

# The Empty Number Line in Dutch Second Grades: *Realistic* Versus *Gradual* Program Design

Anton S. Klein and Meindert Beishuizen,  
*Leiden University, The Netherlands*  
Adri Treffers, *Utrecht University, The Netherlands*

In this study we compare 2 experimental programs for teaching mental addition and subtraction in the Dutch 2nd grade ( $N = 275$ ). The goal of both programs is greater flexibility in mental arithmetic through use of the empty number line as a new mental model. The programs differ in instructional design to enable comparison of 2 contrasting instructional concepts. The Realistic Program Design (RPD) stimulates flexible use of solution procedures from the beginning by using *realistic* context problems. The Gradual Program Design (GPD) has as its purpose a gradual increase of knowledge through initial emphasis on procedural computation followed by flexible problem solving. We found that whereas RPD pupils showed a more varied use of solution procedures than the GPD pupils, this variation did not influence the procedural competence of the pupils. The empty number line appears to be a very powerful model for the learning of addition and subtraction up to 100.

*Key Words:* Addition, subtraction; Children's strategies; Early childhood, K–4; Early number learning; Mental arithmetic; Representations, modeling; Whole numbers

In the *International Handbook on Mathematics Education* Becker and Selter (1996) gave an overview of recent developments in mathematics education in elementary schools. Their central thesis was that “teaching is no longer seen as a treatment and learning as the effect. Learners are people who actively construct mathematics” (p. 511). Mental arithmetic plays a central role in this respect, stimulating not only conceptual understanding and procedural proficiency but also number sense and the understanding of number relations (McIntosh, Reys, & Reys, 1992). Many researchers in the field of mathematics education hold similar viewpoints (cf. Cobb, 1995; Freudenthal, 1991; Gravemeijer, 1994; Hiebert et al., 1996; Müller & Wittmann, 1995; Resnick, Bill, & Lesgold, 1992; Reys, Reys, Nohda, & Emori, 1995). In the Netherlands these ideas have been incorporated in the theoretical framework of Realistic Mathematics Education (RME) (Freudenthal, 1973, 1991; Gravemeijer, 1994; Treffers, 1987, 1991). RME came into prominence after a national evaluation study of mathematics education in the Netherlands in primary schools (Wijnstra, 1988), which

---

Research reported in this article was supported by the Dutch National Science Foundation Grant 575-90-607.

pointed to an unacceptably low level of procedural competency in certain domains; for example, only 55% of the Dutch third graders were capable of solving the subtraction problem  $64 - 28$  correctly. This study also revealed a generally low level of flexibility in using arithmetic strategies. As a result of this evaluation study, Treffers and De Moor (1990) published a “call for reform” proposing a new lower grades curriculum in which mental arithmetic played a central role during the first and second grades. Mental arithmetic was seen as a foundation for the further development of flexible computation and problem-solving strategies (Treffers, 1991) in which calculating could be done not only “in the head” but also by “using one’s head” in that the use of written work was encouraged. This process did not transform mental arithmetic into written arithmetic; in writing, pupils could display the flexible thought processes that are essential to mental arithmetic (Van den Heuvel-Panhuizen, 1996). Because the RME view is now dominant in Dutch schools and textbooks, procedures for written (column) arithmetic are not introduced until the third grade.

#### MODELS AND SOLUTION PROCEDURES IN THE NUMBER DOMAIN 20–100

Treffers and De Moor’s call for reform (1990) proposed *the empty number line* as a new didactic model for addition and subtraction up to 100. A brief summary of previous models can help elucidate why the empty number line was introduced (Beishuizen, 1993; Gravemeijer, 1994). During the 1960s and 1970s multibase arithmetic blocks and Unifix materials were widely in use. Approaching computation through these materials, however, was criticized because the materials provided a strong conceptual but weak procedural representation of operations on numbers (Resnick, 1982). Therefore, Dutch mathematics books of the 1980s turned to the hundred square to model the number system up to 100. The hundred square is a 10-by-10 square with numbers from 1 through 10 in the first row, 11 through 20 in the second row, and so on up to 100. This model embodied not only relations between numbers but also allowed the visualization of addition and subtraction operations by having children draw arrows or jumps (Beishuizen, 1993).

In our research we found empirical evidence that the previously described models differed in their effects on mental computation procedures (Beishuizen, 1993). Arithmetic blocks evoked decomposition or place-value strategies for addition and subtraction, whereas the hundred square stimulated a sequential pattern of counting by tens. We referred to the first procedure as the “split method” or with the acronym *1010* (pronounced *ten-ten*), because the tens and units were split apart and handled separately. We referred to the second procedure as the “jump method” or with the acronym *N10*, because the *tens* were added to or subtracted from the first unsplit *number*. Table 1 provides examples of the use of 1010 and N10 procedures, together with examples of other procedures as categorized in our research (Beishuizen, 1993; Van Mulken, 1992).

Many authors see N10 and 1010 as the two basic strategies for addition and subtraction up to 100 (Cobb, 1995; Fuson, 1992; Jones, Thornton, & Putt, 1994; Reys et al., 1995; Thompson, 1994). For Dutch children N10 is the more effective computation procedure (Wolters, Beishuizen, Broers, & Knoppert, 1990), whereas 1010 causes more errors, especially in subtraction problems requiring regrouping (Beishuizen, 1993; see also Table 1 for a typical error). The procedure we call *10s* can be seen as an adaptation of the 1010 procedure and was designed to overcome the problems found with the 1010 procedure (Beishuizen, Van Putten, & Van Mulken, 1997).

Table 1  
*Mental Computation Procedures for Addition and Subtraction up to 100*

Addition (with regrouping): 45 + 39	Subtraction (with regrouping): 65 - 49, 51 - 49
Sequential procedures:	Sequential procedures:
N10: 45 + 30 = 75; 75 + 5 = 80; 80 + 4 = 84	N10: 65 - 40 = 25; 25 - 5 = 20; 20 - 4 = 16
N10C: 45 + 40 = 85; 85 - 1 = 84	N10C: 65 - 50 = 15; 15 + 1 = 16
A10: 45 + 5 = 50; 50 + 34 = 84	A10: 65 - 5 = 60; 60 - 40 = 20; 20 - 4 = 16
	A10: 49 + 1 = 50; 50 + 10 = 60; 60 + 5 = 65 answer: 1 + 10 + 5 = 16 (adding-on)
	$\curvearrowright^a$ : 51 - 49 = 2 (because 49 + 2 = 51)
Decomposition procedures:	Decomposition procedures:
1010: 40 + 30 = 70; 5 + 9 = 14; 70 + 14 = 84	1010: 60 - 40 = 20; 5 - 9 = 4 (false reversal) 20 + 4 = 24 (false answer)
10s: 40 + 30 = 70; 70 + 5 = 75; 75 + 9 = 84	10s: 60 - 40 = 20; 20 + 5 = 25; 25 - 9 = 16

<sup>a</sup>The Connecting Arc ( $\curvearrowright$ ) can be used only for subtraction problems.

### *The Empty Number Line as a New Didactic Model*

With growing experience during the 1980s, the hundred square, although providing a better model for N10 than the arithmetic blocks, turned out to be an overly complicated learning aid for weaker pupils (Buys, 1988; Cobb, 1995; Treffers & De Moor, 1990). Moreover, the increasing influence of the RME view in our country ran counter to the prestructured character of the hundred square, which left little room for children's informal strategies. Therefore Treffers and De Moor (1990), in their revision of the Dutch primary mathematics curriculum, devised a new format for the old number line: the empty number line up to 100. RME theory provided several didactical and psychological reasons for using the empty number line as a central model for addition and subtraction. The most important of these reasons, discussed in the next paragraphs, served as the theoretical framework for this study.

1. The empty number line is well-suited to link up with informal solution procedures because of the linear character of the number line. Freudenthal (1973)

suggested that a structured number line with marks for every number was a more natural model of children's informal counting strategies than, for instance, arithmetic blocks with their set-representation of numbers (Gravemeijer, 1994). Gravemeijer has said that in the 1970s, experiments with this structured number line failed because of the unwillingness of students to use it in a global, flexible manner. He reasoned that the structured number line was constituted in measurement situations and so was associated with a rigid ruler with fixed, pre-given distances. This use of the structured number line caused counting and passive reading of the answer on the number line, which did not raise the level of the strategies the pupils used to solve a problem. Therefore, following Whitney (1988), Treffers (Treffers & De Moor, 1990) opted for an empty number line on which the pupils can draw marks for themselves. A structured bead string should be used as an introductory model for the empty number line. This bead string has 100 beads, ordered following the ten structure: 10 red beads followed by 10 white beads followed by 10 red beads and so on (see Figure 1). This structure helps students find a given number and familiarizes children with the positioning of numbers up to 100 and the quantities the numbers represent. The tens can serve as a point of reference in two ways: For example, there are 6 tens in 64 and there are almost 7 tens in 69. After children work with the bead string, the number line can be introduced as a model of the bead string (see Figure 1). By using the empty number line, children can extend their counting strategies and raise the sophistication level of their strategies from counting by ones to counting by tens to counting by multiples of ten (see Figure 1).

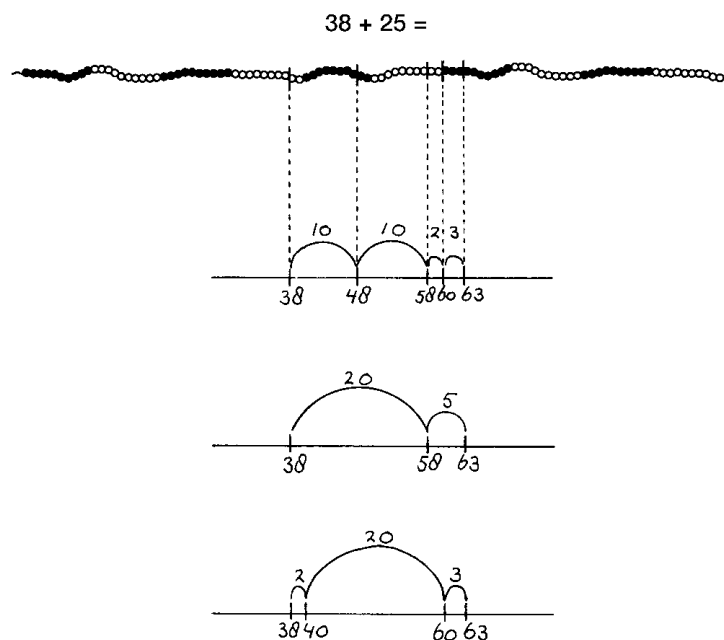


Figure 1. The empty number line as a model of the bead string; the problem  $38 + 25$  is solved in different ways and at different levels.

2. The second reason for using the empty number line is that it provides the opportunity to raise the level of the students' activity (Gravemeijer, 1994). According to the RME view, a model should not only give students freedom to develop their own solution procedures (cf. Selter, 1994), but employing the model should also foster the development of more sophisticated strategies. This progression toward more formal ways of solving a problem is known as the process of *progressive mathematization* and is a key principle in the theory of RME (Freudenthal, 1991; Gravemeijer, 1994; Treffers, 1987, 1991). Another principle of RME is that a model should not only be a *model of a situation* (for instance a context problem) but should also become a *model for representing mathematical solutions* (Gravemeijer, 1994; Streefland, 1991). The empty number line satisfies these requirements; it not only allows students to express and communicate their own solution procedures but also facilitates those solution procedures. Marking the steps on the number line functions as a kind of scaffolding: It shows which part of the operation has been carried out and what remains to be done.

3. The third reason for using the empty number line is its natural and transparent character (Treffers & De Moor, 1990). The empty format stimulates a mental representation of numbers and number operations (addition and subtraction). The model seems to be very suitable for the representation and solution of non-standard context or word problems (Gravemeijer, 1994). The problem representation is more clear and natural on the empty number line than on the hundred square or arithmetic blocks because the row of numbers is not cut off at each ten. Gravemeijer (1994) reported that for subtraction problems with larger numbers, children were in favor of the adding-on strategy. The empty number line is well-suited to making informal solution procedures explicit because of its linear character.

4. Students using the empty number line were cognitively involved in their actions. In contrast, students who use materials such as arithmetic blocks or the hundred square sometimes tend to depend primarily on visualization, which results in a passive "reading off" behavior rather than cognitive involvement in the actions undertaken (Beishuizen, 1993). This lack of involvement is rarely observed with use of the empty number line. Instead, most pupils concurrently solve the computation task within a problem while drawing jumps on the number line. In this way they also keep track of what they are doing, leading to a reduction of the memory load while solving a problem (cf. Baroody, 1987).

## REALISTIC VERSUS GRADUAL PROGRAM DESIGN

We formulated a research proposal to examine and compare two experimental programs in the second grade with the empty number line as a central model for addition and subtraction up to 100: a *Realistic Program Design* and a *Gradual Program Design* (Klein & Beishuizen, 1993). Following Treffers and

De Moor's (1990) curriculum proposal, but also on the basis of other experiences in this field (Cobb, Gravemeijer, Yackel, McClain, & Whitenack, 1995; Klein & Beishuizen, 1993), we decided in both program designs to emphasize N10-like procedures on the number line. The alternative 1010 strategy was introduced in the last months of the second grade only for addition. At that point we used a different base-ten model to underscore the conceptual and procedural differences between the N10 and 1010 strategies. We used, for instance, money (10-guilder notes and 1-guilder coins) to clarify the base-ten structure of numbers. Verbal labels were introduced in both program designs to describe the different solution procedures that were used by the pupils. The labels used were *G* for *Gewoon* (Dutch for *Normal*, meaning N10), *SPV* for *SPring Verder* (Dutch for *Jump Further*, meaning N10C), *S* from *Splitsen* (Dutch for *Splitting both numbers in tens and units*, meaning 1010), and  $\cap$  for the Arc procedure. Table 1 provides explanations for these labels, and Figure 2 shows a worksheet in which some of these labels are introduced. These verbal labels not only facilitated classroom communication about the different solutions but also prompted pupils to first look thoughtfully at a problem before solving it, which is an important issue in RME and in other research on solving



What is your way of solving the problems below?

First, write down the label of the procedure you are planning to use.

Next, write down the steps you used to solve the problem.

	label	Steps
$48 + 25 = 73$	S.	$40 + 20 = 60$ $8 + 5 = 13$ ) $73$
$65 + 19 = 84$	SPV.	$65 + 20 = 85 - 1 = 84$
$95 - 87 = 8$	$\cap$	$95 - 5 = 90 - 3 = 87$
$92 - 69 = 23$	SPV	$92 - 70 = 22 + 1 = 23$

Figure 2. Example of a worksheet. G: N10; S: 1010; SPV: N10C;  $\cap$ : Connecting Arc.

problems (cf. Hiebert et al., 1996). In some cases children were required both to label and to solve problems, and in other cases they were asked only to write down the names of the strategies they would use to solve the problems. To their teachers' surprise, most second graders rather easily learned to use these labels in an adequate way.

### *Realistic Program Design*

The aim of the Realistic Program Design (RPD) was to stimulate flexible use of strategies from the beginning of the program by making connections to children's informal strategies. By "flexible use of arithmetic strategies and computation procedures" we mean choice of the most appropriate and efficient strategy or procedure given the (number) characteristics of the problem at hand (Klein & Beishuizen, 1994). Because the (number) characteristics vary across problems, students have to adjust their strategy use according to the features of the problem. For this reason the students must be able to employ a range of arithmetic strategies and computation procedures among which they can choose flexibly. Flexibility in behavior of the pupils was an important objective of this program. We wanted to make the program manageable by teachers, and at the same time we needed to maintain experimental control in the sense that the same strategies would appear in every classroom in this program. We therefore chose an instructional treatment that differed from the RME and the constructivist ideas about the role of the teacher and how strategy development should take place. We decided that if children did not come up with a particular new strategy after being confronted with an evocative problem, the teacher would introduce this strategy (N10, N10C, A10, or  $\cap$ ; see Table 1). After some practice with a particular strategy, by solving problems on worksheets designed by the researchers, the pupils were free to either use or not use this particular strategy for solving other problems. Compared to RME theory, our experimental RPD was more directive. However, pupils were free to follow their own strategy preferences after practice on strategies introduced to them. In particular, during the last 3 months of the project we emphasized free strategy choice on a variety of problems. Strategy labels and classroom discussion as described earlier were used to stimulate a flexible problem-solving attitude (cf. Hiebert et al., 1996).

Context problems were another important RPD feature. These problems were chosen to reflect everyday experience and to elicit informal and varied solution strategies (Klein & Beishuizen, 1994; Van den Heuvel-Panhuizen, 1996). About one third of each lesson was spent on whole-class instruction and discussion about strategies and context problems. Number sense was stimulated by presenting a variety of exercises using the empty number line and by discussing the different ways of solving a problem. An example of a strategy and problem introduced by the teacher is the so-called Connecting Arc procedure (Treffers & Veltman, 1994), which was introduced as a shortcut strategy for solving subtrac-

tion problems with small differences. Pupils were asked to calculate the difference in price between two toys. In the first problem the toys cost 59 and 61 guilders, respectively. In the second problem two objects cost 36 and 74 guilders, respectively (see Figure 3). It was easier to solve the first problem by “bridging the gap” between 59 and 61 than by subtracting 59 from 61. For this reason a Connecting Arc ( $\cap$ ) was drawn above the 59 and the answer, because  $59 + 2$  equals 61 (see Figure 3). Children became aware of the closeness of the numbers by marking these numbers on the bead string or number line. Similar experiences with the bead string or the empty number line led them to know that the distance between 36 and 74 was much larger than the distance between 59 and 61. It was therefore more efficient to solve the second problem by subtracting 36 from 74.

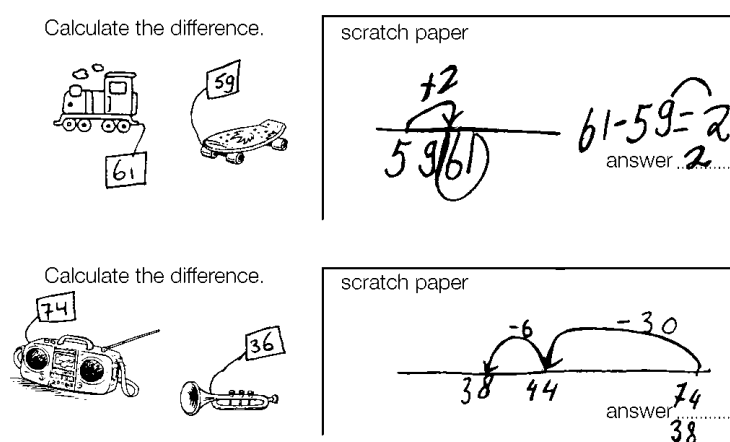


Figure 3. Two pupils solving context difference problems in different ways.

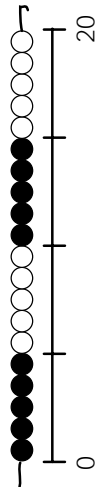
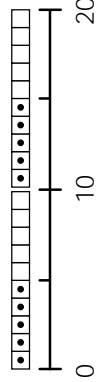
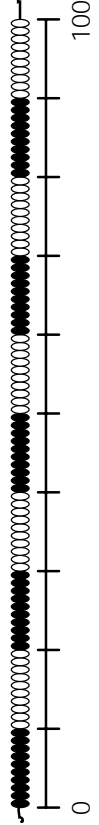
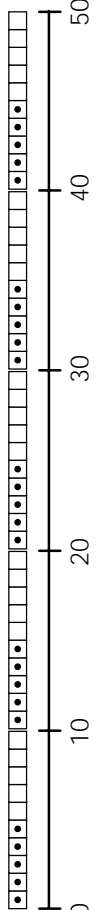
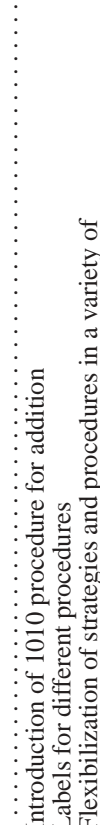
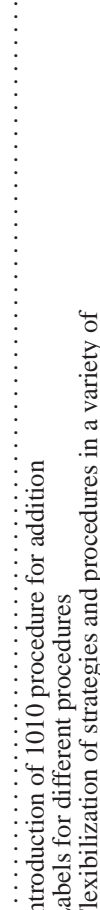
We initiated the RPD at the beginning of the second grade with a bead string with 20 beads following the fives structure (see Table 2). Together with this bead string, a semistructured number line up to 20 was introduced, with marks for the numbers 0, 5, 10, 15, and 20 (see Table 2). After 8 weeks this bead string and number line were replaced by a bead string with 100 beads following the tens structure (see Table 2 and Figure 1) together with a semistructured number line up to 100 with marks for the numbers 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 (see Table 2). Four weeks later the empty number line was introduced.

### *Gradual Program Design*

Many researchers both in our country and abroad agree that instruction based on theories like RME or constructivism appear to be more motivating, exciting, and challenging for children, but perhaps only for average and better students (cf. Ames & Ames, 1989; Gersten & Carnine, 1984; Ruijsenaars, 1994; Van Luit & Van der Rijt, 1996). According to these authors, less capable students would benefit from more structured instruction in



Table 2  
Time Schedule for the Realistic and Gradual Program Designs in the Second Grade

	Realistic Program Design	Gradual Program Design	Tests
Number positioning on bead string as introduction of semistructured number line up to 20	 <p>Sums &lt; 20: 7 + 7, 8 + 7; 14 - 6, 11 - 9 (∩)</p>	 <p>Sums &lt; 20: 7 + 7, 7 + 8; 14 - 6</p>	Arithmetic blocks as introduction of semistructured number line up to 20 .....September
Number positioning on bead string as introduction of semistructured number line up to 100	 <p>Sums &lt; 100: 74 + 8, 93 - 9; 45 + 32, 48 + 36, 51 - 49 *Introduction of the empty number line Various procedures: N10, N10C, A10, ∩ Different strategies: addition, subtraction, adding-on</p>	 <p>Sums &lt; 50: 34 + 8, 43 - 7; 35 + 12, 45 - 23 One procedure: N10 Two strategies: addition, subtraction</p> <p>Sums &lt; 100: 55 + 32, 55 + 37; 85 - 32, 85 - 39 *Introduction of empty number line Various procedures: N10, N10C, A10 Different strategies: addition, subtraction, adding-on</p>	Arithmetic blocks as introduction of semistructured number line up to 50 .....October
Number positioning on bead string as introduction of semistructured number line up to 100	 <p>Introduction of 1010 procedure for addition Labels for different procedures Flexibilization of strategies and procedures in a variety of (context) problems</p>	 <p>Introduction of 1010 procedure for addition Labels for different procedures Flexibilization of strategies and procedures in a variety of (context) problems</p>	.....April

Note. Tests administered in the first and third grade are not mentioned in this schedule; see explanation in text.

which the teacher helped them construct their strategies to solve problems. For this reason these authors have proposed that the early introduction of multiple strategies, as is done in the RPD, could confuse weaker students. Because students are not separated by ability or achievement in Dutch early grades, it seemed appropriate to try out an instructional design that might help these students but that would not have adverse effects on the performance of moderate and better students. Thus, the Gradual Program Design (GPD) might result in better performance for an entire class (in a follow-up study we will look specifically at the learning of weaker and stronger students).

To investigate this assumption we compared the RPD with the more structured GPD. The GPD follows a psychological conceptualization of stagewise knowledge development (cf. Glaser & Bassok, 1989). The sizes of the numbers used increased more gradually over time than in the RPD (cf. Table 2) and addition and subtraction problems that require “passing a ten” (for instance  $48 + 36$ ,  $51 - 49$ ) were introduced later (cf. Table 2). Also, the number line was introduced differently: The RPD used the bead string whereas the GPD used arithmetic blocks in a linear way to introduce the number line (cf. Table 2). From September until January the number line in the GPD went only to 50 and remained semistructured, first with marks for the numbers 0, 5, 10, 15, and 20 and in October with marks for the numbers 0, 10, 20, 30, 40, and 50 (cf. Table 2). As a consequence of this instructional sequence the empty number line was introduced at a later time in the GPD than in the RPD. With respect to the strategies, at first only the N10 procedure was practiced in order to establish a firm procedural knowledge base. Because of the risk of confusing the weaker GPD pupils, we decided that the Connecting Arc procedure would not be introduced in the GPD. During the first 5 month, less time was spent on whole-classroom discussions than in the RPD. Instead, the children in GPD spent more time on procedural paper-and-pencil exercises. Other strategies (N10C, A10) were introduced later, and flexible strategy use was encouraged during the last 3 months of the GPD. In accordance with the stagewise concept of development, more time was needed before children became sensitive to number characteristics and possible shortcut strategies (cf. Baroody & Ginsburg, 1986).

Context problems were less prominent in the GPD than in the RPD. From the beginning of the program GPD pupils were confronted with context problems, but they were presented as applications after the procedures had been learned. Context problems were not used to introduce varied strategies as was done in the RPD. To provide the reader with an understanding of when different problem types, various computation procedures, and the empty number line were introduced in both program designs, a time schedule is given in Table 2.

### *Research Questions*

In order to evaluate the effects and to test the different characteristics of the Realistic and Gradual Program Designs, we formulated hypotheses about test results with respect to procedural and strategic knowledge:

1. We expected that the level of procedural competence would be higher for the GPD pupils than for the RPD pupils. The greater emphasis on proceduralization in the Gradual Program Design was expected to result in a larger number of correctly solved problems. The greater emphasis on different strategies and procedures in the Realistic Program Design was expected to cause more (procedural) confusion for many RPD pupils and therefore result in a smaller number of correctly solved problems.

2. Because of the stronger emphasis on flexible use of strategies in the Realistic Program Design, we expected that halfway through the curriculum the RPD pupils would show a more varied use of solution strategies and computation procedures than the GPD pupils. However, we expected that this difference would disappear by the end of the second grade, because the Gradual Program Design also emphasized flexible strategy use in the last part of the curriculum.

3. Finally, we were interested in how both GPD and RPD pupils would perform on a National Arithmetic Test administered at the end of the second grade. In particular we were interested in how they would perform on problems like  $64 - 28$ , which was correctly solved by only 55% of Dutch third-grade pupils in 1987 (Wijnstra, 1988).

## METHOD

### *Participants*

After 2 years of small-scale curriculum development, try-outs, and revisions, the experimental programs were implemented during the 1994–95 school year in 10 second-grade classes at nine comparable middle-class primary schools in the Netherlands ( $N = 275$ ). To reduce the possibility of differences in arithmetic competence at the beginning of our experiment we administered the National Arithmetic Test at the end of the first grade. Classes with comparable results on this test were matched in five pairs. Within each pair the classes were randomly assigned to the RPD or GPD. There appeared to be no significant differences between the two groups in arithmetic test scores at the start of the experiment.

### *Materials and Procedure*

The second-grade teachers and pupils used experimental lesson materials and teacher guides instead of their regular mathematics textbooks. Experimental materials concerning addition and subtraction replaced about 75% of the regular textbook, *Rekenen & Wiskunde*, which is based on the ideas of RME (Gravemeijer, Van Galen, Kraemer, Meeuwisse, & Vermeulen, 1983). The regular text (25%) was used for instruction on measurement, tables of multiplication, spatial ordering, and telling time. Every fortnight the teachers discussed their experiences with one of the researchers during a visit to the schools.

The development of procedural competence in arithmetic skills was measured with the *Arithmetic Speed Test (AST)* (Klein & Beishuizen, 1995b) administered

five times during the year (see time schedule of tests in Table 2). The pupils had to solve as many number problems as possible within 3 minutes. The test consisted of addition and subtraction exercises with regrouping. We used different number sizes ( $< 20$ ,  $< 50$ ,  $< 100$ ) and categorized problems as adding or subtracting single-digit (SD) numbers ( $8 + 5$ ,  $36 + 6$ ,  $65 + 9$ ;  $12 - 4$ ,  $43 - 6$ ,  $76 - 9$ ) or multidigit (MD) numbers ( $27 + 14$ ,  $57 + 19$ ;  $44 - 26$ ,  $85 - 49$ ). Multidigit number problems were administered only in April and June. Because of space limitations we will report the results for addition and subtraction with single-digit numbers only up to 20 (and not those up to 50 and 100) and for multidigit numbers up to 100 as reported for the April and June testing. These are, respectively, the easiest and the most difficult problems. Because we performed several independent statistical tests on the same data, we accepted only a  $p$ -value of less than .01 as a significant difference.

In the *Arithmetic Scratch-Paper Test (ASPT)* (Klein & Beishuizen, 1995a) administered in April and again in June, the pupils were asked to write down their solution steps in “scratch-paper boxes” (see Figure 3), which appeared beside their answers to the problems, so we could analyze their computation procedures and strategy use. The test reported here consisted of 21 problems. Three addition and five subtraction problem types (i.e., problems selected to elicit certain procedures) were presented with comparable numbers in two formats: as numerical expressions and as context problems. For the context problems we chose problems of the *change* type (e.g., “Marieke has 81 marbles. She loses 79 of them. How many does she have left?”) (Klein & Beishuizen, 1994; Riley, Greeno, & Heller, 1983; Verschaffel & DeCorte, 1990). The third problem type dealt with numbers comparable to those in the subtraction problems, except that the pupils now had to calculate the difference between the two numbers (e.g., difference in weight or price, as illustrated in Figure 3). Five of these *difference* problems were offered in context format. The 21 problems were presented to the pupils in random order to control for set-effects. Table 3 provides an overview of the problem types.

Table 3  
*Addition, Subtraction, and Difference Problem Types and Expected Procedures in ASPT*

Addition	Expected procedure	Subtraction	Expected procedure	Difference	Expected procedure
NE <sup>a</sup> $57 + 36 =$	N10	NE $75 - 36 =$	N10	C 74 and 36	N10 or A10
C <sup>b</sup> $48 + 37 =$	N10	C $84 - 26 =$	N10		
NE $42 + 43 =$	1010	NE $65 - 33 =$	N10	C 65 and 32	N10 or A10
C $33 + 34 =$	1010	C $85 - 42 =$	N10		
NE $54 + 39 =$	N10C	NE $84 - 29 =$	N10C	C 73 and 29	N10C or A10
C $54 + 29 =$	N10C	C $63 - 29 =$	N10C		
		NE $71 - 69 =$	$\cap^c$	C 61 and 59	$\cap^c$
		C $81 - 79 =$	$\cap^c$		
		NE $62 - 48 =$	$\cap^c$	C 82 and 68	$\cap^c$
		C $72 - 58 =$	$\cap^c$		

*Note.* For the context format, only the numbers and not the story and picture of the problem are represented.

<sup>a</sup>NE stands for numerical expression. <sup>b</sup>C stands for context. <sup>c</sup>The Connecting Arc ( $\cap$ ) is expected only for subtraction and difference problems and only for RPD pupils.

The numbers used in the problems were chosen to elicit specific computation procedures. For the addition problems we expected N10, 1010, and N10C procedures (see also Table 1). For the subtraction problems we expected the N10 and N10C procedures. The Connecting Arc ( $\cap$ ) was expected for the last four items (only for RPD pupils) because the differences between the numbers of these subtraction problems are small, and, therefore, bridging the gap between these numbers is more efficient than subtracting the second from the first. For subtraction problems with larger differences (like  $73 - 29$ ), subtracting the second number from the first is more efficient than bridging the gap. Therefore, we did not expect the Connecting Arc for these problems. We did not expect 1010 for subtraction problems because this procedure was introduced only for addition problems (see Table 2). For the difference problems we expected the N10, N10C, and A10 procedures, and for the last two items we expected, again, the Connecting Arc ( $\cap$ ) (only for the RPD pupils). The tests were scored according to the categories of procedures and strategies shown in Table 1.

As an external criterion, the earlier mentioned *National Arithmetic Test* (Janssen, Bokhove, & Kraemer, 1992) was administered not only at the end of the first grade but also at the end of the second grade. As explained before, the scores at the end of the first grade were used as a matching criterion. The scores of both groups on the subtraction problems like  $64 - 28$  are reported here and compared to the results of the national evaluation study described earlier (Wijnstra, 1988).

## RESULTS

### *Procedural Competence*

Paired *t*-tests for the *AST* revealed significant increases in the number of correct answers on each test from September to June for all problem types (see Figure 4). A one-way ANOVA revealed a significant difference in October for addition with single-digit numbers,  $F(1, 265) = 6.9, p < .01$ ; the GPD pupils produced more correct answers than the RPD pupils. In January a significant difference was found for subtraction with single-digit numbers,  $F(1, 260) = 7.1, p < .01$ ; the RPD pupils produced higher scores than the GPD pupils. On the other three occasions there were no significant differences between the two experimental programs on evaluating single-digit numerical expressions. Neither were any significant differences found between the two program designs for multidigit items up to 100. See Klein (1998) for more results from the speed tests.

### *Variation in Use of Solution Procedures and Strategies*

Overall there was a fairly high percentage of correct answers on addition, subtraction, and difference problems on the *ASPT* both in April and June: about 80% for both GPD and RPD pupils. There were no significant differences found

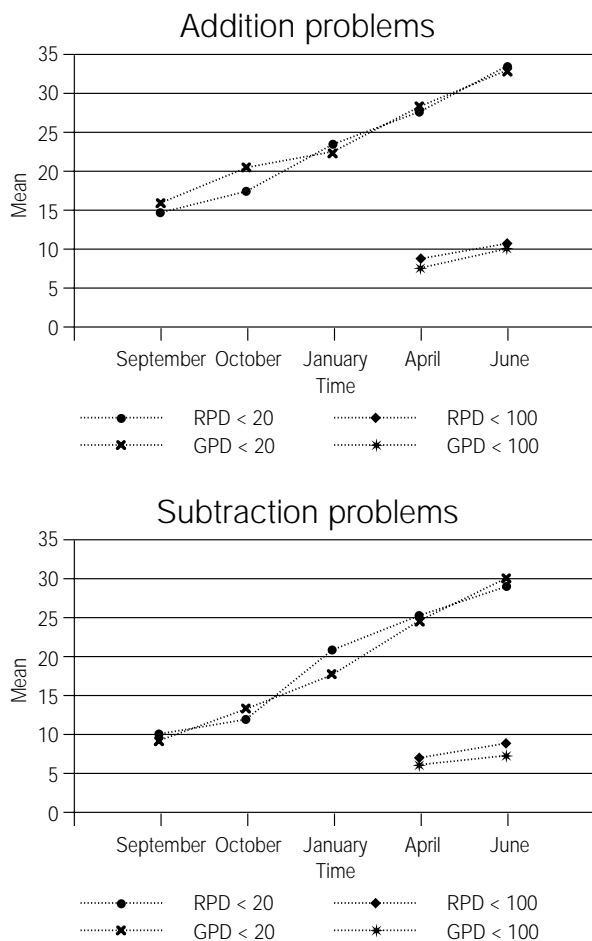


Figure 4. Number of correct answers on the AST on different measurements for RPD pupils ( $n = 139$ ) and GPD pupils ( $n = 136$ ).

between the two program designs. In Klein (1998) a more detailed analysis of *ASPT* results can be found.

Figure 5 shows the solution procedures RPD and GPD pupils used to solve addition problems of the numerical expression (NE) and context (C) types on the *ASPT* in April and June. (The left bar represents the RPD pupils, the right bar the GPD pupils.) Both in April and in June the RPD pupils chose the most efficient procedure for the problem at hand. They switched from the N10 procedure they had used for the  $57 + 36$  problem to the more efficient N10C procedure for the  $54 + 39$  problem. Most of the GPD pupils continued to use the N10 procedure exclusively. The difference in presentation of the problem, whether numerical expression or context, did not seem to influence the pupils' choices of solution procedures. In June we saw increased use of the 1010 procedure by all pupils. This increased use of the 1010 procedure was caused by the introduction of this computation procedure for addition problems in both curricula between April and June (see Table 2).

Figure 6 shows the solution procedures used by the RPD and GPD pupils to solve subtraction numerical expression and context problems in April and June

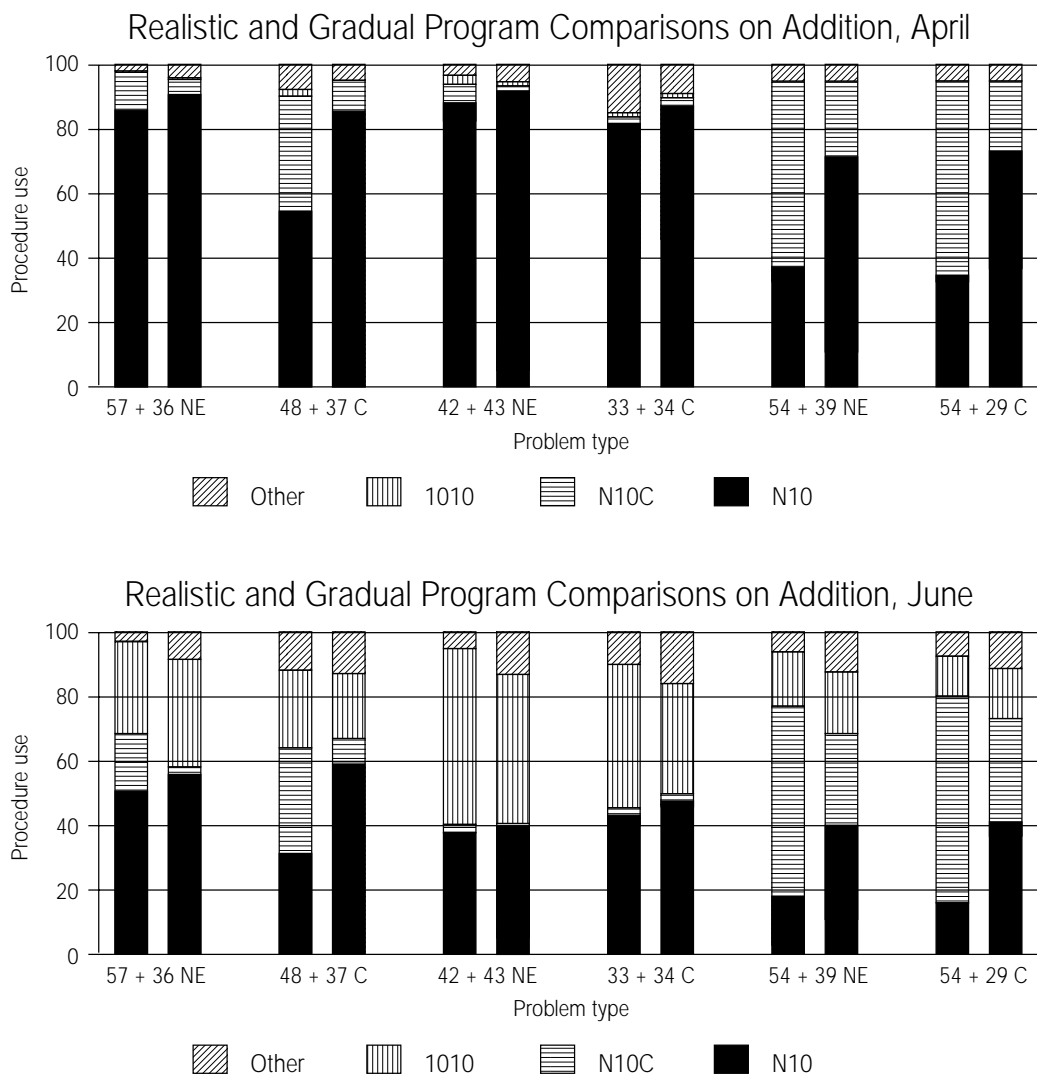


Figure 5. Procedure use on numerical expression (NE) and context (C) addition problems in the April and June ASPT for RPD pupils (left bar,  $n = 139$ ) and GPD pupils (right bar,  $n = 135$ ).

(left bar RPD, right bar GPD). The subtraction problems  $62 - 48$  (NE),  $72 - 58$  (C),  $65 - 33$  (NE), and  $85 - 42$  (C) are not included in Figure 6 because their inclusion would make the figure overly complicated. Their pattern of procedure use resembles the pattern of procedure use for the subtraction problems  $75 - 36$  (NE) and  $84 - 26$  (C). Although there are some changes in procedure use from April to June (especially the use of N10C), the GPD pupils used the N10 procedure much more frequently than the RPD pupils to solve the subtraction problems. The RPD pupils changed their use of procedures according to the characteristics of the problem; for example, they used N10C for the  $84 - 29$  problem, whereas for the problems  $71 - 69$  and  $81 - 79$ , they changed to the Connecting Arc procedure. The Connecting Arc was not introduced to the GPD pupils (see Table 2). However, for a problem with a small difference ( $71 - 69$  or  $81 - 79$ )

about 10% to 20% of the GPD pupils solved the problem by making one jump between the two numbers, probably because they saw the small difference between the two numbers on a mental representation of the empty number line. This solution procedure is categorized as *other*. Just as for the addition problems, the presentation format of numerical expression versus context did not seem to influence the procedures chosen to solve the subtraction problems.

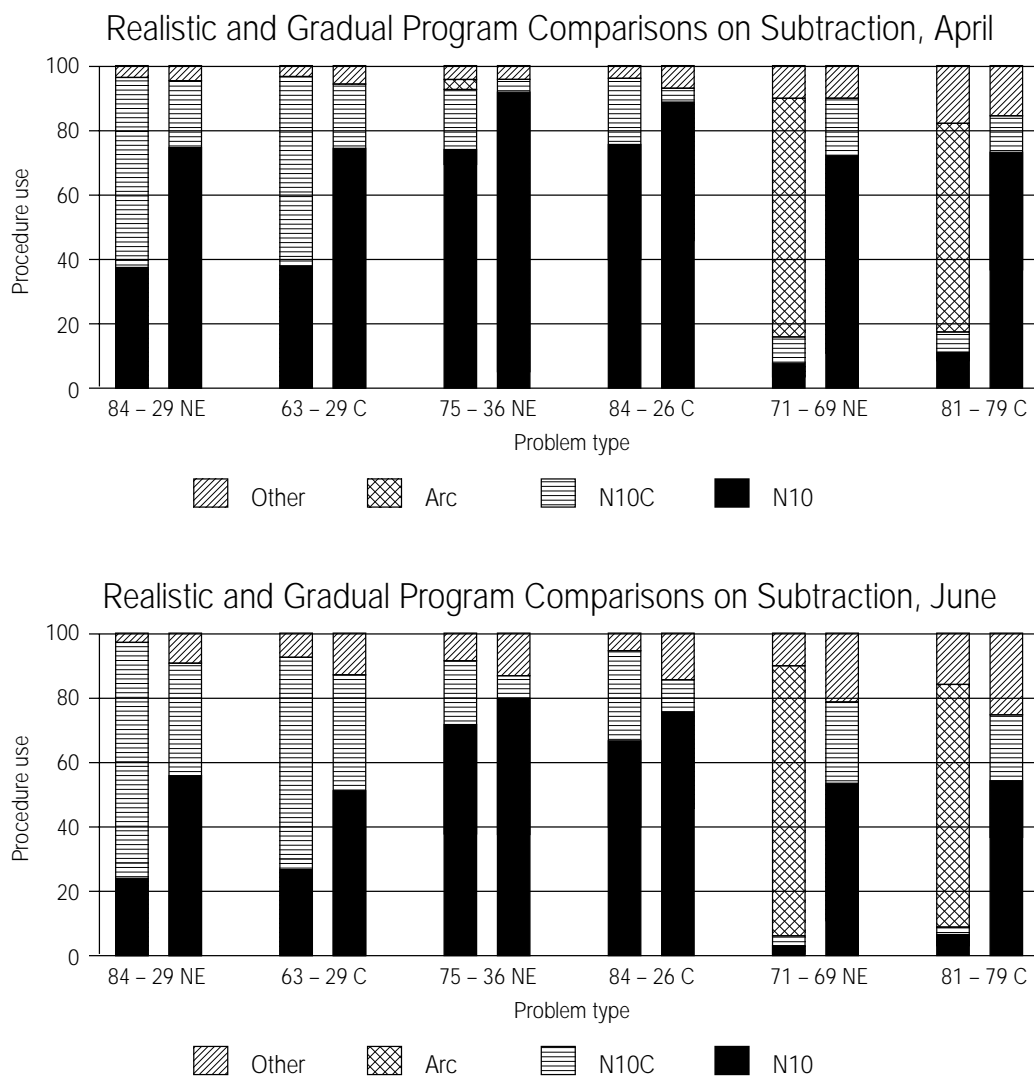


Figure 6. Procedure use on numerical expression (NE) and context (C) subtraction problems in the April and June ASPT for RPD pupils (left bar,  $n = 139$ ) and GPD pupils (right bar,  $n = 135$ ).

Figure 7 shows the solution procedures RPD and GPD pupils used in April and June for solving context problems in which they had to calculate a difference between two numbers. The procedures used by the RPD and GPD pupils



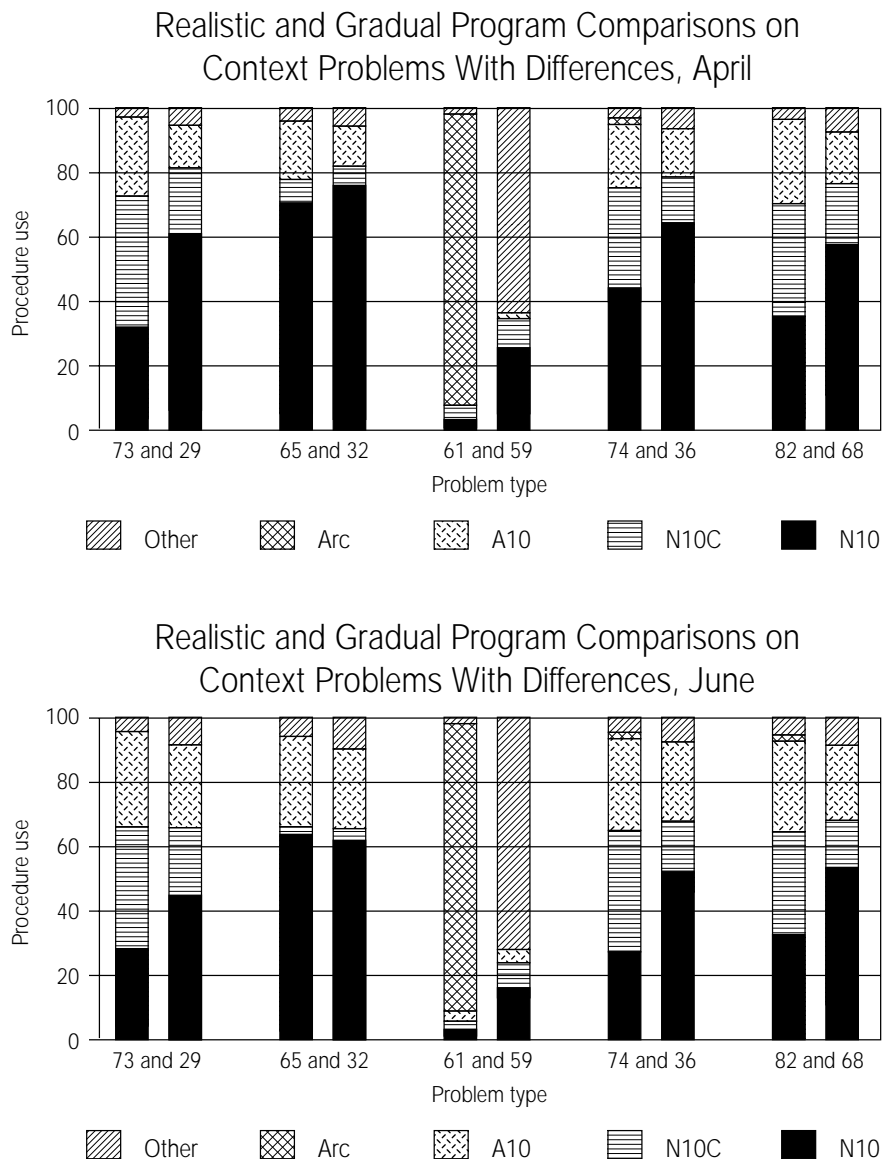


Figure 7. Procedure use on context difference problems in the April and June ASPT for RPD pupils (left bar,  $n = 139$ ) and GPD pupils (right bar,  $n = 135$ ).

for the difference problems were more or less the same. Comparison of the results of the difference problems with the results of the subtraction problems of the change type reveals an increase in use of the A10 procedure (see Table 1) for the difference problems. Also for the GPD pupils the use of the A10 procedure increased from April to June. This use of the A10 procedure is one of the big differences between subtraction and difference problems. The A10 procedure was used primarily as an adding-on strategy (see also second example of the A10 procedure in Table 1). Pupils started adding-on from the smaller to the larger number to calculate the difference between the two numbers. Most RPD

pupils used the Connecting Arc for calculating the difference between 61 and 59. About 25% of the GPD pupils also solved this problem by making one jump from 59 to 61, probably because they saw the small difference between the two numbers on a mental representation of the empty number line. This strategy is again categorized as *other*.

### *External Criterion Test*

The results on the National Arithmetic Test at the end of Grade 2 showed that problems like  $64 - 28$  were correctly solved by 78% of both the RPD and GPD pupils. Compared to the 55% correct for the Dutch third graders, reported in 1988 by Wijnstra, this is a remarkable increase.

## DISCUSSION

### *Procedural Competence*

It was predicted that the GPD pupils would show a higher level of procedural competence than the RPD pupils during the first part of the year but that the RPD pupils could catch up during the second part of the year. As it turned out, there were almost no differences in procedural competence between the two groups of pupils. When significant differences were found, they were mostly in favor of the RPD pupils, especially in the case of subtraction problems. So our first hypothesis was not confirmed. It seems that early introduction of different solution strategies (like N10C and Connecting Arc) did not harm the procedural competence of the RPD pupils. On the contrary, their scores on the *AST* showed a greater increase from October to January than did the scores of the GPD pupils (Figure 4). It seems that stimulating different solution strategies and procedures on the empty number line, as was done in the Realistic Program Design, had a comparable or even larger effect on procedural competence than the greater emphasis on procedural (N10) training in the Gradual Program Design.

### *Flexibility in Using Computation Procedures and Strategies*

In accordance with the difference in emphasis on promoting flexibility of computation procedures in the two program designs, we expected to find more varied use of solution strategies and computation procedures for the RPD pupils halfway through the curriculum. This difference was expected to disappear by the end of the school year. This second hypothesis was partly supported. In April the RPD pupils chose the most efficient procedure from several possible solution procedures. They were especially likely to choose the N10C procedure and the Connecting Arc procedure. The GPD pupils showed a greater tendency to stick to the N10 procedure, and they solved almost every problem in the same way. The June data did not support our second hypothesis. Although the GPD pupils showed increased variation in solution procedures, they still lagged far behind the RPD pupils in flexibility.

With respect to the didactic sequence of the GPD, with its postponed introduction of different solution procedures and strategies, it seems that this sequence had less impact on flexibility than the RPD had. On the one hand, the GPD pupils were already using the N10 procedure successfully and were less sensitive to other, in some cases more efficient, solution procedures like N10C. The RPD pupils, on the other hand, had been confronted with these different procedures in November and January and kept on using them in April and June. This use of different procedures did not, however, result in more incorrect answers. Test performance of both GPD and RPD pupils demonstrated a high level of about 80% correct. The brief introduction of numbers up to 100 and the addition and subtraction problems requiring regrouping of the tens for the RPD pupils (see Table 2) did not create the confusion we expected for many weaker pupils. We saw some temporary confusion in the pupils' worksheets, but this confusion did not persist during the tests (Klein, 1998). To summarize, the RPD pupils attained and sustained a higher level of flexible problem solving than did the GPD pupils. The results provide empirical evidence in support of the RME theory and the empty number line, as stated in the introductory section (Gravemeijer, 1994; Treffers, 1991). The different effects of the Realistic and Gradual Program designs underline that not only the modeling function of the empty number line but also its *more realistic* use contributed to the successful results.

### *The Empty Number Line*

At the end of Grade 2 the success of both RPD and GPD pupils on the difficult subtraction problems in the National Arithmetic Test confirmed that the empty number line is a powerful model for instruction. Pupils' work as well as incidental classroom observations and teachers' experiences provided the following additional clues for the interpretation of its success: During the first half-year in both programs the modeling function of the number line supported both procedural operations and problem representation. Evoking children's own mental activity was also a significant function, and this activity became stronger during the second half-year. In this latter period teachers using the Realistic Program Design, more than those using the Gradual Program Design, created a classroom climate of interactive teaching and discussion about children's different solutions. Students' verbal labeling of strategies and procedures proved to be very useful in this climate. Raising the level of the students' activities from their using the number-line model to their carrying out mental solution steps (cf. Figure 2) occurred more frequently, for instance, in the children's productions on successive tests.

Additional analysis and discussion of these data supported the conclusion that the RPD pupils made better use than the GPD pupils of the different functions of the empty number line (Klein, 1998). In particular, the RPD holds promise for future instruction, and all experimental schools have volunteered to continue

with the empty-number-line program. More fine-grained analyses of the weaker pupils' progress supported the positive outcomes for use of the RPD with these students (Klein, 1998). These results gave us a more complete view of the implications of our research for the learning of addition and subtraction up to 100 in primary school.

## REFERENCES

- Ames, C., & Ames, R. (1989). *Research on motivation in education: Goals and cognitions*. San Diego, CA: Academic Press.
- Baroody, A. J. (1987). *Children's mathematical thinking: A developmental framework for preschool, primary, and special education teachers*. New York: Teachers College Press.
- Baroody, A. J., & Ginsburg, H. P. (1986). The relationship between initial meaningful and mechanical knowledge of arithmetic. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 75–112). Hillsdale, NJ: Erlbaum.
- Becker, J. P., & Selter, C. (1996). Elementary school practices. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook on mathematics education* (pp. 511–564). Dordrecht, The Netherlands: Kluwer.
- Beishuizen, M. (1993). Mental strategies and materials or models for addition and subtraction up to 100 in Dutch second grades. *Journal for Research in Mathematics Education*, 24, 294–323.
- Beishuizen, M., Van Putten, C. M., & Van Mulken, F. (1997). Mental arithmetic and strategy use with indirect number problems up to one hundred. *Learning and Instruction*, 7, 87–106.
- Buys, K. (1988). Schaduwezijden van het honderdveld [Shady sides of the hundred square]. *Tijdschrift voor Nascholing en Onderzoek van het Reken-Wiskunde Onderwijs*, 7(4), 3–10.
- Cobb, P. (1995). Cultural tools and mathematical learning: A case study. *Journal for Research in Mathematics Education*, 26, 362–385.
- Cobb, P., Gravemeijer, K., Yackel, E., McClain, K., & Whitenack, J. (1995, September). *Mathematizing and symbolizing: The emergence of chains of signification in one first-grade classroom*. Paper presented at the sixth conference of the European Association for Research on Learning and Instruction, Nijmegen, The Netherlands.
- Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht, The Netherlands: Reidel.
- Freudenthal, H. (1991). *Revisiting mathematics education*. Dordrecht, The Netherlands: Kluwer Academic.
- Fuson, K. C. (1992). Research on whole number addition and subtraction. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 243–275). New York: Macmillan.
- Gersten, R., & Carnine, D. (1984). Direct instruction mathematics: A longitudinal evaluation of low-income elementary school students. *The Elementary School Journal*, 84, 395–407.
- Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology*, 40, 631–666.
- Gravemeijer, K. (1994). Educational development and developmental research in mathematics education. *Journal for Research in Mathematics Education*, 25, 443–471.
- Gravemeijer, K., Van Galen, F., Kraemer, J. M., Meeuwisse, T., & Vermeulen, W. (1983). *Rekenen & wiskunde [Arithmetic & mathematics]*. Baarn, The Netherlands: Bekadidact.
- Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Human, P., Murray, H., Olivier, A., & Wearne, D. (1996). Problem solving as a basis for reform in curriculum and instruction: The case of mathematics. *Educational Researcher*, 25(4), 12–21.
- Janssen, J., Bokhove, J., & Kraemer, J. -M. (1992). *Leerlingvolgsysteem rekenen-wiskunde 1 [Student monitoring system arithmetic-mathematics 1]*. Arnhem, The Netherlands: National Institute for Educational Measurement.
- Jones, G. A., Thornton, C. A., & Putt, I. J. (1994). A model for nurturing and assessing multidigit number sense among first-grade children. *Educational Studies in Mathematics*, 27, 117–143.
- Klein, A. S. (1995, June). *Een (aangepast) model voor het leren optellen en aftrekken tot honderd*

- [A (revised) model for addition and subtraction up till hundred]. Paper presented at the annual meeting of the Dutch Educational Research Association, Groningen, The Netherlands.
- Klein, A. S. (1998). *Flexibilization of mental arithmetic strategies on a different knowledge base: The empty number line in a realistic versus gradual program design* (Doctoral dissertation, Leiden University). Utrecht, The Netherlands: Freudenthal Institute.
- Klein, A. S., & Beishuizen, M. (1993, September). *New learning routes for addition and subtraction up to one hundred in the Dutch primary math curriculum*. Poster presented at the fifth conference of the European Association for Research on Learning and Instruction, Aix-en-Provence, France.
- Klein, A. S., & Beishuizen, M. (1995a). *Stapjessommen toets 1995* [Arithmetic Scratch-Paper Test 1995]. Unpublished test, Leiden University, The Netherlands.
- Klein, A. S., & Beishuizen, M. (1995b). *Tempotoets rekenen 1995* [Arithmetic Speed Test 1995]. Unpublished test, Leiden University, The Netherlands.
- Klein, T., & Beishuizen, M. (1994). Assessment of flexibility in mental arithmetic. In J. E. H. van Luit (Ed.), *Research on learning and instruction of mathematics in kindergarten and primary school* (pp. 125–152). Doetinchem, The Netherlands: Graviant.
- McIntosh, A., Reys, B. J., & Reys, R. E. (1992). A proposed framework for examining basic number sense. *For the Learning of Mathematics*, 12(3), 2–8, 44.
- Müller, G., & Wittmann, E. C. (1995). *Das Zahlenbuch* [The numberbook]. Stuttgart, Germany: Klett.
- Resnick, L. B. (1982). Syntax and semantics in learning to subtract. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 136–155). Hillsdale, NJ: Erlbaum.
- Resnick, L. B., Bill, V., & Lesgold, S. (1992). Developing thinking abilities in arithmetic class. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development: Implications and applications for education* (pp. 210–230). London: Routledge.
- Reys, R. E., Reys, B. J., Nohda, N., & Emori, H. (1995). Mental computation performance and strategy use of Japanese students in Grades 2, 4, 6, and 8. *Journal for Research in Mathematics Education*, 26, 304–326.
- Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). New York: Academic Press.
- Ruijsenaars, A. J. J. M. (1994). Speciaal rekenen [Special arithmetic]. In M. Dolk, H. van Luit & E. te Woerd (Eds.), *Speciaal rekenen* (pp. 29–41). Utrecht, The Netherlands: Freudenthal Institute.
- Selter, C. (1994). *Eigenproduktionen im Arithmetikunterricht der Primarstufe* [Children's own productions in primary mathematics teaching]. Wiesbaden, Germany: Deutscher Universitäts-Verlag.
- Streefland, L. (1991). *Fractions in realistic mathematics education: A paradigm for developmental research*. Dordrecht, The Netherlands: Kluwer.
- Thompson, I. (1994). Young children's idiosyncratic written algorithms for addition. *Educational Studies in Mathematics*, 26, 323–345.
- Treffers, A. (1987). *Three dimensions. A model of goal and theory description in mathematics education: The Wiskobas Project*. Dordrecht, The Netherlands: Reidel.
- Treffers, A. (1991). Didactical background of a mathematics program for primary education. In L. Streefland (Ed.), *Realistic mathematics education in primary school* (pp. 21–56). Utrecht, The Netherlands: Freudenthal Institute, Utrecht University.
- Treffers, A., & De Moor, E. (1990). *Proeve van een nationaal programma van het reken-wiskunde-onderwijs op de basisschool. Deel 2: Basisvaardigheden en cijferen* [Specimen of a national program for primary mathematics teaching. Part 2: Basic mental strategies and written computation]. Tilburg, The Netherlands: Zwijzen.
- Treffers, A., & Veltman, A. (1994). Relatie-boog als brug tussen bewerkingen [Relation-arc as a bridge between computations]. *Tijdschrift voor Nascholing van het Reken-Wiskunde Onderwijs*, 12, 3, 11–14.

- Van den Heuvel-Panhuizen, M. (1996). *Assessment and realistic mathematics education* (Doctoral dissertation, Utrecht University). Utrecht, The Netherlands: Centrum voor Didactiek van de  $\beta$ -Wetenschappen.
- Van Luit, J. E. H., & Van de Rijt, B. A. M. (1996, December). *Effectiveness of the AEM program for teaching children early mathematics*. Paper presented at Experts Meeting, Leiden, The Netherlands.
- Van Mulken, F. (1992). *Hoofdrekenen en strategisch handelen—Twee grondvormen van optellen en aftrekken tot honderd* [Mental arithmetic and strategic action—Two basic forms of addition and subtraction up to hundred]. Doctoral dissertation, Leiden University, Leiden, The Netherlands.
- Verschaffel, L., & DeCorte, E. (1990). Do non-semantic factors also influence the solution process of addition and subtraction word problems? In H. Mandl, E. DeCorte, N. Bennett, & H. E. Friedrich (Eds.), *Learning and instruction; European research in an international context: Vol. 2.2. Analysis of complex skills and complex knowledge domains* (pp. 415–429). Oxford, England: Pergamon Press.
- Whitney, H. (1988). *Mathematical reasoning, early grades*. Unpublished manuscript, Princeton University, Princeton, NJ.
- Wijnstra, J. M. (Ed.). (1988). *Balans van het rekenonderwijs in de basisschool: Uitkomsten van de eerste rekenpeiling medio en einde basisonderwijs* [Mathematics in primary education: First results of the Dutch national assessment program in primary education in Grades 5 and 8]. Arnhem, The Netherlands: National Institute for Educational Measurement.
- Wolters, G., Beishuizen, M., Broers, G., & Knoppert, W. (1990). Mental arithmetic: Effects of calculation procedure and problem difficulty on solution latency. *Journal of Experimental Child Psychology*, 49, 20–30.

## Authors

---

**Anton S. Klein**, Research Associate, Centre for the Study of Education and Instruction, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands

**Meindert Beishuizen**, Associate Professor of Curriculum Development, Centre for the Study of Education and Instruction, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands; beishuizen@rulfsw.fsw.leidenuniv.nl

**Adri Treffers**, Professor of Mathematics Education, Freudenthal Institute, Utrecht University, Tiberdreef 4, 3561 GG Utrecht, The Netherlands