: *"In this arithmetic task, you can use the space next to each item for calculating the answer. You won't be needing scrap paper apart from this space."* In addition, the experimenter from CITO explicitly stressed that students could use the blank space in their booklets for writing down calculations. Students were free to choose their own solution strategy, including choosing whether or not to make written calculations. In the 1997 assessment, instructions were somewhat less explicit in this respect.

For each student, a measure of general mathematics achievement level (GML) was

cycle	test booklet	item 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	*N*
1997	–	x	x	x	x	x	x	x	x	x	x	x						551
2004	1							x			x		x	x		x		120
2004	2								x	x		x			x			131
2004	3							x								x	x	129
2004	4							x	x						x	x		122
2004	5								x						x		x	123
2004	6									x	x	x	x	x				127
2004	7										x		x	x			x	120
2004	8									x		x					x	123
	***N* per item**	551	551	551	551	551	551	922	927	932	918	932	367	367	376	371	495	

FIGURE 3.3 *Distribution of multiplication items over test booklets, in the 1997 and in the 2004 assessment cyles. Symbol × indicates item was administered.*

computed, based on their performance on all mathematics problems presented to them in their test booklets. In the 1997 cycle, students completed – besides multiplication problems – also other problems from the domain of *numbers and operations.* Using Item Reponse Theory (IRT; e.g., Embretson & Reise, 2000; Van der Linden & Hambleton, 1997, see also below), a latent ability scale was fitted to the responses to all non-multiplication items. Consecutively, each students' position on the latent scale was estimated, and we standardized these estimates in the 1997-sample; range (–3.79, 3.15). For the 2004 cycle, a similar procedure was used, but because the item sampling design was different, students completed different sets of mathematics items from all mathematics domains (*numbers and operations*, *measurement*, and *percentages/fractions/ratios*). A general mathematical ability scale was fitted with IRT, and students' ability estimates were standardized, but now with respect to the 2004-sample; range (–4.52, 3.52). So, the general mathematics level (GML) measure used in the analyses indicates the relative standing of the student compared to the other students in his/her assessment cycle. Three students (one from 1997, two from 2004) with extreme scores (absolute standardized value larger than 3.50) were excluded from the analyses.

Responses

Two types of responses were obtained for each multiplication problem. First, the numerical answer given was scored as correct or incorrect. Skipped items were scored as incorrect. Second, by looking into the students' written work, the strategy used to solve each item was classified. Seven categories were distinguished, see also Figure 3.2. The first strategy (*Traditional*) was the traditional algorithm for multiplication. The second category (*Partitioning both operands*) included strategies in which both the multiplier and the multiplicand were split. An example of this strategy is the RME approach of 'column multiplication' (Van den Heuvel-Panhuizen, 2008). In the third category of strategies (*Partitioning one operand*) only one of the operands was split. The fourth category contained all *Other written* strategies, including only repeated addition. The fifth category (*No Written Working*) consisted of trials (student-by-item combinations) in which nothing was written down except the answer. The sixth category (*Wrong/Unclear*) consisting of erased or unclear strategies, and wrong procedures such as adding the multiplicands. The final category (*Skipped*) contained skipped problems (no written working and no answer).

Solution strategies were coded by 8 independent trained research assistants who each coded a separate part of the data. To assess the reliability of this coding, the work of 256 students was recoded by 2 external independent trained research assistants, and the interrater-reliability coefficient Cohen's κ (Cohen, 1960) was computed. The average κ on categorizing solution strategies was .87, which was more than satisfactory.

Statistical analyses

Several properties of the data set necessitated advanced psychometric modeling. These properties were, first, that the responses *within* each student were not independent, because each student completed several items (i.e., there were repeated observations). Hence, this correlated data structure should be accounted for in the psychometric modeling approach. Second, each of these repeatedly observed responses was bivariate: the item was solved correct or incorrect (dichotomous score variable) and a specific strategy was used (nominal variable). Third, the incomplete design of the data set complicated the comparisons between 1997 and 2004, because different students completed different subsets of items. Analysis on the item level would be justified, but would not take the multivariate aspect of the responses into account, and univariate statistics would be based on different subsets of students. Furthermore, analyses involving changes in performance would be limited to the common items and would therefore not make use of all available information. A final consideration was that it should be possible to include student characteristics as predictor variables in the analysis.

In sum, analysis techniques were needed that can take into account the multivariate aspect of the data and are not hampered by the incomplete design. These demands can elegantly be met by introducing a latent variable. Individual differences are modeled by mapping the correlated responses on the latent variable, while the student remains the unit of analysis. The latent variable can be either categorical or continuous.

Recall that we aimed to analyze changes in students' typical strategy choice, overall differences in accuracy between strategies, and changes in these strategy accuracies. For the analysis of changes in strategy choice, we argue that a categorical latent variable is best suited to model multivariate strategy choice, because individual differences between students are qualitative in this respect. Latent class analysis (LCA) accomplishes this goal, by introducing a latent class variable that accounts for the covariation between the observed strategy choice variables (e.g. Lazarsfeld & Henry, 1968; Goodman, 1974).

The basic latent class model is $f(\mathbf{y}|D) = \sum_{k=1}^{K} P(k) \prod_{i \in D} P(y_i|k)$. Classes run from $k = 1, \ldots, K$, and $\mathbf{y}$ is a vector containing the nominal strategy codes on all items i that are part of the item set D presented to the student given the test design. 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 the probability of solving item i with each particular strategy. So, we search for subgroups (latent classes) of students that are characterized by a specific pattern of strategy use over the items presented. To analyze changes between 1997 and 2004 in the relative frequency of the different strategy classes, year of assessment was inserted as a covariate, so that assessment cycle predicted class membership (Vermunt & Magidson, 2002). All latent class models were fitted with the `poLCA` package (Linzer & Lewis, 2010, 2011) available for the statistical computing program `R` (R Development Core Team, 2009). Because latent class models on variables with 7 different categories were very unstable, we recoded the solution strategies into four main categories: Traditional, Non-Traditional (partitioning both operands, partitioning one operand, other written strategies), No Written Working, and Other (wrong/unclear and skipped).

The second portion of the research question focused on strategy accuracy: how can the strategy used predict the probability of solving an item correctly? We argue that in these analyses a continuous latent variable is appropriate, to be interpreted as (latent) ability or proficiency. The repeatedly observed correct/incorrect scores are the dependent variables, and the nominal strategies take on the role of predictors. The latent variable accounts for the individual differences in proficiency in complex multiplication by explaining the correlations between the observed responses. Item Response Theory (IRT) modeling (e.g., Embretson & Reise, 2000; Van der Linden & Hambleton, 1997) accomplishes this goal. Through the five common items, it was possible to fit one common scale for 1997 and 2004 of proficiency in complex multiplication, based on all 16 items.

In the most simple IRT measurement model (the Rasch model), the probability of a correct response of subject p on item i can be expressed as $P(y_{pi} = 1|\theta_p) = \frac{\exp(\theta_p + \beta_i)}{1+\exp(\theta_p+\beta_i)}$ Latent variable θ_p expresses ability or proficiency, measured on a continuous scale. The item parameters β_i represent the *easiness* of each item. Such descriptive or measurement IRT models can be extended with an explanatory part (Wilson & De Boeck, 2004; Rijmen et al., 2003), meaning that covariates or predictor variables are included of which the effects on the latent scale are determined. These can be (a) item covariates, that vary across items but not across persons, (b) person covariates, that vary across persons but not across items, and (c) person-by-item covariates, that vary across both persons and items. In the present analyses, the strategy chosen on an item was dummy coded and included as person-by-item predictor variables (for further details, see Hickendorff et al., 2009b). Like in the LCA, we used the four main solution strategy categories. Moreover, the category of Other strategies was not of interest in analyzing strategy accuracies, since it was a small heterogeneous category of remainder solution strategies, consisting mainly of skipped items. Therefore, all student-by-item combinations (i.e., trials) solved with an Other strategy were excluded from the explanatory IRT analyses.

All IRT models were fitted using Marginal Maximum Likelihood (MML) estimation procedures within the NLMIXED procedure from SAS (SAS Institute, 2002, see also De Boeck & Wilson, 2004; Rijmen et al., 2003; Sheu et al., 2005). We chose nonadaptive Gaussian quadrature for the numerical integration of the marginal likelihood, with 20 quadrature points, and Newton Raphson as the optimization method.

3.2.2 Results

Strategy choice

Table 3.2 displays proportions of use of the seven strategies, separately for the 1997 and the 2004 assessment. In the first 2 columns, strategy proportions are totaled over the five common items. The traditional algorithm was the most prevalent strategy in both years, but its use decreased markedly between 1997 and 2004. The non-traditional

TABLE 3.2 *Strategy use on multiplication problems in proportions, based on 1997 and 2004 data.*

	common items		all items	
multiplication strategy	1997	2004	1997	2004
traditional	.65	.45	.59	.39
partitioning - both operands	.03	.08	.03	.07
partitioning - one operand	.03	.08	.05	.09
other written strategy	.01	.02	.01	.03
no written working	.17	.25	.23	.31
wrong/unclear	.02	.03	.02	.03
skipped	.08	.09	.07	.09
N observations	2755	1876	6061	3852

strategies (partitioning both operands, partitioning one operand, and other written strategies) each increased in relative frequency of choice: on the common items, from a total of 7% of the trials in 1997 to 18% of the trials in 2004. Furthermore, the frequency of answering without written working also increased between the two cycles. The final two strategy categories (wrong/unclear and skipped items) remained more or less stable in frequency. In the final 2 columns of Table 3.2, strategy proportions are totaled over all items presented in each assessment, so these proportions are based on different item collections for 1997 and 2004. Although these distributions seem slightly different from those based only on the common items, the pattern of shifts between 1997 and 2004 was very similar.

Latent class models on strategy choice, recoded in four main categories, with year of assessment as covariate were fitted with 1 to 6 latent classes. The Bayesian Information Criterion (BIC) was used to select the optimal number of classes. The BIC is a criterion that penalizes the fit (log-likelihood, LL) of a model with the model complexity (the number of parameters; P), and it is computed as -2LL $+P\log(N)$, with N the sample size. Lower BIC-values indicate better models in terms of parsimony. The model with 4 classes showed the lowest BIC-value, and was therefore selected as the best-fitting model. The relative entropy of this latent class model, a measure of classification uncertainty ranging between 0 (high uncertainty) and 1 (low uncertainty) (Dias & Vermunt, 2006), was .84, indicating that the latent classes were well separated.

Figure 3.4 graphically displays the estimated parameters of this 4-class model. These are first the conditional probabilities: for each particular class, the probabilities of choosing each of the four strategies on each of the 16 items. The second parameters were the class sizes in the two assessment years, showing changes over time. First note that the class-specific strategy profiles of the first three classes are more or less dominated by one strategy type chosen on all items. So, apparently students were quite consistent in their strategy choice on this set of multiplication problems.

From these strategy profiles we interpret the latent classes as follows. The first class is dominated by the Traditional algorithm, although item 12 and to a lesser extent item 4 are clear exceptions with a large probability of being solved without written working. Nevertheless, we argue that the best way to summarize this latent class is to label it the Traditional class. In the 1997 assessment, the majority of the students (67%) belonged to this class, while this decreased to less than half (44%) of the students in 2004. The second class is characterized by a very high probability on all items to state the answer without writing down any calculations or solution steps (No Written Working class). This class nearly doubled in size, from 13% in 1997 to 22% in 2004. The third class (Non-Traditional class) is dominated by Non-Traditional strategies, but again items 12 and 4 are exceptions with the modal probability of No Written Working. This class tripled in size, from 7% in 1997 to 22% in 2004. Finally, the fourth class is a mishmash of Other strategies, No Written Working, and Traditional strategies. This Remainder class did hardly change in size between 1997 (13%) and 2004 (12%).

Next, we studied whether the effect of assessment cycle on latent strategy class depended on students' gender, general mathematics level, or socio-economical status. Because inserting these many variables as covariates in latent class analysis would render the model statistically unstable, an alternative approach was used consisting of two steps. First, all students were assigned to the latent class for which they had the highest posterior probability (modal assignment; mean classification error .08). Next, this 4-category class membership variable was used as dependent variable in a multinomial logistic regression model (see for example Agresti, 2002). Fifty-one students were excluded because they had missing or extreme values on at least one of the predictor variables.

The main effects of year (Likelihood Ratio (LR) test[2] = 72.5, $df = 3$, $p < .001$), gender (LR = 59.7, $df = 3$, $p < .001$), GML (LR = 88.8, $df = 6$, $p < .001$), and SES (LR = 27.8,

2 The Likelihood Ratio test can be used to statistically test the difference in fit of two nested models. The test statistic is computed as 2 times the difference between the LL of the general model and the LL of the specific model, and it is asymptotically χ^2-distributed with df the difference in number of estimated parameters between the 2 models.

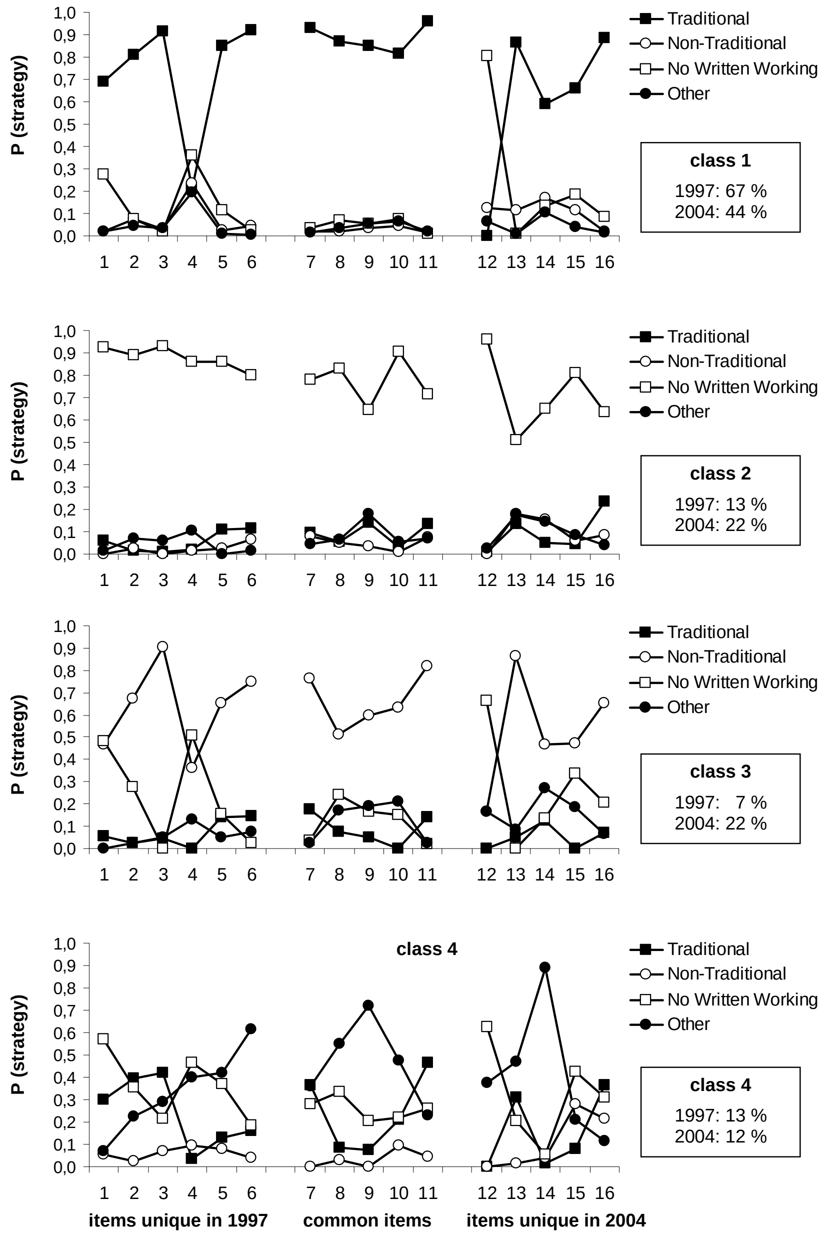

FIGURE 3.4 *Conditional probabilities of strategy choice on multiplication problems of the 4 latent classes model, 1997 and 2004 data.*

TABLE 3.3 *Cross-tabulations of the student background variables general mathematics level, gender, and SES with latent strategy class membership (in proportions); multiplication problems, 1997 and 2004 data.*

		Latent strategy class				
		1 (T)	2 (NWW)	3 (N-T)	4 (R)	*N*
boys	1997	.65	.17	.10	.08	252
	2004	.39	.26	.22	.13	488
girls	1997	.73	.11	.05	.11	274
	2004	.58	.12	.17	.13	481
low GML		.43	.23	.13	.20	480
medium GML		.58	.16	.15	.10	508
high GML		.65	.14	.16	.05	509
SES-1		.58	.14	.16	.12	1306
SES-2		.42	.34	.12	.12	191

Note. T = Traditional class; NWW = No Written Working class; N-T = Non-Traditional class; R = Remainder class.

$df = 3$, $p < .001$) on class membership were all significant. Moreover, the interaction between gender and assessment cycle was also significant (LR $= 8.6$, $df = 3$, $p = .035$), implying that the shift in relative frequency of the strategy choice classes was not the same for boys and girls. The other interaction effects, between GML or SES on the one hand and assessment cycle on the other hand, were not significant (ps $> .05$).

The top portion of Table 3.3 shows the interaction between gender and assessment cycle on latent class membership. Gender differences in overall strategy choice patterns clearly emerge: In both assessment cycles, girls more often typically used the Traditional algorithm than boys, while they were less often classified in the No Written Working or Non-Traditional classes. Interestingly, the shift in strategy choice between 1997 and 2004 was different for boys than for girls. Boys were increasingly classified in the No Written Working class, while the proportion of girls in this class was about stable. Furthermore, the decrease over time in the Traditional strategy class was larger for boys than for girls. Apparently, the shift away from the traditional algorithm towards answering without written working should be mainly attributed to boys. Table 3.3 also shows the main effects of GML (trichotomized based on percentile scores, to facilitate interpretation) and SES. The proportion of students being classified in the Traditional class increased with increasing mathematics achievement level, while the proportion of students being classified in the Remainder class as well as in the No Written Working class decreased with increasing GML. The proportion of students classified in the Non-Traditional class was relatively unaffected by GML. Finally, SES-1 students were more often classified in the Traditional class than SES-2 students, and less often in the No Written Working class.

Strategy accuracy

To evaluate how the found shift in strategy choice should be evaluated with respect to achievement, we investigated whether the multiplication strategies differed in accuracy rate, with (explanatory) IRT models. Starting from the Rasch measurement model without explanatory variables, a model was built with a forward stepwise procedure by successively adding predictor variables and retaining those that had significant effects. All 1,027 trials (student-by-item combinations) involving Other strategies (wrong, unclear, of skipped) were excluded. In total, 1,542 students yielding 8,886 observations were included in the analyses.

First, the null model without any predictor effects was fitted, assuming that the θ_p come from one normal distribution. The 17 parameters were 16 item easiness parameters β_i with estimates ranging between –.86 and 2.19, and the variance of the ability scale θ estimated at 1.35 (the mean of θ was fixed at 0 for identification of the latent scale). Next, the effect of assessment cycle was estimated, which resulted in a substantial decrease in BIC as well as in a significant increase in model fit; $\text{LR} = 42.4$, $df = 1$, $p < .001$. The latent regression parameter of 2004 compared to 1997 was –.64 on the logit scale[3], which was highly significant ($z = -6.52$).

Next, type of strategy used on an item was inserted as a predictor of the probability of solving an item correct. In order to keep the number of parameters manageable and interpretation feasible (see also Hickendorff et al., 2009b), these strategy effects were restricted to be equal for all items. Adding strategy effects yielded a highly significant increase in model fit ($\text{LR} = 393.2$, $df = 2$, $p < .001$). Compared to No Written Working, both using a Traditional strategy (difference on logit scale $\delta_{\text{T vs. NWW}} = 1.53$, $z = 19.58$)

3 The effect of –.64 on the logit scale can be transformed to the odds ratio scale or the probability scale. The odds ratio is computed as $\exp(-.64) = .53$, and implies that the odds of a correct answer for 2004-students is about half the size of the odds for 1997-students. On the probability scale, we can compute that on an item on which 1997 students had a 50% probability to obtain a correct answer, this probability was $\frac{\exp(-.64)}{1+\exp(-.64)} \times 100\% = 35\%$ for students in the 2004 assessment.

and using a Non-Traditional strategy ($\delta_{\text{N-T vs. NWW}} = .93$, $z = 9.74$) yielded a significantly higher probability to obtain a correct answer. Moreover, the Traditional strategy was significantly more accurate than Non-Traditional strategies ($\delta_{\text{T vs. N-T}} = .59$, $z = 6.64$). Clearly, the three main strategy categories differed in accuracy. By accounting for strategy choice shifts between 1997 and 2004, the regression parameter of year decreased to –.43 ($z = -4.49$). Furthermore, the interaction effect of Year and Strategy was not significant (LR = 4.9, $df = 2$, $p = .09$), implying there was a general and equally-sized decrease in success rates from 1997 to 2004 for each of the three strategies.

Subsequently, we tested whether the achievement change over time or the effect of strategy used depended on either gender, general mathematics level (GML), or socio-economical status (SES). Excluding an additional 50 students (201 trials) from the analyses because they had missing or extreme values on one or more of the background variables, these three student characteristic variables were added to the explanatory IRT model, and we tested the interaction effects with year and strategy. None of the two-way interaction effects of the student characteristics with year were significant (year × gender: LR = .2, $df = 1$, $p = .63$; year × GML: LR = .6, $df = 1$, $p = .45$; year × SES: LR = 2.3, $df = 1$, $p = .13$). This implied that the accuracy decrease between assessment cycles was about the same size for boys and girls, for students with low or higher SES, and for students with different mathematics achievement level. By contrast, strategy significantly interacted with gender (LR = 7.2, $df = 2$, $p = .027$) and GML (LR = 37.8, $df = 2$, $p < .001$), but not with SES (LR = 2.3, $df = 1$, $p = .13$), the largest interaction effect being with GML. The strategy-by-gender interaction was no longer significant up and above the interaction between Strategy and GML (LR = 4.6, $df = 2$, $p = .10$); apparently, it was mediated by gender differences in general mathematics achievement level.

Figure 3.5 displays the interaction effects between GML and strategy used on the logit IRT ability scale. It shows that students' general mathematics level was positively related to performance on the multiplication problems, within each particular strategy used. Furthermore, the effect of GML was significantly stronger when the strategy No Written Working was used ($\zeta_{\text{GML in NWW}} = 1.18$, $z = 18.35$) than it was when either the Traditional algorithm ($\zeta_{\text{GML in T}} = .73$, $z = 15.56$) or one of the Non-Traditional strategies ($\zeta_{\text{GML in N-T}} = .84$, $z = 10.02$) was used. The difference between the latter two regression parameters was not significant. Interpreting these effects, it seems that with increasing general mathematics level it became less important which strategy students used on complex multiplication. In particular, for low performers, answering without written

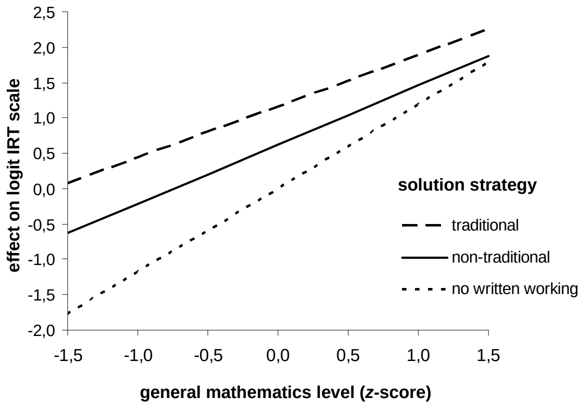

FIGURE 3.5 *Graphical display of interaction effect between strategy used and student's general mathematics level on IRT ability scale, based on multiplication problems in 1997 and 2004 cycles.*

work was much less accurate than using one of the two written strategies; for high performers this difference disappeared.

Importantly, even after accounting for all significant (interaction) effects of student characteristics and strategy used, the performance decrease between 2004 and 1997 remained substantial (−.50) and significant ($z = -5.96$), so shifts in strategy choice only partially accounted for the performance decrease.

3.2.3 Conclusions part I

In the first part of this study, we aimed to get more insight in the lagging and decreasing performance level in multiplication, by analyzing changes in strategy choice and in strategy accuracies between 1997 and 2004. Both descriptive statistics and latent class models showed that strategy choice has shifted from 1997 to 2004: The use of the traditional algorithm decreased, while the use of non-traditional strategies as well as no written working solutions increased, the latter two by approximately the same amount. Moreover, the shift away from typically using the traditional algorithm towards typically answering without written working was observed mainly in boys.

To evaluate how the found shift in strategy choice should be evaluated with respect to accuracy, we investigated whether the multiplication strategies differed in accuracy rate.

Results showed that the traditional algorithm was more accurate than non-traditional strategies, which in turn were more accurate than answering without written working (these differences were smaller for students with higher mathematics achievement level). Consequently, the observed shift in strategy choice – replacing traditional strategies by non-traditional and no written working strategies – can be characterized as unfortunate with respect to achievement, and is one contributor to the general performance decline. However, this did not explain the complete performance decrease: even after accounting for the shift in strategy choice between the two years, still a significant decrease in performance from 1997 to 2004 remained. So, each solution strategy on its own was carried out significantly less accurately in 2004 than it was in 1997.

In conclusion, two changes regarding strategy use appeared to have contributed to the general performance decline on complex multiplication problems: a shift in choice of more accurate to less accurate ones, and a general accuracy decline within each strategy on its own. A relevant next question is what influences students' strategy choice. The effect of student characteristics gender, general mathematics level, and SES were already addressed in the first part of the study. In the next part, we try to get more insight in the effect of teacher's instruction on strategy choice, focusing on differences between complex multiplication and division.

3.3 Part II: Effect of teachers' strategy instruction on students' strategy choice

3.3.1 Method

Sample

The sample used for the second part of this study consisted of the 995 students of the 2004 assessment, who were also part of the sample of part I of this study. These students not only completed the complex multiplication problems, but also problems on complex division.

Material and Procedure

In total, there were 10 complex multiplication problems (see part I of this study) and 13 problems on complex division (see Hickendorff et al., 2009b, Table 1) in the 2004 assessment[4], ranging from the easiest problem $157.50 \div 7.50$ (60.4% correct) to the most difficult one $6.40 \div 15$ (12.6% correct). These 23 problems were administered in an incomplete test design: there were 8 different test booklets containing between 6 and 13 problems. The testing procedure was the same as in part I of this study.

In the schools participating in the 2004-assessment, teachers in grade 4 ($N = 116$), 5 ($N = 115$), and 6 ($N = 118$) filled in a questionnaire about the mathematics curriculum and teaching practices. There were questions included on their approach in teaching multidigit operations (addition, subtraction, multiplication, and division). For each operation, they were asked to choose, from two worked-out examples, which approach best matched the practice in their classroom: (a) the traditional algorithm, or (b) so-called 'column calculation' (Van den Heuvel-Panhuizen, 2008), the RME-alternative to the standard algorithm. Column calculation in multiplication entailed strategies in which both operands were partitioned (see Figure 3.2); in division it entailed repeated subtraction of multiples of the divisor from the dividend (see below). If teachers taught the column calculation procedure first and the traditional algorithm later, they could mark both approaches.

Figure 3.6 shows the distribution of teachers' responses to these questions on multiplication and division. It shows that between grade 4 and grade 6 there is a gradual shift from the RME approach to the traditional approach, in both multiplication and division. However, sixth grade teachers instructed the traditional algorithm much less frequently for solving division problems than for solving multiplication problems. This difference between multiplication and division is in line with the learning/teaching trajectories differences and the mathematics textbooks, which do no longer cover the traditional algorithm for long division (Van den Heuvel-Panhuizen, 2008). In this part of the current study, sixth grade teachers' approach to multiplication problem solving (missing data for 39 students) and to division problem solving (missing data for 57 students) were used as variables predicting students' strategy choice.

4 For reasons of consistency, we used the same item numbers in part II of as in part I of this study for the multiplication problems (7 - 16) and as in Hickendorff et al. (2009b) for the division problems (7 - 19).

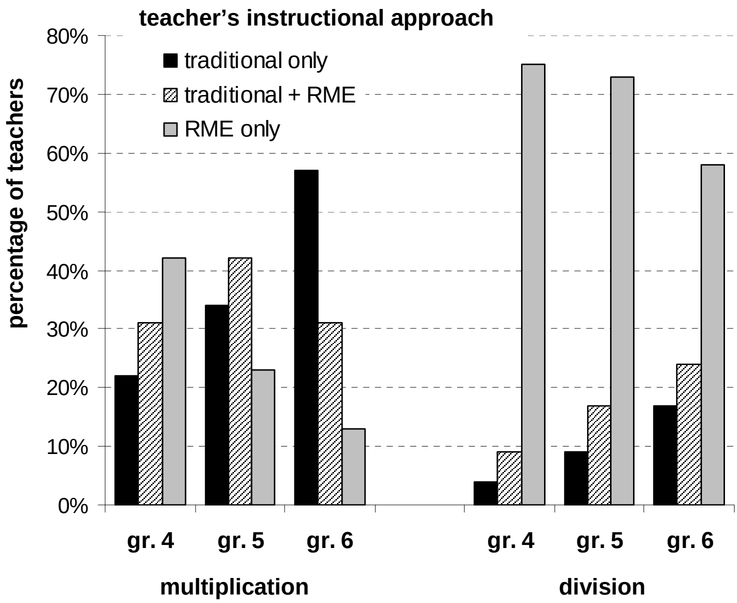

FIGURE 3.6 *Fourth grade, fifth grade, and sixth grade teachers' approach to complex multiplication and division problem solving, as reported in J. Janssen et al. (2005, p. 44).*

Responses

For the multiplication problems, the strategy categorizations from part I of this study were used. For the division problems, we distinguished seven main strategies[5] (see Hickendorff et al., 2009b, 2010 for examples): (a) the traditional algorithm of long division, (b) repeated subtraction of multiples of the divisor from the dividend (the RME alternative to the traditional algorithm; Van den Heuvel-Panhuizen, 2008) (c) repeated addition of multiples of the divisor towards the dividend (multiplying-on), (d) other written strategies, (e) answering without written working, (f) unclear strategies or wrong procedures, and (g) skipping the problem.

5 In Hickendorff et al. (2009b), these 7 categories were recoded into 4 main solution strategies, by combining first the repeated subtraction with the repeated addition strategy and other written strategies into a category labeled Realistic strategies, and second, by combing the last two categories of unclear strategies/wrong procedures and skipped problems into a category labeled Other strategies.

TABLE 3.4 *Strategy use on multiplication and division problems, split by teacher's instructional approach, based on 2004 data.*

	teacher's approach to multiplication			
multiplication strategy	trad. only	trad. + RME	RME only	total
traditional	.43	.34	.31	.39
partitioning - both operands	.08	.11	.09	.07
partitioning - one operand	.04	.10	.16	.09
other written strategy	.02	.05	.02	.03
no written working	.32	.28	.32	.31
wrong/unclear	.03	.03	.02	.03
skipped	.09	.09	.08	.09
number of trials	2242	1084	383	3709
	teacher's approach to division			
division strategy	trad. only	trad. + RME	RME only	total
traditional	.43	.18	.02	.14
repeated Subtraction	.04	.17	.24	.19
repeated Addition	.04	.05	.05	.05
other written strategy	.01	.01	.02	.01
no written working	.39	.41	.47	.44
wrong/unclear	.04	.04	.06	.05
skipped	.07	.14	.13	.12
number of trials	897	1225	2674	4796

3.3.2 Results

Table 3.4 presents the distribution of strategy choice trials on multiplication and division problems, split by the instructional approach of the student's sixth grade teacher. First, it shows that the overall distribution of strategy choice is different for multiplication than for division problem solving. That is, the traditional algorithm for multiplication was used much more frequently (39% of all trials) than the traditional algorithm for division (14% of the trials); while answering without written work was more common on division problems (44% of the trials) than on multiplication (31%).

Second, there was a clear influence of the teacher's approach to problem solving on students' strategy choice, in particular in division.[6] For multiplication, choosing the traditional algorithm increased when the teacher instructed this approach, in particular if it was the only strategy. Moreover, the use of the partitioning-one-operand strategy increased when teachers instructed the RME approach in combination with the traditional algorithm or in particular when it was the only strategy instructed. For division, the use of the traditional algorithm clearly depended on whether the teacher instructed this approach or not. Furthermore, the use of the repeated subtraction strategy increased when teachers instructed it. Finally, both answering without written working and skipping problems appeared to be influenced by the teacher's approach: no written working was most prevalent with teachers instructing only the RME approach to division, while skipping a problem occurred least often when the teacher instructed only the traditional algorithm for division.

3.3.3 Conclusions Part II

In the 2004 assessment, information on the teacher's approach to instruction in multidigit multiplication and division problem solving was available. This variable appeared to affect students' strategy choice on both operations. In multiplication, choice for the traditional algorithm and for partitioning one operand was influenced by the teacher's instructional approach. The effect of instructional approach was particularly strong in division, however: nearly exclusively students whose teacher instructed the traditional algorithm for long division used that algorithm. Moreover, students whose teacher instructed the RME approach to division more frequently used this RME strategy (repeated subtraction), but also more often answered without written working or skipped the problem entirely.

6 Straightforward statistical testing of dependency of rows and columns in Table 3.4 was not possible, however, because observations within cells were not independent. To provide support for the statistical significance of the relation between teacher's approach and students' strategy choice, we tested it using students' latent strategy class membership as dependent variable. For multiplication, strategy choice latent class membership (4 classes) of part I of this study was used, and the effect of teacher's approach to multiplication problem solving was highly significant ($\chi^2(6, N = 956) = 43.2$, $p < .001$). For division, we used strategy choice latent class membership (also 4 classes: mainly Traditional, mainly Non-Traditional, mainly No Written Working, and mainly Other strategies) from p. 340-343 Hickendorff et al. (2009b). The effect of teacher's approach to division problem solving on strategy choice class was even more significant ($\chi^2(6, N = 938) = 251.9$, $p < .001$) than it was in multiplication.

3.4 GENERAL DISCUSSION

In the current study, we aimed to get more insight in Dutch sixth graders' performance level in complex or multidigit multiplication and division, that was found to be decreasing over time and lagging behind educational standards, in a reform-based mathematics learning/teaching trajectory. In secondary analyses of national assessment data we focused on the solution strategies students used as an explanatory mechanism between (change in) instruction and (change in) achievement. In the first part, the relation between solution strategy use and achievement in complex multiplication was investigated to analyze the negative performance trend between 1997 and 2004. Findings showed that two changes regarding solution strategies contributed to the performance decline: a shift in strategy use from a more accurate strategy (the traditional algorithm) to less accurate ones (non-traditional partitioning strategies and answering without written work, the latter shift attributable to boys), as well as a general decline in each strategy's accuracy rate. In the second part, students' strategy choice in multiplication and division in the 2004 assessment appeared to be influenced by the instructional approach held by their teachers, most profoundly in division problem solving. In the following, we discuss the conclusions and the implications in more detail.

3.4.1 Complex multiplication problem solving

Strategy use in complex multiplication shifted between 1997 and 2004. The use of the traditional algorithm decreased, while answering without written work (most likely mental calculation, see also Hickendorff et al., 2010) and the use of non-traditional partitioning strategies increased. An important subsequent question is: How do we evaluate this strategy shift in multiplication? The current findings showed clear differences in accuracy between the strategies, leading us to argue that, first, the increase in mental calculation is a worrisome development, because the success rate of non-written strategies was substantially lower than of the two written strategies. Second, because the non-traditional strategies were less accurate than the traditional algorithm for students of all mathematics levels, an increase in this strategy also does not seem desirable.

Another worrisome – and more difficult to grasp – development is the finding that each of the three main multiplication strategies dropped in accuracy rate between the two assessment cycles. That means that on the same problem with a particular strategy, sixth graders in 1997 had a higher probability to derive the correct answer than sixth graders in 2004 had. It leaves us with another negative trend that needs explanation, that should probably be sought in the educational practice. Potential mechanisms include the amount of time and practice spent on these topics, i.e., opportunity-to-learn (OTL), which has been argued to be the single most important predictor of student achievement (Hiebert & Grouws, 2007). However, the educational assessments do not offer enough information to analyze this rigorously (see also Hickendorff et al., 2009a), and further research is needed. It would also be very informative to carry out an international comparative study between countries that differ in their opportunity-to-learn with respect to multidigit multiplication and division.

Furthermore, there were differences between students in multiplication problem solving. First, there were clear gender differences in strategy choice: girls had a larger tendency than boys to consistently use the traditional algorithm and were less inclined to consistently use non-written strategies. These findings are in congruence with previous findings on gender differences in strategy use, with girls showing a larger reliance on structured strategies and algorithms, and boys having a higher tendency to use more informal, less structured strategies (Carr & Davis, 2001; Carr & Jessup, 1997; Gallagher et al., 2000; Hickendorff et al., 2009b, 2010; Timmermans et al., 2007). These gender differences may be related to the consistent finding that girls have lower levels of confidence with mathematics (Mullis et al., 2008; Timmermans et al., 2007; Vermeer et al., 2000), so they may act more cautiously than boys and therefore choose the safety of using well-structured written strategies. Moreover, gender differences changed over time: between 1997 and 2004, boys and girls showed a strategy shift from the traditional algorithm towards non-traditional strategies, but boys additionally shifted from the traditional algorithm toward answering without written working. The shift towards mental calculation should thus be attributed mainly to boys.

Second, students' mathematics achievement level affected individual differences in strategy choice in multiplication too: the tendency to quite consistently use non-written strategies decreased with higher mathematics achievement level, while the tendency to use written strategies (traditional and non-traditional) increased with higher mathematics level. So, lower performers seem to choose their strategy less adaptively than high performers, congruent with findings of Foxman and Beishuizen (2003), Hickendorff et al. (2010), Torbeyns, De Smedt, et al. (2009b), and Torbeyns et al. (2002, 2004a, 2006). Moreover, for students with lower mathematics achievement level the accuracy gap between written and non-written strategies (see Figure 3.5) was larger, so they seemed to be doubly disadvantaged by choosing a non-written strategy.

These results have theoretical implications for cognitive models of strategy choice, in which it is hypothesized that children choose their strategies adaptively, i.e., they choose the fastest strategy that yields the correct answer (e.g., Shrager & Siegler, 1998). The present findings seem to signal suboptimal strategy choices when students, in particular the lower performing ones, chose non-written strategies. However, cautiousness is called for: because students were free to select their strategies (the so-called choice method; Siegler & Lemaire, 1997) it is likely that selection effects biased strategy accuracy data. For example, the finding that the use of written strategies increased with students' mathematics achievement level may have biased the accuracy of those strategies upwardly, although we were able statistically correct for this. Further research addressing strategy efficiency in an unbiased manner, such as has been done by Hickendorff et al. (2010) in the domain of division, is needed to make firmer conclusions regarding the adaptivity of students strategy choices in the domain of complex multiplication.

3.4.2 Multiplication and division: similarities and differences

The present study shows some remarkable similarities and differences between the domains of multiplication and division problem solving (as reported in part II of the present study and in Hickendorff et al., 2009b).

First, regarding shifts in strategy choice between the two assessment cycles, a similarity was the decrease in use of the traditional algorithm. This is in accordance with a general shift away from algorithmic procedures in mathematics education reform (although it is worth noticing that already in 1997 the large majority of mathematics textbooks used were based on reform principles). A second similarity is an increase in the answering without written working (most likely mental calculation, see also Hickendorff et al., 2010), that was mainly attributable to boys. Although mental calculation plays an important role in mathematics education reform (Blöte et al., 2001; Buijs, 2008) it was not anticipated that this would also affect the way students solve complex arithmetic problems with multidigit numbers, on which the use of paper and pencil was allowed. One would expect that, rather than an increase in mental computation, predominantly the use of non-traditional strategies would increase, because these are part of the learning trajectories (Van den Heuvel-Panhuizen, 2008). Herein also lies a striking difference between multiplication and division. In division instruction, the traditional algorithm has been replaced entirely by the RME approach, but surprisingly, students' behavior does not show an increase of the RME approach. In contrast, in multiplication, students increasingly used the RME strategies, while this was not the end point of the RME learning trajectory. One would expect to find the opposite pattern.

Second, the accuracy differences between the main solution strategy categories in multiplication and division were also characterized by similaries and differences. In both operations, answering without written working was the least accurate strategy, in particular for the lower performers. The accuracy difference between the traditional algorithm and non-traditional strategies, however, depended on the operation. In multiplication, non-traditional partitioning strategies were less accurate than the traditional algorithm, for students of all mathematics levels. In contrast, in division the non-traditional repeated addition/subtraction strategies were equally accurate as the traditional algorithm for most students (although for medium performers the traditional division algorithm was significantly more accurate). A possible explanation for this difference may be that for division, repeated addition/subtraction strategies (the RME approach) are actually the only approach being taught (at least in the mathematics textbooks) and hence serve as a full alternative to the traditional algorithm, while this is not the case for multiplication.

Finally, in the 2004 assessment, information on the teachers' instructional approach to solving multiplication and division problem solving was available from a teacher questionnaire. The teacher's instruction appeared to be quite different for multiplication where the traditional algorithm was still dominant, than it was for division where the RME approach was dominant. Moreover, it appears that in both domains a switch towards an increase in the traditional algorithm has taken place between grade 4 and grade 6. In multiplication, this is in line with the RME learning/teaching trajectory, while that is not the case in division (Van den Heuvel-Panhuizen, 2008). The observation that teachers apparently diverged from the mathematics textbook in division also illustrates the fact that the *enacted* curriculum (the actual instruction taking place in the classroom, e.g. Porter, 2006; Stein et al., 2007) can differ from the *intended* curriculum that is based on written documents such as textbooks and educational standards, and that it is therefore important to take both curricula into account.

The instructional approach of the teacher had substantial effects on the strategy choice of the students. Particularly the approach to solving complex division problems was important: almost exclusively students whose teacher instructed the traditional algorithm for division (as the only strategy or in combination with RME strategies) actually used this strategy. So, only when teachers departed from their textbook and instructed the division algorithm students used it, which is not unexpected since it is probably not a strategy that is easy to self-invent. A further interesting finding is that when teachers instructed the RME approach to division problem solving, the frequency of answering without written working and of skipping problems entirely was higher than when teachers instructed the traditional algorithm. It thus seems that when students were instructed the RME approach to division problem solving, they were less inclined to apply their standard written procedure (the RME approach) and were also less able or confident to attempt solving the problem at all. A tentative explanation for this pattern may be that the RME approach to division is less well structured than the traditional algorithm, so that students know less well how to start and what to do. In addition, it may be that teachers who instruct the traditional algorithm for division, thereby diverging from the mathematics textbook, value standard solution procedures more, affecting their students' behavior.

For multiplication, the relation between teacher's instructional approach and students' strategy choice was less marked than for division. Still, the use of the traditional algorithm was higher in students whose teachers instructed it than it was in students whose teachers instructed only the RME approach. Moreover, students whose teacher instructed the RME approach to multiplication problem solving more often used partitioning strategies. Contrary to division, the frequency of answering without written working was rather unaffected by the teacher's approach.

3.4.3 Educational implications

Regarding multidigit multiplication, the present findings would lead to the educational recommendation that teachers encourage students to use the traditional algorithm. Moreover, for both multiplication and division it seems legitimate to encourage the use of written strategies over non-written strategies for problems with multidigit numbers, in particular for the lower performing students. The current findings give initial support for the idea that changing teacher's strategy instruction may be an effective way to influencing students' strategy choice, although further research is needed. Moreover, one could think of other mechanisms to affect students' problem solving behavior as well, such as for example crediting written solution steps on top of crediting only the correct answer, or changing the appraisal of written compared to mental strategies. Furthermore, the entire domain of multiplicative reasoning performance (the tables, mental arithmetic, and complex arithmetic with the use of paper and pencil allowed) showed a negative trend (J. Janssen et al., 2005), so we plead for vigilance of the educational community regarding the position of this domain in the mathematics curriculum.

All results taken together – an unfortunate shift in strategy choice (mainly in boys), the traditional algorithm being the most accurate strategy (at least in multiplication), the relatively high proportion of lower achievers who answer without written work while for them this is a particularly unsuccessful strategy, and the general decrease in accuracy within each strategy – we argue that reconsideration of several elements of the implementation of the RME approach is called for. These elements are not unique to the Dutch mathematics education reform, so it is also important in an international perspective. For example, students' informal strategies are very important in the reform, which seems not to be without problems. In particular, we think that the transition from informal strategies to the traditional algorithm needs further consideration, such as for example also came forward from findings in the UK (Anghileri et al., 2002). In addition, the idea that students are free to choose how to solve problems may have a negative side-effect in boys, who are on average more inclined to intellectual risk-taking than girls (Byrnes, Miller, & Shafer, 1999) and may overestimate their ability to solve problems without writing down solution steps or intermediate answers. Moreover, like Geary (2003), Torbeyns et al. (2006), and Verschaffel, Luwel, Torbeyns, and Van Dooren (2009) we plead for more research-based evidence into the feasibility of striving for adaptive expertise in mathematics education, especially for the lower performing students who seem to be doubly disadvantaged by making suboptimal strategy choices.

3.4.4 Final considerations

The present study is limited in several ways, since it is based on large-scale educational assessments data (see also Hickendorff et al., 2009a; Van den Heuvel-Panhuizen, Robitzsch, Treffers, & Köller, 2009). Because assessments are surveys, they are descriptive by nature, which has its limitations such as allowing only correlational analyses with explanatory variables. Therefore, our present study is also limited in several ways. For example, it was not possible to study the effect of item characteristics (such as whether a problem was presented in a context or not) on strategy use or accuracy, because these features were not varied in a systematic way. Furthermore, the classroom administration procedure – although making large sample sizes feasible – had the drawback that for gaining insight in solution strategy use we had to revert to students' written work. Therefore, we were left with no further information on the instances in which students did not write down any work but the answer. Presumably, they used mental computation on those trials (as was mostly confirmed by Hickendorff et al., 2010), but we cannot be certain about that. Finally, the results of the present study are limited to the situation in the Netherlands, and the question is to what extent it would generalize to other countries. We believe, however, that the Dutch situation is interesting to study, because of the influential theory of realistic mathematics education gaining international popularity, which contains several elements of the international reform movement. In addition, because there is, unlike the US, a nationally coherent curriculum, trends in mathematics achievement can be linked quite closely to shifts at a national level in instructional approach.

Acknowledging these limitations, we argue that studying solution strategies in data from large-scale assessments was a valuable enterprise, both from a practical educational viewpoint as well as from the perspective of educational psychology. The present findings can be a valuable starting point for evaluating learning outcomes more comprehensively and for raising research questions for further research, and they advanced our insight into performance trends, strategy choice, and strategy accuracy in multidigit multiplication and division in a reform-based educational environment significantly.