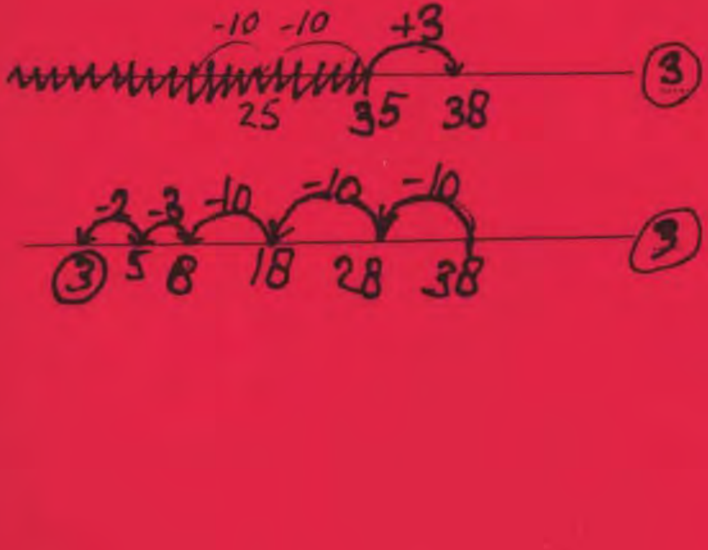
A.S. Klein

Flexibilization of mental arithmetic strategies on a different knowledge base



Flexibilization of mental arithmetic strategies on a different knowledge base A.S. Klein

29

Flexibilization of mental arithmetic strategies on a different knowledge base: the empty number line in a realistic versus gradual program design

CD-β SERIES ON RESEARCH AND MATHEMATICS EDUCATION

1 Didactiek in Perspectief – P.L. Lijnse, W. de Vos, eds.
2 Radiation and Risk in Physics Education – H.M.C. Eijkelhof
3 Natuurkunde-onderwijs tussen Leefwereld en Vakstructuur – R.F.A. Wierstra
4 Een Onverdeelbare Eenheid – M.J. Vogelezang
5 Betrokken bij Evenwicht – J.H. van Driel
6 Relating Macroscopic Phenomena to Microscopic particles: A Central Problem in Secondary Science Education – P.L. Lijnse, P. Licht, W. de Vos, A.J. Waarlo, eds.
7 Kwaliteit van Kwantiteit – H.E. Elzenga
8 Interactieve Video in de Nascholing Rekenen– wiskunde – F. van Galen, M. Dolk, E. Feijs, V. Jonker, N. Ruesink, W. Uittenbogaard
9 Realistic Mathematics Education in Primary Schools – L. Streefland, ed.
10 Ontwikkeling in Energieonderwijs – A.E. van der Valk
11 Methoden in het Reken-wiskundeonderwijs – K. Gravemeijer, M. van den Heuvel-Panhuizen, G. van Donselaar, N. Ruesink, L. Streefland, W. Vermeulen, E. te Woerd, D. van de Ploeg
12 De Volgende Opgave van de Computer – J. Zuidema en L. van der Gaag
13 European Research in Science Education – P.L. Lijnse, ed.
14 Realistic Mathematics Education – K. Gravemeijer
15 De Grafische Rekenmachine in het Wiskundeonderwijs – L.M. Doorman, P. Drijvers, M.Kindt
16 Making sense – Simulation of Research in Organic Chemistry Education – H. van Keulen
17 Perspectives on Research in Chemical Education – O. de Jong, P.H. van Roon, W. de Vos, eds.
18 A Problem – Posing Approach to Teaching the Topic of Radioactivity – C.W.J.M. Klaassen
19 Assessment and Realistic Mathematics Education – M. van den Heuvel– Panhuizen
20 Teaching structures in chemistry – An Educational Structure for Chemical Bonding – G.M. van Hoeve-Brouwer

21 Regulatie en homeostase als onderwijsthema: een biologie–didactisch onderzoek – J. Buddingh

22 Over Natuurkundedidactiek, Curriculumontwikkeling en Lerarenopleiding – P.L. Lijnse en T. Wubbels

23 Integratie en Toepassing van Biologische Kennis – Ontwikkeling en Onderzoek van een Curriculum rond het thema 'Lichaamsprocessen en Vergift' – H. Roebertsen

24 Het thema 'reproductie' in het schoolvak biologie – P.C.F. Reygel

25 Teaching Electrochemical Cells – A study on Teachers Conceptions and Teaching Problems in Secondary Education – J.J.C. Acampo

26 The Role of Contexts and Models in the Development of Mathematical Strategies and Procedures – M. Beishuizen, K.P.E. Gravemeijer, E.C.D.M. van Lieshout (eds.)

27 Thermodynamica leren onderwijzen – Probleemstellende gesprekssituaties voor docent en studenten, leidend tot thermodynamisch begrip en tot een gewijzigd Van Hiele niveauschema – W.H. Kapers

28 Interesse-georiënteerd natuur- en scheikundeonderwijs – Een studie naar onderwijsontwikkeling op de Open Schoolgemeenschap Bijlmer – R.J. Genseberger

The cover illustration is drawn by an RPD student on the Arithmetic Subject Matter test halfway through the curriculum (January). The student had to calculate a difference in price between a pullover that costs 38 guilders and a pair of shorts that costs 35 guilders.

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Klein, Ton

Flexibilization of Mental Arithmetic Strategies on a Different KnowledgeBase: The Empty Number Line in a Realistic versus Gradual Program Design. Ton Klein. Utrecht CD β Press, Center for Science and Mathematics Education, Freudenthal institute, Research Group on Mathematics Education, Utrecht University (CD-β series on research in education; 29)
Thesis Leiden University. -With ref. - With a summary in Dutch

ISBN 90-73346-36-3
Trefw: rekenonderwijs / onderwijsonderzoek / rekenmiddelen / rekenstrategieën / curriculumontwikkeling / evaluatie-onderzoek

The investigations described in this thesis were conducted at the Centre for the Study of Education and Instruction, Faculty of Social and Behavioral Sciences, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands.

Cover: Brouwer Uithof BV
Press: Grafisch bedrijf UFB, Leiden University

Flexibilization of Mental Arithmetic Strategies on a Different Knowledge Base: The Empty Number Line in a Realistic versus Gradual Program Design

Preface

The research project described in this dissertation was financially supported by the Foundation for Behavioral Studies (SGW), which is subsidized by the Dutch Organization for Scientific Research (NWO). The research project could not have been conducted without the cooperation of the primary schools who were willing to participate in the studies: De Dolfijn, Hellevoetsluis; De Kring, Hellevoetsluis; De Spelevaert, Zoetermeer; De Stevenshof, Leiden; De Tjalk, Zoetermeer; De Zwaluw, Leiden; Drie Koningen, De Meern; Mariaschool, Den Hoorn; Marnixschool, Katwijk; St. Josephschool, Heerhugowaard; Woonstede, Den Haag.

Many students assisted in collecting, coding and entering of data, in designing the experimental textbooks and teacher guides, and in coaching the teachers. I would especially like to thank Carlijn Bergmans, Karin Leliveld, Marjan Torn and Erik Hoogenberg for their enthusiasm and cooperation during the research project.

I would like to thank my family, my parents and uncles, for their encouragement and support during all those years. Finally I would like to thank Nanja and Idse for their support and consideration during the years I have invested in this project. They often made clear to me that there are more important things in life than writing a dissertation.

Leiden, January 1998

Contents

1 Mathematics education: cognitive and affective perspectives

1.1 Introduction

After the early Wiskobas projects developed by Freudenthal (1973) and his co-workers textbook authors in the Netherlands began to adopt the ideas of Realistic Mathematics Education (RME). During the 1980s and 1990s many schools replaced their mechanistic maths textbooks by new textbooks based on RME principles. Treffers (1991a) called this curriculum change a "silent revolution" because "not a single innovation expert is heard speaking or writing about it" (p. 11). Several studies were conducted about how this new way of mathematics education influenced the learning outcomes of the students (e.g. Bokhove, Van der Schoot & Eggen, 1996; Gravemeijer et al., 1993; Harskamp, 1988; Harskamp & Suhre, 1986, 1995; Wijnstra, 1988). In this study we compared two different program designs for addition and subtraction up to 100. The Realistic Program Design was based on principles of Realistic Mathematics Education. The Gradual Program Design also has some ideas drawn from Realistic Mathematics Education but follows a psychological conceptualization of stage-wise knowledge development. We were interested how these different program designs effect the development of procedural competence and strategic use of computation procedures. This was investigated both for the whole group of students and for a sample of weaker and better students. Beside these cognitive variables we were also interested in motivational beliefs towards mathematics in general and towards numerical and context problems in particular.

In this chapter two perspectives are presented towards mathematics education: a cognitive perspective (section 1.2) and an affective perspective (section 1.3). The two program designs are described in chapter 2. At the end of this chapter a proponent of the realistic point of view (Treffers) and a proponent of the gradual point of view (Beishuizen) make predictions about the outcomes of this study according to Hofstee's bet-model (1982). In chapters 3, 4 and 5 the empirical study is described, chapter 3 outlines the method of our research, and the results are described in chapters 4 and 5. Chapter 4 provides results on cognitive variables, chapter 5 describes the affective outcomes. Chapter 6 presents the conclusions and discussion of this study.

1.2 Mathematics education: a cognitive perspective

In discussions about the renewal of primary mathematics education there has been a reappraisal of mental computation as "valuable in promoting and monitoring higher-level mathematical thinking strategies" (Reys, Reys, Nohda & Emori, 1995). By doing mental arithmetic in horizontal format children learn to deal with whole numbers "of a piece" (Baroody, 1987) instead of isolated number parts as in column arithmetic. This could make (mental) computation processes more meaningful, stimulating not only conceptual understanding and procedural proficiency but also number sense and the understanding of number relations (McIntosh, Reys & Reys, 1992). Early introduction of written (column) procedures, on the other hand, may lead to algorithmic computation as a series of "concatenated single-digit" operations, which are responsible for many misunderstandings and mistakes (Fuson, 1992).

Although more emphasis on mental arithmetic in the lower grades might foster a better conceptual understanding and prevent some types of procedural errors (Beishuizen, 1993), the acquisition of higher-level thinking strategies or flexibility in mental arithmetic is another matter. For instance, the above-cited authors (Reys et al., 1995) report that the students they interviewed in their study, even those in grades 4 through 8, demonstrated a very narrow range of mental computation strategies: "The use of nonstandard (not taught) strategies was rarely observed" (p. 322).

Similar conclusions were drawn by Becker and Selter (1996) in the International Handbook on Mathematics Education (Bishop, Clements, Keitel, Kilpatrick, & Laborde, 1996) in which an overview is given of recent developments in mathematics education in elementary schools. Their central thesis is that "teaching is no longer seen as a treatment and learning as the effect. Learners are people who actively construct mathematics" (p. 512). This thought is in line with Freudenthal's view on school mathematics (1991) which he sees as "activity" or "hand-made" mathematics. Becker & Selter (1996) describe four different projects in four different countries which make "calls for reform" in the way mathematics has to be taught in primary school according to these ideas. Those projects are the Open-End approach in Japan, Comprehensive School Mathematics Project in the United States of America, Mathe 2000 in Germany (Wittman & Müller, 1995) and Realistic Mathematics Education in the Netherlands. The constructivist view to mathematics education in the U.S. should also be added to this list (Cobb, Yackel & Wood, 1992; Cobb, 1995).

In the Netherlands, as a reaction to the *New Math* movement, and also as a reaction to the *mechanistic approach* which was predominant in the Netherlands in the 60s, the *Wiskobas* project developed the instructional theory of "Realistic Mathematics Education" (RME) (Freudenthal, 1973, 1991; Gravemeijer, 1994; Streefland, 1991a, 1991b; Treffers, 1987; Van den Heuvel-Panhuizen, 1996). The main principle of RME is that formal knowledge can be developed from children's informal strategies (Treffers, 1991b). Children have experience with all kinds of number problems before they come to school. Teaching in school should therefore not be isolated from the *real* world but should relate to that world by using the knowledge children have (cf. Resnick, Bill, & Lesgold, 1992). This process should be natural and the children should contribute to the teaching/learning process as much as possible. To develop knowledge from children's thinking the principle of mathematization is important (Freudenthal, 1968; Resnick, Bill & Lesgold, 1992; Treffers, 1991b). Treffers (1991b) distinguishes between *horizontal* and *vertical* mathematization. In horizontal mathematization the students come up with mathematical tools to help organize and solve a problem located in real-life situation. It leads from the perceived world to the world of symbols. Vertical mathematization, on the other hand, is the process of a variety of reorganizations and operations within the mathematical system itself, the world of symbols.

The outcomes of a national evaluation of mathematics education in the Netherlands halfway through and at the end of the primary school years (Wijnstra, 1988) underscored the need for further reform of mathematics education. Test results pointed not only to an unacceptably low level of procedural competency in certain domains, but also to a generally low level of flexibility in using arithmetic strategies. As a result of this evaluation (Wijnstra, 1988), as well as an elaboration of the national standards for mathematics education, Treffers & De Moor (1990) published a "call for reform" of the Dutch mathematics education from the RME view. They sketch a new lower grades curriculum and propose, amongst other things, *the empty number line* as a new didactic model. The report stresses mental computation not only with smaller numbers under 20, but also with larger numbers up to 100 (in the second grade). Mental arithmetic is not seen as just a stepping stone to (written) column addition and subtraction, but is valued both as a more natural bridge to the informal strategies children bring with them to school. Mental arithmetic is seen as a foundation for the further development of flexible computation and problem-solving strategies (Treffers, 1991b) in which calculating could be done not only *in the head* but rather by *using one's head* in that the use of written work is encouraged. The possibility of writing down one's calculations on paper does not transform mental arithmetic into written arithmetic. In writing, pupils can 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 procedures for mental addition and subtraction up to 100

A brief summary of previous models for mental addition and subtraction can help elucidate why the empty number line was introduced as a new didactic model (Beishuizen, 1993, Gravemeijer, 1994, Treffers & De Moor, 1990). During the 1960s and 1970s multi base arithmetic blocks and Unifix materials were widely in use (see also Figure 1.1). 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 (see Figure 1.1). The hundred square 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).

1960s

1970s

1980s

1990s

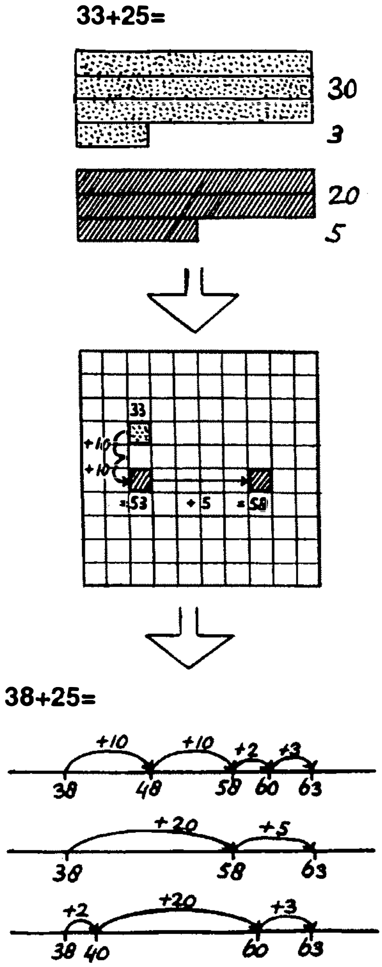


Figure 1.1 Models used for addition and subtraction up to 100 in the Netherlands during the past 40 years

Beishuizen (1989, 1993) found empirical evidence that the above-described models differed in their effects on mental computation procedures. Arithmetic blocks emphasize the conceptual (decimal) structure of numbers as composed of tens and units, and therefore evoked decomposition or place-value strategies for addition and subtraction. The hundred square stimulated a sequential pattern of counting by tens. He referred to the first procedure as the *split method* or with the acronym *1010* (pronounced as ten-ten), because the tens and units were split apart and handled separately (for examples see Table 1.1). He referred to the second procedure as the *jump method* or with the acronym *N10*, because the tens are added to or subtracted from the first unsplit number (for examples see Table 1.1).

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 = 26
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)
	∩[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

[a] The Connecting Arc (∩) can only be used for subtraction problems.

Table 1.1 Mental computation procedures for addition and subtraction up to 100

In U.S. publications 1010-like (mental) computation strategies are predominant, whilst N10 is rarely mentioned (Fuson, 1992; Resnick, 1986), probably because of a greater emphasis in the curriculum on place-value based (column) arithmetic and the use of multi base arithmetic blocks. Many European manuals on the didactics of primary mathematics (Radatz & Schipper, 1983; Treffers & De Moor, 1990; Williams & Shuard, 1982) prefer N10 as the best mental strategy for addition and subtraction up to 100. This does not mean, however, that N10 is frequently used in the lower grades. As research has documented (Fuson, Richards & Briars, 1982) initial acquisition of N10 jumps calls for new knowledge of the numbers up to 100 and therefore is more difficult than acquisition of 1010 (Beishuizen, 1993). This latter procedure is easier to apply because of its strong analogy to already familiar basic number facts (40 + 20 = 60 by analogy to 4 + 2 = 6, cf. Ashcraft, 1985). For instance, Dutch third-graders show a mixed picture, with about one half of them using 1010 and the other half using N10 (Beishuizen, Van Putten & Van Mulken, 1997), and only a minority using both strategies in a flexible way (e.g., 1010 for addition and N10 for subtraction). Extensive use of learning aids like multi base arithmetic blocks or the hundred square might improve this distribution of strategy preferences to some extent, as Beishuizen (1993) found in his study.

Table 1.1 provides an overview of the most important mental computation procedures and their labels, as categorized in our research (for a complete overview of both procedures and types of errors we refer to Appendix A and B). N10 and 1010 can be seen as the two basic strategies; N10 is the more effective computation procedure, while 1010 causes more errors, especially in subtraction problems requiring regrouping (Beishuizen, 1993; Van Mulken, 1992). The procedure we call *10*s can be seen as an adaptation of the 1010 procedure to overcome these problems (Beishuizen, Van Putten, & Van Mulken, 1997). A reaction time study with 3rd-graders (Wolters, Beishuizen, Broers, & Knoppert, 1990) confirmed this procedural view. Again N10 came out as the more efficient mental procedure, also in comparison to experienced and competent use of 1010. Wolters et al. (1990) underline that the greater number of procedural steps with 1010 and the necessity to keep the intermediate steps in mind (cf. Table 1.1) explains why 1010 places a heavier demand on working memory than N10. In addition, Beishuizen, Wolters & Broers (1991) found that less competent students prefer 1010 and more competent pupils prefer N10.

In a review article in 1992, and also later during an experts' meeting in Leiden in 1996, Fuson (1992, 1997) makes the same distinction between two main strategies for addition and subtraction with larger numbers which she calls the "separate-tens" (cf. 1010) and "sequence-tens" (cf. N10) strategy. Further evidence for the generalizibility of these two main categories is provided by some recent research in this domain with larger numbers (Cobb, 1995; Jones, Thornton & Putt, 1994; Thompson, 1994, 1997). For instance, Reys, et al. (1995) discern "group by tens and ones" (cf. 1010) and "hold one addend constant" (cf. N10), and other flexible strategies such as "N10 including compensation" and "adding-on to round tens" (cf. N10C and A10 in Table 1.1). In their research, the latter strategies were, however, seldom used by the students.

The empty number line as a new didactic model

In accordance with actual school practice in the 1980s Buys (1988) but also Treffers and De Moor (1990) reported that the hundred square, although providing a better model of N10 than the arithmetic blocks, is an overly complicated learning aid for weaker pupils. Moreover, the increasing influence of the RME view in our country ran counter to the very pre-structured character of the hundred square, which left little room for children's informal strategies. Therefore Treffers and De Moor (1990), in their "call for reform" of the Dutch primary mathematics curriculum, devised a new format for the old number line: the empty number line up to 100. Earlier Van Gelder (1969) and later Freudenthal (1973) had suggested the number line as a more natural model of children's informal counting strategies. By using the empty number line children could extend their counting strategies and raise the level of their strategies from counting by ones to counting by tens to counting by multiples of tens. Also, the empty number line can be seen as a linear-type representation which is needed to represent counting numbers. This contrasts with manipulatives like Dienes blocks or Multibase Arithmetic Blocks (MAB), with their set representation of numbers. Gravemeijer (1994) mentioned two other reasons why the empty number line should be introduced as a didactical tool for addition and subtraction up to 100. In the first place, the empty number line is very well suited to making informal solution procedures explicit because of its linear-type character of the number line. A lot of informal strategies can be seen as a sophisticated way of counting numbers (Gravemeijer, 1994). Strategies such as counting on and counting down are well documented for children working with small numbers (Carpenter & Moser, 1984). Gravemeijer (1994) reported that children prefer the (easier) adding-on strategy for subtraction problems with larger numbers. He found that second graders find it harder to solve the numerical problem 53 - 45 than the context item where children have to calculate how many beads are left if you have 53 beads in a jar and you need 45 beads to make a necklace (Gravemeijer et al., 1993). He explains this difference in performance, by referring to the use of informal strategies like curtailed counting on, to solve the situated problem. To solve the numerical problem students might have used the more traditional counting down strategy. Both strategies can easily be showed on the empty number line by making jumps as a representation of a sequence of *add-ons* (Gravemeijer, 1994).

The second reason for promoting the empty number line is that it provides the opportunity to raise the level of the student's 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): 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 schematizing* and is a key principle in the theory of RME (Freudenthal, 1991; Gravemeijer, 1994; Treffers, 1987, 1991b). Another principle of RME is that a model should not only be a *model of* situations (for instance a context problem) but should also become a *model for* representing mathematical solutions (Gravemeijer, 1994; Streefland, 1991a). This is true for the empty number line; 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. In Figure 1.1 some examples are given of different ways how a problem can be solved on the empty number line.

Gravemeijer stated (1994) that in the 1970s experiments with the structured number line failed because of the unwillingness of students to use it in a global, flexible manner. He reasoned that the number line was associated with measurement situations in which the number line beared fixed, pre-given distances with a mark for every number (cf. Gilissen & Klep, 1980). This use of the structured number line caused counting and reading-off behavior. This led Treffers (Treffers & De Moor, 1990) to reconsider the use of the number line. He opted for an empty number line on which the pupils can draw marks for themselves. A structured bead string should be introduced as an introductory model for the empty number line (see Figure 1.2). The structure of the bead string (5- or 10-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 six 10s in 64 and there are almost seven 10s in 69. After children work with the bead string, the empty number line can be introduced as a model of the bead string (see Figure 1.2).

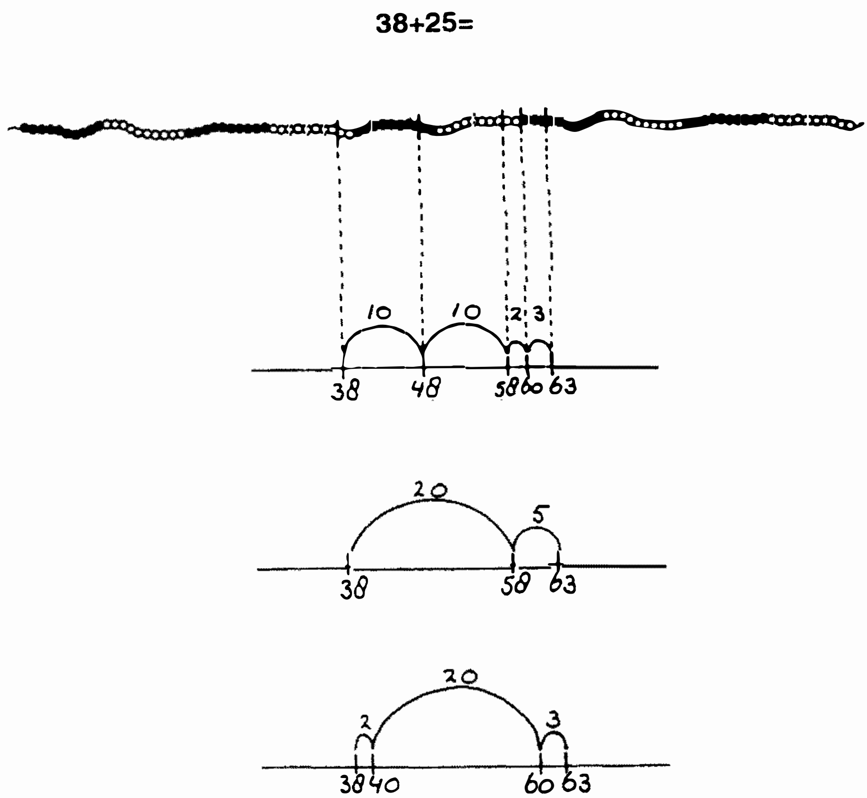


Figure 1.2 The empty number line as a model of the bead string: the problem 38+25 is solved in different ways and levels

Comparison of two program designs for addition and subtraction up to 100

Research in the field of mathematics education showed that stressing the use of the 1010 procedure from the beginning of the second grade was not successful (Harskamp, 1991; Felix, 1992; Klein & Beishuizen, 1989; Klein, 1995; Plunkett, 1979; Radatz, 1987). Both Van Mulken (1992) and Beishuizen et al. (1996) found that children using 1010 had great difficulties solving indirect number problems like 27 + .. = 65. Most of the students using the 1010 procedure changed towards the N10 procedure to solve these problems. Beishuizen et al. (1996) asked themselves if the 1010 procedure should not only be considered as a weaker computation procedure but also as a barrier to strategic problem solving. Beishuizen (1997) also mentioned a possible *misfit* between the 1010 procedure and *sequential* strategies like adding-on to solve a problem. In that respect the N10 procedure seems to be a more flexible procedure which fits into strategies like *adding-on* as well as subtraction. This was one of the reasons for choosing a model which provides a linear representation of numbers. It therefore stimulates the use of N10 and discourages the use of 1010 or other *decadal* procedures.
It is evident that children are very much in favor of using the 1010 procedure, even though it is less successful for the solution of certain problems, because it is a transparent procedure that seems *natural* at first sight (Beishuizen, 1993). We therefore decided to look at this procedure but only after the students had learned to use N10 (like) procedures for different problems. This is in line with Treffers' and De Moor's curriculum proposal (1990) in which they advised to use N10-like strategies from the beginning of the second grade. They do not introduce the alternative 1010 strategy until much later in the second grade, proposing a different base ten number model to clarify the conceptual and procedural differences between the two strategies (cf. Gravemeijer, 1992). Following this line of argument we think that *both* N10 and 1010 should be taught to the students, as this will contribute to a more broadly embedded number sense and greater flexibility of computation strategies. For this reason we adopted the empty number line as a *linear* or *count* type model for addition and subtraction up to hundred at the beginning of the second grade: the empty number line.

In cooperation with Treffers we proposed two experimental programs for the second grade of primary education: a Realistic Program Design (RPD) based on the ideas of RME and a Gradual Program Design (GPD) which has also some ideas drawn from RME but follows a psychological conceptualization of stage-wise knowledge development (cf. Glaser & Bassok, 1989). Both programs use the empty number line as a central model for addition and subtraction up to 100 and aim at greater flexibility in mental arithmetic. Following Treffers and De Moor's (1990) curriculum proposal, and also based on other experiences in this field (Beishuizen, 1997; Cobb, Gravemeijer, Yackel, McClain & Whitenack, 1995; Klein & Beishuizen, 1993), we decided, to emphasize N10-like procedures on the number line in both program designs. The alternative 1010 strategy was introduced in the last months of the second grade, but only for addition. Beside the similarities between the two program designs there are also differences in instructional design between the RPD and GPD. These similarities and differences will be described in chapter 2. At this point in the discussion we only point to the underlying assumptions between the two program designs.

The main aim of the Realistic Program Design was to stimulate flexible strategy use from the beginning of the program largely, by making use of the children's informal strategies. By "flexible strategy use" we mean the choice of the most appropriate and efficient strategy or procedure, given the (number) characteristics of the problem at hand (Klein & Beishuizen, 1994a). Flexibility in pupil behavior was an important objective of this program. Although many researchers in both our country and abroad agree that instruction based on theories like RME or constructivism appear to be more motivating, exciting and challenging for children, some also say that perhaps this is only true for average and better students (cf. Ames & Ames, 1989; Gersten & Carnine, 1984; Ruijssenaars, 1994; Van Luit & Van der Rijt, 1997). According to these authors, less capable students would benefit from more structured instruction in which the teacher helps them to construct their strategies to solve problems. For this reason these authors think that the early introduction of multiple strategies, as is done in the RME, could confuse weaker students. To investigate this assumption we developed the more structured Gradual Program Design. The main differences in instructional design between the RPD and GPD are (1) the increase in the size of the numbers which is more gradually over time in the GPD than in the RPD (cf. Table 3.14), (2) addition and subtraction problems that require passing a ten (for instance 48 + 36, 51 - 49) were introduced later in the GPD (cf. Table 3.14), and (3) the number line featured marks and numbers for a longer period in the GPD than in the RPD. As a consequence the empty number line was introduced at a later time in the GPD than in the RPD.

In the preceding sections we described the differences between the Realistic and Gradual Program Design and what their effects might be on cognitive outcome variables. However, these program designs may also affect the students' reported experiental states and their motivational beliefs.

1.3 Mathematics education: an affective perspective

Until now we emphasized the cognitive aspects of mathematics education. However, there is a growing body of research that advocates that both cognitive and affective variables should be taken into account to describe how students solve mathematics problems (cf. Boekaerts, 1992, 1995, 1997, in preparation; Carr, 1996; McLeod, 1992; Pintrich & De Groot, 1990; Schoenfeld, 1992; Vermeer, 1997). Schoenfeld (1992) and also Nicholls, Cobb, Wood, Yackel, & Patashnick (1990) have argued that doing mathematics can be considered as a social activity, with roots in the cultural and social environment. Environmental variables interact with person variables to shape students' behavior as they work on mathematics tasks.

Person variables may be either cognitive or affective. The relation between *hot and cold cognition* (Boekaerts, in preparation) is described in the model of adaptable learning (Boekaerts, 1992, 1995). Within this model, there are different levels at which motivation can be examined. Following Cantor (1981), Boekaerts (1997) distinguished three levels at which motivational beliefs, just like other aspects of personality, can be studied: the superordinate level, the middle level and the momentary level. In a school context, the superordinate level corresponds with motivational beliefs toward learning and can be contrasted with motivational beliefs toward other activities as sports and leisure activities. To tap these trait-like qualities of students' motivational beliefs towards learning, researchers have developed questionnaires with which a relatively stable person characteristic or students' capacity towards learning can be measured (e.g. need of achievement, fear of failure). At the middle level, the beliefs and attitudes students have toward different school subjects is the main point of interest. With questionnaires researchers try to measure students' tendency to react toward specific content domains as mathematics and history. These two indicators of inclination and tendency should be distinguished from sensitivity to actual curricular tasks. This is measured at the momentary level at which motivation coincides with the quality of the subjective experience within specific learning situations. Various instruments have been constructed to tap the students' task-specific cognitions, affects and intentions elicited before, during and after performing specific assignments. By giving the students different kind of tasks, different cognitions and affects and also their situation-specific willingness to invest and maintain effort can be measured.

Following Lazarus & Folkman (1984), Boekaerts (1997) argued that a major weakness of measurement at the first two levels is that each question is only presented once. The students have to recall specific learning experiences on the basis of a situation description. The judgments students make about the intensity and frequency of their cognitions and affects are based on the activated episodic information. This information may be biased by recent learning experiences and present cognitions, emotions and moods. Boekaerts (1987, 1988) said that when the aim of a study is to explain and predict student motivation in concrete learning situations, it is essential to record the unique ways in which students experience every-day curricular activities. Motivation control plays an important role in these situation-specific forms of motivation. It refers to a self-regulatory skill that students use to appraise events, tasks and activities and to allocate resources.

Model of adaptable learning

To analyze and describe the way students appraise mathematic tasks and activities and the extent to which these appraisals affect effort and task performance, Boekaerts (1992, 1995) presented her Model of the Adaptable Learning Process (see Figure 1.3). This model offers an analytical decomposition of adaptable learning into different self-regulatory skills. The model is hierarchically structured, where affective and cognitive variables, measured at the superordinate and the middle level, are believed to exert an indirect effect on task motivation and task performance through the appraisal processes.

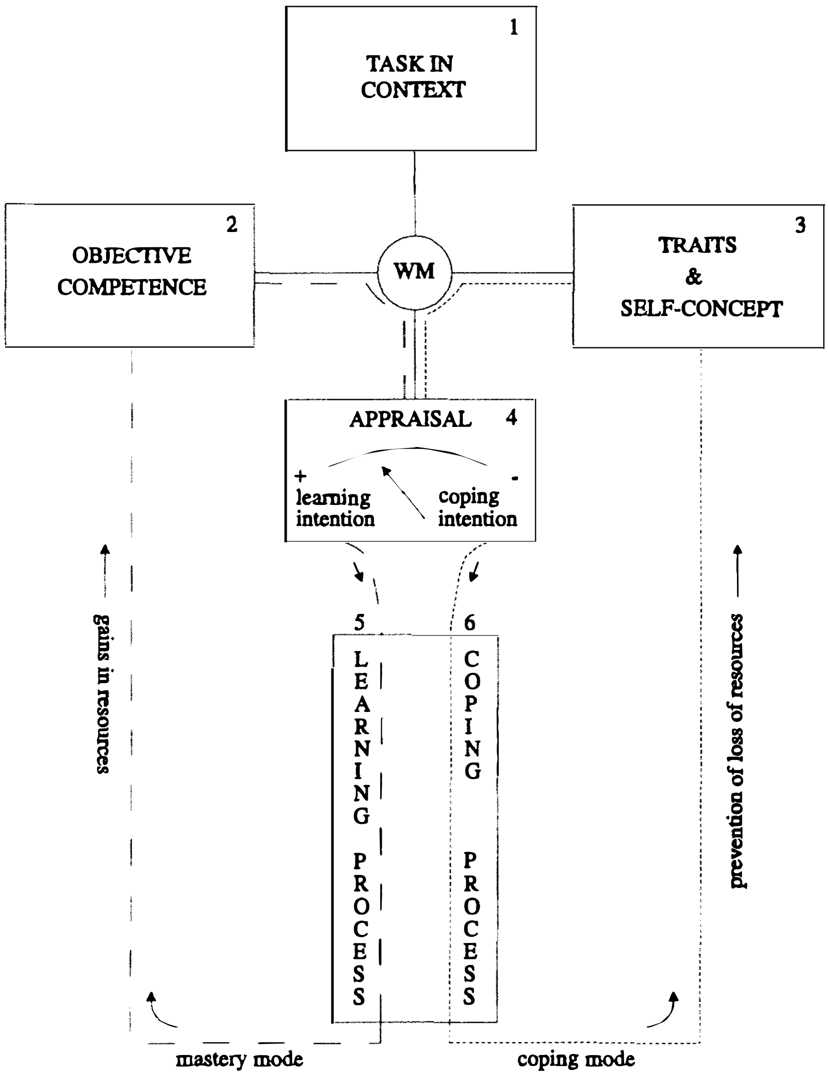


Figure 1.3 The Model of Adaptable Learning

In this model two parallel processing modes are described: the mastery mode and the coping mode. The main idea behind this model is that individuals want to expand their available resources. On the other hand they also want to prevent loss of resources and distortions of well-being. Students have to find a balance between the mastery and coping mode to adapt to the learning process. Appraisals have a central position in the model. They steer and direct the students' attention and energy either to adaptive pay-offs (increase in competence through adequate cognitive self-regulation) or the restoration of well-being (prevention of loss of resources through the use of adequate motivational self-regulation). Positive appraisals motivate students to assemble available learning resources which leads to activity in the mastery mode. Negative appraisals motivate students to assemble available coping resources for the protection of well-being (i.e., a well-being route). The model posits that every learning situation triggers a network of highly specific connotations which may be unique for each person. The working memory (WM) is the main processing unit in this model, in it perceived task demands, objective competence and personal traits, including the self-concept come together. These three sources of information are constantly fed to the appraisals which in turn affect learning and coping intentions (situation specific learning or coping goals).

The social and didactic context in which learning takes place plays an important role in how students judge a learning situation or problem posed to them (Boekaerts, 1997). We therefore expected differences in appraisals between students who worked either with the Realistic or Gradual Program Design. Furthermore, we expected that gender differences would interact with the program design. It is well known that by the end of the primary school differences appear between boys and girls in the way they value mathematics (Boekaerts, 1996; Vermeer, 1997). However, gender differences are not the main issue in this research project and will be reported elsewhere (Klein & Boekaerts, in preparation). In this study, we will focus on the effects that the two program designs may have on students' cognitions and affects differentiating explicitly between numerical and context problems. This will be investigated for students in the second grade, which is earlier than in most research is done. We were interested if the different type of instruction already has an effect at such an early age on motivational beliefs in relation to mathematics as a school subject.

Measurements at the domain-specific level (middle level)

Several studies showed that a major and relatively stable domain-specific personal characteristic is the students' goal orientation (cf. Ames & Archer, 1989; Nicholls, 1989; Seegers & Boekaerts, 1993). Nicholls (1983, 1984) worked out that willingness to invest effort is related to the value that students ascribe to the outcome of the task. He made a distinction between ego-oriented and task-oriented students. Students who are highly task-oriented evaluate their results positively as long as they perceive improved mastery. Their standards for comparison are their own former results and their aspired achievement level. They prefer situations where tasks are challenging and they can expand their knowledge. These students can be contrasted with ego-oriented students who regard performance as reflecting mental abilities (capacity). Their standard is based on comparing their results with their peers' achievement. In their view investment of effort combined with a poor result on a task demonstrates poor capacity. As a consequence, they try to avoid negative public evaluation and social comparisons. In learning situations these students will show less interest in the learning task and demonstrate a more competitive attitude, seeking out situations in which they can demonstrate their abilities.

Pintrich & De Groot (1990) developed a model in which motivational and cognitive aspects towards mathematics are integrated. They adapted the general expectance-value model of motivation and distinguished three motivational components that are related to components of self-regulated learning. The first component is an expectancy component which includes the students' beliefs about their ability to perform a task. The second component they distinguished is a value component which includes the students' goals and beliefs about the importance and interest of the task. The third component is an affective component which refers to the students' emotional reactions to a task. Pintrich & De Groot (1990) designed a questionnaire for seventh graders which was successful in distinguishing between these three components. Following Pintrich & De Groot (1990), Blöte (1993) developed the Mathematics Motivation Questionnaire (MMQ). She adapted the questionnaire to the language use of Dutch children in the age range from 7-10 and reported that the motivational beliefs toward mathematics could be measured in a reliable way (Blöte, 1993; Voogt, 1996). We decided to use the scores on the subscales self-efficacy, valuing and affect towards mathematics as a measurement of motivational aspects at the domain-specific level.

Measurements at the task-specific level

The way students appraise mathematics tasks are crucial in Boekaerts' model of adaptable learning (see Figure 1.3). The model clarifies in what way appraisals affect effort and task performance. To measure students' cognitions and affects in actual learning situations, Boekaerts and her co-workers developed the On-line Motivation Questionnaire (OMQ) (e.g. Boekaerts, 1987; Crombach, Boekaerts, & Voeten, 1994; Seegers & Boekaerts, 1993). This questionnaire is administered just before a student begins with a learning task or homework assignment, and again when it is completed or when the student gives up. The OMQ assesses, amongst other things, the students' appraisals, affects, and learning intention before they begin with a mathematics task, and their reported effort expenditure, affect, and attributions after doing the task. In previous research, three basic appraisals were identified. Two appraisals reflect the students' self-referenced cognitions about the value of a curricular task (task attraction and perceived relevance). The third appraisal taps the students' capacity related beliefs and is called subjective competence and aggregates the students' self-efficacy judgment expressed in relation to the task, their outcome expectation and the perceived difficulty level of the task. In studies with children in the age range from 10-14, Boekaerts (1997) found that students who find a concrete mathematics task personally relevant, attractive, interesting, or challenging experience a positive emotional state and are willing to expend effort to accomplish that task. Boekaerts (in preparation) adapted the OMQ for second grade children in the age range of 7-10 (OMQ 7-10). After several adaptations, especially with respect to the language use, she was able to distinguish five subscales all pertaining to student cognitions and affects before they begin with a mathematics task (self-confidence, task attraction and positive affects, task value and learning intention) and two subscales measuring cognitions and affects after task completion (effortful accomplishment, absence of threat). These scales were used in our research.

Expected differences in motivational beliefs elicitied by the two program design

Boekaerts (1997) suggested that not only individual differences in capacity (IQ-scores or standardized test scores) and inclination (motivational beliefs at the domain-specific level), but also the way in which mathematics teaching compliments these differences, determine students' sensitivity to mathematics in actual learning situations. One of the differences between the two program designs is the role that context problems have in the instructional sequence. In the RPD, meaningful and problem oriented context problems are used as an introduction of a new problem or to introduce a new strategy. The students' attention is explicitly drawn to variations in strategy use and they reflect upon the different solutions by making connection to the task characteristics. The student must take a lot of initiative in the dialogues, visualising the solution strategies they used themselves. The teacher has the role of a guide who enables the students to come up with different solutions and raise the level of the students' solutions. During the first six months of the GPD, the students are trained to be skillful in executing one procedure (N10) with which they can solve every addition and subtraction problem. This is done by solving numerical problems. In this period context problems are used to train students to recognize an operation and to apply the operation (here the N10 procedure) to application situations (Treffers, 1991b). This way of using context problems is often found in structuralistic approaches towards mathematics education (cf. Resnick, 1982). During the last four months of the GPD, attention is paid to flexibility in using different procedures. At that time the function of context problems changes towards the one they have in the RPD and also the role of the students and the teacher changes towards the appearance they have in the RPD.

Boekaerts (in preparation) suggested that students who have learned to solve math problems following the RPD (stimulating the use of different solution strategies and computation procedures from the beginning), will appraise the math problems differently. Compared to the GPD students, to whom only one computation procedure is advocated, RPD students will experience more ambiguity and complexity during the mathematics lessons. It is plausible to suggest that RPD students will develop a tolerance toward uncertainty, usually associated with context problems. Compared to the GPD students, RPD students may have more positive cognitions and affects before beginning with both numerical and context problems and after completing these types of problems. These differences between the RPD and GPD students are expected to be largest half-way through the curriculum (January) since, at that point there are still major differences between the two program designs. At the end of the school year these differences will be reduced or may have disappeared due to the changes introduced in the GPD.

2 Two program designs for addition and subtraction up to 100

This research started with a research proposal (Boekaerts & Beishuizen, 1991), in which the comparison of two program designs was proposed both using the empty number line as a central model for addition and subtraction up to 100. It was decided not to use existing realistic textbooks, with the empty number line as a central model, because then we would not be able to compare programs which contain all the theoretical features we would like to include. By designing two new programs for addition and subtraction in the second grade we did not have to meet with commercial restraints and could make the comparison between the programs as fair as possible. Another issue is the way these programs are implemented. Gravemeijer (1994) distinguishes two paths for implementation of realistic mathematics education: 1) by directly influencing the teachers' views, knowledge insight and skills, or 2) a more directed form of realistic mathematics education in which the textbooks are adapted accordingly. We choose the second path: we rewrote the *realistic* textbook "Rekenen & Wiskunde" (R&W) (Gravemeijer, Van Galen, Kraemer, Meeuwisse, & Vermeulen, 1983) because of the clear structure of both its pupil's worksheets and the teacher's guide. For each day there is a clear description of which exercises a pupil has to make and what the teacher has to do. In this way it is easier to rewrite the teacher guide and the pupil's textbook of R&W than it would be for a textbook like "Wereld In Getallen" (Van de Molengraaf et al., 1981) which is less structured and gives more freedom to the teacher.

The pupil's textbooks and the teacher's guide of the second grade was rewritten because in this period in Dutch primary education, mental arithmetic up to 100 is the main topic. The first grade starts with numbers up to 20 and towards the end of the first grade addition and subtraction problems up to 20 are introduced. In the second grade arithmetic up to 100 is the main topic which includes addition and subtraction and also multiplication. Beside this, subjects like spatial orientation and learning to tell the time also receive attention in the second grade. During the third grade, written or column-wise arithmetic is introduced. For the two program designs we rewrote the part about addition and subtraction which covers about 75% of the exercises in the regular textbook in the second grade. The regular text was used for the other subjects like measurement, tables of multiplication, spatial ordering and telling time.

The teachers and students used the materials of the RPD and GPD instead of their regular mathematics textbooks and teacher guides. For every lesson the instruction for both the whole-class discussions and the worksheets, was written out in the teacher guide. Every fortnight one of the researchers had a meeting with the teachers to discuss their experiences with the program.

In the next paragraphs we will describe the theoretical framework and the most important features of the RPD and GPD. The theoretical framework consists of elements from cognitive psychology (constructivism, Neo-Piagetian theories), literature on word problems and the theory of Realistic Mathematics Education (RME). The two program designs will be characterized by describing the most important features of the RPD and GPD. These include: 1) the way the number line is introduced, 2) the role of mental arithmetic, 3) role of context problems, 4) role of the teacher and 5) a time schedule and instructional sequence.

2.1 Realistic Program Design

Theoretical framework

The Realistic Program Design (RPD) is based on the ideas of RME (Freudenthal, 1973, 1991; Gravemeijer, 1994; Streefland, 1991a; Treffers, 1987; Van den Heuvel-Panhuizen, 1996) and was constructed in cooperation with Treffers from the Freudenthal Institute. One of the main ideas of realistic mathematics education is, that children already have experience with all kinds of number problems before they come to school. Teaching in school should therefore not be isolated from the *real* world but should relate to that world by using the knowledge and informal strategies children already have (cf. Resnick, Bill, & Lesgold, 1992). A second important principle of realistic mathematics education is that children, like in constructivism (cf. Cobb, 1995), should construct their own knowledge and not just apply the strategies and procedures they were taught in their math class. To meet with these two principles informal knowledge should be elicited by starting with a problem that appeals to a child's experience. Children should find the solution to such a problem by constructing their own knowledge (cf. Cobb, 1995). This can be done by oneself or by working together in a small group or by whole-class discussion. For the long-term the learning process within RME should move from concreteness to abstraction: children should be guided from their informal, context-bound methods to formal mathematics. This process of progressive mathematization has two components that are mutually intertwined (Treffers, 1987): *vertical* mathematization where reorganizations and operations within the mathematical system take place (for instance moving from counting, towards using the five-structure on a bead-string), and *horizontal* mathematization where mathematical tools are used to organize and solve problem situations located in reality (for instance: You are reading a book of 61 pages; you are on page 49; How many pages do you still have to read?). In this process of progressive mathematization the teacher plays a crucial role and it is here that a difference arises between constructivism and RME (see also the paragraph about the role of the teacher). This difference between constructivism and RME becomes even more salient in our choice for a more directed form of RME in which the textbooks and the teacher guides were rather prescriptive. The reason for this was that we wanted to make the program manageable by teachers, while 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 decided, together with Treffers, 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

∩ see Table 1.1). After some practice with a particular strategy, by solving problems on worksheets designed by the researchers, the students were free either to use or not use this particular strategy for solving other problems. An example might elucidate our approach in the RPD. To let the children experience which procedure or strategy is the easiest and most efficient at solving a problem, according to the number characteristics of the problem, children were asked to solve problems in two different ways and mark which way they thought was the best. In Figure 2.1 we see how Youri solved two problems in different ways, during an individual interview in October. After he solved the problem in two ways, he was asked to draw a flag beside the way he thought was the best.

Solve each problem in two different ways.
Draw a flag beside the solution way you think is the best one.

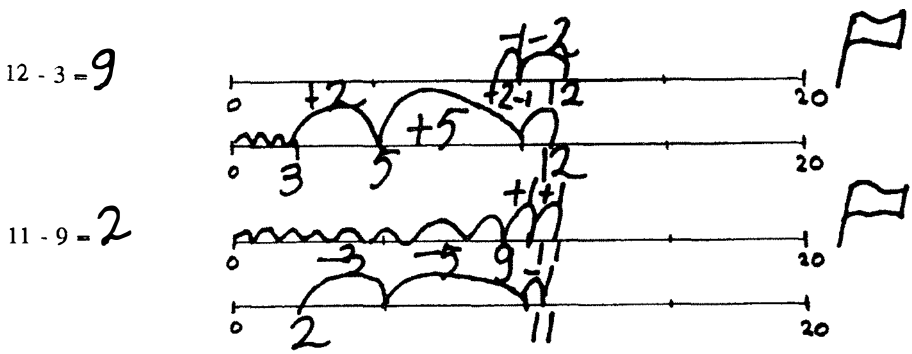


Figure 2.1 Two problems which Youri solved in two ways. He drew a flag beside the way he thought was the best

As you can see, he considered subtracting as the best way to solve 12 - 3; for 11 - 9 the Connecting Arc appeared to be much better. When the interviewer asked how he found out this Connecting Arc procedure, Youri said that he invented this way by himself, long before his teacher introduced this way in the classroom. Youri then said that his teacher calls the Connecting Arc procedure the "Youri way" of solving a problem and that he was very proud of this. Youri frequently explains "his way" to other students but he also adds to it, that it is not always "handy" to use .

This example of Youri illustrates that, compared to RME theory, our experimental RPD was a bit more directive. However students were free to follow their own strategy preferences after practicing strategies they had been introduced to. Before we give a kind of "time schedule" of the program we would like to describe some important features of the RPD.

Introduction of the number line

Because the empty number line is rather abstract, it was far too difficult for the students to start with the empty number line, right from the beginning of the second grade. It was therefore decided to start with a structured number line: first up to 20 and later on up to 100. Both the structured and the empty number line do not feature marks for every number, as that appeared to be not so successful (Gravemeijer, 1994). The empty number line was introduced after 12 weeks working with the RPD. The number line up to 20 had marks for the fives and the tens (see Table 2.1), the number line connected with the bead string up to 100 had only marks for the tens. Note that only the marks and the numbers 0 and 20 or 100 are given. The students can fill in the other numbers if this helps them to solve the problems. After a period of time, these marks were removed; the model became just an empty line and children draw number marks and jumps for themselves (mentally). The number line up to 20 is introduced by a structured bead string up to 20 (see Table 2.1). The bead string contained the five structure to prevent children from counting the beads one by one (cf. Van den Berg & Van Eerde, 1992). Students are trained to "read" the numbers on the bead string by using the five-structure: they learn that 6 beads have 1 group of 5 beads and 1 single bead marked by a different color. After 8 weeks the structured number line up to 100 was introduced by using a bead string with the ten structure (see Table 2.1). The students now learned that 13 beads had 2 groups of 5 beads and 1 group of 3 beads. During the transition between those two bead strings with their different structure (five versus ten structure) some temporary errors are made like seeing 11 beads as 6 beads.

We think that the bead string should primarily be used for showing the structure of numbers and as a concrete representation of the numbers up to 20 and 100 and to prevent children from counting. However for explaining procedures like addition and subtraction we think the bead string is not so suitable. For instance, the direction of subtracting numbers on the bead string is opposite to the direction of subtracting numbers on the number line (see Figure 2.2 and also Willems & Groenewegen, 1997). For this reason addition and subtraction of numbers is hardly done on the bead string but primarily on the number line in the Realistic Program Design.

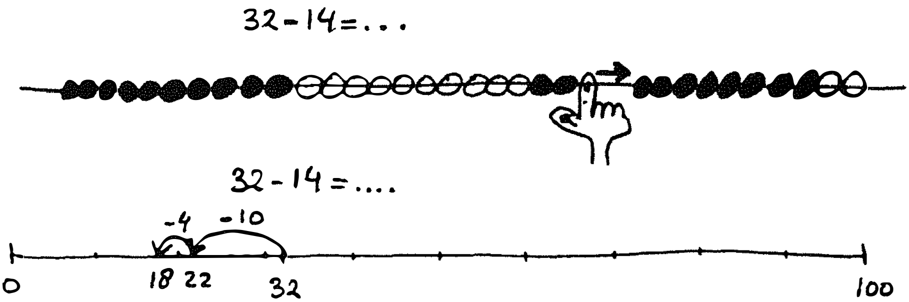


Figure 2.2 Subtraction problem 32 - 14 on a bead string and semi-structured number line.Note the difference in directions of moving the beads (to the right) and making jumps on the number line (to the left)

Mental arithmetic

The fact that the students have to write down their solution steps by use of the number line (and later on by means of the arrow scheme or just solution steps) does not mean that attention is not paid to solving problems by head. The pupil's textbooks also contained exercises in which the students are asked to solve the problems by head and only to write down the answer. Also during the whole-class discussion some time is spent on the rehearsal of number facts. However this is not done by "drill and practice" but by making use of, for instance, double sums which are familiar to most of the children. Other sums like near-doubles can than be derived from these number facts. Also other sums can serve as anchoring points with which it is easy to derive an answer from that sum. If you know, for instance, how much 64 + 10 is (or other 10-jumps) then you can derive easily how much 64 + 9 or 64 + 11 is. In this way children build up a kind of network with which different problems can be solved. During whole-class discussions a lot of attention is paid to building up such a network. Another thing that receives attention during these discussions is making students sensitive to different problem characteristics and experiencing which procedure or strategy is the most efficient one. Students came up with different solutions of a problem and they argued which one is the best. These solutions were not always written on a number line, but also often explained by head.

Role of context problems

In the RME-view context problem types are seen as a means to stimulate mathematical reasoning as a problem solving activity, an approach which contrasts to the emphasis on numerical problems in procedural training in traditional, mechanistic textbooks. In RME two types of context problems are distinguished: context problems as application problems and context problems which also have the function of a model (for instance the "book"-problem as a model for the adding-on strategy). RME also emphasizes the need to make connections to the informal working methods of children before introducing more formal strategies. In the realistic view, and therefore also in the RPD, the instructional sequence should be such that flexibility in strategy use is fostered first (through various types of context problems and models), followed by guidance and practice of the proceduralization (i.e. execution of procedural steps) of number operations (Treffers, 1991b; Van Mulken, 1992).

In cognitive psychology, the research tradition using verbal word problems (Riley, Greeno & Heller, 1983; Verschaffel & DeCorte, 1990) comes close in its intention to the realistic view. Problem representation and the conceptualization of different problem types or cognitive schemes are seen as fundamental to the development of arithmetic competency (Fuson, 1992; Lewis & Mayer, 1987; Stern, 1993). In the first place informal strategies are stimulated through problem-solving activities. This is in contrast to the traditional school practice, which results in rigid proceduralization at an early phase, and may lead to insufficient flexibilization (i.e. the adaptation of procedures in accordance with the demands of a given problem). However, there are differences in the operationalization of verbal problems and context problems, as we will see. Realistic context problems offer more picture-like visualization with a minimum of verbal explanation, and the underlying models or mathematical (semantic) structures are also not always the same (Van den Heuvel-Panhuizen, 1996; Van den Heuvel-Panhuizen & Gravemeijer, 1991). On the other hand we notice that in the RME research tradition, not so much attention is paid to the literature on word-problems. Especially the work by Verschaffel & DeCorte (199●, 1996) gives us useful information about the influence of the semantic structure of a story on the strategy choice of the students. In our research project we tried to combine the knowledge presented in the literature on problem solving and RME in constructing the items in both worksheets and tests (cf. Klein & Beishuizen, 1994a).

Role of the teacher

Beside the pupil's textbooks and teacher guides, the teacher plays a crucial role in RME. This can also be seen as an Achilles' heel of the RME (cf. Gravemeijer et al., 1993; Hiebert, Carpenter, Fennema, Fuson, Human, Murray, Olivier & Wearne, 1996). In the Netherlands more than 75% of the arithmetic textbooks used in primary education are based on the concepts of RME (Treffers, 1991b). However this does not always mean that all the teachers use the textbooks in the way the authors wanted them to use them (Gravemeijer et al., 1993; Gravemeijer & Ruesink, 1992). The role of the teacher in RME differs from the role the teacher had while using traditional textbooks. The teachers should not just show the children how they should solve a problem, but they should more or less play the role of a coach: they should encourage children to look back and reflect on the learning/teaching process. They should also provoke and reinforce a succession of changes of perspective which are necessary for a successful learning process: they should guide the re-invention of different solution strategies by the students (Freudenthal, 1991; Streefland, 1991a, 1991b). This may cause problems with respect to authority in the classroom (cf. De Lange, 1992). No longer is one procedure (the one showed by the teacher) the correct procedure but there are several ways of solving a problem. One of these might even be a possibility the teacher did not think of. On the other hand the teacher must have advance knowledge of the prospects for the future of each strategy used by a pupil. The teacher should discourage the use of strategies which may hinder progression later in the course. In order to be able to make this decision, teachers must discover which strategy or procedure have prospects for the future and which have not. In that respect RME can also be very tough (Klein, Beishuizen & Treffers, in press) and more prescriptive than, for instance, constructivism (cf. Cobb et al., 1995). An example may clarify that it is not commendable to accept every strategy that a pupil comes up with. When we designed our pilot version of the RPD we had many discussions with Treffers about how a "realistic" program should look like. Initially we stimulated every informal strategy that children came up with. For instance a problem like 9 + 6 is not always solved by adding up 1 with 9 to get to 10. Some children see 6 as 3 + 3 and therefore add 9 + 3 + 3 (see also Figure 2.3).

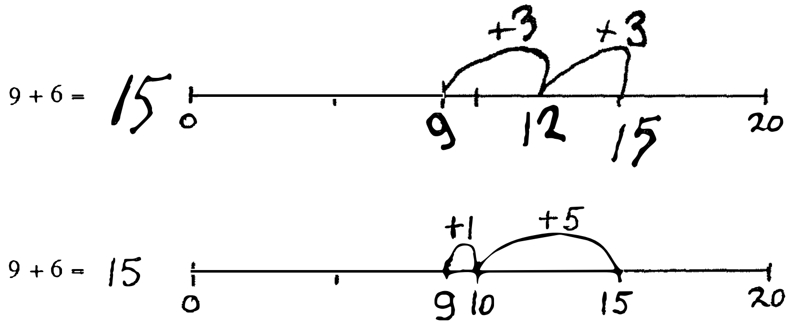


Figure 2.3 Examples of two ways of solving the problem 9 + 6

Treffers made clear that it is not the purpose of RME to focus on this type of strategy mainly because it does not have right perspectives for problems with larger numbers. It is better to concentrate on the strategy in which you complete towards the next 10 (CS strategy) which has more benefits for the future.

Time schedule and instructional sequence

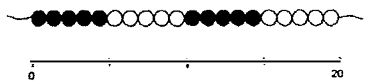
The time schedule for the moments at which different subjects are introduced in the RPD is given in Table 2.1. (see next page) The dotted lines indicate the moments when testing took place.

Week 1- week 8: arithmetic up to 20

During the first eight weeks we started with addition and subtraction up to 20. This subject was introduced at the end of the first grade, and was partly rehearsed. A lot of attention was paid to the positioning of the numbers on the bead string and the number line: 9 is closer to 10 than to 5 and also the distance between 9 and 11 is smaller than the distance between 2 and 11. Games were also played in which children had to recognize quickly how many beads the teacher showed. Here the children could use the five-structure to recognize a number pattern. Beside the number line and the bead string up to 20, also the bus model (Gravemeijer et al., 1983; Van den Brink, 1974) and the double-decker were used (see Figure 2.4).

Tests .... September

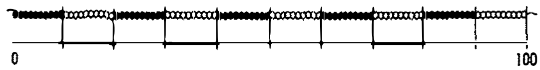
Number positioning on bead string as introduction of semi-structured number line up to 20

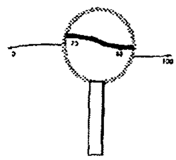


Sums <20: 7 + 7, 7 + 8; 14 - 6, 11 - 9

Tests .... October

Number postioning on bead string as introduction of semi-structured number line up to 100





Introduction of the empty number line

Practicing of "10-jumps" : + 10, 20, 30 and -10, 20, 30

Sums<100: 74 + 8, 93 - 9; 45 + 32, 48 + 36, 45 - 23

Context problems as starting point to discuss procedures like N10, N10C, A10, Connecting Arc

Tests .... January

Sums <100: 85 - 32, 85 - 39; 81 - 79, 81 - 19 (subtraction of two-digit number that require regorouping)

Context problems as starting point to discuss procedures like N10, N10C, A10, Connecting Arc

Tests .... April

'Money' context to discuss the 1010 procedure for addition problems: 33 + 34, 38 + 35

Labels for different procedures

Tests .... June

Table 2.1 Time schedule for the Realistic Program Design

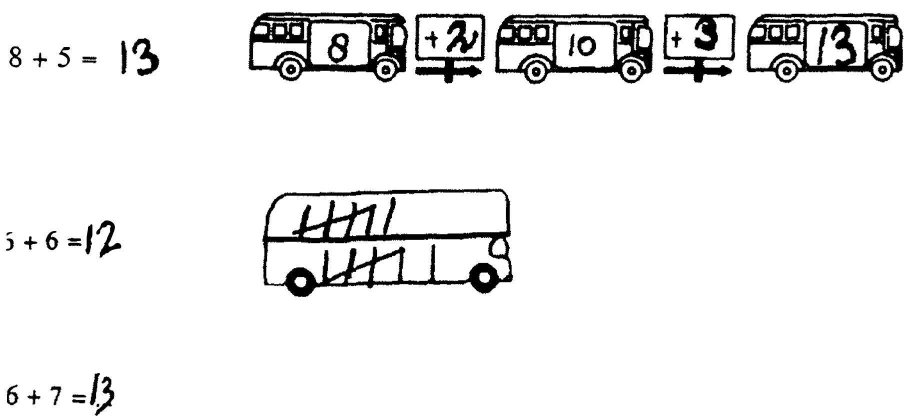


Figure 2.4 Examples of the bus model and the double-decker with tally-marks in the upper and lower level of the bus

makes clearer to children what addition and subtraction means: people getting on or getting off the bus. The double-decker was used to show the advantage of using doubles and near-doubles. Children were asked to solve a double sum by drawing tally-marks in the upper and lower level of the double-decker (see Figure 2.4). A near-double sum was presented below this problem, and children were asked to solve this problem by deriving the answer from the double sum presented above the sum.

In week 5 we started with the introduction of problems which require crossing tens and are not so easily solved by deriving the answers from doubles. For addition problems the children learned to complete to the ten and than add what was left of the addend (see Figure 2.5). For subtraction problems children first learned to go down to the ten and than subtract what was left of the subtrahend (see Figure 2.5). Both procedures are referred to as Complementary Structuring (CS). With larger numbers they can be a part of the earlier mentioned N10 procedure (see also Figure 2.5).

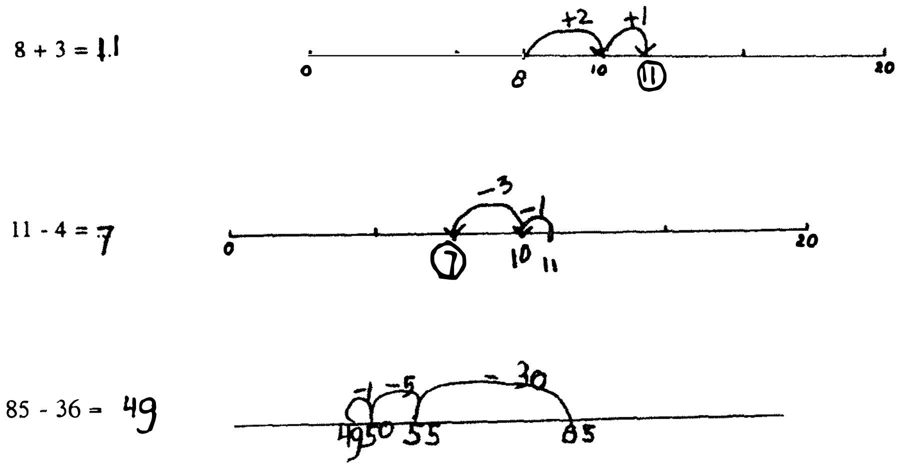


Figure 2.5 Examples of Complementary Structuring (CS) for the problems up to 20 and for problems with larger numbers as part of the N10 procedure

Another strategy that was introduced in this period for subtraction problems, which required regrouping was the Connecting Arc (Treffers, 1995; Treffers & Veltman, 1994; see also Table 1.1). First we introduced the Connecting Arc as a "strategy afterwards": a way to check the answer of subtraction problems: 12 - 4 = 8 because 8 + 4 = 12. This resembles the way you can check division problems: 36 : 6 = 6 because 6 x 6 = 36. When students had checked the answer of a subtraction problem they could draw an arc above the 8 and 4. Later the use of the Connecting Arc was transformed to a "strategy before". When, for instance, students had checked the answer of two sums like 11 - 2 = 9 and 11 - 9 = 2 they would probably notice that it is easier to solve 11 - 9 by *bridging the gap* between 9 and 11 than subtracting 9 from 11 (see Table 1.1). For a subtraction problem like 11 - 2 it is probably more efficient to subtract 2 from 11. The number characteristics of the two problems can be shown on the number line or on the bead string. Children will see that 9 and 11 are much *closer to each other* than 11 and 2. It is therefore more efficient to solve 11 - 9 with an adding-on strategy, like the Connecting Arc, and 11 - 2 with a subtraction strategy. For this latter problem the Connecting Arc can still be used to check the answer and remains a "strategy afterwards" while for the former problem the Connecting Arc becomes a "strategy before". To give the reader an impression of the way this procedure was introduced in the RPD, we refer to Figure 2.1.

Week 9 - week 16: introduction arithmetic up to 100 and the empty number line

In this period we started with addition and subtraction of numbers up to 100. This was done by the introduction of the bead string up to 100 (see Table 2.1). This bead string has a ten-structure of beads with two different colors. Together with this bead string, the number line up to 100 was introduced with marks at every ten (see Table 2.1). In week 14 the empty number line, without any marks, was introduced (see Table 2.1). This empty number line causes more mental activation than the structured number line, because the students had to decide for themselves in which number range they want to operate. The empty number line can therefore be seen as a more abstract way of representing the number space up to 100. After the introduction of the empty number line, at the end of this period, the so called arrow scheme was introduced (see Figure 2.6). This arrow scheme can be seen as a

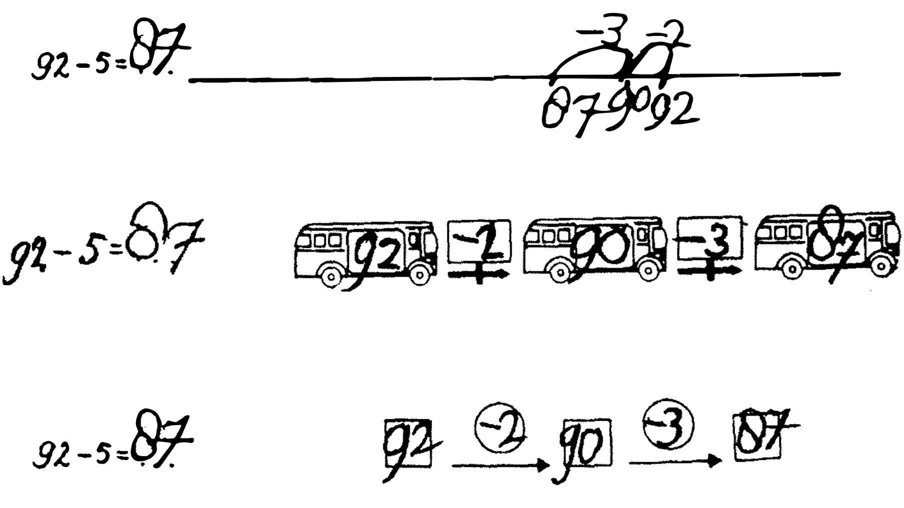


Figure 2.6 Three ways of writing down the solution steps a student used to solve the problem: the empty number line, bus model and arrow scheme

further abstraction of the bus model with which students can show how they solved a problem. The arrow scheme is also considered as a more abstract scheme than the empty number line (cf. Moerlands, 1992).

As with the introduction of numbers up to 20, a lot of attention was paid to developing children's number sense in the area of numbers up to 100. Children had to count aloud in the classroom forwards or backwards from one number to another number. Solving problems with larger numbers was prepared by making 10-jumps on the number line forwards or backwards. These 10-jumps were practiced by counting aloud the jumps with all the students in the classroom. Number sense was also developed by using all kinds of games such as looking at the number of beads the teacher shows, and writing down the number of beads as soon as possible, using the ten structure of the bead string. Also games like "***Raad mijn Getal***" (Guess my number) were played. The teacher (or a pupil) thinks of a number and the students have to guess this number by asking "Is it more than.....?" or "Is the number less than ....?". Students use a number line to cross-out the area, in which the number cannot be located. The idea is that you try to guess the number with as few questions as possible. Another exercise which simulates number sense (Klein & Beishuizen, 1993) was making children jump on the number line from one number to a second number (see Figure 2.7). It is not necessary that the children calculate how much is in between those numbers, they just have to make more than one correct jump from one number to the other.

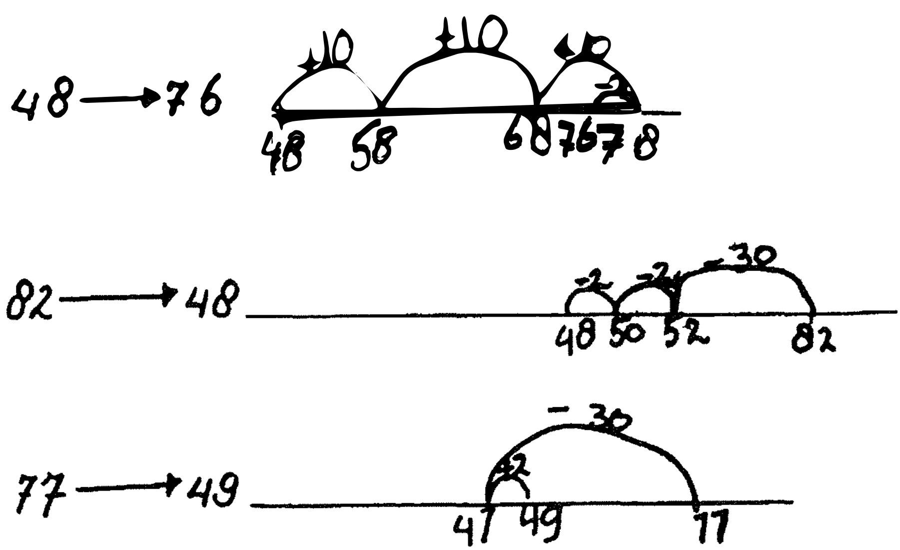


Figure 2.7 Examples of exercices in which the student is asked to jump from one number to another number

It appeared that children liked these exercises a great deal: they use different ways to solve the problem. It was also apparent that these exercises stimulated the use of different solution procedures to solve addition and subtraction problems (Klein & Beishuizen, 1993).

Addition and subtraction problems up to 100 were introduced by sums where a single-digit number had to be added or subtracted from a multi-digit number (for example 57 + 5, 64 - 7). To *go through the ten* in these problems with larger numbers (57 + 3 + 2, 64 - 4 - 3), the CS strategy mentioned previously (completion to the nearest ten) was also used. Beside this procedure, the N10C procedure (see Table 1.1) was introduced. This was done by presenting context problems in which students had to solve the problem in two ways (see also Figure 2.8). We let the children experience which procedure was the most efficient according to the number characteristics of the problem (cf. Figure 2.1).

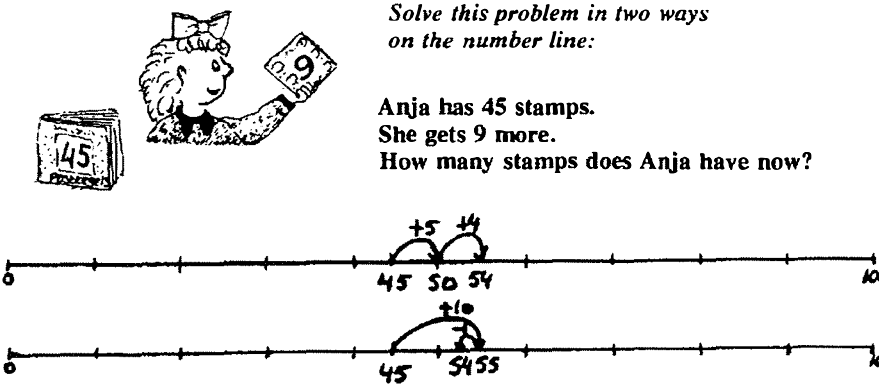


Figure 2.8 Example of how the N10C procedure is introduced in the RPD

During this period the Connecting Arc was also introduced for subtraction problems with larger numbers: 71 - 69 = ... Because the students had not yet dealt with solving these problems by subtracting 69 from 71, they will be more sensitive to add-on from 69 to 71, than if we had introduced this procedure before.

Week 17 - week 24: further elaboration of arithmetic up to 100

In the previous period the empty number line as well as the more abstract arrow scheme had been introduced. In this period attention was also paid to just writing down the solution steps a pupil had used to solve a problem. This can be done on a piece of scratchpaper, which was often depicted beside the problem (see Figure 2.10). The students were free to use either the number line, arrow scheme or just writing down the steps. In this way it was possible to differentiate between students: some students did not need the number line anymore while others were still using the bead string to solve a problem. The teacher should encourage children to use more abstract models (vertical mathematization) and bring them to higher levels. On the other hand this must not be done too quickly because then there is a risk that children will no longer be able to solve a problem and lose self-confidence. For most problems students had to write down the solution steps which they used to solve the problem. When the Connecting Arc was used as a *strategy before*, they only had to draw an arc. The students knew the answer immediately so it would be artificial to have them write down solution steps.

After the introduction of addition and subtraction up to 100 with single-digit numbers (57 + 6, 64 - 8) we came to the problems in which two-digit numbers had to be added or subtracted (57 + 26, 64 - 38). Together with the N10 procedure (see Table 1.1), also the N10C and A10 procedure were introduced by using context problems. Also the difference between the Direct Subtraction and Adding On strategy was made clear by the use of context problems. This could be done by using certain context problems like the so called book problem (Vuurmans, 1991) (see Figure 2.9). For problems in which a difference had to be calculated (see Figure 2.9), the Adding On strategy also seems to be more natural than the Direct Subtraction strategy (Beishuizen, 1997; Klein & Beishuizen, 1994).

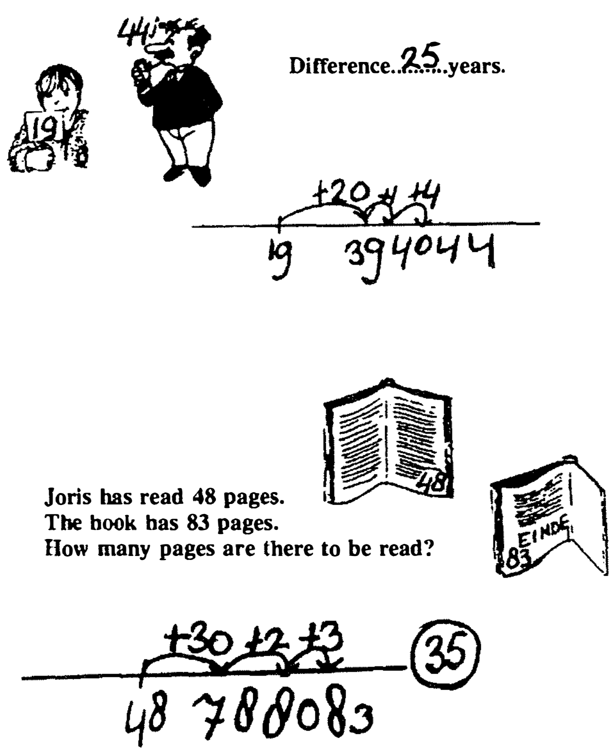


Figure 2.9 Examples of two context problems: a difference problem (calculate the difference in age) and the book-problem

Figure 2.9 Examples of two context problems: a difference problem (calculate the difference in age) and the book-problem

Another way of showing the difference in efficiency between using the Direct Subtraction or Adding On strategy, was using certain number characteristics in both context problems and formula sums. By using two numbers which were close to each other, you hope these will elicit more use of the Adding On strategy, while using numbers with a great difference will evoke Direct Subtraction. This can also be done for the N10C procedure by using addends or diminuends with an 8 or a 9 at the end of the number. To make children sensitive to these different number characteristics, almost half of the time was spent on whole-class discussion. In this period time was also spent on the rehearsal of number facts by using doubles and near-doubles.

During the last two weeks of this period the introduction of the 1010 procedure is prepared by using the context of money. Children have to draw how many 10 guilder bills and guilders a certain object costs.

Week 25 - week 32: introduction of 1010 for addition, labels for different procedures

In this period we introduced the 1010 procedure for addition problems since the use of this procedure for subtraction problems with regrouping may cause serious confusion in the children. Because at this time the students have learned different N10-like procedures to solve all kind of problems it is easier to let the children experience what the disadvantages are of the 1010 procedure, especially for subtraction problems which require regrouping. After these disadvantages were recognized by the children, the teacher indicated that the 1010 procedure could only be used for addition problems.[1]

Because all problem types for addition and subtraction up to 100 had been introduced by now, we had time to make the students more aware of the number characteristics and the structure of different problems. We also had the opportunity

1 This may seem a bit prescriptive and far more pre-structured than you would expect from a program design based on the ideas of RME. However we decided to do so because our purpose was to improve the learning of addition and subtraction up to 100 in the second grade. At the time we started with the introduction of 1010, most of the students could deal with almost all addition and subtraction problems up to 100. We thought that if you also wanted to explain how 1010 could be used for subtraction problems with regrouping you needed more than three months, and three months was all that was left at the moment the 1010 procedure was introduced. Another possibility was to continue the program in the third grade. However, we did not have the time (and money) to continue the program for so long. If we had introduced the 1010 procedure for subtraction problems without sufficient time being available, we expected that most of the students would get confused and make all kind of mistakes they had not made previously.

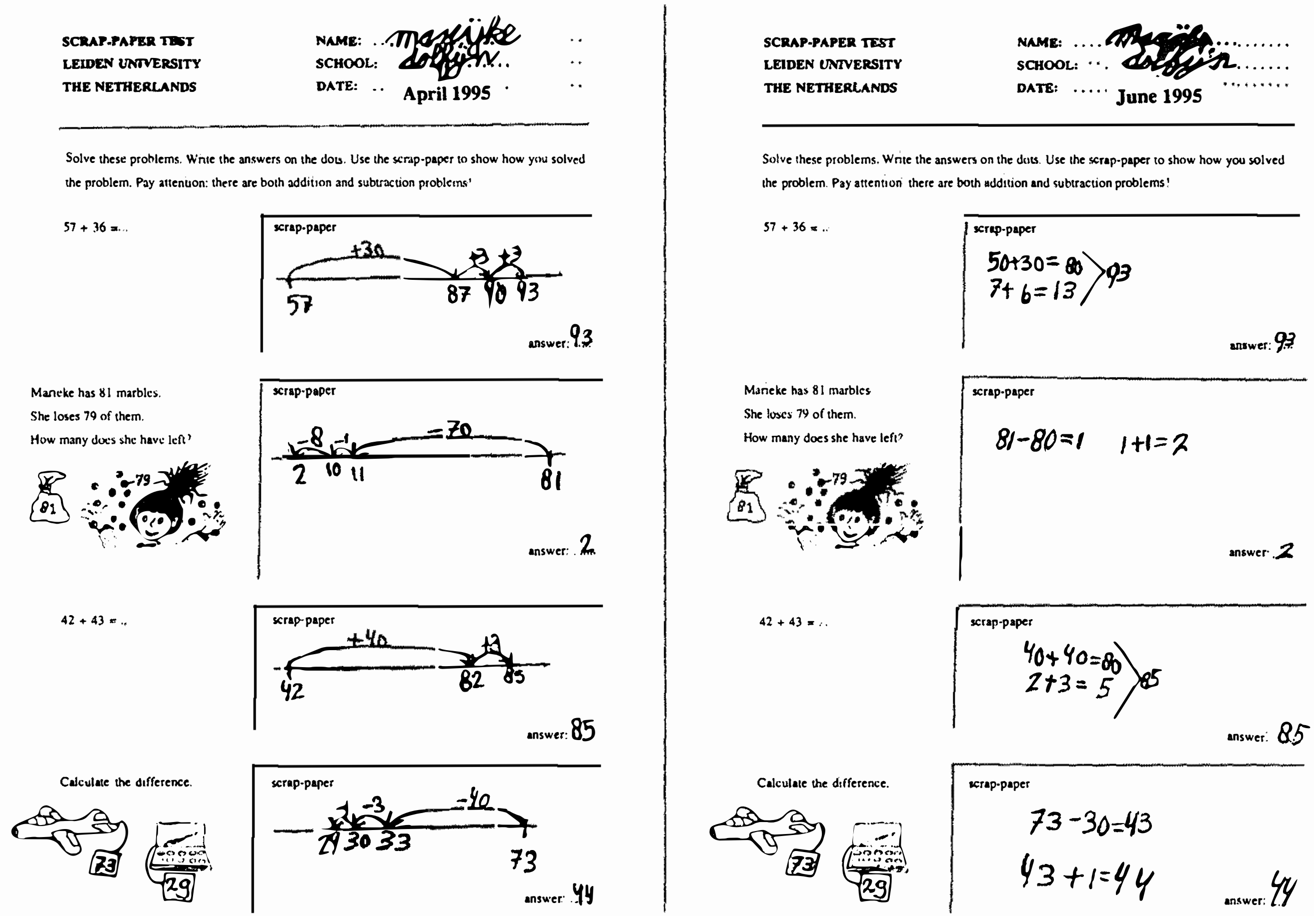


Figure 2.10 Example of Marijke who solved the first page of the Arithmetic Scratch Paper Test in April (left) and June (right)

students were using the more abstract arrow scheme and solution steps instead of the number line, to solve the different problems. An example of how this was done is shown in Figure 2.10.

In Figure 2.10 we see how Marijke solved some problems on one of the tests that were administered in April and June (see also chapter 3). In April she used the number line to solve the problems. For addition problems the N10 procedure is used in April. In June she used the arrow scheme to write down her solution steps and then she also used the 1010 procedure to solve addition problems. For subtraction problems she kept using the N10 procedure.

In the previous periods we paid a lot of attention to relating procedure use to number characteristics of the problems. This was done for bare formula addition and subtraction sums as well as for context problems. To facilitate communication

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
37 + 37 = ...	S	30+30 = 60, 7+7 = 14 (74)
48 + 25 = ...	G	48+20 = 68+2+3=(73)
64 - 29 = ...	SPV	64-30= 34 + 1 =(35)
92 - 87 = ...	∩	87+5=92 (5)

Figure 2.11 Example of how the different labels are introduced for the different procedures. G: N10; S: 1010; SPV: N10C; ∩: Connecting Arc

about these different procedures and strategies, we introduced labels for each procedure during in this period of the curriculum. The way how this was done is shown in Figure 2.11.

On the worksheet we showed different children solving problems in different ways and we introduced different labels for these solutions. The students were then asked how they solved the problem. To make students aware of looking first at a problem before solving it, the students were asked to write down the label of the procedure they would use to solve the problem. 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 name of the strategies they would use to solve the problems. Both to our and to their teachers' surprise, most second graders learned quite easily to use these labels in an adequate way.

2.2 Gradual Program Design

Theoretical framework

Compared to the Realistic Program Design, the Gradual Program Design has a more traditional psychological view towards knowledge acquisition and instruction (cf. Glaser & Bassok, 1989). However in the GPD there is more emphasis on the different aspects of processing and use of solution strategies, than the one-sided task analytic approach which dominated instructional psychology in the 1970s (cf. Gagné, 1977, Resnick, 1983). An important difference compared to the RPD, and also with recent cognitive psychological theories like constructivism, is that the GPD does not use students' informal strategies as a starting point. Instead more emphasis is laid on the procedures students need as a prerequisite to learn addition and subtraction up to 100. This prerequisite knowledge is introduced gradually or stage-wise as in Neo-Piagetian theories about development of knowledge (cf. Case, 1992; Case & Griffin, 1989; Demetriou et al., 1992). Such a design principle was translated into GPD as that the sizes of the numbers should increase more gradually over time and that addition and subtraction problems that require crossing tens (for instance 48 + 36, 51 - 49) were introduced later than in the RPD (cf. Table 3.14). Prerequisite knowledge and conditional relations are seen as very important and also the limited capacity of the working memory plays a central role. Trying to avoid too much cognitive demand on the working memory (cf. Baroody & Ginsburg, 1986) is seen as a driving mechanism for children to choose for a certain solution strategy or procedure. This point of view is also advocated in the Gradual Program Design.

Model of the GPD

Initially (Boekaerts & Beishuizen, 1991) it was proposed that the GPD should be designed according to the developmental model for addition and subtraction up to 100 which was developed in earlier research (Beishuizen, Felix, & Beishuizen, 1990; Felix, 1992). To teach the children the N10 procedure for arithmetic up to 100 a model was formulated to describe the development of addition and subtraction strategies up to 100. Beishuizen et al. (1990) proposed a hierarchical order of the 1010, 10s and N10 procedure. According to these authors, the 10s procedure can be seen as a *by-pass* to reach the N10 procedure. As we said before (cf. Table 1.1) it is difficult to solve subtraction problems with regrouping by using the 1010 procedure. To deal with those problems you need a intermediate step in the 1010 procedure: 62 - 28 = ..; 60 - 20 = 40; 40 + 2 = 42; 42 - 8 = 34. Because of its sequential nature, the 10s procedure approximates to the N10 procedure.

The spontaneous use of the 10s procedure was observed with second grade students by Beishuizen (1993) but recently also by Fuson et al. (1997) and Carpenter (1997). Not every pupil abbreviated this procedure towards the N10 procedure. In those cases, according to Beishuizen et al. (1990) the 10s procedure can be seen as a final stage and a continuation of the 1010 procedure. However, at the moment that this abbreviation towards the N10 procedure does take place, the 10s procedure can be seen as a *by-pass* to learning the N10 procedure.

Beishuizen et al. (1990) have further developed the model by describing the different stages children can go through while developing their arithmetic procedures up to 100. In a longitudinal research project they found empirical evidence for this developmental model. Analyses of students' protocols, who had been working with a computer program that could diagnose the procedures students used while solving different problems (Felix, 1992; Klein & Beishuizen, 1989), showed that the students were very consistent in using the procedure they had chosen. Also the solution behavior of a majority of the students could be characterized according to the different stages that could be distinguished within the developmental model. For a more extensive overview we refer to Beishuizen et al. (1990).

Felix (1992) used this developmental model to teach a group of students the N10 procedure via the 10s procedure by using a computer coach. A second group of students was taught the N10 directly by using a different computer coach. To his surprise the direct teaching of the N10 procedure appeared to be at least as effective as the teaching of the N10 procedure via the *by-pass* of the 10s procedure. This last type of instruction even caused confusion and a relapse towards the 1010 procedure for a number of students.

The developmental model, as described above, served as a starting point for the outline of the first version of the Gradual Program Design (Torn & Ruyters, 1992) which was implemented and evaluated in two second grade classes of a primary school in Leiden. For an extensive overview of this study, we refer to Torn and Ruyters (1992) and Klein (1995). Here only brief mention will be made of the most important outcomes of this first try-out.

During the course year 1991/1992 62 second graders of a primary school in Leiden used the Gradual Program Design, based on the developmental model described earlier for arithmetic up to 100, instead of their normal arithmetic textbooks. At the beginning of the second grade the 1010- en the 10s procedure was introduced for solving addition and subtraction problems. After several months the transition towards the N10 procedure took place. The introduction of the 10s procedure, as an adaptation of the 1010 procedure for subtraction problems with regrouping, caused a lot of numerical and conceptual errors (Klein, 1995). Similar results were found by Fuson et al. (in press) and Radatz (1993). The 1010 procedure seemed to link up well with the previous knowledge children had about the structure of numbers (tens and units). On the other hand, the N10 procedure also links up with informal strategies like counting and ordering strategies, which children have at the beginning of the second grade for the number domain 0-20. With the use of the earlier described developmental model (Beishuizen et al., 1990) as an instruction model in the GPD it was not sufficiently taken into account that the N10 procedure requires an substantially different approach (ordinal versus cardinal) to addition and subtraction up to 100 (cf. Cobb et al., 1995; Gravemeijer, 1994; Greeno, 1992; Lawler, 1990). The introduction of the N10 procedure in the GPD passed off much easier than expected. One of the reasons for this was that the number line appeared to be a very powerful instruction model for learning the use of the N10 procedure. Together with sufficient training in orientation on the (empty) number line in the number domain 0-100, many problems which are frequently mentioned by other authors (cf. Fuson, 1992) did not occur. This made us decide to change the outline of the Gradual Program Design. This new version of the GPD was compared to the earlier described RPD. In the next paragraph we will describe the new outline as well as the most important features of the revised GPD.

Introduction of the number line

As in the RPD, we started at the beginning of the second grade in the GPD with a structured number line up to 20. After 8 weeks a structured number line up to 50 was introduced. This differed from the RPD which already started with a structured number line up to 100 at that time. The empty number line was introduced after 18 weeks working with the GPD, which was 6 weeks later than in the RPD. The use of marks on the number line in the GPD also differed from the RPD. The GPD number line had up to 20 featured marks and numbers for the fives and the tens where the RPD number line only had marks and numbers for the 0 and 20 (for the five and tens only marks are given). We saw the same pattern with the GPD number line up to 50: marks and numbers for every ten where the RPD number line up to 100 only had marks and numbers for the 0 and 100 (for the other tens only the marks are given). In sum the marking on the number lines in the GPD was more prescribed and structured than in the RPD where the students could decide for themselves whether they write down the numbers below the number line or not. However, the empty number line in the GPD had the same format as in the RPD: it did not feature any marks so that the children could draw number marks and jumps for themselves. In the GPD the structured number line up to 20 and 50 were introduced with manipulatives in a number track (see also Figure 2.12).

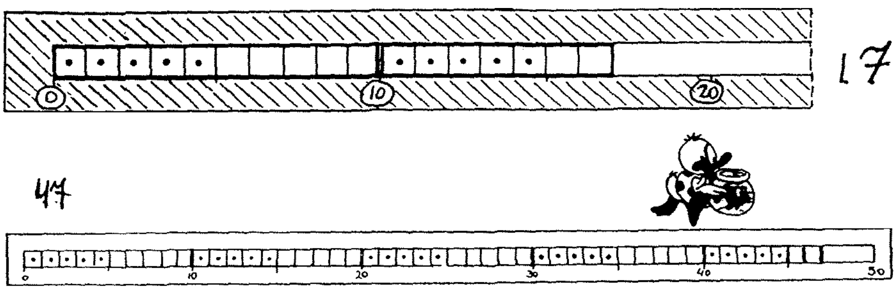


Figure 2.12 Example of the manipulatives with the five-structure in a number track up to 20 and 50

The well known MAB-rods were used in a linear way by using a ruler with a slot in the middle in which the MAB-rods can be put one after the other. MAB rods used in this way will not enhance the 1010 procedure as Beishuizen found in earlier research (Beishuizen, 1993). The MAB rods as used in the GPD differed from the genuine MAB rods. We painted the rods according to the *family* they belong to. Using colors helped the children to recognize the different numbers (cf. Cuisenaire rods). We distinguished the *orange family* (2, 4 and 8), the *yellow family* (3, 6 and 9), the *dark-brown* 7 and the *light-brown* 5 which belongs to the *white family* of 1 and 10. As with the bead string we also used the five-structure to keep children from counting one by one. We put five black dots on the rods of five, six, seven, eight, nine and ten (see also Figure 2.12). This five-structure was also used when we started with adding and subtracting up to 50. Here the bead string in the RPD changed toward the 10-structure. The bead string also went up to 100 where the manipulatives in the number track only went up to 50.

We made a choice for the manipulatives, and not for the bead string, because we hypothesized that the children would keep on counting the beads one-by-one and would not make the step towards recognizing whole numbers (Boekaerts & Beishuizen, 1991). We also thought that the manipulatives offered more structure and are more concrete (every number has a different rod) than the bead string. In this way the use of the manipulatives in a number track comes closer to the character of the GPD.

Mental arithmetic

As in the RPD, mental arithmetic also has a central place in the GPD. Mental arithmetic is seen as *using one's head* instead of doing arithmetic *in your head*. In this way students were also allowed to write down their solution steps by using the number line and later on the arrow scheme or just solution steps.

During the first half of the GPD much emphasis was put on practicing the N10 procedure. Compared to the RPD less time was spent on talking about different solution strategies and developing number sense by games like "*Raad mijn Getal*" (Guess my number). Instead more time was spent on written number exercises like splitting up the 10 in two different numbers (for a more detailed overview see chapter 3). Compared to the RPD less emphasis was laid on doubling strategies and no attention was paid to strategies other than N10 like, for instance, the Connecting Arc and N10C. You could say that mental arithmetic during the first half of the GPD came closer to *drill and practice* than in the RPD. In this way a firm procedural knowledge base was established for sums up to 50.

During the second half of the GPD (from January on) the perspective on mental arithmetic changed and came closer to the way mental arithmetic is seen in the RPD. More time could be spend on talking about different solution strategies for numbers up to 100. Like in the RPD, labels were introduced for the different solution strategies to facilitate communication and make students aware of which procedure is the most efficient according to the number characteristics of the problem. From January till April mental arithmetic in the GPD was primarily done by solving addition and subtraction problems with paper and pencil where in the RPD talking about different strategies and developing number sense remained more crucial (for a more detailed overview see chapter 3). From April onwards the GPD and RPD were more or less the same.

Role of the context problems

The role of context problems in the GPD differed from the role played by context problems in the RPD. In the RPD context problems were used to elicit informal strategies of the students. These problems served as a starting point of the mathematization process. Since in the GPD informal strategies did not have the same function as in the RPD (at least during the first half of the GPD), context problems were also used differently. First the children had to practice the N10 procedure with numerical problems. After sufficient practice they could apply this procedure in real-life context problems. In this respect context problems were used more traditionally and in a more pre-structured way than in the RPD. During the second half of the GPD more attention was paid to talking about different solution strategies. From April on context problems were used at the introduction of a lesson, to elicit different strategies. You could say that the role of context problems in the GPD changed from traditional towards *realistic*.

Role of the teacher

The role of the teacher in the GPD changed over time. During the first half year of the GPD the teacher was the one who decided what the children should do. Compared to the RPD there was less time for interaction and whole-class instruction because more time was spent on paper and pencil work. During the second half of the program the role of the teacher changed towards that of a coach. For the description of this role we refer to the paragraph about the role of the teacher in the RPD. This transition was not immediately clear to every GPD teacher in January. This was illustrated by one of the GPD teachers who said to one of the researchers in January: "Now that the children know the standard procedure, I can leave them more gradual than in the RPD where we immediately went up to 100. The structured number line up to 100 and the empty number line were introduced in the RPD in this period.

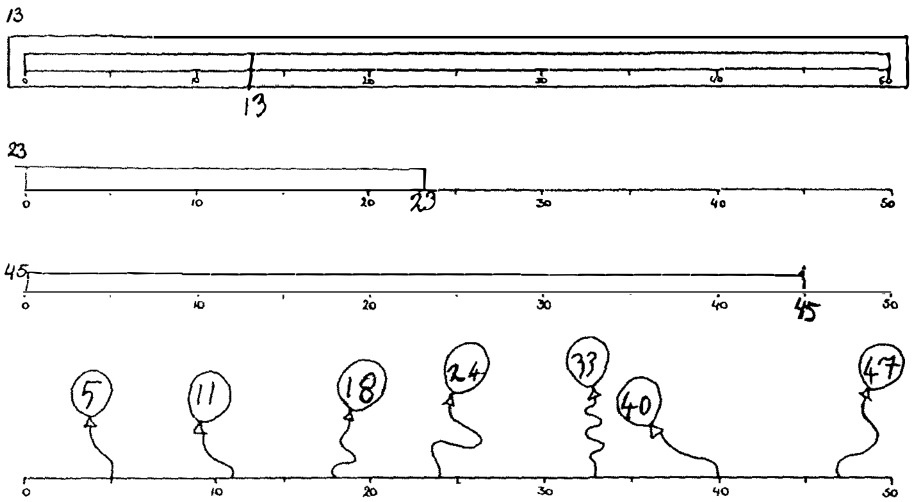


Figure 2.14 Transition of manipulatives in 50-number track towards the semistructured number line up to 50

We started with the positioning of numbers on the number line up to 50. The students were asked to count aloud forwards and backwards from one number to another number and to make 5- and 10-jumps (15, 25, 35, ...). Knowing these 10-jumps by heart was a prerequisite for successfully using the N10 procedure. As in the RPD games like "*Raad mijn Getal*" (Guess my number) were played with numbers from 0 to 50 (for more information about this game see RPD, week 10-17). The exercise of jumping from one number to another number on the number line (see Figure 2.7) appeared to be very useful in stimulating number sense in this number area (Klein & Beishuizen, 1993).

The introduction of addition and subtraction problems up to 50 started with sums where single-digit numbers had to be added and subtracted (for example 22 +5 , 37 + 6, 37 - 3, 32 - 8). For the sums that go through the ten the earlier mentioned CS strategy (completing up or down until the nearest ten) was taught. For sums with multi-digit numbers up to 50 only the N10 procedure was introduced. This was first done with sums like 25 + 10, 25 + 20, 45 - 10, 45 - 20. Later problems like 25 + 13 and 45 - 23 were introduced. The difficulty of crossing a ten was postponed until week 18. During this period there was more emphasis too on the training of written procedures than talking about different solution strategies. The splitting of numbers was practiced very frequently in exercises on the worksheets.

Week 17 - week 24: introduction arithmetic up to 100 and empty number line

In this period we began with addition and subtraction problems up to 100. Students had to add and subtract tens (55 + 30, 85 - 20) but soon the most difficult sum types with regrouping had to be solved (47 + 35, 62 - 37). At the beginning of this period the structured number line up to 100 was introduced. After some weeks the empty number line was introduced as in the RPD (cf. Table 2.2) followed by the arrow scheme as a more abstract way of writing down the solution steps of a problem. The final stage was just writing down the solution steps. At the end of this period the students were free to choose one of these notation forms.

During the first 4 weeks of this period, only the N10 procedure was taught for the solution of the problems. In week 21 we started for the first time with the introduction of another procedure apart from the N10 procedure. Firstly the N10C procedure and later the A10 procedure was introduced. The introduction of these procedures did not take place through the use of context but with numerical problems (see Figure 2.15). As explained before, context problems in the GPD were used after the procedures had been learned and not to introduce new strategies.

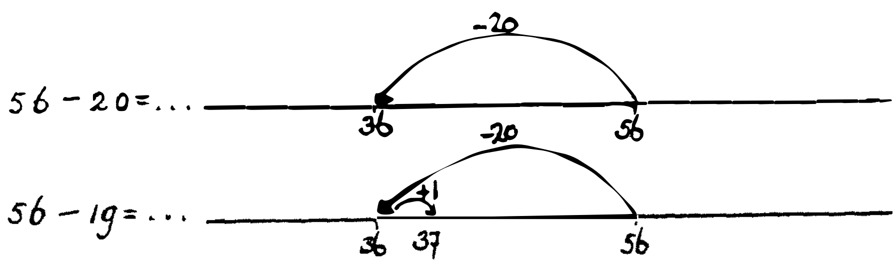


Figure 2.15 Example of how N10C is introduced in the GPD

The difference between the Direct Subtraction strategy and the Adding-On strategy was explained in this period. This was done in the same way as in the RPD which meant that context problems were used in which students had to calculate a difference (cf. Figure 2.9). We thought that it would be too artificial to explain this difference in strategy by starting with numerical problems. During this period whole-class discussions became more and more important but accounted for less time than in the RPD. Conversely, the worksheets contained more paper-and-pencil sums and exercises than the worksheets used in the RPD condition. During the last two weeks of this period the introduction of the 1010 procedure was prepared by using the context of money. This was done in the same way as in the RPD.

Week 25 - week 32: introduction of 1010 for addition, labels for different procedures

The last eight weeks of the GPD both the worksheets and the teacher manuals were almost identical to the RPD. That means that much time was spent on making the students sensitive to the number characteristics and the structure of different problems, not only with worksheets but also by whole classroom discussions. The earlier mentioned labels were introduced for the different procedures and strategies (cf. Table 1.1). The only difference with the RPD was that the Connecting Arc was not introduced to the GPD students and therefore no label for this procedure was introduced. This meant that most GPD students used the N10 procedure for solving a problem like 71 - 69 where most of the RPD students would chose the Connecting Arc. However there were some GPD students who invented this procedure by themselves (see Figure 2.16). To what extent this was done by the GPD students will be shown in the next chapters.

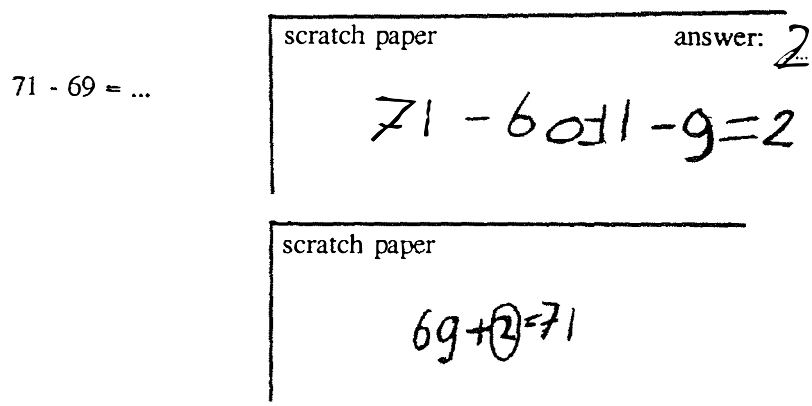


Figure 2.16 Two different GPD students solving a problem on the Arithmetic Scratch Paper Test in June

2.3 Research questions

In the preceding sections we described the theoretical backgrounds of this study as well as the contents of the Realistic and Gradual Program Designs. With these diferent settings in mind we can formulate hypotheses about the results of these two program designs. In formulating these hypotheses we used the Hofstee's (1982) *bet-model*. According to this model a scientific comparison is made between two or more hypotheses or predictions. Here hypotheses have been formulated by two parties, each of them representing a different point of view. The two parties have to bind themselves to their predictions before the experiment starts. After the experiment is over, these predictions are tested towards *a reality*, in our case the outcome of the experiment. Hofstee (1982) claimed that in this way a control-group is no longer indispensable. The outcome for the RPD condition was formulated by Treffers, one of the proponents of RME designs, while Beishuizen formulated hypotheses for the GPD. To make the bet as clear as possible, some predictions were formulated rather extremely. The hypotheses were written down in the research proposal before the experiment started (Boekaerts & Beishuizen, 1991). These hypotheses can be subdivided into 3 main clusters: 1) a cluster of hypotheses formulated from a realistic point of view by Treffers, 2) a cluster of hypotheses from a gradual point of view by Beishuizen, and 3) post hoc questions formulated by Klein.

Treffers and Beishuizen formulated hypotheses for the results of the two program designs concerning development of procedural and strategic knowledge (hypotheses 1-4), results for weaker and better students (hypotheses 5-8), development of motivational processes (hypotheses 9-12). Because of the differences in instructional sequence for the two program designs they formulated hypotheses for half-way through the program, in January, and the end of the program in June.

The predictions made by Treffers and Beishuizen are the most important to be answered. However other questions arose during the research project. Three of them are formulated by Klein as post hoc questions. These questions concern motivational processes for weaker and stronger students (hypothesis 13), possible transfer from what the students learned for addition and subtraction in the number domain of 0-100 to the number domain 0-1000 (hypothesis 14) and retention of the strategies and procedures some months after they have worked with the experimental program (hypotesis 15). Table 2.3 gives an overview of hypotheses formulated by Treffers, Beishuizen and Klein.

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
Development of procedural and strategic knowledge	
1. Half-way through the program there will be no differences in the number of correctly solved problems for RPD and GPD students. However there will be a difference in the type of errors students make: RPD students will make less procedural or conceptual mistakes than GPD students, due to a better insight and adaptation of the solution strategy or computation procedure towards the structure or number characteristic of a problem. RPD students will make more non-procedural mistakes (due to slovenliness) than GPD students. The GPD students will solve more numerical problems in a limited amount of time than RPD students. GPD students will only use the N10 procedure in a proceduralized way where the RPD students use different strategies and procedures. RPD students will adapt their strategy use to the characteristics of a problem and therefore, both the flexibility in use of computation procedures and solution strategies and the number of correctly solved context problems will be higher for RPD than for GPD students.	2. Half-way through the program there will be a higher level of procedural competence in numerical problems shown by the GPD rather than the RPD students which will be reflected by a higher number of correctly solved problems. The GPD students will solve more of these problems in a limited amount of time (speed test) as a result of more procedural practice in solving such problems. With respect to context problems the GPD students will solve fewer problems correctly than the RPD students. With respect to flexible use of different strategies and procedures, the GPD students will be more rigid than the RPD students. The GPD students will stick to the N10 procedure where the RPD students will also use other procedures as well.

Table 2.3 Hypotheses formulated by Treffers, Beishuizen and Klein about the results of this study

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
Development of procedural and strategic knowledge (continued)	
3. At the end of the program the RPD students will show a higher level of procedural competence in solving the most difficult sum type (subtraction problems which require regrouping) than the GPD students. This is the result of a better understanding of these problems and a more adequate adaptation of the strategy used to solve these problems. For the other sum types there will be no differences in the number of correctly solved problems between the two groups of students. At the end of the second grade the GPD students lag behind the RPD students with respect to the flexible use of solution strategies and computation procedures, both for numerical and context problems.	4. At the end of the program the GPD students will still have a higher level of procedural competence than the RPD students with respect to standard number problems. This will be reflected by a higher number correctly solved problems and a higher number of correctly solved number problems within a limited amount of time (speed-test). The GPD students will also solve correctly as many context problems as the RPD students, because more attention is paid to these problems during the last months of the GPD. In the GPD flexible strategy use is also emphasized in the last part of the program and therefore the use of these solution strategies and computation procedures will be less rigid than half-way through the program. At this time we expect there will be no differences between the GPD and RPD students, with respect to the flexible use of different solution strategies and computation procedures.

Table 2.3 (Continued) Hypotheses formulated by Treffers, Beishuizen and Klein about the results of this study

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
Results for weaker and better students	
5. Half-way through the program there will be significant differences between weaker and better RPD students, however, this will be more on a strategic than on a procedural level. Weaker RPD students will often use solution strategies and computation procedures in a non-abbreviated or inefficient way, but these will bring them to the correct answers. The weaker GPD students will use the N10 procedure in a proceduralized way without understanding what they are doing.	6. The expected higher level of procedural competence for the GPD students, half-way through the second grade, will mainly be caused by the relatively better scores of the weaker students. They will have benefited of the gradual and structured approach of the GPD. The differences between the better and weaker students will be significant larger in the RPD. In the RPD emphasis is laid on flexibility from the start of the second grade, which will cause many inadequate inventions or combinations of solution strategies and computation procedures (for instance confusion in the execution of the N10C procedure for addition and subtraction problems). So weaker RPD students will be more flexible in using different strategies and procedures, but, compared to the weaker GPD students, this use will be of a lower quality in both a strategic and a procedural sense half-way through the program.
7. At the end of the program the quality of strategy and procedure use will have increased for the weaker RPD students. The situation will be the same for the weaker GPD students. They will not be amenable to the adoption of new solution strategies or computation procedures, because they will stick to the use of the N10 procedure.	8. At the end of the program the results for the weaker and better students will be the same as predicted from this point of view half-way through the second grade.

Table 2.3 (Continued) Hypotheses formulated by Treffers, Beishuizen and Klein about the results of this study

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
Development of motivational processes	
9. The RPD can be characterized with a process oriented view. The answer to a problem is not the main issue, but the way you solved the problem is more important. The use of self-invented and informal strategies is encouraged. This will make the RPD students feel involved and give them pleasure in arithmetic which will result in a higher motivation towards mathematics than the GPD students. RPD students will have more favorable cognitions and affects towards both number and context problems than GPD students, half-way through the program.	10. Half-way through the program, GPD students will have more favorable cognitions and affects towards numerical problems than RPD students. This is expected because during the first half year of the GPD, much attention is paid to the procedural competence of addition and subtraction up to 100. Emphasis is put on solving numerical problems. Context problems are less frequent in this period and only used to apply the learned procedure in a real-life situation. Therefore the opposite will be true with respect to context problems: RPD students will have more favorable cognitions and affects towards these problems than GPD students. With respect to motivation for mathematics in general (domain-specific) there will be no differences between RPD and GPD students.
11. At the end of the program the RPD students will have a higher motivation towards arithmetic, but the differences between the RPD and GPD students will be smaller than half-way through the program. Secondly, the RPD students will still have more favorable cognitions and affects towards numerical and context problems compared to the GPD students. The differences between the two groups of students will be smaller than half-way through the program.	12. At the end of the program there will be no significant differences between the two programs with respect to favorable cognitions and affects towards numerical problems and context problems. This accounts also for motivation for mathematics in general. During the second half year of the GPD, more emphasis is paid towards context problems and discussing different solution strategies and computation procedures to solve these problems. From April on, the GPD and RPD are more or less the same.
KLEIN (post hoc questions)	
13. Do weaker and better RPD and GPD students differ in their cognitions and appraisals towards arithmetic as a school subject and more specific towards numerical and context problems? 14. Do we see any transfer from what the students learned for addition and subtraction in the number domain of 0-100 to the number domain 0-1000? 15. What is the retention of the strategies and procedures RPD and GPD students have learned for addition and subtraction up to 100, some months after they have worked with the experimental program?	

Table 2.3 (Continued) Hypotheses formulated by Treffers, Beishuizen and Klein about the results of this study

3 Method

3.1 Subjects

Schools which had at least two years of experience with the realistic mathematics textbook *Rekenen & Wiskunde* (Gravemeijer et al., 1983) were invited to participate in our project. According to this criterium, the publisher of the textbook *Rekenen & Wiskunde* provided us with a list of 60 schools in the south-west part of Holland. From the schools which were willing to participate we selected 10 classes in 9 schools according to three criteria (1) less than 25% immigrant students, (2) homogeneous classes with all students about the same grade instead of mixed-age classes, and (3) one teacher instead of two part-time teachers. These schools had a total of 275 students in the second grade (7-8 years).

From the total sample, 100 students were selected to test the predictions concerning the results of the weaker and better students. We selected 49 weaker and 51 better students (25 for each program design) based on the students' scores on the National Arithmetic Test administered at the end of the first grade (CITO LVS E3). The National Arithmetic Test distinguishes 5 levels of competence: A (25% best scoring students), B (25% students scoring just above the national mean), C (25% students scoring just below the national mean), D (15% students scoring well below the national mean), and E (10% students with the lowest scores). Students with a D- or E-score were assigned to the group of weaker students, students with an A- or B-score were assigned to the group of better students. We had difficulties in finding enough D- and E students (probably due to our selection criteria). Therefore we had to select students with a C score to get a group of 25 weaker students for each program design. As a consequence of this we selected more A students for the group of better students so that the difference in arithmetic competence between the two groups was maintained. Table 3.1 shows the number of A-, B-, C-, D- and E students that were assigned to the different groups within the RPD and GPD.

	RPD	GPD
Better students (n=51)	14 A students 11 B students	11 A students 15 B students
Weaker students (n=49)	13 C students 9 D students 3 E students	11 C students 11 D students 2 E students

Table 3.1 Number of A-, B-, C-, D- and E students that were assigned to the groups of better and weaker students for each program design

During the school year some students moved to other schools. At the end of the second grade complete data were available for 23 better and 25 weaker RPD students and 26 better and 23 weaker GPD students.

3.2 Instruments

RPD and GPD textbooks and teacher guides

Teachers and pupils used experimental teacher guides and textbooks instead of their regular mathematics textbooks. Both the RPD and GPD textbooks and teacher guides were developed during 2 years of try-out and revision on a small scale (Klein & Beishuizen, 1993, 1994b). Experimental materials of both the RPD and the GPD concerning addition and subtraction up to 20 and 100 replaced about 75% of the mathematics textbook *Rekenen & Wiskunde*. The regular text was used for instruction on other aspects of the curriculum, such as measurement, tables of multiplication, spatial ordering, and telling the time. The RPD and the GPD differed in many ways accordant with the different theoretical bases on which the two programs were founded. One of the differences between the RPD and GPD is the number of context problems and the way these problems were practiced. Another difference between the RPD and GPD is the amount of time spent on wholeclass discussions in which for instance context problems are discussed. This is reflected by the number of whole class exercises in the RPD and GPD. Table 3.2 provides an overview of the number of context problems and whole class exercises in both program designs during the first and second half of the curricula.

		RPD	GPD
week 1 - week 18	context problems	70	32
	whole class exercises	147	105
week 19 - week 32	context problems	133	117
	whole class exercises	90	84

Table 3.2 Number of context problems and whole class exercises during the first half and the second half of the RPD and GPD

In accordance with the ideas behind the two program designs, the RPD classes spent more time on whole class discussions and exercises than the GPD classes. However, during the second half of the school year these differences between the two program designs became less pronounced.

During the school year teachers had to fill in a concise log-book in which they recorded how much time they spent on their arithmetic teaching and what their experiences were. On average both the RPD and GPD teachers spent 50 minutes on arithmetic teaching each day. For that reason we can conclude that the GPD students spent more time on written exercises to build a firm procedural knowledge base for addition and subtraction up to 100. The RPD students spent more time on whole class discussions.

Test and questionaires

Table 3.3 gives an overview of which tests and questionnaires were administered at the different moments. In the following sections the content of these instruments is described more extensively.

Moments in time	Tests
End of the first grade	CITO LVS E3, Analogies & Categories SON-R
September second grade	AST single-digit addition and subtraction problems < 20
October second grade	AST single-digit addition and subtraction problems < 20 ASMT October, MMQ
January second grade	AST single-digit addition and subtraction problems < 20, <50, <100 ASMT January, MMQ, OMQ (7-10), CITO LVS M4
April second grade	AST single-digit addition and subtraction problems < 20, <50, <100 AST multi-digit addition and subtraction problems <50, <100 ASMT April, ASPT, MMQ, OMQ (7-10)
June second grade	AST single-digit addition and subtraction problems < 20, <50, <100 AST multi-digit addition and subtraction problems <50, <100 ASMT June, ASPT, ATT, MMQ, OMQ (7-10), CITO LVS E4
November third grade	AST multi-digit addition and subtraction problems <50, <100 ASPT, ATT, OMQ (7-10)

Table 3.3 Overview of tests administered to the RPD and GPD students

Tests for abstract reasoning ability

Two subtests from the Snijders-Oomen Non-Verbal Intelligence Test (SON-R) (Laros & Tellegen, 1991) were administered to measure non-verbal intelligence. Because reasoning tests form the core of most intelligence tests we administered two subtests for abstract reasoning: Categories and Analogies. In the Category test the students had to classify objects into categories. In the test three related objects were given. Two related objects had to be chosen from five other objects. The test consisted of 21 problems and three parallel versions were used. The reliability coefficient (internal consistency) for the three versions was .78. In the Analogy test a pair of related geometrical figures was given. From a number of alternatives, a second pair analogous to the given pair must be formed. The task consisted of 24 problems and three parallel versions were used. The reliability coefficient (internal consistency) for the three versions was .84.

Tests for procedural competence

The development of procedural competence in arithmetic skills was measured with an *Arithmetic Speed Test* (AST) which was developed for this project (Klein & Beishuizen, 1995a). 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 with single-digit (SD) numbers (8 + 5, 36 + 6, 65 + 9; 12 - 4, 43 - 6, 76 - 9) and with multi-digit (MD) numbers (27 + 14, 57 + 19; 44 - 26, 85 - 49). We constructed different subtests for addition and subtraction problems with SD numbers and MD numbers. Reliability coefficients (internal consistency) for all these tests were always higher than .85.

Tests for strategic competence

The *Arithmetic Subject Matter Tests* (ASMT) were also developed for this project (Klein & Beishuizen, 1995b). These tests were used to investigate the development of the flexible use of computation procedures and solution strategies and were comparable for the two program designs. However, they differed over time because the subject matter of the preceding 8 weeks was the content of these tests (see Table 3.14 for what was taught in the preceding weeks). The ASMT always consisted of comparable numerical addition and subtraction problems which had to be solved under three conditions: (1) by head, (2) using the number line, and (3) in a non-standard context format with use of the number line. In this way we could investigate how capable students were in solving addition and subtraction problems by head. By comparing these results with the outcomes of the problems that could be solved using the number line we were able to detect the effect of using the number line while solving a problem. In January we took into account the size of the numbers (<50 and <100) since here the RPD and GPD differed in the number size of the problems that were introduced to the students. The format of the number line differed at this moment between the two program designs (semi-structured versus empty) which was reflected in the January test (cf. Table 3.14). In April the ASMT was the same for both program designs. In June the addition and subtraction problems of the test administered in April were tested again. The numbers used in the problems were chosen to elicit specific computation procedures. Besides the spontaneous use of different computation procedures and solution strategies, we were interested in what procedures and strategies students would use if they were asked to solve some context problems in two different ways. This *flexibility on demand* was only requested during the ASMT in June. Tables 3.4, 3.5, 3.6 and 3.7 give an overview of the most important problem types, their different formats and expected procedures of the ASMT in October, January April and June[1].

[1] We only mention the problem types that are important for the hypotheses mentioned in chapter 2, concerning the RPD and GPD. For a complete overview of the tests and the analyses of all the different problem types we refer to Klein (1997).

By Head	Semi-structured number line	Expected procedure
8 + 4 and 12 - 3	7 + 5 and 11 - 4	Complementary Structuring/Complementary Structuring
7 + 9 and 14 - 8	6 + 9 and 12 - 9	Complementary Structuring/∩[a]
9 + 3 and 13 - 4	8 + 3 and 13 - 5	Complementary Structuring/Complementary Structuring
6 + 8 and 11 - 9	4 + 9 and 11 - 8	Complementary Structuring/∩[a]

a The Connecting Arc (∩) is only expected here for subtraction problems and only for RPD students

Table 3.4 Most important addition and subtraction problem types in the Arithmetic Subject Matter Test in October

By head <50	By head <100	Semi-structured number line <50[a]	Semi-structured number line <100[b]	Expected procedures
37 + 4 and 42 - 3	79 + 3 and 71 - 3	36 + 5 and 41 - 4	68 + 3 and 51 - 4	CS CS
19 + 6 and 33 - 6	67 + 6 and 84 - 6	28 + 7 and 22 - 7	74 + 7 and 62 - 7	CS CS
25 + 9 and 24 - 9	46 + 9 and 73 - 9	34 + 9 and 33 - 9	55 + 9 and 75 - 9	N10C N10C
26 + 22 and 45 - 43	54 + 23 and 87 - 85	24 + 21 and 37 - 35	62 + 22 and 75 - 73	N10 ∩[c]
18 + 25 and 31 - 28	67 + 26 and 71 - 68	17 + 26 and 41 - 39	58 + 24 and 62 - 59	N10 ∩[c]

Note. CS stands for Complementary Structuring (cf. Figure 2.5), other labels are explained in Table 1.1 The problems that had to be solved on the empty number line are not mentioned here because the empty number line was not yet introduced to the GPD students.
a Only for GPD students because otherwise the test would become too long for RPD students
b Only for RPD students because problems with numbers >50 were not introduced in the GPD in January (cf. Table 3.14)
c The Connecting Arc (∩) is only expected for subtraction and difference problems and only for RPD students

Table 3.5 Numerical addition and subtraction problems in the Arithmetic Subject Matter Test in January

Context problems	Expected procedure	Difference problems	Expected procedure
36 + 8 and 46 - 7	CS/CS	7 and 43	CS
9 + 28 and 43 - 41	CS/∩[a]	38 and 35	∩[a]
12 + 34	N10		

Note. CS Stands for Complementary Structuring (cf. Figure 2.5), other labels are explained in Table 1.1 For these problems only the numbers and not the story and picture of the context problem are presented
a The Connecting Arc (∩)is only expected for subtraction and difference problems and only for RPD students

Table 3.6 Non-standard context problems in the Arithmetic Subject Matter Test in January

By Head	Empty Number Line	Context problems	Expected procedure	Context problems: 2 ways[a]	Expected procedures
58 + 34 and 82 - 35	47 + 36 and 92 - 34	58 + 33 and 73 - 25	1010 N10	34 + 32 and 84 - 65	1010, N10 N10, A10
6 + 78 and 71 - 68	8 + 76 and 71 - 68	7 + 68 and 81 - 78	CS ∩[b]	55 + 19 and 73 - 39	N10C, N10 N10C, N10
45 + 29 and 63 - 29	55 + 19 and 53 - 19	43 + 29 and 72 - 39	N10C N10C	16 + 78 and 74 - 56	N10, A10 N10, A10
15 + 67 and 94 - 75	16 + 57 and 74 - 56	15 + 67 and 84 - 65	N10 N10		

Note. CS Stands for Complementary Structuring (cf. Figure 2.5), other labels are explained in Table 1.1. For the context problems only the numbers and not the story and picture of the problem are presented.
a These problems were only administered in June
b The Connecting Arc (∩)is only expected for subtraction and difference problems and only for RPD students

Table 3.7 Numerical addition and subtraction problems in the Arithmetic Subject Matter Tests in April and June

The procedures could be detected for the problems in which the students were allowed to use the number line to solve the problem. For the problems that had to be solved by head only the answer and type of error could be detected. Reliability coefficients (internal consistency) for the ASMT were always higher than .70.

The *Arithmetic Scratch-Paper Test* (ASPT) was also developed for this project (Klein & Beishuizen, 1995c). The major difference with the ASMT was that the students were now free in choosing a way to write down their solution steps in *scratch-paper boxes* (see Figure 2.10). These scratch-paper boxes 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 and the reliability coefficient (internal consistence) was both in April and in June .79. Three addition and five subtraction problem types were presented with comparable numbers in two formats: as numerical and as context problems. For the context problems we chose problems of the *change* type (Klein & Beishuizen, 1994a; 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). These 5 *difference problems* were offered in context format. The 21 problems were presented to the pupils randomly to control for set effects. Table 3.8 provides an overview of the different problem types.

Addition	Expected Procedure	Subtraction	Expected Procedure	Difference	Expected Procedure
N[a] 57 + 36 C[b] 48 + 37	N10 N10	N[a] 75 - 36 C[b] 84 - 26	N10 N10	C[b] 74 and 36	N10 or A10
N[a] 42 + 43 C[b] 33 + 34	1010 1010	N[a] 65 - 33 C[b] 85 - 42	N10 N10	C[b] 65 and 32	N10 or A10
N[a] 54 + 39 C[b] 54 +2 9	N10C N10C	N[a] 84 - 29 C[b] 63 - 29	N10C N10C	C[b] 73 and 29	N10C or A10
		N[a] 71 - 69 C[b] 81 - 79	∩[c] ∩[c]	C[b] 61 and 59	∩[c]
		N[a] 62 - 48 C[b] 72 - 58	∩[c] ∩[c]	C[b] 82 and 68	∩[c]

Note. For the context format, only the numbers and not the story and picture of the problem are represented.
a N stands for numerical; b C stands for context; c The Connecting Arc (∩) is only expected for subtraction and difference problems and only for RPD pupils.

Table 3.8 Addition and subtraction problem types and expected procedure in the Arithmetic Scratch Paper Test

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 1.1). For the subtraction problems we expected the N10 and N10C procedures. The Connecting Arc (∩) was expected for the last two items (only for RPD pupils) because the difference between the numbers of these subtraction problems is 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 3.14). For the difference problems we expected the N10, N10C, and A10 procedures, and for the last two items we expected, again, the Connecting Arc (∩) (only for the RPD students).

Test for transfer

The *Arithmetic Transfer Test* (ATT) (Klein & Beishuizen, 1995d) was developed to investigate if there was any transfer from what the students had learned in the domain of addition and subtraction up to 100 to the domain of addition and subtraction up to 1000. The test consisted of 3 numerical addition problems and 3 numerical subtraction problems for which the students were asked to write down their solution steps in a *scratch-paper box* as with the ASPT. Two non-standard context problems were administered, which were based on the *Kino* problems developed by Selter (1994). Table 3.9 gives an overview of the problems of the Arithmetic Transfer Test.

Numerical addition	Numerical Subtraction	Non-standard context
330 + 200	301 - 298	235 + 124[a]
450 + 110	404 - 395	285 seats 143 occupied
225 + 124	368 - 234	

Note. For the context problems, only the numbers and not the story and the picture are represented

Table 3.9 Numerical addition and subtraction and non-standard context problems of the Arithmetic Transfer Test

External criterion tests

As an external criterion for the students' performance, *the Student's Monitoring Tests for Arithmetic and Mathematics* (LVS E3, M4, and E4)) (Janssen, Bokhove, & Kraemer, 1992) developed by the National Institute for Educational Measurement (CITO) were administered.

Questionnaires for measuring domain specific motivational beliefs

Students' motivation with respect to arithmetic was assessed by the Mathematics Motivation Questionnaire (MMQ) (Blöte, 1993; Voogt, 1996) which consisted of 3 subscales: affect towards mathematics, self-concept of mathematics ability and effort you are willing to invest in doing mathematics. All items were Likert-type scales. Example items of each of the three subscales, number of items, and Cronbach's alphas are printed in Table 3.10.

1. Affect - I think arithmetic is very boring (boring, not boring, not boring at all) - I like doing sums very much (much, not much, not at all)	7 items α .94
2. Self-concept - In arithmetic, I am doing much better (better, as well, less) than the other children in my class - I know heaps (a lot, a little, not so much) about arithmetic - When I am doing sums I know really well (well, not so well, not at all) what I have to do	7 items α .85
3. Effort - When I am doing sums I am working very hard (hard, not so hard, not hard at all) - During arithmetic lessons I very often (often, sometimes, never) day-dream - I never (sometimes, often, always) check the outcome of my sums	7 items α .65

Note. The Cronbach alphas are the average alphas for the MMQ administered in October, January, April and June

Table 3.10 The three subscales of the Mathematics Motivation Questionnaire (Blöte, 1993; Voogt, 1996), example items and Cronbach's alphas

Questionnaires measuring task specific cognitions and affects

The On-line Motivation Questionnaire (OMQ) was developed by Boekaerts (1987) to obtain students' perceptions about relevant aspects of the learning situation during actual learning tasks. The OMQ was initially developed for sixth grade students. Boekaerts (in preparation) adapted the OMQ to the phenomenological world and language use of second and third grade students. This instrument was labeled OMQ (7-10) and consisted of 3 scales that measure the students' cognitions and affects *before* they start on a curricular task and 2 scales that measure their cognitions and affects *after* task completion. Similar to the original OMQ, all items are Likert-type scales. We were interested if the RPD and GPD students differed in their cognitions and appraisals toward numerical problems and non-standard context problems. Therefore the students were given two assignments separated by approximately two weeks. The first assignment (administered in the same week as the tests mentioned before) was a set of three context problems. The second assignment was a set of six numerical problems. The students had to complete the OMQ (7-10) before and after solving these two types of problems but they had a chance to glimpse at the problems so that they knew what type of assignment they had to do. Example items of each of the five scales, number of items, and Cronbach's alphas are printed in Table 3.11.

1. Self-confidence - How difficult is this kind of task for you? - Do you think you can solve this problem? - How well can you do this kind of task?	6 items α .81
2. Task attraction and positive affects - How much do you like this kind of task? - How happy do you feel now?	4 items α .86
3. Task value and learning intention - How important is it to learn to solve these problems? - How much effort are you going to put in? - How well did you do on the problem?	5 items α .64
4. Effortful accomplishment - How happy do you feel now that you have done the problem? - How much did you like working on the problems? - How much effort did you put in?	5 items α .70
5. Absence of threat - How tired are you now? - How worried do you feel now? - How difficult did you find the problems?	5 items α .51

Note. The Cronbach alphas are the average alphas for the OMQ administered in January, April, and June

Table 3.11 The five subscales of the OMQ (7-10), example items and Cronbach's alphas (taken from Boekaerts, in preparation)

3.3 Procedures

Assigning classes to RPD and GPD

To reduce the possibility of differences in arithmetic competence at the beginning of the experiment we administered the CITO LVS E3 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. In Table 3.12 the mean number of correct answers and standard deviations are given for the two groups of classes, after they were assigned to one of the two programs. The maximum score for this test was 53. There appeared to be no significant differences in arithmetic test scores between the two groups at the beginning of the experiment. We also checked if their were differences in level of abstract reasoning for the RPD and GPD classes. In Table 3.12 also the scores on the two subtests of the SON-R for the RPD and GPD classes are given. There appeared to be no significant differences between the two groups.

	CITO LVS E3		SON-R Categories		SON-R Analogies	
	mean	s.d.	mean	s.d.	mean	s.d.
RPD classes (n=139)	38.3	12.5	10.8	3.5	12.1	5.1
GPD classes (n=135)	37.6	14.3	10.4	3.8	11.8	5.0

Table 3.12 Mean number of correct answers and standard deviations on CITO LVS E3 and the Category and Analogy tests (maximum scores 21 and 24 respectively) of the SON-R

Controlling the selection of weaker and better students

We controlled the selection of weaker and better students by looking at their scores on the CITO LVS E3 and the two subtests of the SON-R. The results are presented in Table 3.13.

	CITO LVS E3		SON-R Categories		SON-R Analogies	
	mean	s.d.	mean	s.d.	mean	s.d.
Better RPD students (n=25)	46.2	2.4	11.7	3.1	14.8	3.8
Weaker RPD students (n=25)	31.3	5.6	9.0	4.0	8.4	4.8
Better GPD students (n=26)	45.1	2.0	12.3	3.4	13.6	5.0
Weaker GPD students (n=24)	31.3	9.1	9.0	3.8	8.9	3.8

Table 3.13 Mean number of correct answers and standard deviations on CITO LVS E3 and the Category and Analogy tests (maximum scores 21 and 24 respectively) of the SON-R

For both the CITO LVS E3 and the subtests of the SON-R, the main effect of Type of Program Design is not significant. The main effect of Arithmetic Competence is significant for the scores on the National Arithmetic Test, Categories and Analogies: $F(1, 96) = 160.0, p < .000$; $F(1, 97) = 17.7, p < .000$ and $F(1, 97) = 38.9, p < .000$ respectively. The interaction effect Type of Program Design x Arithmetic Competence was not significant.

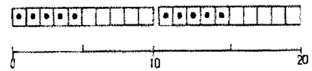
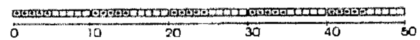
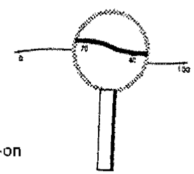
The mean scores for the better and weaker students on the CITO LVS E3 can be translated in the CITO competence levels mentioned before. The mean score of the weaker RPD and GPD students fell between the ranges of the D level. The mean score of the better RPD and GPD students fell between the ranges of the A level.

Implementation of the RPD and GPD

The schools were provided with teacher guides, students' textbooks, and bead strings or manipulative tracks to be used instead of their regular mathematics textbooks, teacher guides, and additional materials. The schools did not have to pay for these materials. The teachers had to fill in a concise log-book in which they described the amount of time they spent on arithmetic teaching and what their experiences were. Every fortnight the teachers discussed their experiences with one of the researchers during a visit to the schools. The researcher sketched what they could expect in the forthcoming period. The most important results of the different tests were also discussed with the teacher. To give the reader an impression of what was taught in the different program designs at the time the tests were administered, we refer to Table 3.14. The CITO LVS E3 that was administered before the experiment started, and the retention tests, administered when the students were in the third grade are not mentioned in this table.

Data collection

At the end of the first grade, before the experiment started, the CITO LVS E3 and the two subscales of the SON-R were administered to all students. During the second grade, tests were administered to the RPD and GPD students in September, October, January, April and June. In the third grade, when the students had returned to their regular textbook *Rekenen & Wiskunde* some tests were administered as a retention test. Table 3.3 gives an overview of which tests where administered at the different moments. All tests were administered by one of the researchers of the project. To investigate how consistent students were in solving the ASPT problems (cf. Van der Heijden, 1993), we administered the ASPT twice in April and twice in June, each time with a one week interval. We tested the consistency by using 0/1 scores for the main procedures that were used to solve the problems (Blöte, Klein & Beishuizen, in preparation). Per category two analyses were performed, one for the two occasions in April and one for the two occasions in June. MANOVAs with a repeated-measures design revealed no significant effects in either month (Blöte, Klein & Beishuizen, in preparation). We can therefore conclude that the students were very consistent in the way they solved the problems on the ASPT.

Realistic Program Design	Tests	Gradual Program Design
	September	
Number positioning on bead string as introduction of semi-structured number line up to 20 		Arithmetic blocks as introduction of semi-structured number line up to 20 
Sums < 20: 7 + 7, 8 + 7; 14 - 6, 11 - 9 (∩)		Sums < 20: 7 + 7, 7 + 8; 14 - 6
	October	
Number positioning on bead string as introduction of semi-structured number line up to 100 		Arithmetic blocks as introduction of semi-structured number line up to 50 
Sums < 100: 74 + 8, 93 - 9; 45 + 32, 48 + 36, 51 - 49		Sums < 50: 34 + 8, 43 - 7; 35 + 12, 45 - 23
*Introduction of the empty number line		
Various procedures: N10,N10C,A10,∩; Different strategies: addition,subtraction, adding-on		One procedure: N10; Two strategies: addition, subtraction
	January	
Sums <100: 85 - 32, 85 - 39; 81 - 79, 81 - 19		Sums <100: 55 + 32, 55 + 37; 85 - 32, 85 - 39 
		*Introduction of the empty number line
Various procedures like N10, N10C, A10, ∩		Various procedures like N10, N10C, A10;
Different strategies: addition, subtraction, and adding-on		Different strategies:addition, subtraction, and adding-on
	April	
Introduction of 1010 procedure for addition; Labels for different procedures		Introduction of 1010 procedure for addition; Labels for different procedures
Flexibilization of strategies and procedures in a variety of (context) problems		Flexibilization of strategies and procedures in a variety of (context) problems
	June	

Note. Context problems and tests administered in the first and third grade are not mentioned in this schedule, cf. explanation in text

Table 3.14 Time schedule with moments when tests were administered for the Realistic and Gradual Program Design

General scoring procedures

Answers on all the arithmetic tests were scored as correct or incorrect. These scores served as an indication of the procedural competence of these sum types. Beside the problems that had to be solved by head (AST and part of the ASMT), the computation procedures used to solve the problems were scored and labeled by one of the researchers according to the categories of procedures and strategies shown in Table 1.1. For the problems on the ATT with numbers > 100 we translated the labels of the computation procedures and arithmetic strategies in Table 1.1 (see Appendix C). The problems were scored and labeled in the same way as the ASPT and the Arithmetic Subject Matter Tests.

Procedure for analyzing computation procedures

The computation procedures the students used to solve the problems were analyzed in two different ways. First we looked at the computation procedures and solution strategies that were used by the whole group of RPD and GPD students. However, we were also interested to what extent a student changed his solution behavior across items. To be more specific, we explored whether the students were rigid in using one procedure across the different problems or they chose a solution procedure according to the characteristics of the problem. To analyze the data for the whole group we used MANOVA with repeated measures. We used 0/1 scores for either or not using the CS procedure for single-digit problems and N10 for multi-digit problems on the ASMT half-way through the curriculum and the ASMT and ASPT at the end of the curriculum (cf. Blöte, Klein, & Beishuizen, in preparation). We used this score as an indication of the student's flexibility in use of computation procedures.

To analyze the pattern of computation procedures for each students across the different problems, we collected the patterns of solution procedures across the most important items of the ASMT in January and June and the ASPT in June. This resulted in a large number of different profiles which had to be reduced. We started therefore with the reduction of the patterns of solution procedures for numerical addition problems, context addition problems, numerical subtraction problems and context subtraction problems. The solution procedures to solve context problems with differences were not analyzed in this way because that would make the analysis too complicated. This resulted in four lists of profiles for each test. To reduce the number of solution patterns we used some decision rules. We first formulated a *flexible profile* according to the number characteristics of the different problems (see also Tables 3.4, 3.5, 3.6, and 3.7). Students who matched this profile were labeled as flexible. A second rule was based on the principle that if a student used a certain procedure in the majority of the problems (3 or more out of 4 problems or 2 or more out of 3 problems) the student was labeled as a user of that procedure. Students who could not be categorized according to these rules were put in the category "else". This resulted in one label for each of the four types of problems. We then put these four labels together, which gave us a new list of profiles of solution procedures for the different tests in January and June. Here again we applied the rule that if a student had a certain label in a majority of the 4 problem types, he was given that label. If a student had two labels, both used for 2 problem types, the student was given a *mixed label*. For the three tests this resulted in the following 8 profiles:

Flexible:	the students were labeled "flexible" when they were labeled flexible for at least three out of the four problem types;
Half-Flexible:	the students were labeled "half flexible"when they were labeled flexible for two problem types. For the other two problem types they used a different procedure that did not fit with the flexible profile;
CS/N10[2]:	the students were labeled "CS/N10" when they used the CS procedure to solve the single-digit problems and the N10 procedure for the multi-digit problems;
Else/N10[2]:	the students were labeled "Else/N10" when they used other procedures than the CS procedure for the single-digit problems (they solved these problems for instance in one step) and the N10 procedure for multi-digit problems;
N10[3]:	when the students used the N10 procedure for at least three of the four problem types, they were labeled "N10";
N10C[3]:	when the students used the N10C procedure for at least three of the four problem types, they were labeled "N10C";
1010/N10[3]:	the students were labeled "1010/N10" when they used the 1010 procedure for addition problems and the N10 procedure for subtraction problems;
Else:	when the students could not be categorized according to one of the before mentioned labels, they were labeled "Else".

For the problems that had to be solved in two ways on the ASMT in June we analyzed the pattern of solution procedures in the same way (see Table 3.6 for expected procedures). This resulted in the following categories: Flexible, Half-Flexible, N10, 1010/N10 and Else.

The distribution of the different profiles were compared for the different program designs and statistically tested with a chi-square test. Beside the answers and solution procedures, we also distinguished different types of errors which could be

---

2 This profile appears only for the ASMT in January because the CS procedure is only used for problems in which a single-digit had to be added or subtracted. In June these type of problems were not administered any more.

3 This profile appears only for the tests in June because in these tests only multi-digit problems were administered.

divided into two main categories: procedural and non-procedural errors. With procedural errors, the procedure is not carried out in the right way. Often this is caused by a serious misconception about how to operate on numbers. These errors are often more persistent than non-procedural errors (Felix, 1992). An example of a procedural error is the so-called *smaller from larger bug* which often occurs when the 1010 procedure is used to solve a subtraction problem with regrouping (cf. Table 1.1). A non-procedural error is often less serious and can be considered as a slovenliness. An example of a non-procedural error is when a student has counted one more or less than is necessary given the number of units the student has to add or subtract. For a complete overview of all the procedures, strategies and types of errors we refer to Appendices A and B.

4 Results: cognitive variables

In this chapter we will describe the results on the cognitive arithmetic tests regarding the development of procedural and strategic knowledge. Procedural knowledge was measured in two ways: Fluency in solving numerical problems in a limited amount of time (Arithmetic Speed Test) and the number of correct answers on the Arithmetic Subject Matter Test, the Arithmetic Scratch Paper Test and an external criterion test (CITO LVS). Strategic knowledge was measured by looking at solution procedures the RPD and GPD students used to solve numerical and context problems on the number line (Arithmetic Subject Matter Test) and on a piece of scratch paper, which was depicted beside the problem (Arithmetic Scratch Paper Test). The analyses of the solution procedures used were performed for the whole group of RPD and GPD students. We also looked at the pattern of computation procedures across the different problems of the test for each student.

We will begin by describing the results regarding the first hypotheses (1-4) for procedural competence and strategic knowledge for the whole group of RPD and GPD students half-way through (chapter 4.1) and at the end of the curriculum (chapter 4.2). Then we will describe the results regarding the hypotheses (5-8) for the sample of weaker and better RPD and GPD students half-way through (chapter 4.3) and at the end of the curriculum (chapter 4.4). Finally we will describe the results regarding two of three post-hoc questions (14-15) on the transfer and retention tests for the whole group of RPD and GPD students (chapter 4.5)[1]. Before describing the results we will summarize the relevant hypotheses (for a more extensive description of the hypotheses we refer to chapter 2). Conclusions based on these results will be drawn and discussed in chapter 6.

[1] For a complete overview of the data on the Arithmetic Subject Matter Tests we refer to Klein (1997).

4.1 Procedural and strategic knowledge: RPD versus GPD half-way through the curriculum

We will discuss the results concerning procedural competence, type of errors and strategic knowledge of the RPD and GPD students half-way through the curriculum.

Procedural competence on researcher designed and external criterion tests

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
1. Half-way through the program there will be no differences in the number of correctly solved problems for RPD and GPD students. The GPD students will solve more numerical problems in a limited amount of time than RPD students. RPD students will solve more context problems correctly than GPD students.	2. Half-way through the program there will be a higher level of procedural competence in numerical problems shown by the GPD rather than the RPD students which will be reflected by a higher number of correctly solved problems. The GPD students will solve more of these problems in a limited amount of time (speed test). With respect to context problems the GPD students will solve fewer problems correctly than the RPD students.
Outcome variables: • number of correct answers on Arithmetic Speed Test • number of correctly solved problems on Arithmetic Subject Matter Test • number of correctly solved problems on external criterion test (CITO LVS M4)	

Tables 4.1 and 4.2 show the mean number of correctly solved single-digit problems and standard deviations on the Arithmetic Speed Test in January for RPD and GPD students.

	single-digit < 20		single-digit < 50		single-digit < 100	
	mean	s.d.	mean	s.d.	mean	s.d.
RPD students (n=139)	23.6	10.0	15.7	6.7	6.6	13.4
GPD students (n=136)	22.6	10.8	12.4	6.1	5.4	8.4

Table 4.1 Mean number of correctly solved single-digit addition problems on the Arithmetic Speed Test in January for RPD and GPD students

	single-digit < 20 mean	s.d.	single-digit < 50 mean	s.d.	single-digit < 100 mean	s.d.
RPD students (n=139)	20.9	9.4	14.2	7.5	11.8	7.3
GPD students (n=136)	17.7	9.6	10.7	6.0	8.1	5.4

Table 4.2 Mean number of correctly solved single-digit subtraction problems on the Arithmetic Speed Test in January for RPD and GPD students

MANOVA with repeated measures showed a significant effect for type of program, $F(1, 259) = 16.8$, $p < .001$, and for the interaction type of program x type of problem, Pillais F (2, 258) = 6.7, $p < .01$. For single-digit addition problems ANOVAS revealed significant differences between the RPD and GPD students for problems with numbers < 50, $F(1, 261) = 16.4$, $p < .01$, and problems with numbers < 100, $F(1, 261) = 45.2$, $p < .01$. For single-digit subtraction problems, ANOVAS showed significant differences between the two groups of students for problems with numbers < 20, $F(1, 260) = 7.1$, $p < .01$; for problems with numbers < 50, $F(1, 260) = 17.3$, $p < .01$; and for problems with numbers < 100, $F(1, 260) = 22.2$, $p < .01$. For all these types of problems the RPD students solved a greater number of problems than the GPD students.

Procedural competence was also measured by looking at the number of correctly solved problems with and without use of the semi-structured number line on the Arithmetic Subject Matter Test in January. Tables 4.3, 4.4, and 4.5 show the mean number of correctly solved problems and standard deviations for the RPD and GPD students. For the numerical problems that had to be solved by head the maximum score was 5.

	By head addition < 50 mean	s.d.	By head addition < 100 mean	s.d.	By head subtraction < 50 mean	s.d.	By head subtraction < 100 mean	s.d.
RPD students (n=139)	4.4	.79	4.1	1.2	4.2	1.1	4.0	1.3
GPD students (n=136)	4.4	.88	3.9	1.3	4.0	1.1	3.7	1.4

Table 4.3 Mean number of correctly solved numericals by head on the Arithmetic Subject Matter Test in January for RPD and GPD students

MANOVAS with repeated measures revealed a significant effect for type of problem [$F(3, 214) = 19.0$, $p < .001$]. The main effect type of program appeared not to be significant. The mean number of correctly solved problems did not differ between the RPD and GPD students.

Table 4.4 gives an overview of comparable addition and subtraction problems in which the students could use the semi-structured number line. The maximum score for these problems was 5. Since the GPD students were not yet used to problems with numbers greater than 50, they did not solve the problems with numbers > 50. These problems were only solved by the RPD students.

	Number line addition < 50		Number line addition < 100		Number line subtraction < 50		Number line subtraction < 100	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
RPD students (n=139)	-	-	4.5	1.6	-	-	4.4	2.6
GPD students (n=136)	4.4	1.0	-	-	4.2	2.9	-	-

Table 4.4 Mean number of correctly solved numericals using the number line on the Arithmetic Subject Matter Test in January for RPD and GPD students

MANOVA with repeated measures showed a significant effect for type of problem: $F(1, 241) = 25.5, p < .001$. However, there were no significant differences between the two groups of students in the mean number of correctly solved problems using the number line.

Table 4.5 gives an overview of the number of correctly solved context addition and subtraction problems and context problems in which the students had to calculate a difference between two numbers. The maximum score for context addition problems was 3, for context subtraction problems and context problems with differences it was 2.

	Context addition < 50		Context subtraction < 50		Difference < 50	
	mean	s.d.	mean	s.d.	mean	s.d.
RPD students (n=139)	2.6	.66	1.5	.67	1.4	.76
GPD students (n=136)	2.5	.78	1.7	.58	1.4	.72

Table 4.5 Mean number of correctly solved context addition and subtraction problems and context problems with differences using the number line by RPD and GPD students on the Arithmetic Subject Matter Test in January

Here too, we did not find significant differences between the two groups of students in the number of correctly solved problems.

Beside the procedural competence on researcher designed tests we also looked at the scores on a more objective external criterion test half-way through the second grade. Figure 4.1 shows the percentage of correctly solved problems on the external criterion test by the RPD and GPD students for the different subscales of the CITO LVS M4.

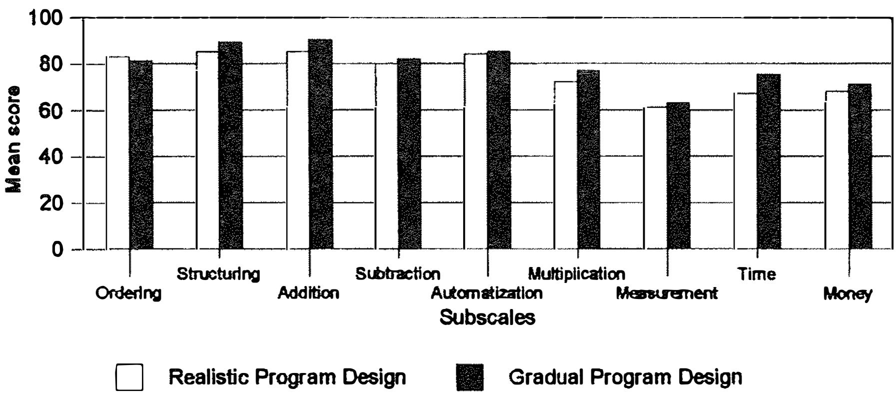


Figure 4.1 Percentages of problems correctly solved by RPD and GPD students on an external criterion test (CITO LVS M4) half-way through the second grade

A MANOVA revealed only a significant effect for type of problem [$F(13, 198) = 20.8, p < .001$]. No significant effect was found for type of program: There were no differences between the RPD and GPD students with respect to the number of correct answers on the whole test. ANOVAS for the different subscales revealed significant differences for the subscale structuring of numbers, $F(1,196) = 7.9, p < .01$, in favor of the GPD students.

Type of errors on the arithmetic subject matter test

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
1. Half-way through the curriculum there will be a difference in the type of errors students make: RPD students will make less procedural or conceptual mistakes than GPD students. RPD students will make more non-procedural mistakes (due to slovenliness) than GPD students.	2. No explicit predictions were made about the type of errors students would make.
Outcome variables: • Type of errors on Arithmetic Subject Matter Tests	

For the Arithmetic Subject Matter Tests we categorized the different types of errors as procedural and non-procedural errors (see section 3.3). Table 4.6 shows the number of procedural and non-procedural errors for the RPD and GPD students. The problems which had to be solved using the empty number line, were only administered to the RPD students, and therefore not taken into account in Table 4.6.

	non-procedural errors	procedural errors
RPD students (n=139)	402	153
GPD students (n=136)	450	146

Table 4.6 Number of procedural and non-procedural errors on the Arithmetic Subject Matter Test in January for RPD and GPD students

Chi-square analyses showed no significant differences in the number of procedural or non-procedural errors.

Strategic use of computation procedures

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
1. RPD students will adapt their strategy use to the characteristics of a problem and therefore the RPD students will be more flexible in their use of solution procedures than GPD students. Half-way through the curriculum GPD students will only use the N10 procedure in a proceduralized way where the RPD students use different strategies and procedures.	2. With respect to flexible use of different strategies and procedures, the GPD students will be more rigid than the RPD students. Half-way through the curriculum the GPD students will stay to the N10 procedure where the RPD students will also use other procedures as well.
Outcome variables: • Computation procedures used on Arithmetic Subject Matter Test	

Figure 4.2 shows the solution procedures RPD and GPD students used to solve numerical addition problems on the semi-structured number line. Because the GPD only introduced problems with numbers up to 100 after January, the GPD students solved problems with numbers up to 50 and the RPD students solved comparable problems with numbers up to 100.

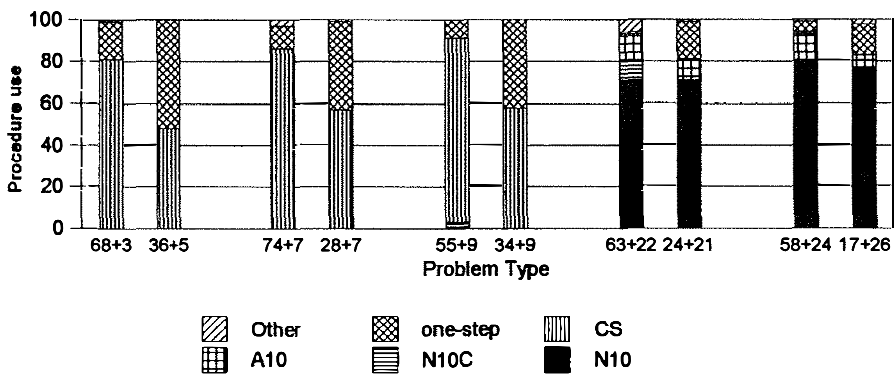


Figure 4.2 Solution procedures used by PRPD students (left bar) and GPD students (right bar) used to solve numerical addition problems by using the semi-structured number line on the Arithmetic Subject Matter Test in January

For the single-digit addition problems the RPD students tended to use the Complementary Structuring (CS) procedure more frequently than the GPD students who solved most of the problems by making one jump on the number line.

For the single-digit addition problem in which a 9 had to be added, hardly any pupil used the N10C procedure to solve this problem. To solve the multi-digit addition problems, both groups of students mainly used the N10 procedure. The only difference between these two groups of students was the use of the A10 procedure which was more frequently used by the RPD students to solve the multi-digit addition problems on the semi-structured number line.

Figure 4.3 shows the solution procedures RPD and GPD students used to solve numerical subtraction problems on the semi-structured number line. Here too we compared problems with numbers up to 50 with comparable problems with numbers up to 100.

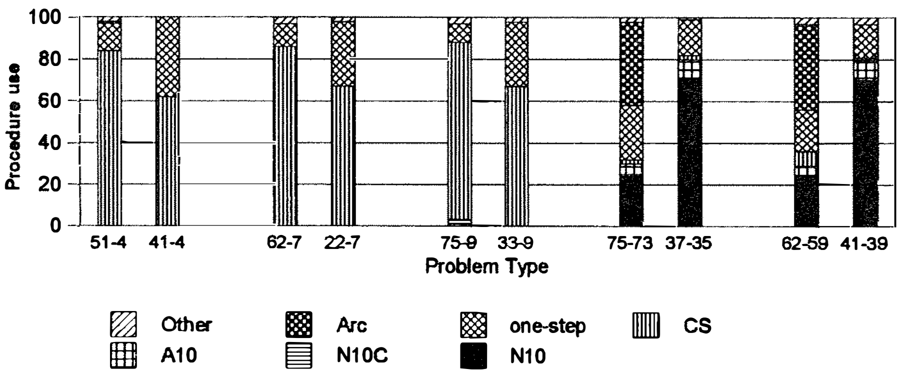


Figure 4.3 Solution procedures RPD students (left bar) and GPD students (right bar) used to solve numerical subtraction problems by using the semi-structured number line on the Arithmetic Subject Matter Test in January

For the single-digit subtraction problems we also see that the RPD students tended to use the CS procedure more frequently than the GPD students who tended to solve comparable problems by making one jump on the semi-structured number line. For the multi-digit problems with a small difference the RPD students used the Connecting Arc procedure most frequently. Since this procedure had not yet been introduced to the GPD students, they used mainly the N10 procedure. However, about 20% of the GPD students solved these two problems by making one jump on the number line which indicates that they had *seen* the difference between the two numbers.

Figure 4.4 gives an overview of the procedures RPD and GPD students used to solve the addition and subtraction context problems.

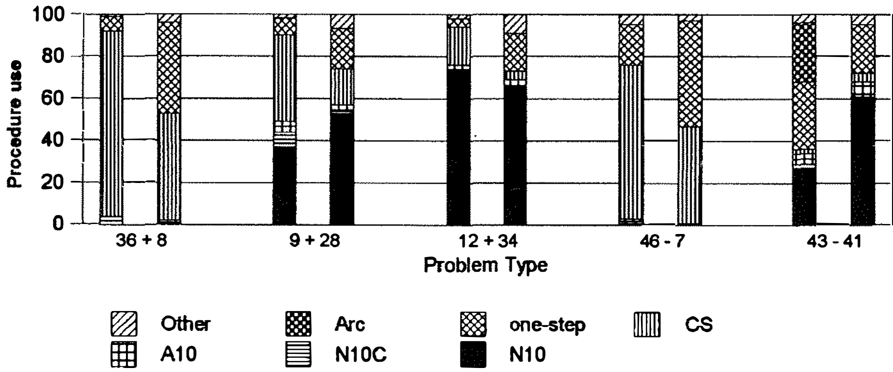


Figure 4.4 Solution procedures RPD students (left bar) and GPD students (right bar) used to solve context adddition and subtraction problems by using the semi-structured number line on the Arithmetic Subject Matter Test in January

For the context problems 36 + 8 and 46 - 7 we see the same pattern as for the numerical problems that had to be solved on the number line: The RPD students tended to use the CS procedure more frequently than the GPD students, who solved these problems more frequently by making one jump on the number line. For the problem 9 + 28 we see that the GPD students used the N10 procedure more frequently than the RPD students who used the CS procedure instead. This may be caused by the fact that the RPD students reversed the order of the numbers into 28 + 9, and started with adding up 9 by regrouping the 9 into 2 and 7 (CS). The GPD students instead solved the problem 9 + 28 by adding 20 first and than adding the 8 (N10). For the context addition problem 12 + 34 both RPD and GPD preferred the N10 procedure to solve this problem. However, for this problem the RPD students also tended to use the CS procedure more frequently than the GPD students who preferred to solve the problem by making one jump on the number line.

For the remaining context subtraction problem with a small difference (43 - 41), we see that the RPD students preferred the Connecting Arc procedure where the GPD students preferred the N10 procedure. However, GPD students also solved this problem by making one jump on the number line, which suggests that they *saw* the difference between the two numbers at once.

Figure 4.5 shows the computation procedures the RPD and GPD students used to solve the two problems in which the students had to calculate the difference between two numbers.

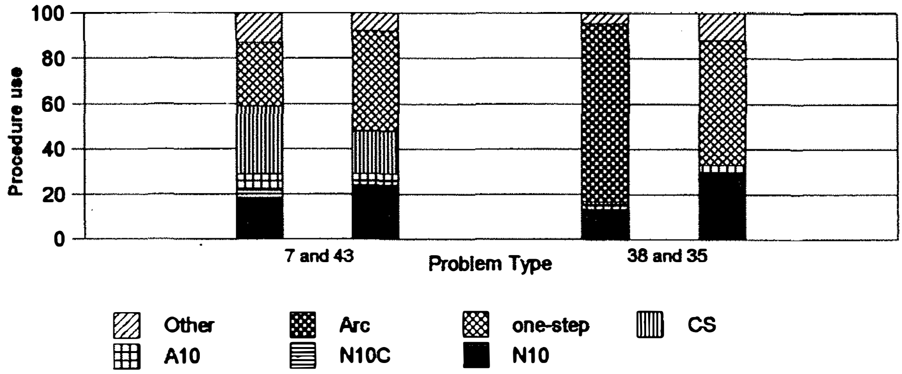


Figure 4.5 Solution procedures RPD students (left bar) and GPD students (right bar) used to solve context problems with differences by using the semi-structured number line on the Arithmetic Subject Matter Test in January

The pattern of solution procedures between the two groups of students resembled what we have seen with the subtraction context problems: The RPD students preferred the CS and Connecting Arc procedure respectively where the GPD students tended to solve the problems by making one jump on the number line or by using the N10 procedure. However, the differences between RPD and GPD students in type of solution procedure used are smaller than for the context subtraction problems.

MANOVAS for strategy use for RPD and GPD students

Since the N10 procedure was hardly used in January, we tested the differences between the RPD and GPD students in their use of the CS procedure for single-digit problems by using 0/1 scores for either using this procedure or not. MANOVA with repeated measures revealed a significant effect for type of program [$F(1, 247) = 47.7$, $p < .001$] and type of program x problem type [Pillais $F(3, 245) = 9.9$, $p < .001$] . ANOVAS showed that the RPD students used the CS procedure more frequently for both numerical and context addition and subtraction problems: $F(1, 247) = 36.2$, $p < .001$; $F(1, 248) = 44.5$, $p < .001$; $F(1, 247) = 20.4$, $p < .001$. $F(1, 247) = 19.5$, $p < .001$.

Consistency of solution procedures across problems

Until now we have analyzed the procedures used by the whole group of RPD and GPD students for each problem. These analyses do not reveal how many RPD and GPD students changed their solution procedure according to the number characteristics of the problems. In order to gain insight into this issue, we analyzed the consistency of use of solution procedures across the numerical and context addition and subtraction problems on the Arithmetic Subject Matter Test in January. We distinguished five different profiles of solution procedures used: Flexible, Half-

Flexible, CS/N10, Else/N10 and Else (for further information see section 3.3). The results of this analysis are presented in Table 4.7.

	Flexible	Half Flexible	CS/N10	Else/N10	Else
RPD students (n=122)	6	37	20	0	59
GPD students (n=117)	0	0	33	12	72

Table 4.7 Profiles of solution behavior of RPD and GPD students on Arithmetic Subject Matter Test in January

Chi-square analyses revealed a significant difference between the RPD and GPD with respect to the distribution across the different profiles: Chi^2 (4, N=239) = 80.7, $p < .0001$. Inspection of the cells showed that the number of pupils who could be categorized as Flexible or Half-Flexible was larger for the RPD than for the GPD. As can be seen many students were categorized as Else.

4.2 Procedural and strategic knowledge: RPD versus GPD at the end of the curriculum

We will discuss the results concerning procedural competence, type of errors and strategic knowledge of the RPD and GPD students at the end of the curriculum.

Procedural competence on researcher designed and external criterion tests

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
3. At the end of the program the RPD students will show a higher level of procedural competence in solving the most difficult sum type (subtraction problems which require regrouping) than the GPD students. For the other sum types there will be no differences in the number of correctly solved problems between the two groups of students.	4. At the end of the program the GPD students will still have a higher level of procedural competence than the RPD students with respect to standard number problems, especially on the speed test. The GPD students will also solve correctly as many context problems as the RPD students, because more attention is paid to these problems during the last months of the GPD.
Outcome variables: • number of correct answers on Arithmetic Speed Test • number of correctly solved problems on Arithmetic Subject Matter Test • number of correctly solved problems on Arithmetic Scratch Paper Test • number of correctly solved problems on external criterion test (CITO LVS E4)	

Tables 4.8 and 4.9 show the mean number of correctly solved single-digit problems and standard deviations on the Arithmetic Speed Test in June for RPD and GPD students.

	single-digit < 20 mean	s.d.	single-digit < 50 mean	s.d.	single-digit < 100 mean	s.d.
RPD students (n=133)	33.6	10.5	23.2	7.3	19.4	7.7
GPD students (n=119)	33.2	15.6	20.2	8.8	15.4	8.1

Table 4.8 Mean number of correctly solved single-digit addition problems on the Arithmetic Speed Test in June for RPD and GPD students

	single-digit < 20 mean	s.d.	single-digit < 50 mean	s.d.	single-digit < 100 mean	s.d.
RPD students (n=133)	29.1	11.7	21.9	8.9	17.5	7.7
GPD students (n=119)	30.1	13.5	18.8	8.7	15.5	8.1

Table 4.9 Mean number of correctly solved single-digit subtraction problems on the Arithmetic Speed Test in June for RPD and GPD students

MANOVA with repeated measures for the single-digit problems revealed a significant effect for type of problem [Pillais $F(5, 244) = 203.5, p < .001$] but no significant effect for type of program was found. The interaction effect type of program x type of problem appeared to be significant: Pillais, $F(5, 244) = 8.1, p < .001$. ANOVAS revealed significant differences between the RPD and GPD students for single-digit addition problems < 50, $F(1, 250) = 9.1, p < .01$, and single-digit addition problems < 100, $F(1, 250) = 16.2, p < .001$. The RPD students solved correctly more problems than the GPD students. For single-digit addition problems < 20 no significant differences were found between the two groups of students. For single-digit subtraction problems significant differences between the RPD and GPD students were found for problems with numbers < 50, $F(1, 253) = 7.9, p < .01$. The number of correctly solved problems was highest for the RPD students. For the other two sum types no significant differences were found.

During the administration of the Arithmetic Speed Test in June we also administered multi-digit addition and subtraction problems, with and without regrouping the units, with numbers less than 50 and less than 100. Table 4.10 shows the mean number of correctly solved multi-digit problems < 50 and standard deviations on the Arithmetic Speed Test in June for RPD and GPD students.

	addition < 50 without regrouping mean	s.d.	addition < 50 with regrouping mean	s.d.	subtraction < 50 without regrouping mean	s.d.	subtraction < 50 with regrouping mean	s.d.
RPD students (n=133)	19.0	10.8	12.6	8.6	11.3	10.0	9.3	8.8
GPD students (n=119)	17.9	9.7	11.4	5.6	11.7	8.0	8.3	6.1

Table 4.10 Mean number of correctly solved multi-digit problems with numbers < 50 on the Arithmetic Speed Test for RPD and GPD students in June

MANOVA with repeated measures for multi-digit problems with numbers < 50 only showed a significant effect for type of problem [Pillais $F(3, 245) = 130.5$, $p < .001$]. No significant main effect for type of program or interaction effect type of problem x type of program was found. There were no differences between the RPD and GPD students in the number of correctly solved multi-digit problems with numbers < 50 on the Arithmetic Speed Test in June.

Table 4.11 shows the mean number of correctly solved multi-digit problems with numbers < 100 and standard deviations on the Arithmetic Speed Test in June for RPD and GPD students.

	addition < 100 without regrouping mean	s.d.	addition < 100 with regrouping mean	s.d.	subtraction < 100 without regrouping mean	s.d.	subtraction < 100 with regrouping mean	s.d.
RPD students (n=133)	18.6	8.7	10.9	6.0	10.9	7.1	9.0	6.2
GPD students (n=118)	16.1	9.3	10.2	6.4	9.5	6.5	7.4	5.3

Table 4.11 Mean number of correctly solved multi-digit problems with numbers < 100 on the Arithmetic Speed Test for RPD and GPD students in June

MANOVA with repeated measures revealed a significant main effect for type of program [$F(1, 255) = 4.0$, $p < .05$] and for type of problem [Pillais $F(3, 253) = 138.2$, $p < .001$]. The interaction effect type of program x type of problem appeared also to be significant: Pillais $F(3, 253) = 2.9$, $p < .05$. Separate ANOVAS showed significant differences between the RPD and GPD students for addition problems without regrouping, $F(1, 255) = 4.6$, $p < .05$, and for subtraction problems with regrouping $F(1, 255) = 5.1$, $p < .05$. For both problem types the RPD students solved a greater number of problems than the GPD students on the Arithmetic Speed Test in June.

In June, procedural competence was also measured by analyzing the number of correctly solved problems on the Arithmetic Subject Matter Test and the Arithmetic

Scratch Paper Test. Tables 4.12 and 4.13 show the mean number of correctly solved problems and standard deviations on the Arithmetic Subject Matter Test in June for the RPD and GPD students. For the addition and subtraction problems that had to be solved in one way the maximum score was 5, for the problems that had to be solved in two ways the maximum score was 6.

	By head addition < 100		Number line addition < 100		Context addition < 100		Context: 2 ways addition < 100	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
RPD students (n=139)	3.6	.83	3.7	.57	3.6	.68	5.6	.78
GPD students (n=136)	3.6	.71	3.6	.79	3.6	.70	5.6	.95

Table 4.12 Mean number of correctly solved addition problems on the Arithmetic Subject Matter Test for RPD and GPD students in June

MANOVA with repeated measures did not show a significant effect for type of program for the addition problems.

	By head subtraction < 100		Context subtraction < 100		Number line subtraction < 100		Context: 2 ways subtraction< 100	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
RPD students (n=139)	3.3	1.0	3.6	.74	3.7	.62	5.0	1.5
GPD students (n=136)	3.2	1.1	3.4	.88	3.3	.95	4.7	1.5

Table 4.13 Mean number of correctly solved subtraction problems on the Arithmetic Subject Matter Test for RPD and GPD students in June

MANOVA with repeated measures for the first three problem types showed a significant effect for the type of program [$F(1, 244) = 7.8, p < .01$] and for problem type [Pillais $F(2, 243) = 7.3, p < .01$]. The interaction effect type of program x type of problem appeared also to be significant: Pillais $F(2, 243) = 3.1, p < .05$. ANOVAS for the different problem types revealed significant differences between RPD and GPD students for numerical subtraction problems that were solved with use of the empty number line, $F(1, 239) = 12.3, p < .01$. The RPD students solved correctly more problems than the GPD students. This difference is caused by the two subtraction problems 81 - 78 and 72 - 39. The number of GPD students who solved correctly these two subtraction problems was significantly lower than the number of correct answers for the RPD students [$F(1, 251) = 26.6, p < .01$; $F(1, 250) = 7.1, p < .01$]. Most answers were incorrect because they were one or two units or tens more or less than the correct answer.

Table 4.14 shows the mean number of correctly solved problems (and standard deviations) with use of either the number line, arrow schema or solution steps for the Arithmetic Scratch Paper Test in June for the RPD and GPD students. The maximum score for the numerical and context addition problems was 3, for the numerical and context subtraction problems 5, and for the context problems with differences also 5.

	Addition N		Addition C		Subtraction N		Subtraction C		Difference	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
RPD students (n=139)	2.7	.54	2.5	.69	4.5	.81	4.5	.86	4.1	1.3
GPD students (n=136)	2.6	.60	2.6	.64	4.3	1.2	4.4	.95	3.9	1.3

Table 4.14 Mean number of correctly solved numerical (N) and context (C) addition problems, numerical (N) and context (C) subtraction problems, and difference problems on the Arithmetic Scratch Paper Test for RPD and GPD students in June

MANOVA with repeated measures did not reveal significant differences between the two groups of students for the different scales of the ASPT.

Beside the procedural competence on researcher designed tests we also looked at the scores on a more objective external criterion test at the end of the second grade. Figure 4.6 shows the percentage of correctly solved problems by the RPD and GPD students for the different subscales of the CITO LVS E4.

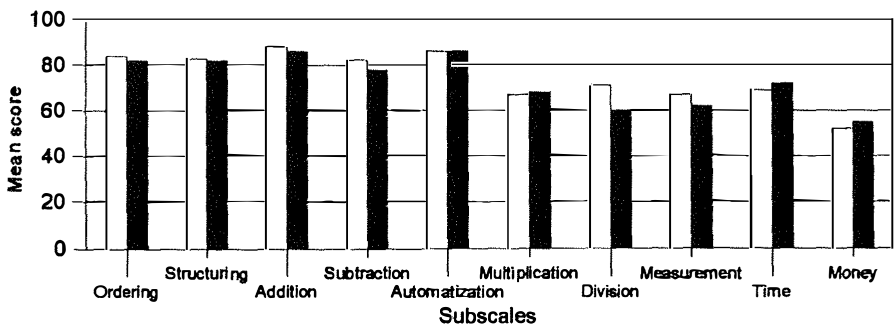


Figure 4.6 Percentages of problems correctly solved by RPD and GPD students on an external criterion test (CITO LVS E4) at the end of the second grade

MANOVA with repeated measures revealed a significant effect for type of problem [Pillais $F(9, 191) = 43.9$, $p < .001$] but did not reveal a significant effect for type of program. No differences were found between the RPD and GPD students with respect to the total number of correct answers on this test.

Type of errors on arithmetic subject matter test and arithmetic scratch paper test

With respect to the type of errors made at the end of the curriculum, no explicit hypotheses were formulated for the RPD and GPD students. For the Arithmetic Subject Matter and Scratch Paper Test we categorized the different types of errors into procedural and non-procedural errors (see section 3.3). Tables 4.15 and 4.16 show the number of procedural and non-procedural errors for the RPD and GPD students on the two tests.

	non-procedural errors	procedural errors
RPD students (n=139)	251	111
GPD students (n=136)	267	122

Table 4.15 Number of procedural and non-procedural errors on the Arithmetic Subject Matter Test for RPD and GPD students in June

	non-procedural errors	procedural errors
RPD students (n=139)	137	181
GPD students (n=136)	150	170

Table 4.16 Number of procedural and non-procedural errors on the Arithmetic Scratch Paper Test for RPD and GPD students in June

For both the Arithmetic Subject Matter Test and the Arithmetic Scratch Paper Test in June Chi-square analyses showed no significant differences between RPD and GPD students in number of procedural and non-procedural errors. As for the test in January, we see that for the Arithmetic Subject Matter Test in June the number of non-procedural errors exceeds the number of procedural errors. However, for the Arithmetic Scratch Paper Test the opposite is true: more procedural than non-procedural errors were made, both by RPD and GPD students. This is probably caused by the fact that for the Arithmetic Scratch Paper Test addition and subtraction problems were presented in a mixed order where there was a separate addition and subtraction part (administered at different moments) for the Arithmetic Subject Matter Test. Therefore more mistakes were made in choosing between adding or subtracting. These errors were categorized as procedural errors.

Strategic use of computation procedures

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
3. At the end of the second grade the GPD students lag behind the RPD students with respect to the flexible use of solution strategies and computation procedures, both for numerical and context problems.	4. At the end of the second grade we expect there will be no differences between the GPD and RPD students, with respect to the flexible use of different solution strategies and computation procedures.
Outcome variables: • Computation procedures used on Arithmetic Subject Matter Test • Computation procedures used on Arithmetic Scratch Paper Test • Notation forms used on Arithmetic Scratch Paper Test	

Figure 4.7 shows the solution procedures RPD and GPD students used to solve numerical addition problems on the empty number line.

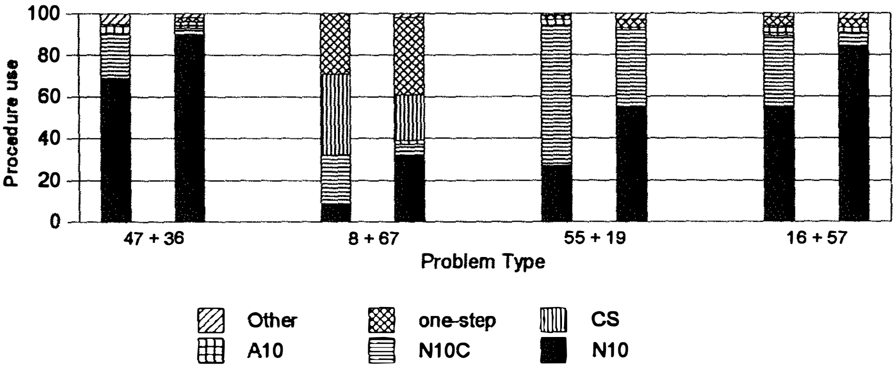


Figure 4.7 Solution procedures RPD students (left bar) and GPD students (right bar) used to solve numerical addition problems by using the empty number line on the Arithmetic Subject Matter Test in June

For the multi-digit numerical addition problems the GPD students tended to stick to the N10 procedure to solve these addition problems. One-third of the GPD students used the N10C procedure only for the addition problem 55 + 19. The RPD students tended to use the N10C procedure not only for this problem, but they also used the N10C procedure more frequently than the GPD students to solve the other addition problems. For the single-digit addition problem 8 + 67 the RPD students used the CS procedure more frequently than the GPD students who preferred to solve this problem in one step.

The solution procedures that were used by the RPD and GPD students to solve context addition problems in one way, on the empty number line, are shown in Figure 4.8

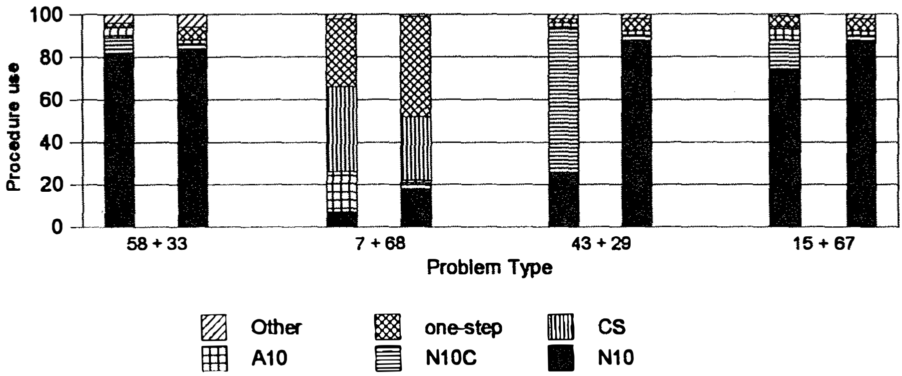


Figure 4.8 Solution procedures RPD students (left bar) and GPD students (right bar) used to solve context addition problems by using the empty number line on the Arithmetic Subject Matter Test in June

The behavioral differences between RPD and GPD students with respect to solving addition problems of the context type are comparable to the solving behavior displayed with the previous addition problems of the numerical type. On the other hand we generally see that, compared to the numerical problems, the context problems are more frequently solved using the N10 procedure instead of the N10C. This may be caused by the semantic structure of the context problem which, compared to numerical problems, leaves little room for manipulating the numbers (Linssen, 1996; Van Lieshout, 1997; Verschaffel, 1997; Verschaffel & De Corte, 1990).

In Figure 4.9 the solution procedures are shown that were used by the RPD and GPD students to solve context addition problems on the empty number line in two ways.

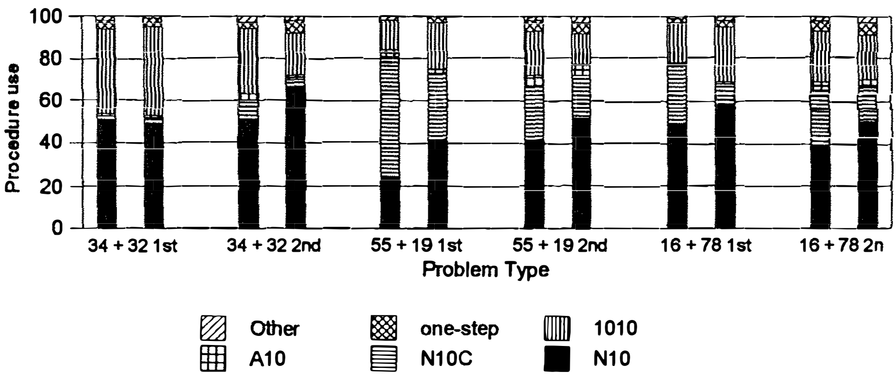


Figure 4.9 Solution procedures RPD students (left bar) and GPD students (right bar) used to solve context addition problems in two ways on the Arithmetic Subject Matter Test in June

For the RPD students we see that the differences between the first and second method of solving this problem are greatest for the context addition problem 55 + 19: First time round the N10C and the second time the N10 procedure was the most frequently used procedure. For the GPD students the change in procedures is most salient for the addition problem 34 + 32. They changed from using the 1010 procedure the first time to using the N10 procedure the second time, although on both occasions the N10 procedure remained the most frequently used procedure.

Figure 4.10 shows the solution procedures RPD and GPD students used to solve numerical subtraction problems on the empty number line.

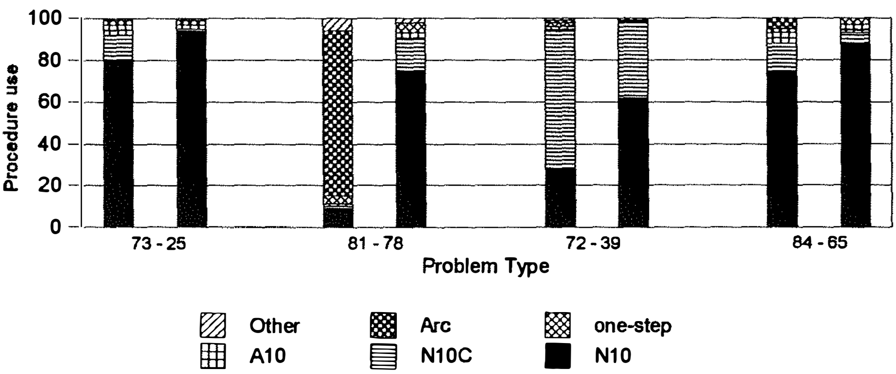


Figure 4.10 Solution procedures RPD students (left bar) and GPD students (right bar) used to solve numerical subtraction problems by using the empty number line on the Arithmetic Subject Matter Test in June

The RPD students more frequently altered their procedure use to the number characteristics of the problem. For the subtraction problem 81 - 78 they used the Connecting Arc procedure and for the problem 72 - 39 they changed to the N10C procedure. For this last problem the GPD students also changed to using the N10C procedure but to a lesser extent than the RPD students (36% GPD students versus 66% RPD students).

The solution procedures that were used by the RPD and GPD students to solve context subtraction problems on the empty number line in one way are shown in Figure 4.11.

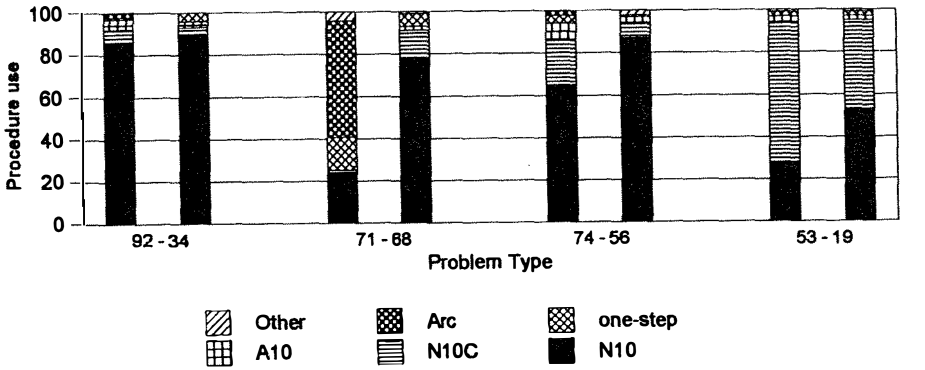


Figure 4.11 Solution procedures RPD students (left bar) and GPD students (right bar) used to solve context subtraction problems by using the empty number line on the Arithmetic Subject Matter Test in June

For the subtraction problems of the context type the picture is similar to the subtraction problems of the numerical type: The RPD students were more flexible in their procedure use than the GPD students. The overall differences between the type of presentation, context versus numerical, are smaller than for addition problems.

In Figure 4.12 the solution procedures are shown that were used by the RPD and GPD students to solve context subtraction problems on the empty number line in two ways.

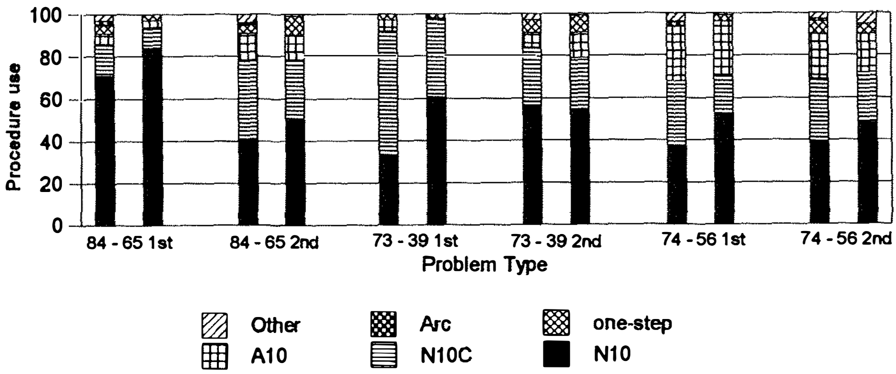


Figure 4.12 Solution procedures RPD students (left bar) and GPD students (right bar) used to solve context subtraction problems in two ways on the Arithmetic Subject Matter Test in June

For the RPD students the major changes in procedure use occurred for the context subtraction problems 84 - 65 and 73 - 39. For the first problem the RPD students changed from N10 on the first occasion to N10C on the second occasion. For the second problem N10C was the most frequently used procedure on the first occasion and N10 procedure was the most frequently used procedure on the second occasion. For the GPD students the change in procedure use was largest for the context subtraction problem 84 - 65 where the N10C procedure was more frequently used on the second occasion. For the subtraction problem 73 - 39 some GPD students also preferred the N10C procedure but this pattern remained the same on the second occasion. For the GPD students, the N10 procedure remained the most favored solution procedure to solve the different subtraction problems.

Figure 4.13 shows the solution procedures RPD and GPD pupils used to solve addition problems of the numerical (N) and context (C) type at the end of the curriculum.

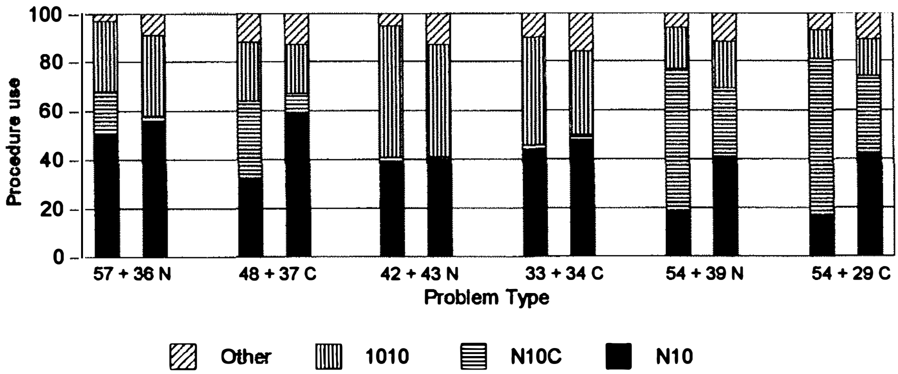


Figure 4.13 Solution procedures RPD students (left bar) and GPD students (right bar) for numerical (N) and context (C) addition problems on the Arithmetic Scratch Paper Test in June

Compared to the GPD students, the RPD students chose the most efficient procedure for the problem at hand. They switched from the N10 procedure for the 57 + 36 problem to the more efficient N10C procedure for the 54 + 39 problem. The GPD pupils mainly keep to their N10 procedure. The difference in presentation of the problem, numerical versus context, did not seem to influence the pupils' choices of solution procedures.

Figure 4.14 shows the solution procedures used by the RPD and GPD pupils to solve subtraction numerical and context problems at the end of the curriculum.

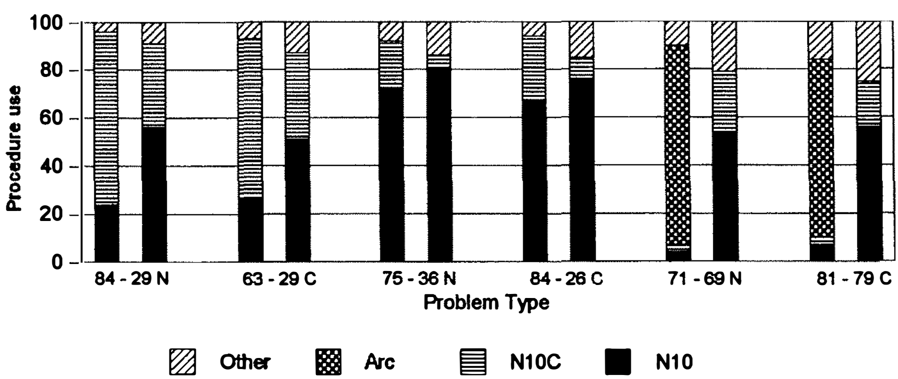


Figure 4.14 Solution procedures RPD students (left bar) and GPD students (right bar) for numerical (N) and context (C) subtraction problems on the Arithmetic Scratch Paper Test in June

The subtraction problems 62 - 48 (N), 72 - 58 (C), 65 - 33 (N), and 85 - 42 (C) are not included in Figure 4.14 because this would make the figure overly complicated. Their pattern of procedure use resembled the pattern of procedure use for the subtraction problems 75 - 36 (N) and 84 - 26 (C). Similarly to the addition problems, 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 number characteristics of the problem; for example, they used N10C for the 84 - 29 problem, while 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 3.14). However, for the problems with a small difference (71 - 69 and 81 - 79) about 10 to 20% of the GPD pupils solved the problem by making one jump between the two numbers. This is 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". Like the addition problems, the presentation format of numerical versus context did not seem to influence the procedures chosen to solve the subtraction problems.

Figure 4.15 shows the solution procedures RPD and GPD pupils used at the end of the curriculum for solving context problems in which they had to calculate a difference between two numbers.

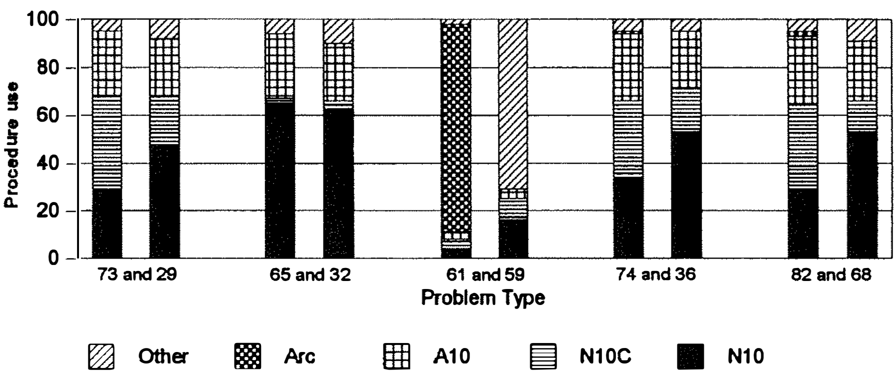


Figure 4.15 Solution procedures RPD students (left bar) and GPD students (right bar) for context problems with differences on the Arithmetic Scratch Paper Test in June

The procedures used by the RPD and GPD pupils 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, revealed an increase in use of the A10 procedure (see Table 1.1) for the difference problems. 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 1.1). Pupils started adding-on from the smallest to the largest 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. As in Figure 4.14, this is categorized as "other".

MANOVAS for strategy use by RPD and GPD students

We tested the differences between the RPD and GPD students in their use of the N10 procedure at the end of the curriculum by using 0/1 scores for either using this procedures or not. MANOVA with repeated measures on the addition and subtraction problems solved in one way on the Arithmetic Subject Matter Test in June revealed a significant effect for type of program [$F(1, 239) = 35.21, p < .001$] and for problem type [Pillais $F(3, 237) = 7.5, p < .001$]. ANOVAS for the different problem types revealed significant differences in use of the N10 procedure: numerical addition problems, $F(1, 239) = 44.4, p < .001$; context addition problems, $F(1, 239) = 16.4, p < .001$; numerical subtraction problems, $F(1, 239) = 24.5, p < .001$; and context subtraction problems, $F(1, 239) = 18.9, p < .001$. For all four of these problem types the GPD students stuck more to the N10 procedure whereas the RPD students adapted their solution procedure according to the number characteristics of the problems.

For the Arithmetic Scratch Paper Test the same type of analyses were done. MANOVA with repeated measures revealed a significant effect for type of program: $F(1,239) = 15.94, p < .001$. ANOVAS for the different problem types revealed significant differences in use of the N10 procedure for context addition problems, $F(1, 236) = 11.9, p < .01$; for numerical subtraction problems, $F(1, 236) = 10.7, p < .01$; and for context subtraction problems, $F(1, 236) = 22.3, p < .001$. No significant difference was found for numerical addition problems. For the other three problem types the GPD students kept using the N10 procedure, where the RPD students changed their procedure according to the number characteristics of the problem.

Consistency of solution procedures across problems

With regard to the tests in June, we were also interested in how many RPD and GPD students adapted their use of solution procedures across the different problems. Table 4.17 shows the profiles of solution procedures used to solve the numerical and context addition and subtraction problems on the Arithmetic Subject Matter Test in June. Students were labeled as Flexible, Half-Flexible, N10, N10C and Else (see section 3.3 for further information).

	Flexible	Half Flexible	N10	N10C	Else
RPD students (n=129)	38	27	26	10	28
GPD students (n=112)	22	17	56	3	14

Table 4.17 Profiles of solution behavior of RPD and GPD students on Arithmetic Subject Matter Test in June

Chi-square analyses revealed a significant difference between the RPD and GPD with respect to the distribution across the different profiles: Chi^2 (4, N = 241) = 24.9, $p < .0001$. Inspection of the cells shows that the number of pupils who could be categorized as Flexible or Half-Flexible was larger for the RPD than for the GPD. Many GPD students stuck to the use of the N10 procedure for all problems. Table 4.18 shows the profiles of solution procedures used to solve addition and subtraction problems in two ways. We categorized the students, using the same criteria as for the problems that had to be solved in one way (Table 4.17). The only difference is the profile 1010/N10 (see section 3.3 for further information).

	Flexible	Half Flexible	N10	1010/N10	Else
RPD students (n=139)	39	34	2	3	30
GPD students (n=136)	17	30	2	16	33

Table 4.18 Profiles of solution behavior of RPD and GPD students on Arithmetic SubjectMatter Test in June on problems that had to be solved in two ways

Chi-square analyses revealed a significant difference between the RPD and GPD with respect to the distribution across the different profiles: Chi^2 (4, N = 206) = 17.5, $p < .01$. Inspection of the cells shows that the number of pupils who could be categorized as flexible was larger for the RPD than for the GPD. With regard to the profile 1010/N10 the opposite is true, there are more GPD than RPD students in this category. There is hardly any difference between the number of RPD and GPD students that are flexible in half of the times.

Table 4.19 shows the patterns of solution procedures used to solve the different problems on the Arithmetic Scratch Paper Test in June.

	Flexible	Half Flexible	N10	N10C	1010/N10	Else
RPD students (n=129)	37	15	17	17	7	39
GPD students (n=112)	17	6	37	4	8	35

Table 4.19 Profiles of solution behavior of RPD and GPD students on Arithmetic Scratch Paper Test in June

Chi-square analyses revealed a significant difference between the RPD and GPD with respect to the distribution across the different profiles: Chi^2 (5, N = 241) = 24.7, $p < .001$. Inspection of the cells shows that the number of pupils who could be categorized as flexible or half-flexible was larger for the RPD than for the GPD. Many GPD students stuck to the use of the N10 procedure for all problems. We see that some RPD students also used the N10C procedure for all problems, even when this procedure was not the most efficient one.

Notation forms used on Arithmetic Scratch Paper Test in June

On the Arithmetic Scratch Paper Test the students were free to choose a notation form to write down their solution to the problem. This could be the number line, arrow scheme, or just solution steps to write down the answer. We were interested in which notation form RPD and GPD students would choose to solve these problems. Table 4.20 shows the frequencies of type of notation form that was used by the RPD and GPD students.

		Addition NE	Addition C	Subtraction NE	Subtraction C	Difference
RPD students (n=139)	NL	33%	32%	44%	41%	51%
	Arrow	3%	3%	5%	5%	5%
	Steps	62%	63%	51%	53%	43%
	Other	2%	2%	-	1%	1%
GPD students (n=136)	NL	3%	5%	7%	7%	11%
	Arrow	5%	2%	3%	2%	3%
	Steps	91%	92%	88%	89%	84%
	Other	1%	1%	2%	2%	2%

Note. NE stands for Numerical, C stands for Context, NL stands for Number Line, Arrow stands for Arrow Scheme

Table 4.20 Kind of notation form used by RPD and GPD students to solve problems on the Arithmetic Scratch Paper Test in June

We see that the RPD students used the number line more frequently than the GPD students who preferred to write down the steps they used to solve the problem.

Until now we have described the results for the whole group of RPD and GPD students. However, we were also interested what the effects of the RPD and GPD would be on a sample of weaker and better students. Fifty weaker and fifty better students were selected according to their scores on the National Arithmetic Test at the end of the first grade (CITO LVS E3, for more details on the selection procedure see chapter 3). The results for these groups of students are described in the next sections.

4.3 Procedural and strategic knowledge: Weaker and better RPD and GPD students half-way through the curriculum

We will discuss the results concerning procedural competence, type of errors and strategic knowledge of the weaker and the better students half-way through the curriculum. Since the students were selected on the basis of their competence level we will not mention all the univariate tests concerning the level of competence. For a complete overview of the data we refer to Klein (1997).

Procedural competence on researcher designed and external criterion tests

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
5. Half-way through the program there will be significant differences between weaker and better RPD students, however, this will be more on a strategic than on a procedural level. Weaker RPD students will often use solution strategies and computation procedures in a non-abbreviated or inefficient way, but these will bring them to the correct answers.	6. The expected higher level of procedural competence for the GPD students, half-way through the second grade, will mainly be caused by the relatively better scores of the weaker students. The differences between the better and weaker students will be significant larger in the RPD.
Outcome variables: • number of correct answers on Arithmetic Speed Test • number of correctly solved problems on Arithmetic Subject Matter Test • number of correctly solved problems on external criterion test (CITO LVS M4)	

Tables 4.21 and 4.22 show the mean number of correctly solved single-digit problems and standard deviations on the Arithmetic Speed Test in January for better and weaker RPD and GPD students.

	single-digit < 20 mean	s.d.	single-digit < 50 mean	s.d.	single-digit < 100 mean	s.d.
Better RPD students (n=23)	27.8	16.2	18.3	10.3	16.0	9.2
Better GPD students (n=25)	22.3	6.6	12.3	4.3	9.8	4.2
Weaker RPD students (n=25)	18.1	6.8	11.8	5.5	9.6	5.4
Weaker GPD students (n=23)	17.7	6.5	10.8	4.9	5.1	3.2

Table 4.21 Mean number of correctly solved single-digit addition problems on the Arithmetic Speed Test in January for better and weaker students

	single-digit < 20 mean	s.d.	single-digit < 50 mean	s.d.	single-digit < 100 mean	s.d.
Better RPD students (n=23)	25.6	13.6	17.4	10.9	15.0	10.6
Better GPD students (n=25)	20.9	9.4	13.0	5.8	10.7	5.5
Weaker RPD students (n=23)	15.5	5.7	8.6	5.5	7.0	4.5
Weaker GPD students (n=23)	17.7	9.6	6.9	4.1	5.2	3.9

Table 4.22 Mean number of correctly solved single-digit subtraction problems on the Arithmetic Speed Test in January for better and weaker students

MANOVA with repeated measures on both the addition and subtraction problems showed a significant effect for type of program [$F(1, 92) = 8.4, p < .01$] and for competence level [$F(1, 92) = 24.1, p < .001$]. The interaction-effect type of program x competence level appeared not to be significant.

When we looked at the differences between better RPD and GPD students, ANOVAS revealed significant differences for single-digit addition problems with numbers up to 50 and up to 100: $F(1, 46) = 7.2, p < .01$; $F(1, 46) = 9.3, p < .01$. For single-digit addition problems with numbers < 20 no significant differences were found between better RPD and GPD students. For single-digit subtraction problems we only found significant differences between better RPD and GPD students for problems with numbers < 20: $F(1, 46) = 4.4, p < .05$. For the other number sizes no significant differences were found.

For the differences between weaker RPD and GPD students, ANOVAS only revealed significant differences for single-digit addition problems with numbers < 100, $F(1, 46) = 11.8, p < .01$, and single-digit subtraction problems with numbers < 20, $F(1, 46) = 4.5, p < .05$. For the other problems no significant differences were found between weaker RPD and GPD students.

Procedural competence was also measured by looking at the number of correctly solved problems with and without use of the number line on the Arithmetic Subject Matter Test in January. Tables 4.23, 4.24, and 4.25 show the mean number of correctly solved problems and standard deviations for the weaker and better RPD and GPD students. The maximum score for all the numerical problems was 5. The maximum score for the context addition problems was 3, for the context subtraction problems, and the context problems with a difference it was 2.

	By head addition < 50 mean	s.d.	By head addition < 100 mean	s.d.	By head subtraction < 50 mean	s.d.	By head subtraction < 100 mean	s.d.
Better RPD students (n=22)	4.7	.48	4.5	.96	4.4	.91	4.7	.63
Better GPD students (n=23)	4.6	.58	4.4	.73	4.6	.58	4.0	.88
Weaker RPD students (n=20)	4.2	.89	3.8	1.3	3.7	1.4	3.0	1.7
Weaker GPD students (n=17)	4.1	1.1	3.4	1.5	3.2	1.4	2.6	1.5

Table 4.23 Mean number of correctly solved numericals by head on the Arithmetic Subject Matter Test in January for weaker and better RPD and GPD students

MANOVA with repeated measures revealed a significant effect for competence level [$F(1, 78) = 33.42, p < .001$] and for type of problem [Pillais $F(3, 76) = 12.3$, $p < .001$] but not for type of program. The interaction effect competence level x type of problem appeared also to be significant [Pillais $F(3, 76) = 5.3, p < .01$] but all the other interaction effects did not reach the level of significance.

	Number line addition < 50 mean	s.d.	Number line addition < 100 mean	s.d.	Number line subtraction < 50 mean	s.d.	Number line subtraction < 100 mean	s.d.
Better RPD students (n=22)			4.5	.91			4.4	1.2
Better GPD students (n=25)	4.8	.41	-	-	4.5	.71	-	-
Weaker RPD students (n=23)	-	-	4.0	1.5	-	-	3.8	1.3
Weaker GPD students (n=21)	4.1	1.0	-	-	3.3	1.2	-	-

Table 4.24 Mean number of correctly solved numericals using the number line on the Arithmetic Subject Matter Test in January for RPD and GPD students

MANOVA with repeated measures revealed a significant main effect for competence level [$F(1, 87) = 15.2, p < .001$], type of problem [$F(1, 87) = 10.8, p < .01$] but not for type of program. The interaction effects were also not significant.

	Context addition < 50		Context subtraction < 50		Difference < 50	
	mean	s.d.	mean	s.d.	mean	s.d.
Better RPD students (n=23)	2.8	.42	1.7	.56	1.7	.62
Better GPD students (n=25)	2.7	.46	1.8	.47	1.7	.56
Weaker RPD students (n=23)	2.6	.73	1.0	.80	1.0	.73
Weaker GPD students (n=23)	2.0	1.0	1.4	.7	1.1	.81

Table 4.25 Mean number of correctly solved context addition, context subtraction, and difference problems using the number line on the Arithmetic Subject Matter Test in January for RPD and GPD students

MANOVA with repeated measures showed a significant effect for competence level ($F(1, 90) = 21.8, p < .001$) but the effect of type of program was not significant. The interaction effect of type of program x competence level was also not significant.

Beside the procedural competence on researcher designed tests we also looked at the scores on a more objective external criterion test half-way through the second grade. Figure 4.16 shows the percentage of correctly solved problems by the weaker and better RPD and GPD students for the different subscales of the CITO LVS M4.

MANOVA revealed a significant main effect for competence level [Pillais $F(9, 64) = 7.4, p < .001$] and for type of problem [Pillais $F(13, 60) = 12.0, p < .001$]. The main effect type of program and the interaction effects were not significant.

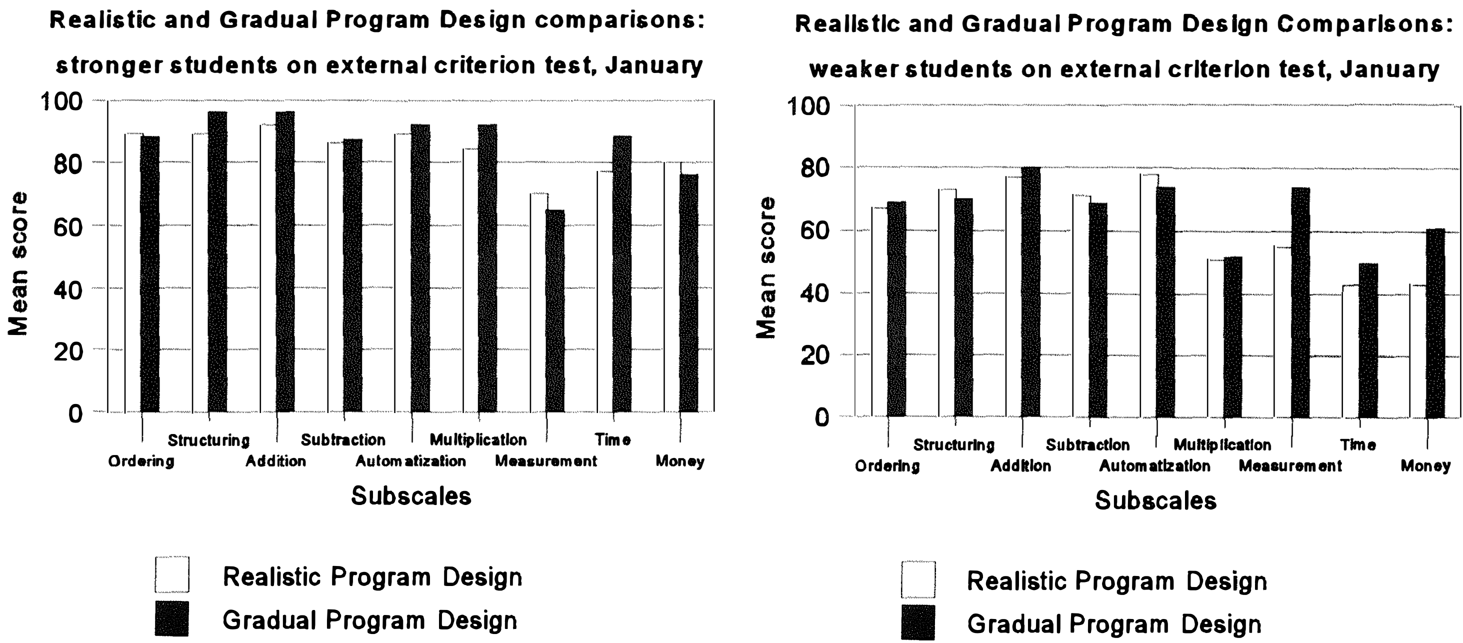


Figure 4.16 Percentages of problems correctly solved by better and weaker RPD and GPD students on external criterion test (CITOLVS M4) half-way through the second grade

Type of errors on the arithmetic subject matter tests

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
5. No explicit predictions were made about the type of errors students would make.	6. In the RPD emphasis is laid on flexibility from the start of the second grade, which will cause many inadequate inventions or combinations of solution strategies and computation procedures (for instance confusion in the execution of the N10C procedure for addition and subtraction problems).
Outcome variables: • Type of errors on Arithmetic Subject Matter Tests	

Table 4.26 shows the number of procedural and non-procedural errors for weaker and better RPD and GPD students.

	non-procedural errors	procedural errors
Better RPD students (n=23)	50	20
Better GPD students (n=26)	59	14
Weaker RPD students (n=25)	131	37
Weaker GPD students (n=24)	129	36

Table 4.26 Number of procedural and non-procedural errors on the Arithmetic Subject Matter Test in January for weaker and better RPD and GPD students

Separate Chi-square analyses for better RPD and GPD students and weaker RPD and GPD students revealed no significant differences in numbers of procedural or non-procedural errors.

Strategic use of computation procedures

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
5. Weaker RPD students will often use solution strategies and computation procedures in a non-abbreviated or inefficient way, but these will bring them to the correct answers. The weaker GPD students will use the N10 procedure in a proceduralized way without understanding what they are doing.	6. Weaker RPD students will be more flexible in using different strategies and procedures, but, compared to the weaker GPD students, this use will be of a lower quality in both strategic as procedural sense half-way through the program.
Outcome variables: • Computation procedures used on Arithmetic Subject Matter Test	

Figure 4.17 shows the solution procedures better RPD and GPD students used to solve numerical addition problems on the semi-structured number line. Figure 4.18 shows the same for weaker RPD and GPD students. Since problems with numbers up to 100 were only introduced in the GPD after January, the GPD students solved problems with numbers up to 50 and the RPD students solved comparable problems with numbers up to 100.

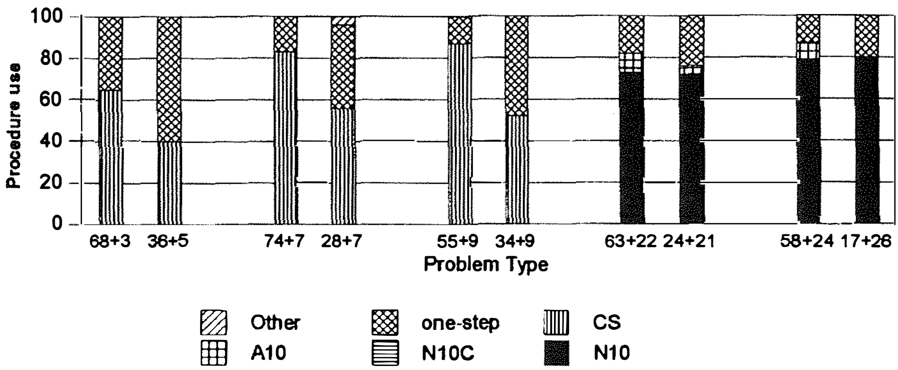


Figure 4.17 Solution procedures better RPD students (left bar) and GPD students (right bar) used to solve numerical addition problems by using the semi-structured number line on the Arithmetic Subject Matter Test in January

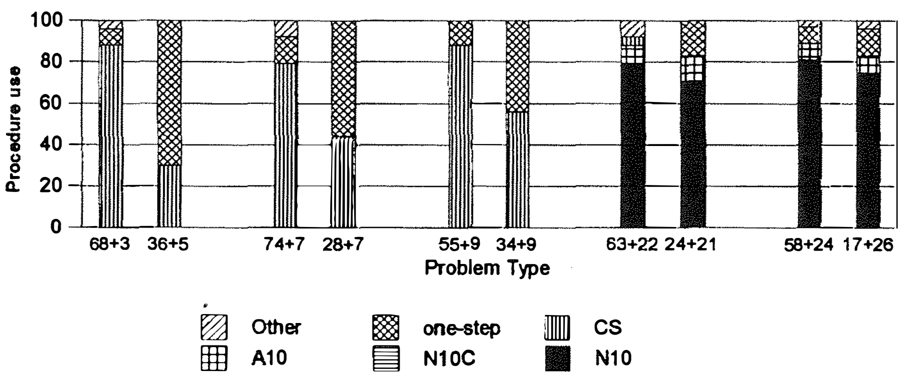


Figure 4.18 Solution procedures weaker RPD students (left bar) and GPD students (right bar) used to solve numerical addition problems by using the semi-structured number line on the Arithmetic Subject Matter Test in January

For the single-digit numerical addition problems both the better and the weaker RPD students tended to use the Complementary Structuring (CS) procedure more frequently than the GPD students, who preferred to solve these problems by making one jump on the number line. These differences are greatest for the weaker RPD and GPD students. The N10C procedure was not used by any of the students. Some students used the A10 procedure to solve the multi-digit addition problems but this was done both by RPD and GPD students.

Figures 4.19 and 4.20 show the solution procedures weaker and better RPD and GPD students used to solve numerical subtraction problems on the semi-structured number line. Again the GPD students solved problems with numbers up to 50 and RPD students solved problems with numbers up to 100.

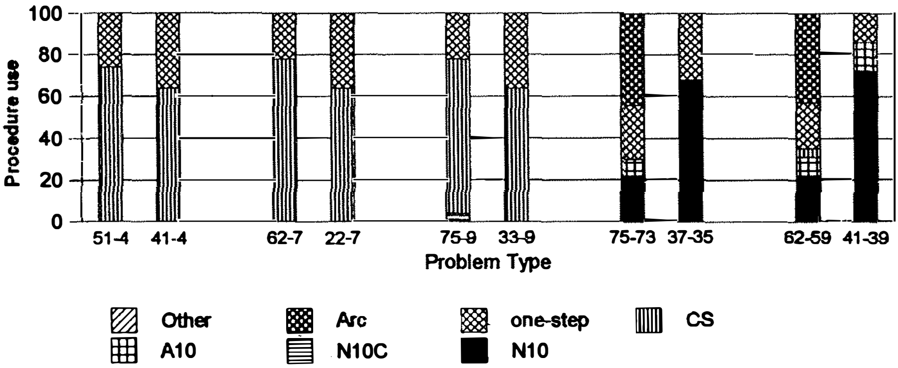


Figure 4.19 Solution procedures better RPD students (left bar) and GPD students (right bar) used to solve numerical subtraction problems by using the semi-structured number line on the Arithmetic Subject Matter Test in January

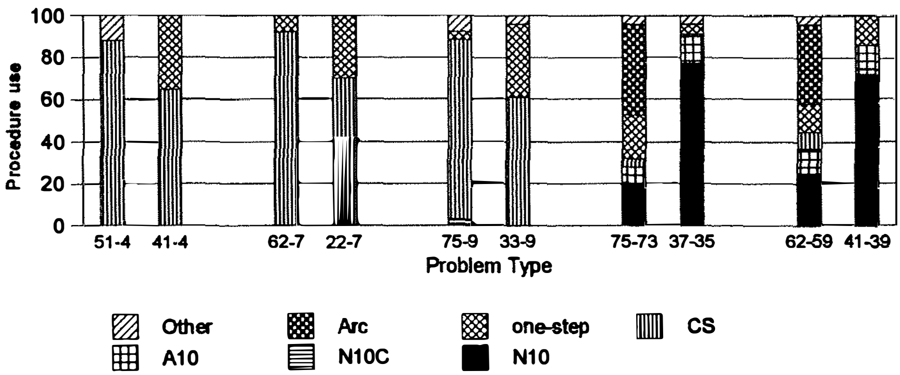


Figure 4.20 Solution procedures weaker RPD students (left bar) and GPD students (right bar) used to solve numerical subtraction problems by using the semi- structured num ber line on the Arithmetic Subject Matter Test in January

For the single-digit numerical subtraction problems we see the same pattern as for single-digit addition problems: Both the weaker and better RPD students used the CS procedure much more frequently than the weaker and better GPD students. They still solved these problems more frequently by making one jump on the number line. However, the weaker GPD students used the CS procedure more frequently for single-digit numerical subtraction problems than for single-digit numerical addition problems. For the multi-digit subtraction problems both the weaker and better RPD students used the Connecting Arc procedure most frequently. Since this procedure was not introduced to the GPD students, they mainly used the N10 procedure. With the better RPD students we see that they also solved this problem quite frequently by making one jump on the number line which may indicate that they had seen the difference between the two numbers.

Figures 4.21 and 4.22 show which solution procedures weaker and better RPD and GPD students used to solve addition and subtraction context problems.

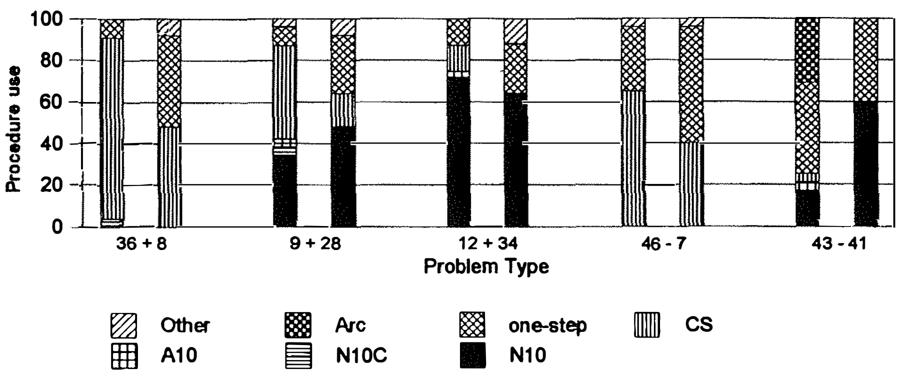


Figure 4.21 Solution procedures better RPD students (left bar) and GPD students (right bar) used to solve context addition and subtraction problems by using the semi-

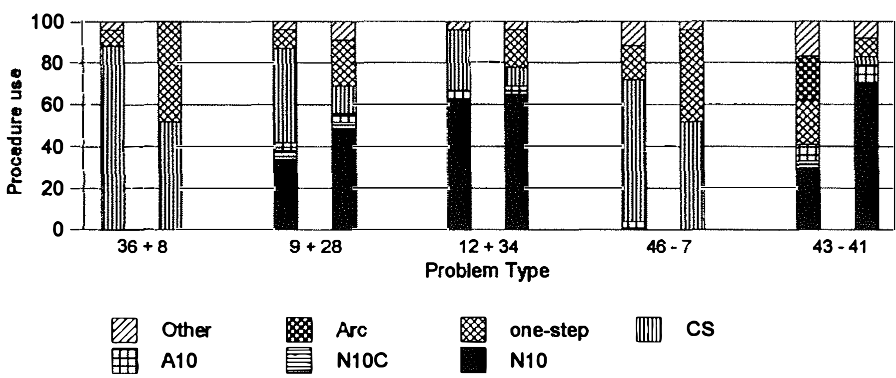


Figure 4.22 Solution procedures weaker RPD students (left bar) and GPD students (right bar) used to solve context addition and subtraction problems by using the semi-structured number line on the Arithmetic Subject Matter Test in January

For the context problems 36 - 8 and 46 - 7 the same pattern shows up as for the numerical problems: Both the weaker and better RPD students preferred the CS procedure for solving these problems whereas the weaker and better GPD students preferred solving these problems by making one jump on the number line. To solve the context addition problem 12 + 34, most weaker and better RPD and GPD students used the N10 procedure. There were also some weaker and better RPD students who solved the inversed problem 34 + 12 by using the CS procedure. When solving the context subtraction problem 43 - 41 the better, but more especially the weaker, GPD students preferred the N10 procedure. For the weaker and better RPD students the distribution across the different solution procedures was more diverse. The Connecting Arc procedure was used several times by weaker and better students but the problem was also frequently solved in one step. The better GPD students also used this last procedure several times which may indicate that they *saw* the difference between the two numbers at once.

Figure 4.23 shows how the weaker and better RPD and GPD students solved the two problems in which the students had to calculate the difference between two numbers.

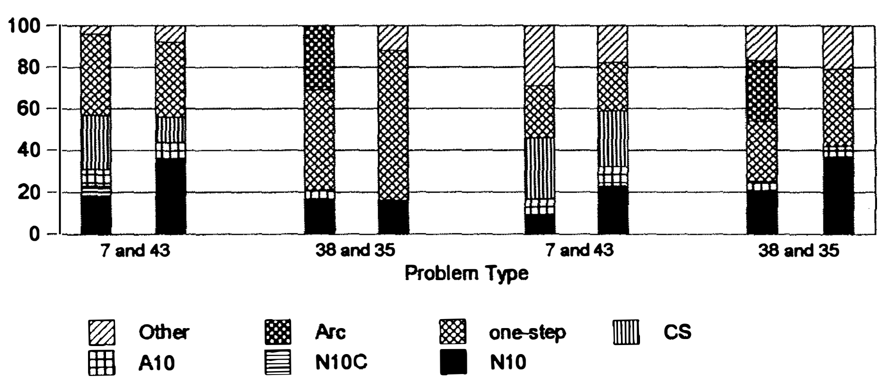


Figure 4.23 Solution procedures better (left-half) and weaker (right-half) RPD (left bar) and GPD (right bar) students used to solve context problems with differences by using the semi-structured number line on the Arithmetic Subject Matter Test in January

The patterns of solution procedures used for the difference problems resemble the patterns of the used solution procedures for the context subtraction problems: Better and weaker RPD students preferred the CS and Connecting Arc procedure respectively to solve these problems, whilst the better and weaker GPD students solved these problems in one step or by using the N10 procedure. However, the differences between the RPD and GPD students in solution procedures used to solve the difference problems were smaller than for the procedures used to solve the context subtraction problems.

MANOVAS for strategy use for weaker and better RPD and GPD students

We tested the differences between the weaker and better RPD and GPD students in their use of the CS procedure for single-digit problems by using 0/1 scores for use or non-use of this procedures. MANOVA with repeated measures revealed a significant main effect for type of program ($F(1, 91) = 19.2$, $p < .001$), for type of problem [Pillais $F(3, 89) = 112.5$, $p < .001$] but not for competence level. The interaction effect type of problem x type of program appeared to be significant [Pillais $F(3, 89) = 8.1$, $p < .001$] but the other interaction effects appeared not to be significant. This means that both the weaker and better RPD students used the CS procedure more frequently than the weaker and better GPD students. However, the weaker students (especially the weaker RPD students) did not use the CS procedure more frequently than the better students (especially the better RPD students).

Consistency of solution procedures across problems

Table 4.27 shows the patterns of solution procedures used by better and weaker RPD and GPD students to solve the numerical and context addition and subtraction problems on the Arithmetic Subject Matter test in January.

	Flexible	Half Flexible	CS/N10	Else/N10	Else
Better RPD students (n=22)	0	9	3	0	10
Better GPD students (n=25)	0	0	7	3	15
Weaker RPD students (n=23)	0	6	4	0	13
Weaker GPD students (n=21)	0	0	5	1	15

Note. CS stands for Complementary Structuring (see also Figure 2.5). Half Flexible means that in two of the four problem types, the students adopted their strategy use according to the number characteristics of the problem.

Table 4.27 Profiles of solution behavior of weaker and better RPD and GPD students on Arithmetic Subject Matter Test in January

Chi-square analyses were not reliable because many cells have an expected frequency of less than 5. However, we see that for both the better and the weaker RPD students there are some students considered as flexible in half of the number of problems while none of the GPD students were given this label.

4.4 Procedural and strategic knowledge: Weaker and better RPD and GPD students at the end of the curriculum

We will discuss the results concerning procedural competence, type of errors and strategic knowledge of the weaker and the better students half-way through the curriculum.

Procedural competence on researcher designed and external criterion tests

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
7. At the end of the program the quality of strategy and procedure use will have increased for the weaker RPD students. Therefore the differences in procedural competence between weaker and better RPD students will have become smaller.	8. At the end of the program the results for the weaker and better students will be the same as predicted from this point of view half-way through the second grade. Weaker GPD students will have relatively better scores and the differences between the better and weaker students will be significant larger in the RPD.
Outcome variables: • number of correct answers on Arithmetic Speed Test • number of correctly solved problems on Arithmetic Subject Matter Test • number of correctly solved problems on Arithmetic Scratch Paper Test • number of correctly solved problems on external criterion test (CITO LVS E4)	

Tables 4.28 and 4.29 show the mean number of correctly solved single-digit problems and standard deviations on the Arithmetic Speed Test in June for better and weaker RPD and GPD students.

	single-digit < 20 mean	s.d.	single-digit < 50 mean	s.d.	single-digit < 100 mean	s.d.
Better RPD students (n=23)	36.0	13.1	25.8	9.7	23.1	11.0
Better GPD students (n=23)	30.4	9.9	19.2	6.3	15.0	5.3
Weaker RPD students (n=24)	30.3	10.2	19.7	6.2	15.4	5.6
Weaker GPD students (n=23)	23.3	9.4	14.4	5.3	10.2	5.2

Table 4.28 Mean number of correctly solved single-digit addition problems on the Arithmetic Speed Test in June for better and weaker students

	single-digit < 20 mean	s.d.	single-digit < 50 mean	s.d.	single-digit < 100 mean	s.d.
Better RPD students (n=23)	34.3	15.5	25.1	10.8	20.2	9.4
Better GPD students (n=24)	28.2	10.3	19.5	6.3	15.8	5.7
Weaker RPD students (n=24)	24.6	9.9	17.8	6.5	13.5	5.4
Weaker GPD students (n=22)	23.3	9.6	13.2	6.8	10.6	5.4

Table 4.29 Mean number of correctly solved single-digit subtraction problems on the Arithmetic Speed Test in June for better and weaker students

MANOVA with repeated measures for single-digit problems showed significant main effects for type of program [$F(1, 88) = 10.3$, $p < .01$]; for competence level [$F(1, 88) = 15.9$, $p < .001$]; and for type of problem [Pillais $F(5, 84) = 100.8$, $p < .001$]. The interaction effect type of program x type of problem appeared also to be significant [Pillais $F(5, 84) = 2.6$, $p < .05$]. The other interaction effects appeared not to be significant.

For the better RPD and GPD students and weaker RPD and GPD students, separate ANOVAS were performed for each type of addition and subtraction problem. When we compared better RPD and GPD students we found significant differences for single-digit addition problems with numbers < 50, $F(1, 44) = 7.7$, $p < .01$, and for single-digit addition problems with numbers < 100, $F(1, 44) = 10.1$, $p < .01$. For single-digit subtraction problems we only find a significant difference between better RPD and GPD students for problems with numbers < 50: $F(1, 45) = 4.9$, $p < .05$. For all these problem types the better RPD students solved correctly more problems within three minutes than the GPD students.

For the weaker RPD and GPD students we found significant differences for single-digit addition problems with numbers < 20, < 50 and < 100: $F(1, 45) = 5.8$, $p < .05$; $F(1, 45) = 9.7$, $p < .01$ and $F(1, 45) = 10.6$, $p < .01$ respectively. For the single-digit subtraction problems we only found significant differences for problems with numbers < 50: $F(1, 44) = 5.6$, $p < .05$. For these problems the weaker RPD students also solved more answers correctly than the weaker GPD students.

	addition < 50 without regrouping mean	s.d.	addition < 50 with regrouping mean	s.d.	subtraction < 50 without regrouping mean	s.d.	subtraction < 50 with regrouping mean	s.d.
Better RPD students (n=23)	20.8	12.4	13.6	7.2	13.4	11.7	10.5	7.0
Better GPD students (n=24)	17.6	7.4	10.7	4.2	12.8	9.3	8.0	5.7
Weaker RPD students (n=24)	14.6	4.6	10.3	5.0	7.9	5.2	7.1	4.4
Weaker GPD students (n=21)	11.5	5.2	8.3	3.5	7.6	5.5	6.2	4.5

Table 4.30 Mean number of correctly solved multi-digit problems with numbers < 50 on the Arithmetic Speed Test in June for weaker and betterRPD and GPD students

	addition < 100 without regrouping mean	s.d.	addition < 100 with regrouping mean	s.d.	subtraction < 100 without regrouping mean	s.d.	subtraction < 100 with regrouping mean	s.d.
Better RPD students (n=23)	21.3	12.2	13.0	8.6	12.2	10.1	10.7	8.7
Better GPD students (n=24)	16.3	7.0	10.1	5.5	10.3	6.8	8.3	5.6
Weaker RPD students (n=24)	14.0	6.1	9.3	5.5	8.7	5.5	7.0	5.7
Weaker GPD students (n=22)	9.4	5.0	6.1	3.4	5.4	4.1	4.6	3.5

Table 4.31 Mean number of correctly solved multi-digit problems with numbers < 100 on the Arithmetic Speed Test in June for RPD and GPD students

MANOVA with repeated measures for multi-digit addition and subtraction problems revealed significant main effects for type of program [$F(1, 88) = 4.2$, $p < .05$], for competence level [$F(1, 88) = 12.9, p < .01$], and for type of problem [Pillais $F(7, 82) = 26.8, p < .001$]. The interaction effect competence level x type of problem appeared also to be significant: Pillais $F(7, 82) = 2.3, p < .05$] appeared not to be significant.

For both the better and the weaker RPD and GPD students, separate ANOVAS were carried out for the different types of problems. For the better RPD and GPD students we did not find significant differences for any of the multi-digit addition and subtraction problems.

For the weaker RPD and GPD students we found significant differences for multi-digit addition problems without regrouping with numbers < 50, $F(1, 43) = 4.7$, $p < .05$, and for the same problems with numbers < 100, $F(1, 44) = 7.6$, $p < .01$. For multi-digit addition problems with regrouping, we found significant differences for problems with numbers < 100: $F(1, 44) = 5.2, p < .05$. No significant differences were found between weaker RPD and GPD students for the same kind of problems with numbers < 50. For the multi-digit subtraction problems, significant differences between weaker RPD and GPD students were only found for subtraction problems without regrouping with numbers < 100: $F(1, 44) = 5.1$, $p < .05$. For all these significant differences accounted that the weaker RPD students correctly solved more problems than the weaker RPD students.

In June, procedural competence was also measured by analyzing the number of correctly solved problems on the Arithmetic Subject Matter Test and the Arithmetic Scratch Paper Test. Tables 4.32 and 4.33 show the mean number of correctly solved problems and standard deviations on the Arithmetic Subject Matter Test in June for the RPD and GPD students. The maximum score for the addition and subtraction problems that had to be solved in one way was 5. For the addition and subtraction problems that had to be solved in two ways the maximum score was 6.

	By head addition < 100		Number line addition < 100		Context addition < 100		Context: 2 ways addition < 100	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
Better RPD students (n=23)	3.7	.69	3.6	.50	3.7	.76	5.7	.71
Better GPD students (n=23)	3.8	.39	3.8	.52	3.7	.45	5.8	.49
Weaker RPD students (n=24)	3.4	.78	3.8	.65	3.5	.67	5.5	.90
Weaker GPD students (n=20)	3.6	.68	3.4	.77	3.6	.60	5.1	1.2

Table 4.32 Mean number of correctly solved addition problems on the Arithmetic Subject Matter Test in June for weaker and better RPD and GPD students

MANOVA with repeated measures did not reveal any significant main effects or two-way interaction effects for the first three types of addition problems of the Arithmetic Scratch Paper Test in June. However, the three-way interaction type of program x competence level x type of problem appeared to be significant: Pillais $F(2, 82) = 5.6$, $p < .05$. Separate ANOVAS for the different problem types revealed a significant interaction effect for type of program x competence level: $F(1, 81) = 2.0$, $p < .05$]. A separate ANOVA for context addition problems that had to be solved in two ways revealed a significant effect for competence level: $F(1, 81) = 5.0$, $p < .05$

	By head subtraction < 100		Context subtraction < 100		Number line subtraction < 100		Context: 2 ways subtraction< 100	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
Better RPD students (n=23)	3.7	.69	3.7	.54	3.6	.50	5.2	1.5
Better GPD students (n=24)	3.5	.78	3.3	.82	3.6	.72	5.1	1.1
Weaker RPD students (n=24)	3.1	1.1	3.7	.57	3.5	.90	5.2	1.3
Weaker GPD students (n=22)	2.6	1.3	3.2	1.1	3.4	.88	3.9	1.7

Table 4.33 Mean number of correctly solved subtraction problems on the Arithmetic Subject Matter Test in June for weaker and better RPD and GPD students

MANOVA with repeated measures for the first three problem types revealed significant main effects for type of program [$F(1, 85) = 5.2$, $p < .05$], and competence level [$F(1, 85) = 6.3$, $p < .05$]. The main effect type of problem appeared not to be significant. The interaction effect competence level x type of problem appeared also to be significant: Pillais $F(2, 84) = 3.9$, $p < .05$. The other interaction effects did not reach the level of significance. Separate ANOVAS for the different problem types revealed a significant difference between better and weaker students for subtraction problems that had to be solved mentally [$F(1, 72) = 4.0$, $p < .05$]. For the subtraction problems that had to be solved in two ways a significant main effect of type of program was found for the weaker students: $F(1, 33) = 4.6$, $p < .05$.

Table 4.34 shows the mean number of correctly solved problems (and standard deviations) on the Arithmetic Scratch Paper Test in June for weaker and better RPD and GPD students. The maximum score for numerical and context addition problems was 3. For numerical and context subtraction problems and context problems with a difference, the maximum score was 5.

	Addition NE		Addition C		Subtraction NE		Subtraction C		Difference	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
Better RPD students (n=23)	2.8	.42	2.7	.54	4.3	.88	4.6	.80	4.5	1.2
Better GPD students (n=22)	2.8	.39	2.9	.35	4.6	.59	4.6	.58	4.4	.78
Weaker RPD students (n=25)	2.7	.56	2.6	.65	4.4	1.0	4.3	.95	3.3	1.7
Weaker GPD students (n=22)	2.6	.58	2.2	.93	3.9	1.2	3.6	1.7	3.0	1.8

Table 4.34 Mean number of correctly solved numerical (N) and context (C) addition problems, numerical (N) and context (C) subtraction problems and difference problems on the Arithmetic Scratch Paper Test in June for weaker and better RPD and GPD students

MANOVA with repeated measures showed a significant effect for competence level: $F(1,84) = 11.24, p < .01$. In particular the difference problems seemed to be more difficult for weaker students than for better students. The main effect type of program and the interaction effect type of program x competence level appeared not to be significant.

Beside the procedural competence on the researcher designed tests we also looked at the scores on a more objective external criterion test at the end of the curriculum. Figure 4.24 shows the percentage of correctly solved problems by the weaker and better RPD and GPD students for the different subscales of the CITO LVS E4.

MANOVA with repeated measures revealed a significant main effect for competence level [Pillais $F(1, 67) = 85.3, p < .001$], and for type of problem [Pillais $F(9, 59) = 28.1, p < .001$]. The main effect type of program appeared not to be significant. The interaction effect type of program x competence appeared also to be significant: $F(1, 67) = 5.7, p < .05$. ANOVAS for the different subscales only revealed a significant interaction effect of type of program x competence level for division problems: $F(1, 68) = 8.1, p < .01$. For the other subscales no significant differences were found.

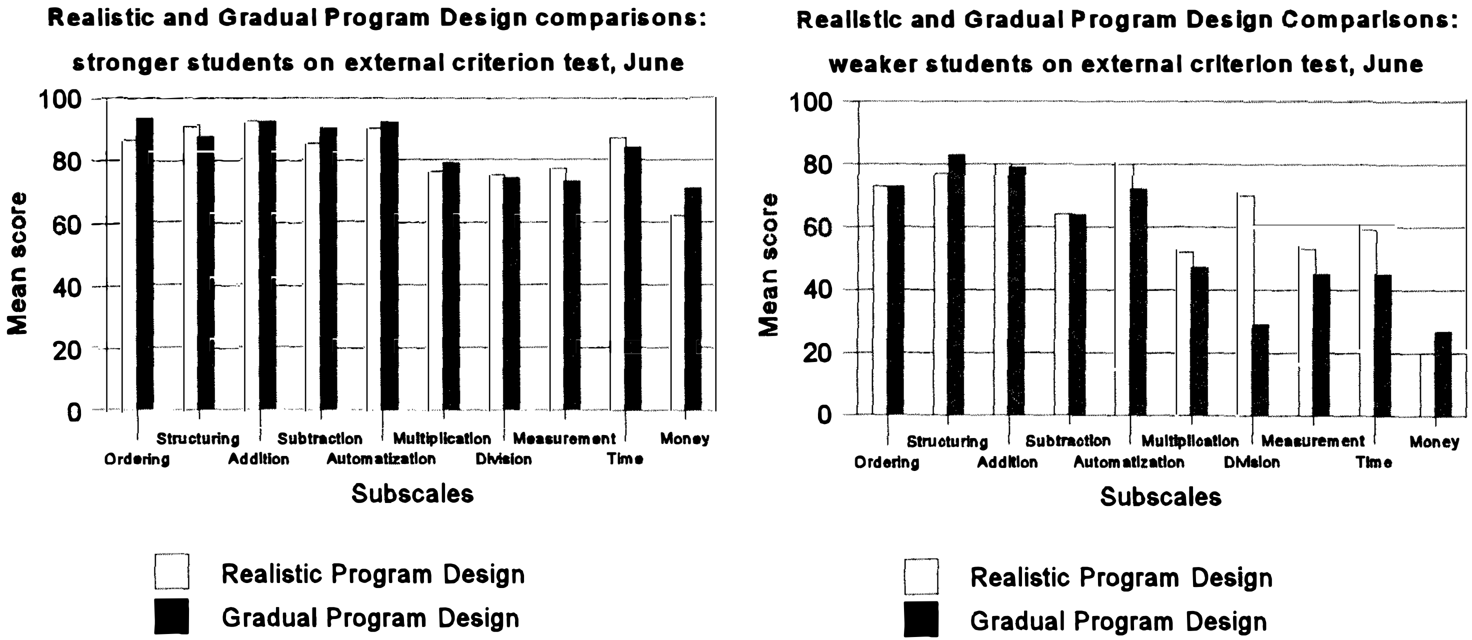


Fig 4.24 Percentages of problems correctly solved by better and weaker RPD and GPD students on external criterion test (CITO LVSM4) at the end of the second grade

Type of errors on the arithmetic subject matter tests and arithmetic scratch paper test

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
5. No explicit predictions were made about the type of errors students would make.	6. At the end of the program the results for the weaker and better students will be the same as predicted from this point of view half-way through the second grade. Therefore the weaker RPD students will still use many inadequate inventions or combinations of solution strategies and computation procedures (for instance confusion in the execution of the N10C procedure for addition and subtraction problems).
Outcome variables: • Type of errors on Arithmetic Subject Matter Tests	

Tables 4.35 and 4.36 show the number of procedural and non-procedural errors for the weaker and better RPD and GPD students on these two tests.

	non-procedural errors	procedural errors
Better RPD students (n=23)	28	21
Better GPD students (n=25)	32	18
Weaker RPD students (n=25)	58	12
Weaker GPD students (n=24)	71	27

Table 4.35 Number of procedural and non-procedural errors on the Arithmetic Subject Matter Test in June for weaker and better RPD and GPD students

	non-procedural errors	procedural errors
Better RPD students (n=23)	25	23
Better GPD students (n=25)	30	13
Weaker RPD students (n=25)	33	51
Weaker GPD students (n=24)	54	54

Table 4.36 Number of procedural and non-procedural errors on the Arithmetic Scratch Paper Test in June for weaker and better RPD and GPD students

Chi-square analyses revealed no significant differences between weaker and better RPD and GPD students in distribution among procedural and non-procedural errors for both tests.

Strategic use of computation procedures

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
7. At the end of the program the quality of strategy and procedure use will have increased for the weaker RPD students. Compared to hlaf-way through the curriculum the situation will be the same for the weaker GPD students. They will not be amenable to the adoption of new solution strategies or computation procedures, because they will stick to the use of the N10 procedure.	8. At the end of the program the results for the weaker and better students will be the same as predicted from this point of view half-way through the second grade. Weaker RPD students will still be more flexible in using different strategies and procedures, but, compared to the weaker GPD students, this use will be of a lower quality in both a strategic and a procedural sense.
Outcome variables: • Computation procedures used on Arithmetic Subject Matter Test • Computation procedures used on Arithmetic Scratch Paper Test • Notation forms used on Arithmetic Scratch Paper Test	

Figures 4.25 and 4.26 show the solution procedures weaker and better RPD and GPD students used to solve numerical addition problems on the empty number line.

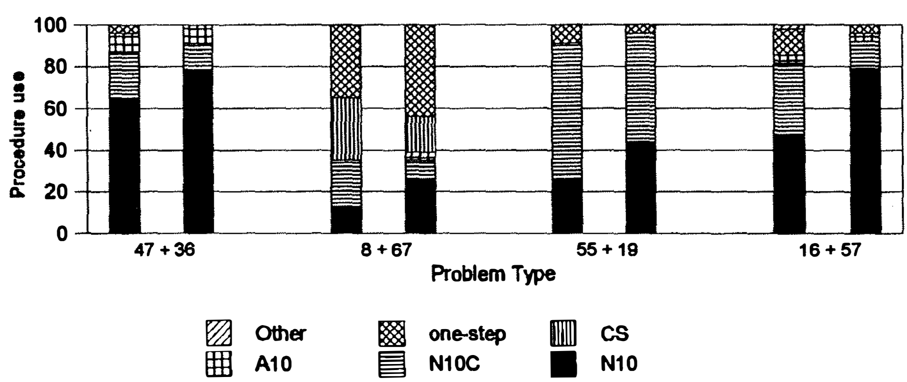


Figure 4.25 Solution procedures better RPD students (left bar) and GPD students (right bar) used to solve numerical addition problems by using the empty number-line on the Arithmetic Subject Matter Test in June

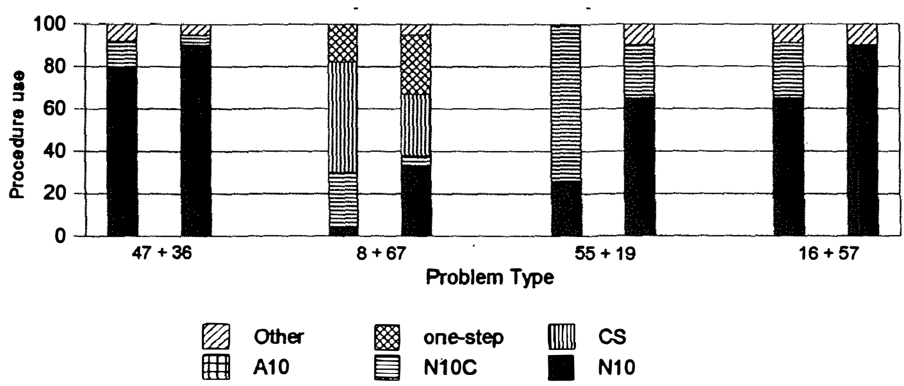


Figure 4.26 Solution procedures weaker RPD students (left bar) and GPD students (right bar) used to solve numerical addition problems by using the empty number line on the Arithmetic Subject Matter Test in June

In general both the better and the weaker GPD students preferred the N10 procedure to solve numerical addition problems where both the weaker and better RPD students also used other procedures like N10C. For the numerical addition problem 55 + 19 we see that the weaker and better RPD students preferred the N10C procedure. A small majority of the better GPD students also preferred the N10C procedure but 65% of the weaker GPD students kept using the N10 procedure to solve this problem.

Figures 4.27 and 4.28 show the solution procedures weaker and better RPD and GPD students used to solve context addition problems in one way on the empty number line.

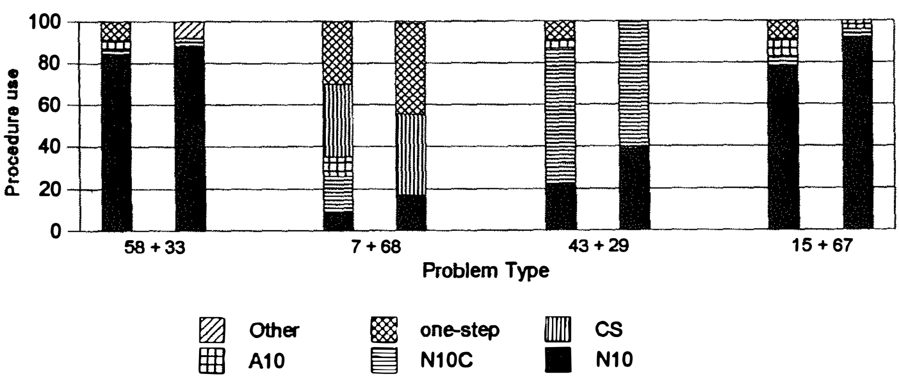


Figure 4.27 Solution procedures better RPD students (left bar) and GPD students (right bar) used to solve context addition problems by using the empty number line on the Arithmetic Subject Matter Test in June

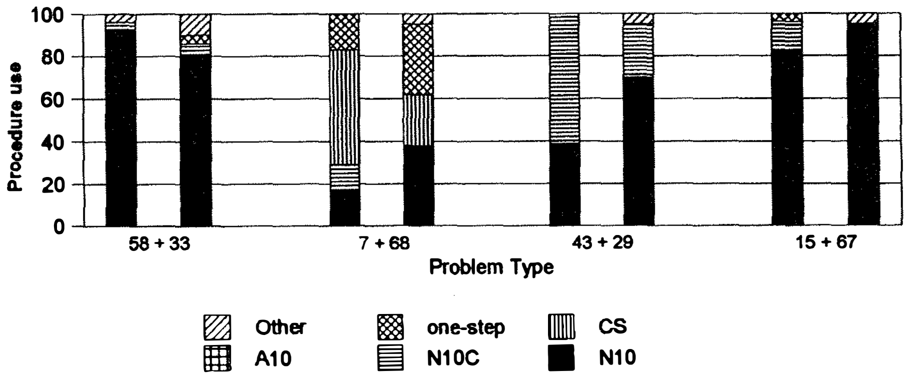


Fig 4.28 Solution procedures weaker RPD students (left bar) and GPD students (right bar) used to context addition problems by using the empty number line on the Arithmetic Subject Matter Test in June

For the context addition problems the weaker and better GPD students also kept using the N10 procedure most of the times. However, the differences with the weaker and better RPD students seems to narrow. The weaker RPD students, in particular, used the N10 procedure much more frequently for context addition problems than for the numerical addition problems. For instance, the numerical problem 55 + 19 was solved with the N10 procedure by 26% of the weaker RPD students whereas 39% of these pupils used the N10 procedure to solve the context addition problem 43 + 29.

The solution procedures that were used by the weaker and better RPD and GPD students to solve context addition problems in two ways on the empty number line are shown in Figures 4.29 and 4.30.

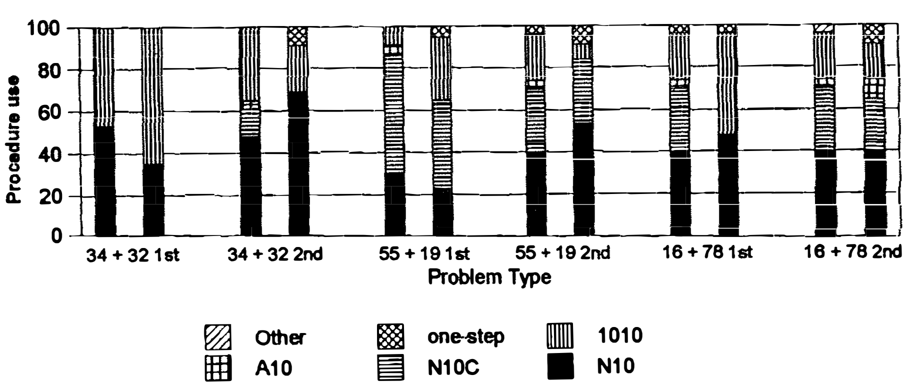


Figure 4.29 Solution procedures better RPD students (left bar) and GPD students (right bar) used to solve context addition problems in two ways on the Arithmetic Subject Matter Test in June

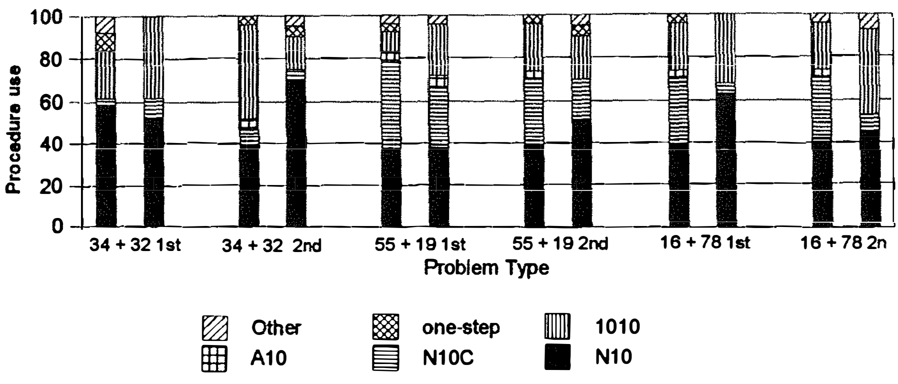


Figure 4.30 Solution procedures weaker RPD students (left bar) and GPD students (right bar) used to solve context addition problems in two ways on the Arithmetic Subject Matter Test in June

For the weaker and better RPD students we see that most of the students changed to a different procedure the second time they had to solve a problem, especially for the context addition problems 34 + 32 (1010 and N10) and 55 + 19 (N10C and N10). Most of the better GPD students also used a different solution procedure the second time they had to solve a problem. Only the weaker GPD students kept using the N10 procedure, even when they had to solve a problem a second time.

Figure 4.31 and Figure 4.32 show the solution procedures weaker and better RPD and GPD students used to solve numerical subtraction problems on the empty number line.

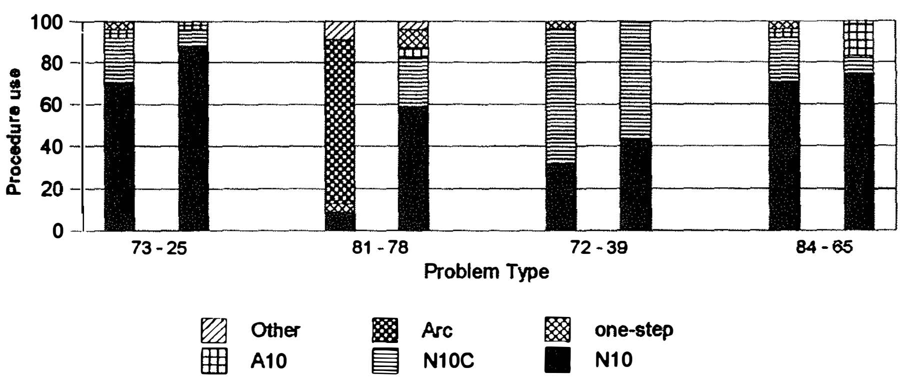


Figure 4.31 Solution procedures better RPD students (left bar) and GPD students (right bar) used to solve numerical subtraction problems by using the empty number-line on the Arithmetic Subject Matter Test in June

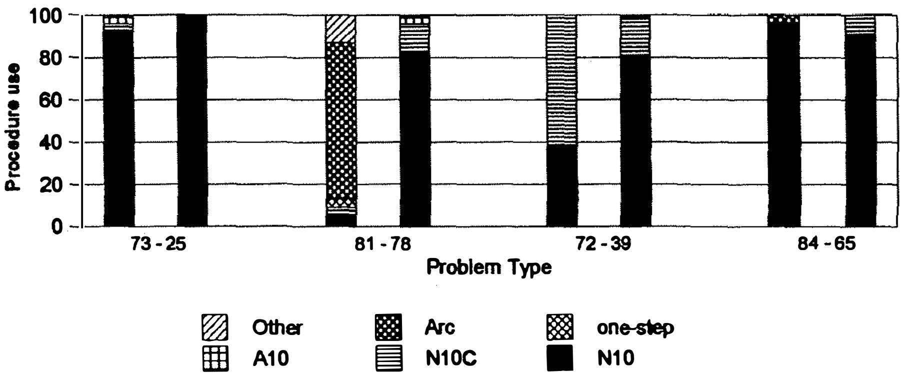


Figure 4.32 Solution procedures weaker RPD students (left bar) and GPD students (right bar) used to solve numerical subtraction problems by using the empty number line on the Arithmetic Subject Matter Test in June

Here we see the same pattern occurred as for the addition problems that were solved in two ways: The weaker and better RPD students and the better GPD students changed their solution procedure according to the number characteristics of the problems. Only the weaker GPD students stuck to the N10 procedure for all four numerical subtraction problems.

The solution procedures that were used to solve context subtraction problems on the empty number line are shown in Figures 4.33 and 4.34.

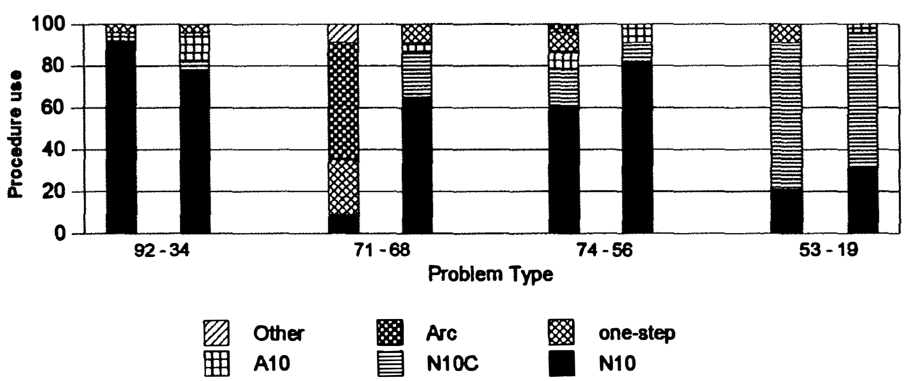


Figure 4.33 Solution procedures better RPD students (left bar) and GPD students (right bar) used to solve context subtraction problems by using the empty number line on the Arithmetic Subject Matter Test in June

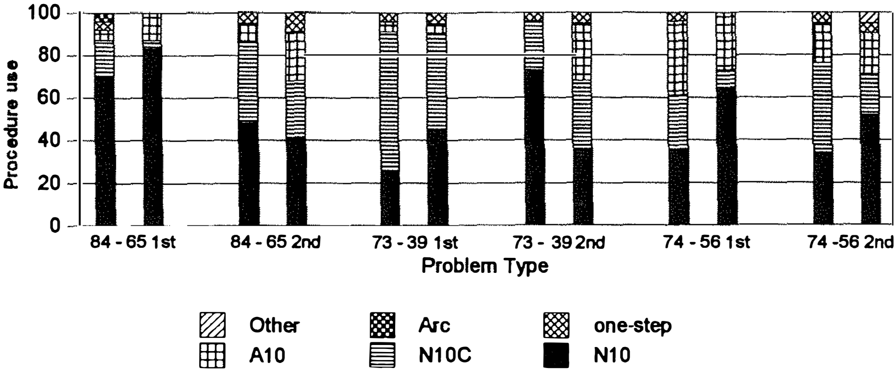


Figure 4.34 Solution procedures weaker RPD students (left bar) and GPD students (right bar) used to solve context subtraction problems by using the empty number line on the Arithmetic Subject Matter Test in June

The results for the context subtraction problems are similar to the use of solution procedures for solving numerical subtraction problems: The weaker and better RPD students and the better GPD students appeared to be more flexible in using different solution procedures than the weaker GPD students. Overall, the solution procedures used to solve numerical or context subtraction problems differred less than the solution procedures that were used to solve numerical or context addition problems. For context addition problems the N10 procedure was more frequently used than for numerical addition problems. This was not the case for context subtraction problems.

Figures 4.35 and 4.36 show the solution procedures that were used by weaker and better RPD and GPD students to solve context subtraction problems in two ways.

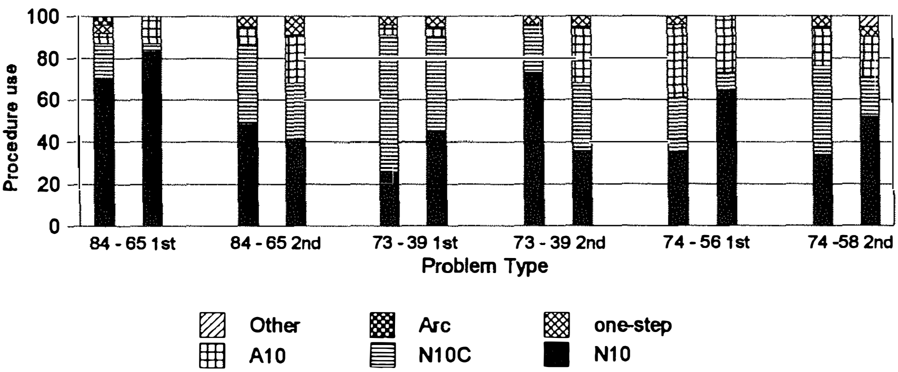


Figure 4.35 Solution procedures better RPD students (left bar) and GPD students (right bar) used to solve context subtraction problems in two ways on the Arithmetic Subject Matter Test in June

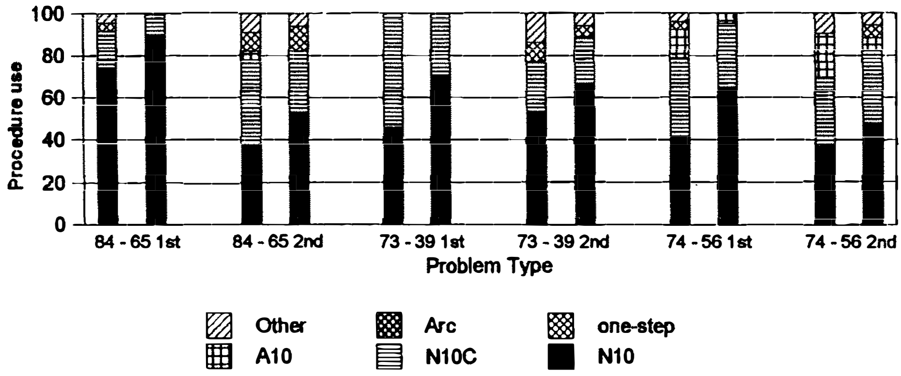


Figure 4.36 Solution procedures weaker RPD students (left bar) and GPD students (right bar) used to solve context subtraction problems in two ways on the Arithmetic Subject Matter Test in June

Most weaker and better RPD students used different procedures on the different occasions they had to solve context subtraction problems. The procedures they mainly used were the N10, N10C and A10 procedure. Some weaker RPD students also used the 1010 procedure during the second time they were asked to solved the problems. This always resulted in a wrong answer. Most better GPD students also used a different solution procedure the second time they were asked to solve a context subtraction problem, although to a lesser extent than the RPD students. Most weaker GPD students kept using the N10 procedure, even the second time they were asked to solve the problem.

Compared to the weaker RPD students, the weaker GPD students also made a lot of mistakes in solving the context subtraction problems. On average the context subtraction problem 73 - 39 was solved correctly by 85% of the RPD students whilst only 65% of the weaker GPD students solved this problem correctly. For the context subtraction problem 74 - 56 only 45% of the weaker GPD students solved the problem correctly whereas 80% of the weaker RPD students arrived at the correct answer.

Figures 4.37 and 4.38 show the solution procedures weaker and better RPD and GPD students used to solve addition problems of the numerical (N) and context (C) type on the Arithmetic Scratch Paper Test in June.

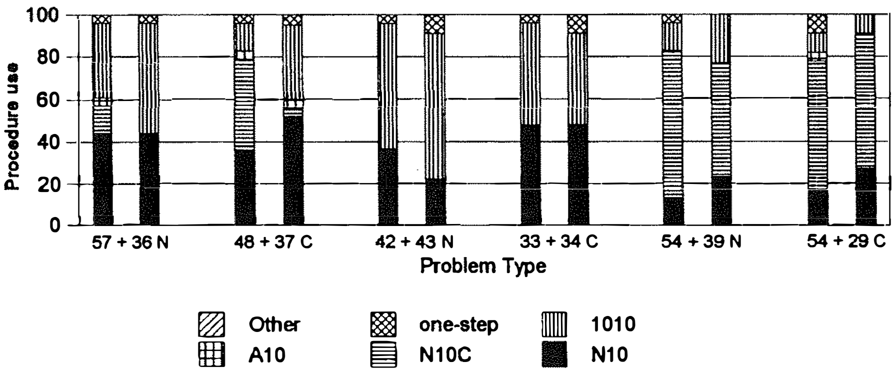


Figure 4.37 Solution procedures stronger RPD students (left bar) and GPD students (right bar) used to solve numerical (N) and context (C) addition problems on the Arithmetic Scratch Paper Test in June

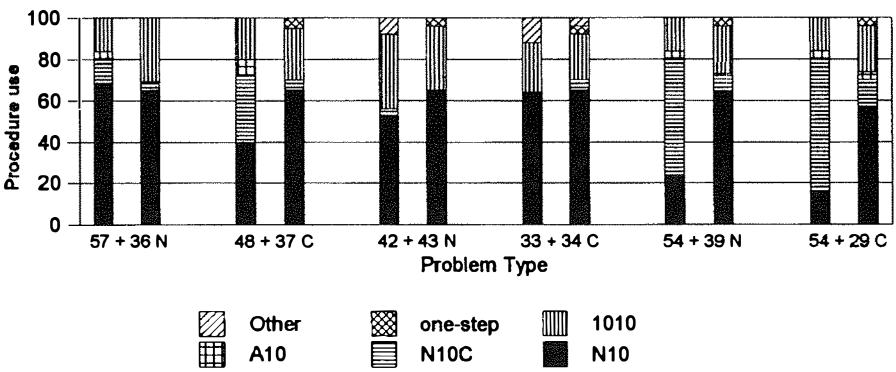


Figure 4.38 Solution procedures weaker RPD students (left bar) and GPD students (right bar) used to solve numerical (N) and context (C) addition problems on the Arithmetic Scratch Paper Test in June

We see that the weaker and better RPD students and the better GPD students changed their procedure use according to the number characteristics of the problem. We see that both groups of better students also preferred the 1010 procedure the most. This is true for numerical addition problems even more than for context addition problems. The weaker GPD students were not so flexible in changing their solution procedure according to the number characteristics of the problem: They preferred the N10 procedure to solve addition problems.

The solution procedures weaker and better RPD and GPD students used to solve numerical subtraction problems (N) and context (C) problems are shown in Figures 4.39 and 4.40.

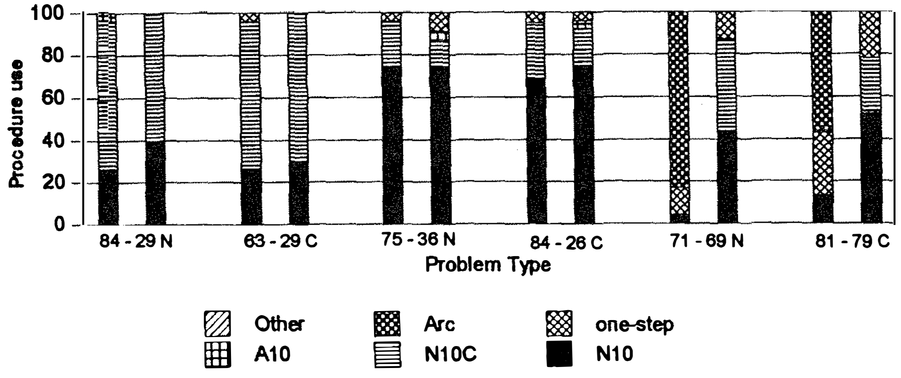


Figure 4.39 Solution procedures betterr RPD students (left bar) and GPD students (right bar) used to solve numerical (N) and context (C) subtraction problems on the Arithmetic Scratch Paper Test in June

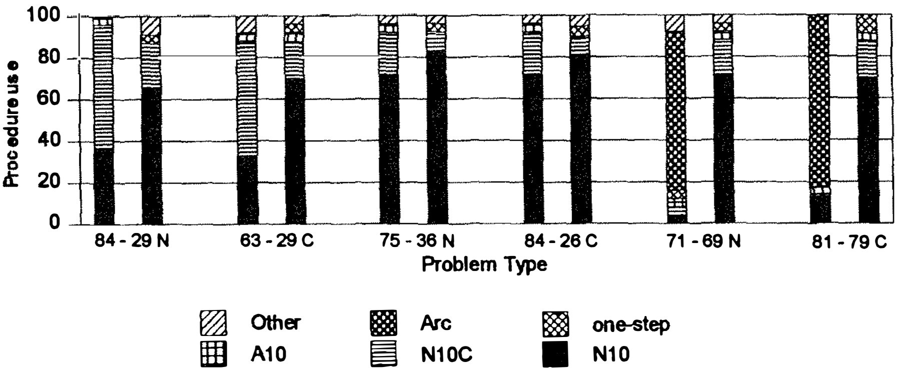


Figure 4.40 Solution procedures weaker RPD students (left bar) and GPD students (right bar) used to solve numerical (N) and context (C) subtraction problems on the Arithmetic Scratch Paper Test in June

As for the whole group of RPD and GPD students, the subtraction problems 62 - 48 (N), 72 - 58 (C), 65 - 33 (N), and 85 - 42 (C) are not included in Figures 4.39 and 4.40 because this would make the figures overly complicated. The patterns of procedures used to solve these problems resembled the patterns of procedures used for the subtraction problems 75 - 36 (N) and 84 - 26 (C). Similarily to the addition problems, the weaker and better RPD students and also the better GPD students changed their solution procedures according to the number characteristics of the subtraction problems. The weaker GPD students stuck to the N10 procedure to solve the subtraction problems. An example may clarify the difference in solution behavior between weaker and better GPD students. Since the Connecting Arc procedure was not introduced to the GPD students they did not use this procedure for problems like 71 - 69. We see that most of the weaker GPD students

(74%) solved this problem by using the N10 procedure. About half of the better GPD students (44%) also used the N10 procedure to solve this problem. However, 43% of the better GPD students used the N10C procedure to solve this problem which is also an efficient solution procedure for this kind of problems.

Figures 4.41 and 4.42 show the solution procedures weaker and better RPD and GPD students used to solve context problems in which they had to calculate a difference between two numbers.

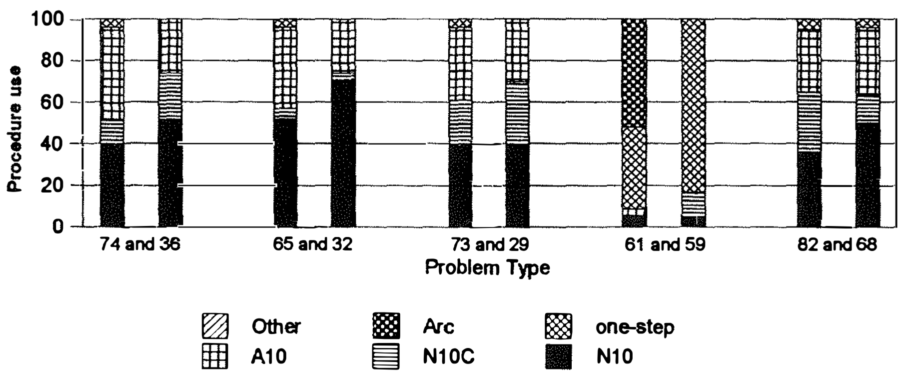


Figure 4.41 Solution procedures better RPD students (left bar) and GPD students (right bar) used to solve context problems with differences on the Arithmetic Scratch Paper Test in June

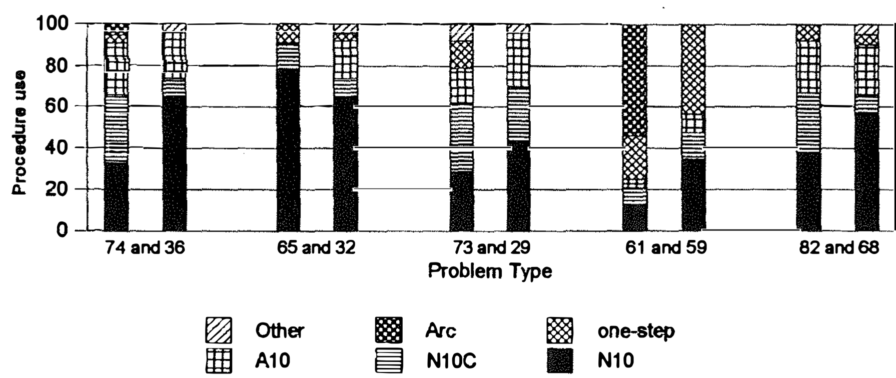


Figure 4.42 Solution procedures weaker RPD students (left bar) and GPD students (right bar) used to solve context problems with differences on the Arithmetic Scratch Paper Test in June

The procedures weaker and better RPD and GPD students used did not differ much. We see an increased use of the A10 procedure for solving difference problems.

Many RPD students *bridged* the difference between 61 and 59 by using the Connecting Arc, while most better GPD students saw the difference between 61 and 59 at once and used a one-step jump to solve the problem.

MANOVAS for strategy use by weaker and better RPD and GPD students

We tested the differences in respect of the whole group of RPD and GPD students between the weaker and better RPD and GPD students for the use of the N10 procedure by using 0/1 scores for either using this procedure or not. MANOVA with repeated measures on the procedures used to solve numerical and context addition and subtraction problems solved in one way on the Arithmetic Subject Matter Test in June revealed a significant effect for type of program [$F(1, 82) = 8.1$, $p < .01$]; for competence level [$F(1, 82) = 8.3$, $p < .01$]; and for type of problem [Pillais $F(3, 80) = 4.7$, $p < .01$]. The interaction effect of type of program x competence level appeared not to be significant. ANOVAS on the use of the N10 procedure for the different problem types revealed a significant effect for type of program for numerical addition problems [$F(1, 82) = 11.1$, $p < .01$] and for numerical subtraction problems [$F(1, 82) = 4.5$, $p < .05$]. For these problem types the GPD students stuck more to the N10 procedure than the RPD students. The main effect competence level was significant for numerical subtraction problems [$F(1, 82) = 12.0$, $p < .01$] and context subtraction problems [$F(1, 82) = 7.8$, $p < .01$]. For these problem types the weaker students used the N10 procedure more frequently than the better students. ANOVA failed to reveal a significant interaction effect for any problem type.

For the Arithmetic Scratch Paper Test the same type of analyses were done. MANOVA with repeated measures revealed a significant effect for type of program [$F(1, 85) = 4.6$, $p < .05$] and for competence level [$F(1, 85) = 6.7$, $p < .05$]. The interaction effect type of program x competence level appeared not to be significant. ANOVAS for the different problem types revealed a significant effect in use of the N10 procedure for type of program for context addition problems [$F(1, 85) = 6.1$, $p < .05$]. For the other problem types no significant effect was found for type of program. Competence level was only significant for numerical addition problems [$F(1, 85) = 9.9$, $p < .01$]. ANOVAS did not reveal a significant interaction effect type of program x competence level for any of the different problem types.

Consistency of solution procedures across problems

Until now we have looked at the solution procedure use by the whole group of weaker and better RPD and GPD students. However, we were also interested in how many students changed their solution procedure according to the number characteristics of the problems. Table 4.37 shows the profiles of solution procedures, weaker and better RPD and GPD students used to solve addition and subtraction problems in one way, on the Arithmetic Subject Matter Test in June. The same categories and criteria were used as for the analyses of the whole group of students.

	Flexible	Half Flexible	N10	N10C	Else
Better RPD students (n=23)	6	7	4	1	5
Better GPD students (n=22)	8	3	6	2	3
Weaker RPD students (n=23)	7	4	6	0	6
Weaker GPD students (n=18)	1	4	12	0	1

Table 4.37 Profiles of solution behavior of weaker and better RPD and GPD students on Arithmetic Subject Matter Test in June

Because many cells have an expected frequency of less than 5, chi-square analyses were not reliable. When we look at Table 4.37 we see that the largest difference is between the weaker RPD and GPD students. The number of weaker RPD students that can be considered as flexible or half-flexible is twice as large as the number of weaker GPD students with the same label. Most of the weaker GPD students used the N10 procedure to solve the problems on the Arithmetic Subject Matter Test in June.

Table 4.38 shows the patterns of solution procedures used to solve addition and subtraction problems in two ways on the Arithmetic Subject Matter Test in June.

	Flexible	Half Flexible	N10	1010/N10	Else
Better RPD students (n=20)	7	7	1	0	5
Better GPD students (n=20)	5	8	0	2	5
Weaker RPD students (n=20)	6	9	1	0	4
Weaker GPD students (n=15)	2	4	1	2	6

Table 4.38 Profiles of solution behavior of weaker and better RPD and GPD students on Arithmetic Subject Matter Test in June on problems that had to be solved in two ways

Since too many cells have an expected frequency of less than 5, chi-square analyses were not reliable. When we look at Table 4.38, we see that the differences between the weaker RPD and GPD students are the largest. Both the better RPD and GPD students chose a different procedure the second time they had to solve a problem. This was also true for most of the weaker RPD students, but the weaker GPD students were less flexible in their use of a procedure other than the N10 procedure.

Table 4.39 shows the patterns of used solution procedures by weaker and better RPD and GPD students, to solve the problems of the Arithmetic Scratch Paper test in June.

	Flexible	Half Flexible	N10	N10C	1010/N10	Else
Better RPD students (n=22)	6	1	3	4	2	6
Better GPD students (n=22)	7	3	5	0	1	6
Weaker RPD students (n=25)	7	2	3	4	2	7
Weaker GPD students (n=20)	0	1	10	1	1	7

Table 4.39 Profiles of solution behavior of weaker and better RPD and GPD students on Arithmetic Scratch Paper Test in June

Again the chi-square analyses were not reliable because of too many cells with expected frequencies of less than 5. When we look at Table 4.39 we see that the distribution of the better GPD students across the different categories is almost equal to the distribution of the better RPD students. There are even more better GPD students who could be classified as flexible or half-flexible. With respect to the weaker RPD and GPD students we see that most of the weaker GPD students used the N10 procedure to solve the problems on the Arithmetic Scratch Paper Test. About one-third of the weaker RPD students can be regarded as flexible or half flexible in solving these problems.

4.5 Transfer and retention

At the end of the second grade (June) we also asked the students to solve problems with numbers > 100. This was done as a kind of transfer test. The students had only very little experience with problems with numbers > 100. This test was repeated at the beginning of the third grade (November). We administered the Arithmetic Scratch Paper Test again at this time to see how the students solved these problems when they no longer worked with one of the two experimental programs.

Transfer at the end of the second grade

KLEIN (post hoc questions)
14. Do we see any transfer from what the students learned for addition and subtraction in the number domain of 0-100 to the number domain 0-1000?
Outcome variables: • number of correct of correct answers and solution procedures used on Arithmetic Transfer Test at the end of the second grade

Table 4.40 shows the number of correct answers for the different problems on the Arithmetic Transfer Test at the end of the second grade (for examples of the problems see Table 3.11). The maximum score numerical addition and subtraction problems was 3. For context problems the maximum score was 2.

	Numerical addition		Numerical subtraction		Context problems	
	mean	s.d.	mean	s.d.	mean	s.d.
RPD students (n=126)	2.4	.90	2.1	.93	1.2	.75
GPD students (n=109)	2.6	.71	1.8	1.2	1.3	.75

Table 4.40 Number of correct answers by RPD and GPD students for numerical addition and subtraction problems and context problems on the Arithmetic Transfer Test in June

MANOVA with repeated measures did not show a significant effect for type of program. ANOVAS for the different problem types revealed a significant type of program effect for numerical subtraction problems: $F(1, 201) = 4.5$, $p < .05$. The RPD students solved more problems correctly (70%) than the GPD students (60%). For the other type of problems no significant effects were found.

For the numerical problems with numbers up to 1000 we can conclude that both groups of students showed quite a high level of competence in solving these problems. Because we had not paid much attention to these kind of problems during the second grade, we can conclude that there was a positive transfer effect from addition and subtraction in the domain of numbers up to 100 towards addition and subtraction with numbers up to 1000. For the context problems this transfer effect is less: This is partly caused by the large number of incorrect answers for the second context problem (Kino problem, developed by Sundermann and Selter, 1995).

As for the other tests, we also looked at the results of the weaker and better students. We will only mention here that no significant interaction effects for type of program x competence level were found. For a complete overview we refer to Klein (1997).

The solution procedures used by the RPD and GPD students to solve these problems are depicted in Figures 4.43, 4.44 and 4.45. Because the numbers of the problems were up to 1000 we also adapted the names of the procedures according to the size of the numbers. However, the nature of the different procedures resembles the same. For an overview of the different solution steps for problems with numbers up to 1000 we refer to Appendix C.

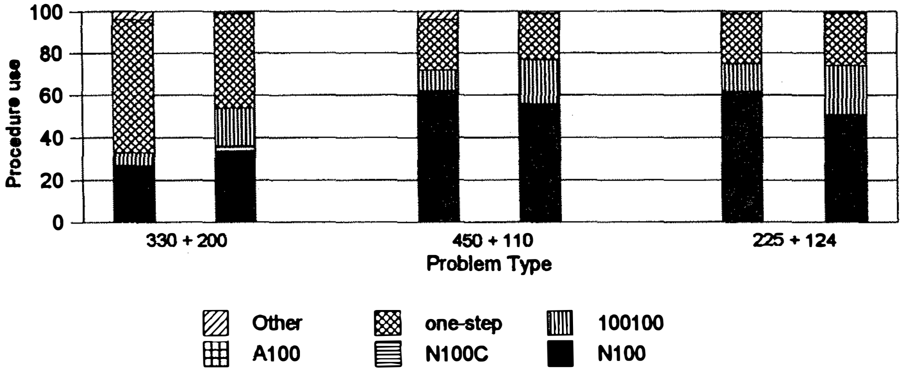


Figure 4.43 Solution procedures RPD students (left-bar) and GPD students (right-bar) used to solve numerical addition problems on the Arithmetic Scratch Paper Test in June

For the numerical addition problems with numbers > 100 there are small differences between the two program designs in use of the N100 procedure and 100100 procedure: RPD students tended to use the N100 procedure more frequently and the GPD students preferred the 100100 more frequently[2].

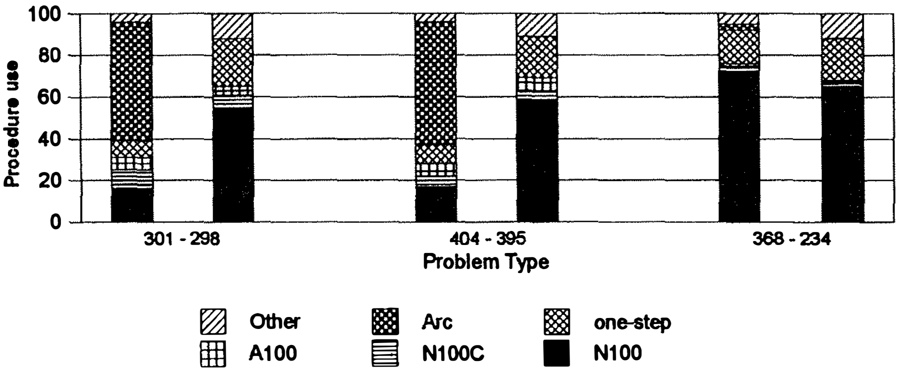


Figure 4.44 Solution procedures RPD students (left bar) and GPD students (right-bar) used to solve numerical subtraction problems on the Arithmetic Transfer Test in June

For numerical subtraction problems with numbers >100 the major difference between RPD and GPD students is the use of the Connecting Arc procedure, which had only been introduced to the RPD students. The GPD students used the N100

2 For the first numerical addition problem 330 + 200 it was difficult to distinguish between the solution procedures N100 and one-step. If the student made two jumps of 100 we scored this as N100, if the student made one jump of 200, we scored this as one-step.

procedure instead of the Connecting Arc procedure. The subtraction problem 368 - 234 was solved almost identically by RPD and GPD students.

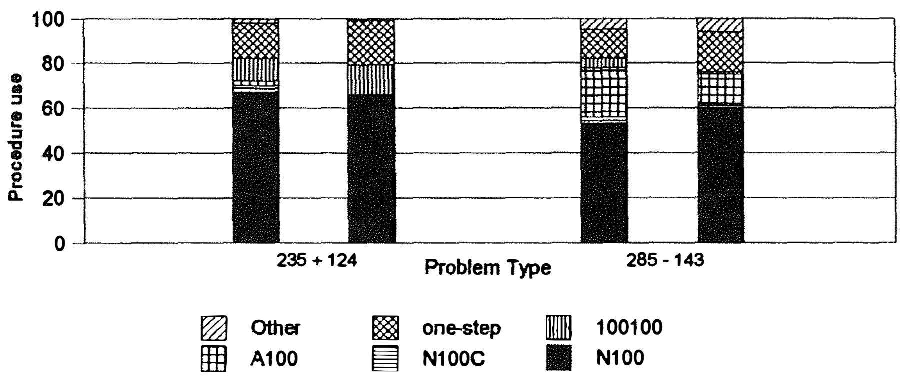


Figure 4.45 Solution procedures RPD students (left-bar) and GPD students (right-bar) used to solve context problems on the Arithmetic Transfer Test in June

The way the two context problems were solved by RPD and GPD students was not so different. For both problems the N100 procedure was the most popular solution procedure.

Transfer and retention at the beginning of the third grade

KLEIN (post hoc questions)	
14. Do we see any transfer from what the students learned for addition and subtraction in the number domain of 0-100 to the number domain 0-1000? 15. What is the retention of the strategies and procedures RPD and GPD students have learned for addition and subtraction up to 100, some months after they stopped working with the experimental program?	
Outcome variables:	• number of correct answers and solution procedures used on Arithmetic Transfer Test at the beginning of the third grade • number of correct answers, solution procedures and notation forms used on Arthmetic Scratch Paper Test at the beginning of the third grade

Table 4.41 shows the number of correct answers for the different problems on the Arithmetic Transfer Test at the beginning of the third grade. The maximum score numerical addition and subtraction problems was 3. For context problems the maximum score was 2.

	Numerical addition		Numerical subtraction		Context problems	
	mean	s.d.	mean	s.d.	mean	s.d.
RPD students (n=126)	2.6	.71	1.8	1.1	1.3	.70
GPD students (n=102)	2.7	.66	1.8	1.1	1.4	.69

Table 4.41 Number of correct answers by RPD and GPD students for numerical addition and subtraction problems and context problems on the Arithmetic Transfer Test at the beginning of the third grade (November)

MANOVA with repeated measures did not show a significant effect for type of program, nor did ANOVAS for the different problem type reveal a significant type of program effect. Overall we see that both RPD and GPD students solved about 90% of the numerical addition problems correctly and about 60% of the numerical subtraction problems. These percentages are comparable with the percentages of correctly solved problems on the Arithmetic Transfer Test in June. We still see that the context problems, and especially the *Kino* problem with 50% correct answers, were the most difficult ones. As for the Arithmetic Transfer test in June, we also looked for an effect of competence level. Here again we did not find an interaction effect between type of program and competence level.

The frequencies of the solution procedures the RPD and GPD students used to solve the problems on the Arithmetic Transfer Test at the beginning of the third grade are almost the same as at the end of the second grade. The only difference is that the especially the RPD students tended to use the 100100 procedure for numerical addition problems more frequently in November (23% in November versus 6% in June). For the GPD students this percentage remained more or less the same (23% in November versus 18% in June). The RPD students tended to use the Connecting Arc less frequently in November (58% in June, versus 40% in November). Because of these minor differences in procedure use we will not give the complete figures of the way RPD and GPD students solved these problems. For an complete overview we refer to Klein (1997).

At the start of the third grade we administered the Arithmetic Scratch Paper Test again. The number of correct answers is given in Table 4.42. The maximum scores for numerical and context addition problems was 3. For numerical and context subtraction problems, and context problems with a difference, the maximum score was 5.

	Addition NE		Addition C		Subtraction NE		Subtraction C		Difference	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
RPD students (n=130)	2.7	.57	2.7	.63	4.4	.88	4.2	.96	3.7	1.4
GPD students (n=111)	2.6	.67	2.5	.75	3.9	1.3	4.1	1.3	3.7	1.6

Table 4.42 Mean number of correctly solved numerical (N) and context (C) addition problems, numerical (N) and context (C) subtraction problems and difference problems on the Arithmetic Scratch Paper Test in November for RPD and GPD students

MANOVA with repeated measures revealed a significant effect for type of program: $F(1, 232) = 5.0, p < .05$. ANOVAS for the different problem types revealed significant differences between RPD and GPD students for context addition problems [$F(1, 237) = 6.6, p < .05$] and for numerical subtraction problems [$F(1, 237) = 10.8$, $p < .01$]. For these two problem types the RPD students solved more answers correctly than the GPD students.

When we compare these results with the results on the Arithmetic Scratch Paper Test in June (Table 4.14) than we see that the number of correctly solved problems in November was somewhat lower than in June. However, in November about 80% of the problems are also solved correctly. We can conclude that both groups of students were still competent in solving addition and subtraction problems with numbers up to 100. We will also analyze if this is true for the use of solution procedures and the way they wrote down their solution steps.

The way the students solved the problems on the Arithmetic Scratch Paper Test in the beginning of the third grade is shown in Figures 4.46, 4.47 and 4.48.
Figure 4.46 shows the solution procedures RPD and GPD students used to solve numerical and context addition problems at the beginning of the third grade (November).

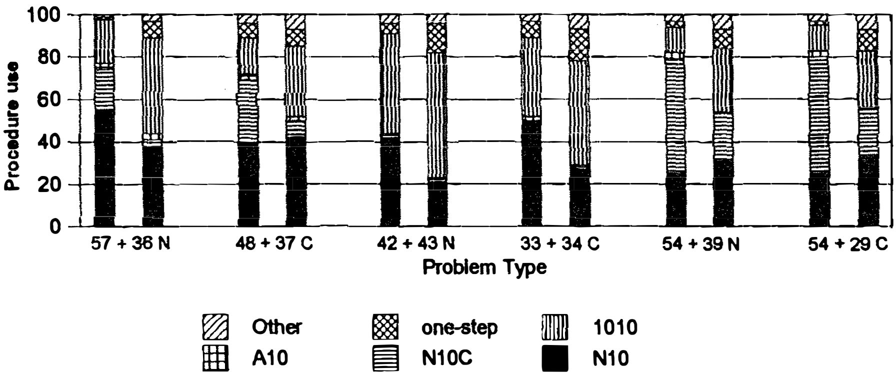


Figure 4.46 Solution procedures RPD students (left-bar) and GPD students (right-bar) used to solve numerical (N) and context (C) addition problems on the Arithmetic Scratch Paper Test, at the beginning of the third grade (November)

For the numerical and context addition problems we see that the GPD students used the 1010 procedure more frequently than the RPD students. The RPD students used the N10C procedure more frequently especially for the numerical problem 54 + 39 and the context problem 54 + 29. When we compare these results with the solution procedures that were used in June on the ASPT (Figure 4.13) then we see that the GPD students use the 1010 procedure for addition problems more frequently instead of the N10 procedure at the beginning of the third grade. The pattern of used solution procedures for the RPD students remains more or less the same.

Figure 4.47 shows the solution procedures the RPD and GPD students used to solve numerical and context subtraction problems on the ASPT at the beginning of the third grade.

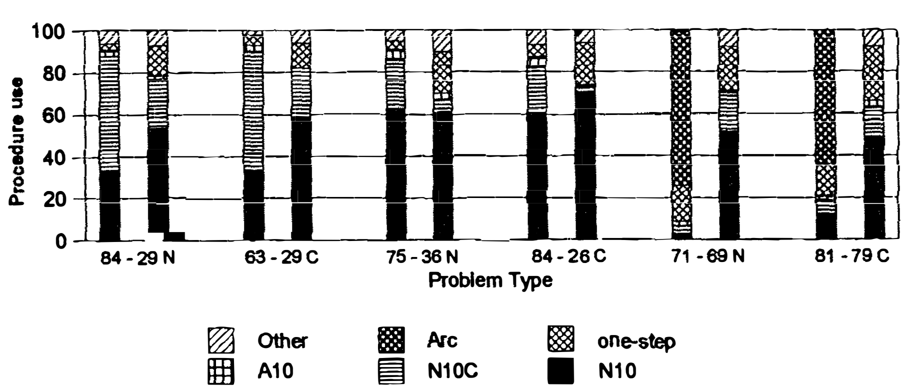


Figure 4.47 Solution procedures RPD students (left-bar) and GPD students (right-bar) used to solve numerical (N) and context (C) subtraction problems on the Arithmetic Scratch Paper Test, at the beginning of the third grade (November)

When we look at Figure 4.47 we see that more GPD than RPD students solved subtraction problems in one step. As with the addition problems the RPD students used the N10C procedure more frequently than the GPD students, especially for the subtraction problems 84 - 29 and 63 - 29. The RPD students still used the Connecting Arc procedure to solve the problems 71 - 69 and 81 - 79. When we compare these results with the solution procedures that were used in June to solve these problems (Figure 4.14) then we see that at the beginning of the third grade more GPD students solve subtraction problems in one step. Compared to June results on the ASPT, the RPD students used the N10C procedure less frequently in November.

The solution procedures used by the RPD and GPD students to solve context problems with a difference on the ASPT in the beginning of the third grade are shown in Figure 4.48.

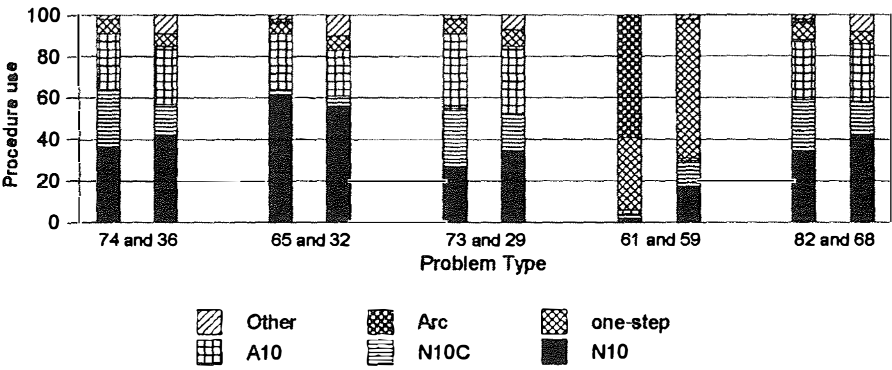


Fig 4.48 Solution procedures RPD students (left-bar) and GPD students (right-bar) used to solve context with differences on the Arithmetic Scratch Paper Test, at the beginning of the third grade (November)

Compared to the addition and subtraction problems we see that for context problems with a difference, both the RPD and GPD students used the A10 procedure more frequently. The differences in procedure used by both groups of students also differed less for these problems than for the other problems of the ASPT. When we compare these results with the solution procedures that were used in June to solve these context problems with a difference (Figure 4.15), then we see that the RPD students used the N10C procedure less frequently at the beginning of the third grade. The frequency of the use of the A10 procedure was almost the same at both moments.

Notation forms used on Arithmetic Scratch Paper Test in November

As on the Arithmetic Scratch Paper Test in June in November too the students were free to choose a notation form to write down their solution to the problem. This could be the number line, arrow scheme, or just solution steps to write down the answer. We were interested in what kind of notation forms the students would use, three months after they had worked with one of the two program designs. Table 4.43 shows the frequencies of type of notation form that was used by the RPD and GPD students.

		Addition NE	Addition C	Subtraction NE	Subtraction C	Difference
RPD students (n=130)	NL	35%	38%	46%	44%	56%
	Arrow	7%	4%	6%	6%	5%
	Steps	54%	54%	47%	46%	38%
	Other	4%	4%	1%	4%	1%
GPD students (n=111)	NL	3%	1%	3%	4%	16%
	Arrow	3%	2%	5%	4%	3%
	Steps	89%	90%	89%	90%	76%
	Other	5%	7%	3%	2%	5%

Note. NE stands for Numerical, C stands for Context, NL stands for Number Line, Arrow stands for Arrow Scheme

Table 4.43 Kind of notation form used by RPD and GPD students to solve problems on the Arithmetic Scratch Paper Test at the beginning of the third grade (November)

We see that the RPD students still used the number line more frequently than the GPD students. They preferred to write down their solution steps without using the number line or arrow scheme.

MANOVAS for strategy use by RPD and GPD students

Like the tests that were administered during the second grade, we also tested the differences between the RPD and GPD students in use of the N10 procedure at the beginning of the third grade. We used 0/1 scores for use or non-use of this procedure. MANOVA with repeated measures on the addition and subtraction problems revealed no significant main effect for type of program. The separate ANOVAS for the different problem types revealed only a significant difference for numerical addition problems: $F(1, 233) = 5.1, p < .05$. For these problems the RPD students used the N10 procedure more frequently than the GPD students. For the other problem types no significant differences were found between RPD and GPD students in use of the N10 procedure.

Because the GPD students used the 1010 procedure more frequently for addition problems at the beginning of the third grade, we also ran a MANOVA with repeated measures with 0/1 scores for use or non-use of the 1010 procedure. This analysis revealed a significant main effect for type of program [$F(1, 233) = 12.0$, $p < .01$]. Separate ANOVAS for the different problem types revealed significant differences for both numerical and context addition problems: $F(1, 233) = 16.1$, $p < .001$ and $F(1, 233) = 9.8$, $p < .01$ respectively. For both problem types the GPD students used the 1010 procedure more frequently than the RPD students.

Consistency of solution procedures across problems

Just as we had done for the tests during the second grade, we also looked at the consistency or flexibility of each student in using solution procedures across the different problems. Table 4.44 shows the profiles of solution procedures used to solve the different problems on the Arithmetic Scratch Paper Test in November. The same profiles and criteria were used as for the solution procedure used on the ASPT in June. (Table 4.19)

	Flexible	Half Flexible	N10	N10C	1010/N10	Else
RPD students (n=126)	29	7	23	19	5	43
GPD students (n=109)	11	4	26	2	18	48

Table 4.44 Profiles of solution behavior of RPD and GPD students on Arithmetic Scratch Paper Test at the beginning of the third grade (November)

Chi-square analyses revealed a significant difference between the RPD and GPD with respect to the distribution across the different profiles: $Chi^2 (5, N = 235) = 29.4$, $p < .001$. Inspection of the cells shows that the number of pupils who could be categorized as flexible or half-flexible was larger for the RPD than for the GPD. Also the number of students that used the N10C procedure for most of the problems was larger for the RPD than for the GPD. The number of students that used the 1010 procedure for addition and N10 for subtraction, was larger for the GPD students than for the RPD students.

5 Results: affective variables

In the previous chapter the results on the cognitive arithmetic tests were described. In this chapter data collected regarding motivational processes will be reported. Data were collected at the domain-specific level by asking students what their motivational beliefs were toward mathematics as a school subject. At the task-specific level we asked how the students appraised numerical and context problems.

First the answers on the questionnaires administered half-way through the second grade will be described regarding the hypotheses 9 and 10 (chapter 5.1). The same will be done for the answers given at the end of the second grade regarding the hypotheses 11 and 12 (chapter 5.2). Finally the answers of the sample of weaker and better students on these questionnaires will be described to give an answer to hypothesis 13 (chapter 5.3). At the beginning of each section the relevant hypotheses are summarized (for a more extensive description of the hypotheses the reader is referred to chapter 2). Conclusions based on these results will be drawn and discussed in chapter 6.

5.1 Motivational aspects at the domain-specific level

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
Predictions about motivational beliefs	
9. In the RPD the use of self-invented and informal strategies is encouraged. This will make the RPD students feel involved and give them pleasure in arithmetic which will result in a higher motivation towards mathematics than the GPD students.	10. Half-way through the program, there will be no differences between RPD and GPD students with respect to motivation for mathematics in general.
11. At the end of the program the RPD students will have a higher motivation towards arithmetic, but the differences between the RPD and GPD students will be smaller than half-way through the program.	12. At the end of the program there will be no significant differences between the two programs with respect to motivation for mathematics in general.
Outcome variables: • answers on the Mathematics Motivation Questionnaire half-way through and at the end of the curriculum	

We measured three aspects of motivation at the domain-specific level: affect towards mathematics, self-concept of mathematics ability and intended effort in doing mathematics. Table 5.1 shows the mean and standard deviations for these three motivational aspects half-way through the second grade (January). The scores for the three subscales range from 1 (low) to 4 (high).

	Affect		Self-Concept		Effort	
	mean	s.d.	mean	s.d.	mean	s.d.
RPD students (n=134)	3.1	.88	2.9	.61	3.3	.50
GPD students (n=123)	3.3	.83	3.1	.47	3.3	.50

Table 5.1 Means and standard deviations for the RPD and GPD students on the Mathematics Motivation Questionnaire in January

MANOVA with repeated measures revealed a significant effect for type of program $F(1, 251) = 3.8, p < .05$. Separate ANOVAS for the different subscales revealed significant differences for Affect [$F(1, 255) = 4.1, p < .05$] and for Self-Concept [$F(1, 253) = 8.0, p < .01$]. The GPD students had a more positive affect towards mathematics and a higher self-concept of mathematical ability than the RPD students half-way through the curriculum. No significant differences were found for intended effort between RPD and GPD students.

At the end of the curriculum (June) we administered the Mathematics Motivation Questionnaire again to the RPD and GPD students. The scores for affect towards mathematics, self-concept of mathematical ability and intended effort are shown in Table 5.2.

	Affect		Self-Concept		Effort	
	mean	s.d.	mean	s.d.	mean	s.d.
RPD students (n=134)	3.2	.79	3.1	.45	3.3	.40
GPD students (n=123)	3.1	.84	2.9	.46	3.3	.42

Table 5.2 Means and standard deviations for the RPD and GPD students on the Mathematics Motivation Questionnaire in June

MANOVA with repeated measures showed a significant main effect for type of program: $F(1, 217) = 3.92, p < .05$. ANOVAS for the different subscales showed a significant difference between RPD and GPD students with respect to self-concept of mathematical ability: $F(1, 227) = 4.5, p < .05$. The RPD students had a higher self-concept of mathematical ability than the GPD students. For the subscales affect towards mathematics and intended effort no significant differences were found between RPD and GPD students at the end of the second grade.

5.2 Motivational aspects at the task-specific level

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
Predictions about motivational beliefs	
9. RPD students will have more favorable cognitions and affects towards both number and context problems than GPD students, half-way through the program.	10. Half-way through the program, GPD students will have more favorable cognitions and affects towards numerical problems than RPD students. The opposite will be true with respect to context problems: RPD students will have more favorable cognitions and affects towards these problems than GPD students.
11. At the end of the program the RPD students will still have more favorable cognitions and affects towards numerical and context problems compared to the GPD students. The differences between the two groups of students will be smaller than half-way through the program.	12. At the end of the program there will be no significant differences between the two programs with respect to favorable cognitions and affects towards numerical problems and context problems.
Outcome variables: • answers on the On-line Motivation Questionnaire (7-10) half-way through and at the end of the curriculum.	

At the task-specific level we looked at motivational aspects towards numerical problems and context problems. The RPD and GPD students were asked to answer questions before and after solving these problems. The motivational aspects that could be measured in a reliable way before doing the task were self-confidence, task attraction, and task value/learning intention. After doing the task the effortful accomplishment and absence of threat could be measured in a reliable way. Table 5.3 show the scores of the RPD and GPD students on these motivational aspects for context and numerical problems half-way through the second grade (January). We also examined whether the context and numerical problems were correctly solved (cf. Table 4.14). Since the numbers of context and numerical problems were not the same (3 context problems and 6 numerical problems), we calculated percentages of correctly solved problems. These percentages are also shown in Table 5.3.

		confidence		attractiveness		task value		correctly solved problems	effortful ac-complishment		absence of threat	
		mean	s.d.	mean	s.d.	mean	s.d.	percentages	mean	s.d.	mean	s.d.
RPD students (n=127)	Context	3.4	.60	3.1	.87	3.6	.53	60%	3.4	.58	3.9	.42
	Numerical	3.5	.53	3.2	.81	3.5	.57	87%	3.5	.56	4.0	.20
GPD students (n=114)	Context	3.3	.52	3.0	.82	3.6	.53	71%	3.3	.64	3.9	.39
	Numerical	3.4	.49	3.0	.77	3.5	.63	87%	3.3	.63	3.9	.44

Table 5.3 Means, standard deviations for each of the motivational aspects and percentages correctly solved problems by RPD and GPD students on the On-line Motivation Questionnaire(7-10) in January

MANOVA with repeated measures were performed for each motivational aspect with type of problem as the within subjects effect and type of program as the between subjects effect. No significant main effect for type of program was found for any of the motivational aspects. The main effect type of problem (context versus numerical) was significant for self-confidence [$F(1, 239) = 19.9, p < .001$] and task value/learning intention [$F(1, 244) = 4.71, p < .05$]. Both RPD and GPD students were more self-confident about numerical problems than context problems. The task value and learning intention to solve problems was higher for context than for numerical problems for both RPD and GPD students. The interaction effect type of problem x type of program was significant for confidence [$F(1, 239) = 7.3, p < .01$]; task attractiveness [$F(1, 246= 5.6, p < .05$]; and absence of threat [$F(1, 244) = 5.2, p < .05$]. Separate ANOVAS revealed that RPD students' self-confidence for numerical problems was higher than the self-confidence of the GPD students for these problems [$F(1, 253) = 5.1, p < .05$]. The task attractiveness for numerical problems was also higher for RPD students than for GPD students [$F(1, 253) = 4.5, p < .05$]. Regarding absence of threat, the RPD students scored higher for numerical problems than the GPD students [$F(1, 254) = 5.7, p < .05$].

MANOVA with repeated measures for the percentage of correctly solved problems revealed a significant main effect for type of program [$F(1, 248) = 4.7, p < .05$]: The GPD students solved correctly more context and numerical problems than the RPD students. The main effect type of problem appeared also to be significant [$F(1, 248) = 94.8, p < .001$]: The percentage of correctly solved numerical problems was higher than the percentage of correctly solved context problems. The interaction effect type of program x type of problem also appeared to be significant [$F(1, 248) = 7.3, p < .01$]. Separate ANOVAS revealed a significant difference for the percentage of correctly solved context problems [$F(1, 256) = 9.0, p < .01$]: The percentage of correctly solved context problems was higher for the GPD than for the RPD students.

At the end of the second grade we administered the OMQ(7-10) again to the RPD and GPD students. The results are presented in Table 5.4.

		confidence		attractiveness		task value		correctly solved problems	effortful ac-complishment		absence of threat	
		mean	s.d.	mean	s.d.	mean	s.d.	percentages	mean	s.d.	mean	s.d.
RPD students (n=110)	Context	3.6	.38	3.0	.90	3.3	.64	81%	3.3	.56	3.7	.57
	Numerical	3.6	.41	3.0	.90	3.3	.70	86%	3.2	.58	3.8	.40
GPD students (n=114)	Context	3.5	.42	2.8	.92	3.2	.68	77%	3.3	.55	3.7	.54
	Numerical	3.5	.41	2.8	.94	3.3	.72	84%	3.2	.61	3.8	.51

Table 5.4 Means, standard deviations for each of the motivational aspects and percentages of correctly solved problems by RPD and GPD students on the On-line Motivation Questionnaire (7-10) in June

MANOVA with repeated measures for each motivational aspect with type of problem as the within subjects effect and type of program as the between subjects effect showed significant main effects for type of program for self-confidence [$F(1, 222) = 7.1, p < .01$] and task attractiveness [$F(1, 227) = 3.9, p < .05$]. The RPD students had higher scores on these two aspects for both numerical and context problems. The other main effect, numerical versus context problems, appeared to be significant for effortful accomplishment [$F(1, 219) = 4.9, p < .05$] and absence of threat [$F(1, 229) = 8.0, p < .01$] after completion of the task. Effortful accomplishment was higher for context problems where absence of threat was higher for numerical problems. The interaction effect of type of problem x type of program appeared not to be significant for any of the five motivational aspects.

MANOVA with repeated measures for the percentage of correctly solved context and numerical problems revealed a significant main effect for type of problem [$F(1, 226) = 12.4, p < .01$]: The percentage of correctly solved problems was significantly higher for context problems than for numerical problems. The main effect type of program, the interaction effect and also the separate ANOVAS appeared not to be significant.

Again we were interested in how the weaker and better RPD and GPD students would score on these five motivational aspects. In section 5.3 their cognitions, appraisals and results for context and numerical problems are described.

5.3 Motivational aspects for weaker and better RPD and GPD students

Klein (post hoc questions)	
13. Do weaker and better RPD and GPD students differ in their cognitions and appraisals towards arithmetic as a school subject and more specific towards numerical and context problems?	
Outcome variables:	• answers on the Mathematics Motivation Questionnaire and the On-line Motivation Questionnaire (7-10) half-way through and at the end of the curriculum

Motivational aspects at the domain-specific level

Table 5.5 shows the answers of the weaker and better RPD and GPD students on the Mathematics Motivation Questionnaire half-way through the curriculum.

	Affect		Self-Concept		Effort	
	mean	s.d.	mean	s.d.	mean	s.d.
Better RPD students (n=22)	3.3	.65	3.1	.60	3.4	.39
Better GPD students (n=23)	3.4	.77	3.2	.43	3.4	.37
Weaker RPD students (n=25)	3.2	.86	2.6	.60	3.2	.52
Weaker GPD students (n=23)	3.3	.74	2.8	.39	3.2	.61

Table 5.5 Means and standard deviations for the better and weaker RPD and GPD students on the Mathematics Motivation Questionnaire in January

MANOVA with repeated measures showed a significant main effect for competence level [F (1, 87) = 6.2, $p < .05$]. The main effect type of program and the interaction affect type of program x competence level appeared not to be significant. Separate ANOVAS for better RPD and GPD students and weaker RPD and GPD students revealed no significant differences for any of the three aspects.

Table 5.6 shows the scores of the weaker and better RPD and GPD students on the Mathematics Motivation Questionnaire at the end of the curriculum.

	Affect		Self-Concept		Effort	
	mean	s.d.	mean	s.d.	mean	s.d.
Better RPD students (n=23)	3.1	.78	3.2	.44	3.4	.40
Better GPD students (n=22)	3.2	.51	2.9	.28	3.3	.33
Weaker RPD students (n=25)	3.2	.90	2.9	.42	3.3	.47
Weaker GPD students (n=23)	3.0	.90	2.6	.52	3.2	.46

Table 5.6 Means and standard deviations for the better and weaker RPD and GPD students on the Mathematics Motivation Questionnaire in June

MANOVA with repeated measures showed no significant main effects for type of program or competence level. The interaction effect also appeared to be non-significant. Again separate ANOVAS for better and weaker RPD and GPD students were done. Both weaker and better RPD students appeared to have a significantly higher self-concept of mathematical ability than the weaker and better GPD students [$F(1, 39) = 5.4$, $p < .05$; $F(1, 41) = 6.1$, $p < .05$]. No significant differences were found for affect towards mathematics and intended effort.

Motivational aspects at the task-specific level

Table 5.7 shows the scores of the weaker and better RPD and GPD students on the five motivational aspects for context and numerical problems half-way through the second grade. The percentages correctly solved problems are also mentioned in this table.

		confidence		attractiveness		task value		correctly solved problems	effortful ac-complishment		absence of threat	
		mean	s.d.	mean	s.d.	mean	s.d.	percentages	mean	s.d.	mean	s.d.
Better RPD students (n=22)	Context	3.4	.50	3.1	.73	3.6	.38	68%	3.3	.54	3.9	.23
	Numerical	3.6	.43	3.2	.66	3.5	.48	88%	3.4	.45	4.0	.21
Better GPD students (n=22)	Context	3.5	.40	3.3	.53	3.7	.42	83%	3.4	.61	3.9	.31
	Numerical	3.6	.39	3.3	.62	3.6	.46	95%	3.4	.46	3.9	.33
Weaker RPD students (n=24)	Context	3.3	.58	3.2	.81	3.8	.35	57%	3.7	.33	3.8	.48
	Numerical	3.7	.46	3.4	.80	3.8	.35	87%	3.7	.35	4.0	.20
Weaker GPD students (n=19)	Context	3.2	.60	3.2	.87	3.7	.41	55%	3.4	.71	4.0	.15
	Numerical	3.3	.48	3.2	.74	3.6	.52	84%	3.5	.61	3.9	.35

Table 5.7 Means, standard deviations for each of the motivational aspects and percentages correctly solved problems by weaker and better RPD and GPD students on the On-line Motivation Questionnaire (7-10) in January

MANOVA with repeated measures for each motivational aspect, with type of problem as the within subjects effect and type of program and competence level as the between subjects effect revealed no significant main effects or interaction effect. Separate ANOVAS revealed only a significant difference for weaker RPD and GPD students with respect to self-confidence about numerical problems [$F(1, 44) = 6.9$, $p < .05$]: Weaker RPD students were more self-confident about solving numerical problems than weaker GPD students. For all the other motivational aspects we can conclude that there were no significant differences between better and weaker RPD and GPD students with respect to the motivatial aspect measured with the OMQ(7-10) for context and numerical problems.

MANOVA with repeated measures for percentage of correctly solved context and numerical problems revealed a significant main effect for competence level

[$F(1, 86) = 13.0$, $p < .05$]: The better students solved more problems correctly than the weaker students. The main effect type of problem appeared also to be significant [$F(1, 86) = 34.1$, $p < .001$]: The percentage of correctly solved numerical problems was higher than the percentage of correctly solved context problems. The interaction effect competence level x type of problem appeared also to be significant. All the other main and interaction effects, and also separate ANOVAS for better and weaker RPD and GPD students did not reach the level of significance.

The results of the weaker and better RPD and GPD students on the OMQ(7-10) at the end of the second grade are presented in Table 5.8

		confidence		attractiveness		task value		correctly solved problems	effortful accomplishment		absence of threat	
		mean	s.d.	mean	s.d.	mean	s.d.	percentages	mean	s.d.	mean	s.d.
Better RPD students (n=20)	Context	3.6	.40	3.2	.71	3.4	.64	82%	3.3	.57	3.8	.30
	Numerical	3.6	.33	3.1	.75	3.4	.61	86%	3.3	.57	3.9	.33
Better GPD students (n=22)	Context	3.5	.34	2.8	.84	3.1	.63	78%	3.1	.52	3.8	.66
	Numerical	3.5	.34	2.5	.81	3.0	.60	90%	2.9	.49	3.8	.65
Weaker RPD students (n=19)	Context	3.7	.36	3.3	.94	3.6	.44	65%	3.6	.62	3.7	.75
	Numerical	3.7	.37	3.3	.90	3.8	.46	75%	3.4	.70	4.0	.12
Weaker GPD students (n=23)	Context	3.3	.43	3.0	.82	3.3	.67	74%	3.5	.49	3.7	.52
	Numerical	3.4	.45	3.0	.81	3.4	.67	75%	3.4	.67	3.8	.44

Table 5.8 Means, standard deviations for each of the motivational aspects and percentages correctly solved problems by weaker and better RPD and GPD students on the On-line Motivation Questionnaire(7-10) in June

MANOVA with repeated measures, for each motivational aspect, with type of problem as within subjects variables and type of program and competence level as between subjects variables, revealed a significant main effect for type of program for self-confidence, task attractiveness and task value/learning intention: $F(1, 80) = 9.3$, $p < .05$; $F(1, 81) = 6.2$, $p < .05$; $F(1, 80) = 8.1$, $p < .05$. For both numerical and context problems the better and weaker RPD students scored higher on these three motivational aspects than the better and weaker GPD students. The main effect competence level was found to be significant for task value/learning intention [$F(1, 80) = 6.2$, $p < .05$], and effortful accomplishment [$F(1, 79) = 6.5$, $p < .05$]. For both aspects the weaker RPD and GPD students scored higher than the better RPD and GPD students. The main effect type of problem appeared only to be significant for absence of threat [$F(1, 81) = 6.9$, $p < .05$]. The absence of threat appeared to be higher for numerical problems than for context problems. The interaction effect type of program x competence level appeared only to be significant for self-confidence [$F(1, 80) = 4.0$, $p < .05$].

Separate ANOVAS for better RPD and GPD students revealed a significant effect for type of program for task attractiveness of numerical problems [$F(1, 41) =$

5.5, $p < .05$], task value/learning intention for numerical problems [$F(1, 40) = 4.7$, $p < .05$]; and effortful accomplishment after solving numerical problems [$F(1, 41) = 6.7$, $p < .05$]. For all these aspects the better RPD students had higher scores than the better GPD students. Separate ANOVAS for weaker RPD and GPD students showed significant differences for self-confidence about solving context and numerical problems [$F(1, 41) = 12.2$, $p < .01$; $F(1, 40) = 7.7$, $p < .01$], and task value/learning intention to solve context problems [$F(1, 41) = 4.5$, $p < .05$]. For these motivational aspects the weaker RPD students had higher scores than the weaker GPD students.

MANOVA with repeated measures for the percentage of correctly solved context and numerical problems for weaker and better RPD and GPD students showed a significant main effect for competence level [$F(1, 81) = 7.0$, $p < .05$]. The better students solved more problems correctly than weaker students. The main effect type of problem appeared also to be significant [$F(1, 81) = 4.2$, $p < .05$]. The percentage of correctly solved problems appeared to be higher for numerical problems than for context problems. The main effect type of program, the interaction effects and also the separate ANOVAS were not significant.

6 Discussion

In chapter 2 Treffers (realistic point of view) and Beishuizen (gradual point of view) formulated hypotheses about the results of the RPD and GPD students regarding procedural and strategic knowledge in the domain of addition and subtraction up to 100. They also hypothesized about the results of the weaker and better RPD and GPD students regarding these cognitive processes. Predictions were made for the whole group of RPD and GPD students, concerning motivational aspects towards mathematics as a school subject (domain-specific level) and cognitions and affects towards context and numerical problems (task-specific level). In this chapter we will compare the hypotheses with the outcomes of the experiment, and conclusions will be drawn. Klein also formulated post hoc questions concerning motivational processes for weaker and better RPD and GPD students, retention of what was learned in the second grade at the beginning of the third grade, and transfer from the domain of addition and subtraction up to 100 towards problems with numbers up to 1000. In discussing the results, we will first look at the results from the cognitive tests of the whole group of RPD and GPD students (sections 6.1 and 6.2), followed by the results of the weaker and better students (sections 6.3 and 6.4). Then we will look at the transfer (section 6.5) and retention (section 6.6) effects of the two program designs. This is followed by a discussion of the results from the affective questionnaires for both the whole group of RPD and GPD students (section 6.7), as well as the weaker and better students (section 6.8). Finally we will discuss these results in general and their implications for future research (section 6.9).

6.1 Development of procedural and strategic knowledge: RPD versus GPD students half-way through the curriculum

Conclusions about procedural knowledge

TREFFERS (realistic point of view)	**BEISHUIZEN (gradual point of view)**
1. Half-way through the program there will be no differences in the number of correctly solved problems for RPD and GPD students. The GPD students will solve more numerical problems in a limited amount of time than RPD students. RPD students will solve more context problems correctly than GPD students.	2. Half-way through the program there will be a higher level of procedural competence in numerical problems shown by the GPD rather than the RPD students which will be reflected by a higher number of correctly solved problems. The GPD students will solve more of these problems in a limited amount of time (speed test). With respect to context problems the GPD students will solve fewer problems correctly than the RPD students.

It turned out that, for five of the six problem types on the Arithmetic Speed Test in January, the RPD students solved more numerical problems correctly than the GPD students. For the number of correct answers on the Arithmetic Subject Matter Test no significant differences were found between the RPD and GPD students. This means that for context problems too, there were no differences between RPD and GPD students in the number of correctly solved problems. The same was true for the overall score on the external criterion test CITO LVS M4. The GPD students solved more problems correctly than the RPD students when it came to the subscale structuring of numbers.

Generally we can say that the procedural competence and fluency of the RPD students, regarding numerical problems, was higher than that of the GPD students, half-way through the curriculum. This is remarkable since the RPD put relatively less emphasis on written exercises with numerical problems. Instead, in this condition, more time was spent on interactive teaching and oral solution of such problems during whole-class teaching. These results indicate that the early introduction of different solution strategies (like N10C and the Connecting Arc), and the rapidly increasing size of the numbers during the first four months of the RPD (cf. Table 3.14), did not harm but stimulated the procedural competence of the RPD students. With respect to context problems, no significant differences were found between RPD and GPD students in level of procedural competence. This means that neither of the hypotheses from the Realistic and Gradual point of view were confirmed: For numerical problems the opposite was found of what was expected to be true and for the context problems no differences were found.

Conclusions about type of errors

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
1. Half-way through the curriculum there will be a difference in the type of errors students make: RPD students will make less procedural or conceptual mistakes than GPD students. RPD students will make more non-procedural mistakes (due to slovenliness) than GPD students.	2. No explicit predictions were made about the type of errors students would make half-way through the curriculum.

With respect to the errors made on the Arithmetic Subject Matter Test in January we saw that the total number of non-procedural errors was larger than the number of procedural errors. However, this was true for both RPD and GPD students. No differences in distribution across these two types of errors were found between RPD and GPD students. This means that half-way through the curriculum, there was no difference in level of insight between RPD and GPD students in the domain of addition and subtraction up to 100.

Conclusions about strategic use of computation procedures

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
1. Half-way through the program the RPD students will adapt their strategy use to the characteristics of a problem and therefore the RPD students will be more flexible in their use of solution procedures than GPD students. Half-way through the curriculum GPD students will only use the N10 procedure in a proceduralized way where the RPD students use different strategies and procedures.	2. With respect to flexible use of different strategies and procedures half-way through the curriculum, the GPD students will be more rigid than the RPD students. The GPD students will stick to the N10 procedure where the RPD students will also use other procedures as well.

On the Arithmetic Subject Matter Test in January their was not much difference between the RPD and GPD students with respect to the use of the N10 procedure. This is probably caused by the type of problems: Most of the problems in January, were not suited to using the N10 procedure because a single-digit had to be added or subtracted. For these problems it was expected that the Complementary Structuring (CS) procedure would be the most adequate procedure to solve these problems (cf. Table 3.5 and 3.6). It turned out that the RPD students used this procedure more frequently than the GPD students who preferred to solve these problems by making one jump on the number line. This is illustrated in Figure 6.1 in which an RPD and a GPD student solve the same problem on the Arithmetic Subject Matter Test in a different way.

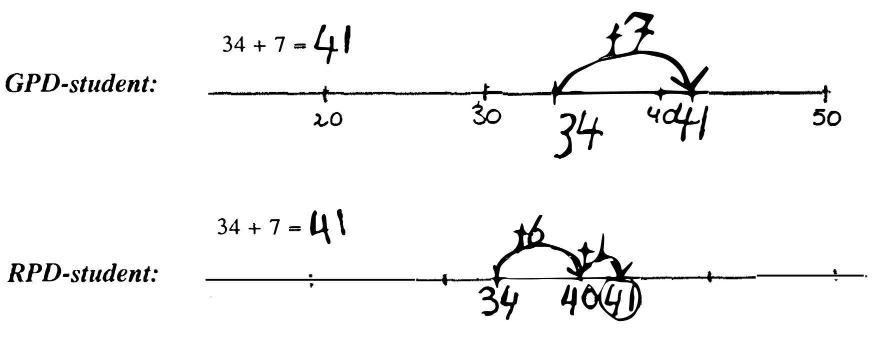


Figure 6.1 A GPD student solving the problem 34+7 in one jump and an RPD student solving the problem with the CS procedure

It was expected that the RPD students would use the N10C procedure to solve the problems 55 + 9, 75 - 9, but they also preferred to solve these problems by using the CS procedure. To solve the multi-digit subtraction and context problems with a small difference between the two numbers, the RPD students predominantly used the Connecting Arc procedure where the GPD students kept on using the N10 procedure instead.

In general we could say that half-way through the curriculum, the RPD students are more flexible in their procedure use than the GPD students, which was predicted from both points of view. This conclusion was confirmed by the analysis of the consistency of use of solution procedures across problems, which showed that more RPD students could be categorized as flexible or half-flexible than GPD students. However, in January this difference in the number of flexible students is mainly the result of the use of the CS procedure and the Connecting Arc. The N10C procedure was not used as much as had been expected. The fact that the GPD students did not use the Connecting Arc and N10C procedure can be explained by the content of the GPD where these procedures had not yet been introduced half-way through the curriculum (cf. Table 3.14). However, the lack of use of the CS procedure by the GPD students cannot be explained in this way: the RPD and GPD both pay equal amounts of attention to this procedure. One could argue that it might be preferable to use the CS procedure for solving single-digit problems, or to solve these problems in one step. The CS procedure is a semi-abbreviated sequential way of crossing tens, offered to raise the level of counting towards structuring in small steps (using tens or decades as turning points). From a didactic point of view, crossing tens in one step in January is much too early, especially for weaker students. This interpretation was confirmed by some of the tests we scored. We saw that some pupils draw one jump on the number line, but that dots were put under the jump which indicates that they may have counted to solve the problem. This phenomenon is illustrated in Figure 6.2.

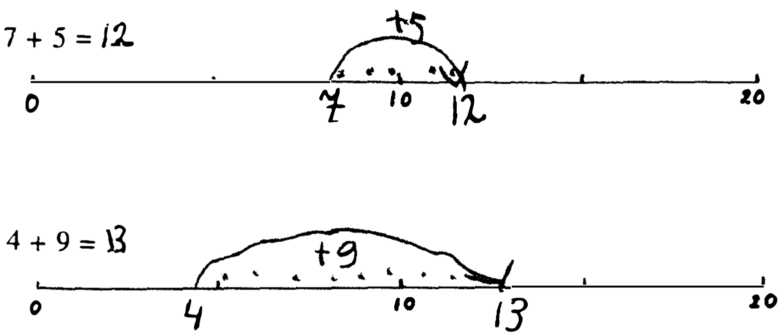


Figure 6.2 A GPD student solving the problem in one jump but counted numbers. To keep track he puts dots onder the jump

Probably the weaker GPD students did not use the number line in the appropriate mental way. Individual analyses and interviews suggest that students calculated the answer to the problem before they made their jumps on the number line (see Beishuizen, Treffers and Klein, in preparation). We will return to this difference in use of the procedures CS or one-step for crossing tens, when we discuss the results of the better and weaker students.

6.2 Development of procedural and strategic knowledge: RPD versus GPD students at the end of the curriculum

Conclusions about procedural knowledge

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
3. At the end of the program the RPD students will show a higher level of procedural competence in solving the most difficult sum type (subtraction problems which require regrouping) than the GPD students. For the other sum types there will be no differences in the number of correctly solved problems between the two groups of students.	4. At the end of the program the GPD students will still have a higher level of procedural competence than the RPD students with respect to standard number problems, especially on the speed test. The GPD students will also solve correctly as many context problems as the RPD students, because more attention is paid to these problems during the last months of the GPD.

The analyses of the most difficult problems on the Arithmetic Speed Test (multi-digit subtraction problems with regrouping with numbers < 100) showed that the RPD students solved more problems correctly than the GPD students. For four of the other thirteen types of multi- and single-digit problems significant differences were found in favor of the RPD students: They solved more problems correctly than the GPD students. No significant differences were found between the two groups of students for the other problem types. This was also not the case for the results on the National Arithmetic Test (CITO LVS E4) and for the Arithmetic Scratch Paper Test. Using the Arithmetic Subject Matter Test we found significant differences between the RPD and GPD students with respect to numerical subtraction problems that were solved with the use of the empty number line. The RPD students solved more problems correctly than the GPD students. This was caused by the two problems, 81 - 78 and 72 - 39. A majority of the RPD students solved these problems in a non-standard way by using the Connecting Arc procedure and the N10C procedure respectively. Most of the GPD students used the N10 procedure to solve these problems. This is illustrated in Figure 6.3. It seems that the use of the Connecting Arc and N10C procedures by RPD students resulted in a higher number of correct answers for these problems.

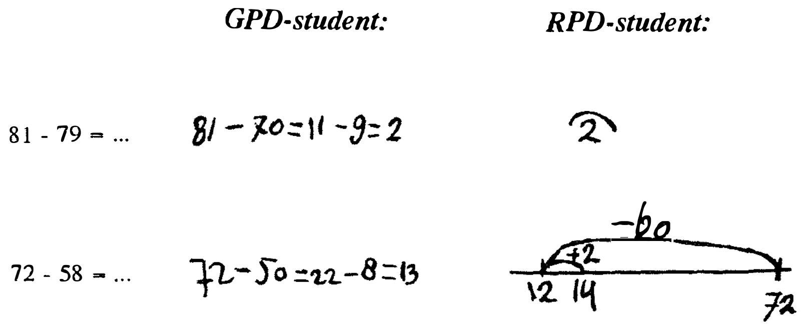


Figure 6.3 A GPD and RPD student solving two problems on the Arithmetic Subject Matter Test in June

With respect to the procedural competence shown by the students at the end of the curriculum we can conclude that Treffers' hypothesis was confirmed. The RPD students still performed better on the most difficult problems in the Arithmetic Speed Test. The results for the other problem types were at least comparable, and sometimes also better for the RPD students. This is a remarkable outcome, given that there were fewer written exercises in the RPD condition. In general Beishuizen's hypothesis was not confirmed. It was only on the Arithmetic Subject Matter Test and Arithmetic Scratch Paper Test that the test scores of RPD and GPD students did not differ significantly. However these tests are less sensitive to procedural speed than the Arithmetic Speed Test. This means that a small part of Beishuizen's hypothesis was confirmed namely that GPD students would catch up on RPD students on non-standard context problems. However, the general trend in these tests pointed in the direction of higher scores for the RPD students.

Conclusions about strategic use of computation procedures

TREFFERS (realistic point of view)	**BEISHUIZEN (gradual point of view)**
3. At the end of the second grade the GPD students lag behind the RPD students with respect to the flexible use of solution strategies and computation procedures, both for numerical and context problems.	4. At the end of the second grade we expect there will be no differences between the GPD and RPD students, with respect to the flexible use of different solution strategies and computation procedures.

In order to test these hypotheses we made a distinction between *spontaneous* flexibility and flexibility *on demand*. Spontaneous flexibility was not demonstrated by most of the GPD students. They preferred to use the N10 procedure. The RPD students spontaneously used various procedures like N10, N10C and the Connecting Arc. Some of the context problems had to be solved in two different ways (flexibility on demand). Here we noticed that the differences between RPD and GPD students in flexibility became smaller. Whilst more RPD students could be categorized as flexible students for these problems, the number of GPD students that was flexible for one of the two problem types had increased. It appeared that especially on the addition problems, the GPD students used the 1010 procedure instead of the N10 procedure. This increase in flexibility shown by the GPD students for the latter type of problems may have been caused by the fact that the former problems had to be solved in one way on the empty number line. Since it is almost impossible to show the 1010 procedure on the empty number line, the GPD students could not use this procedure for these problems. For the problems that had to be solved in two ways the students were free to choose how to write down their solution steps. They used the scratch paper box to show how they solved the problem. This way of writing down one's solution procedure could also be used for the problems of the Arithmetic Scratch Paper Test. It was noted that both the RPD and GPD students used the 1010 procedure to solve the numerical and context problems.

However, in general, for both addition and subtraction problems, the flexibility of the GPD students lagged far behind the flexibility in strategy use of the RPD students. The latter students adapted their computation procedures more often to the number characteristics of the problems (N10C, Connecting Arc, see also Figure 6.3). The discrepancy in flexibility between RPD and GPD students only became less pronounced for the context problems with a difference. Quite a number of A10 solutions were noted (see also Figure 6.4). This specific outcome highlights the didactic power of *open* context problems: even the GPD students solve the difference problems in a more varied way than they do with the other problems on the Arithmetic Scratch Paper Test (cf. Beishuizen, 1997). Finally it was noted that in June more RPD students (about 40%) chose to solve the ASPT problems on the empty number line, while only 6% of the GPD students did so.

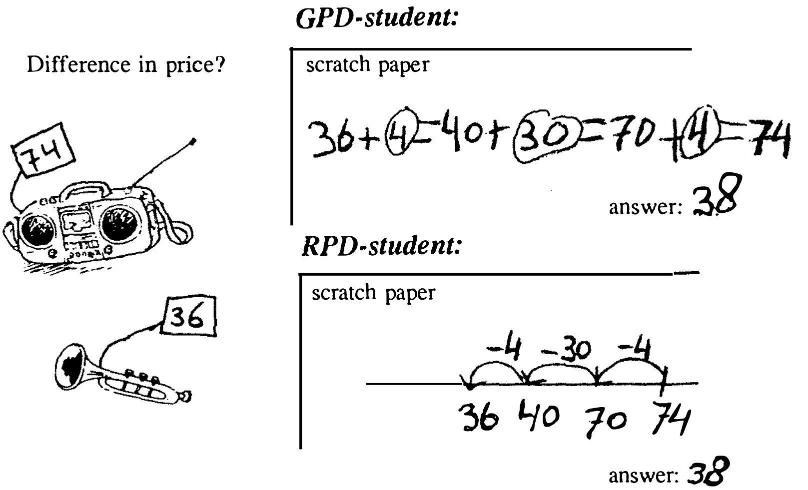


Figure 6.4 A GPD and RPD student solving a difference problem on the Arithmetic Scratch Paper Test in June

The results of the Arithmetic Subject Matter Test and the Arithmetic Scratch Paper Test lead to the conclusion that the RPD students were more flexible in adapting their procedures according to the number characteristics of the problem than the GPD students. It should be realized that the use of the Connecting Arc procedure was not introduced in the GPD. However this does not account for the other procedures, since these procedures were also introduced in the GPD during the second half of the school year. This conclusion about differences in flexibility between RPD and GPD students is confirmed by the analysis of the profiles of solution behavior: More RPD students could be categorized as flexible or half-flexible on both the Arithmetic Subject Matter and Arithmetic Scratch Paper Test. The MANOVAS on the use of the N10 procedure showed that the GPD students used this procedure more frequently than the RPD students. Treffers' hypothesis concerning strategic knowledge at the end of the curriculum is thus confirmed, and Beishuizen's hypothesis is rejected.

6.3 Development of procedural and strategic knowledge: weaker and better RPD and GPD students half-way through the curriculum

Conclusions about procedural competence

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
5. Half-way through the program there will be significant differences between weaker and better RPD students, however, this will be more on a strategic than on a procedural level. Weaker RPD students will often use solution strategies and computation procedures in a non-abbreviated or inefficient way, but these will bring them to the correct answers.	6. The expected higher level of procedural competence for the GPD students, half-way through the second grade, will mainly be caused by the relatively better scores of the weaker students. The differences between the better and weaker students will be significant larger in the RPD.

The results of the Arithmetic Speed Test in January showed that there was no significant overall effect for competence level x type of program. Separate analyses for each type of problem revealed significant differences for the better students on three of the six problem types. The better RPD students solved more answers correctly that the better GPD students. No significant differences were found between better RPD and GPD students for the number of correct answers given in the Arithmetic Subject Matter Test and the National Arithmetic Test (CITO LVS M4). No differences were found between the weaker RPD and GPD students either. We conclude that with respect to the procedural knowledge of the weaker students Beishuizen's hypothesis was falsified and Treffers' hypothesis was confirmed. There were no differences between weaker RPD and GPD students with respect to procedural competence half-way through the second grade. Support for this conclusion was found in the analysis of the types of errors made by the better and weaker students. Here, too, no differences were found in distribution of procedural and non-procedural errors made by weaker RPD and GPD students.

Conclusions about strategic use of computation procedures

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
5. Weaker RPD students will often use solution strategies and computation procedures in a non-abbreviated or inefficient way, but these will bring them to the correct answers. The weaker GPD students will use the N10 procedure in a proceduralized way without understanding what they are doing.	6. Weaker RPD students will be more flexible in using different strategies and procedures, but, compared to the weaker GPD students, this use will be of a lower quality in both strategic as procedural sense half-way through the program.

Half-way through the curriculum the better and weaker RPD and GPD students did not differ much with respect to the use of the N10 procedure. This is probably due to the types of problems that had to be solved. We therefore examined the use of the Complementary Structuring (CS) procedure, because we thought that this procedure would be the most adequate to solve addition and subtraction problems like 36 + 5 and 41 - 4 (cf. Table 3.5 and 3.6). It appeared that both the better and weaker RPD students used the CS procedure more frequently than the better and weaker GPD students to solve single-digit addition and subtraction problems. Most of the GPD students solved these problems in one step (see Figure 6.1), which solution behavior we have to mistrust as not using the most adequate procedure on the number line. For the multi-digit addition problems the N10 procedure is the most frequently used procedure by both groups of RPD and GPD students in January. For multi-digit subtraction problems both better and weaker RPD students used the Connecting Arc procedure to solve subtraction problems with a small difference between the two numbers of the problem. For the other multi-digit problems they preferred the N10 procedure. Most of the weaker and better GPD students used the N10 procedure for all multi-digit problems. With respect to flexibility in using different solution procedures, it appeared that none of the GPD students could be classified as flexible or half-flexible, whilst 25% or more of the weaker and better RPD students could be categorized that way.

In general we can conclude that both the better and weaker RPD students were more flexible in using different solution procedures than the better and weaker GPD students. We also saw that the RPD students used the CS procedure more frequently, where the GPD students solved these problems in one-step. With respect to the hypotheses formulated about the strategic knowledge of using solution procedures we are able to conclude that Treffers' hypothesis was confirmed and that Beishuizen's hypothesis was falsified: The weaker RPD students were not confused by the introduction of different solution procedures and did not invent inadequate combinations of different solution procedures. Unfortunately, the mistaken use of the N10C procedure (N10 with compensation), that had been predicted, could not be tested yet, since the RPD students hardly used this procedure. We could conclude that the hypothesis from the Realistic point of view was confirmed.

6.4 Development of procedural and strategic knowledge: weaker and better RPD and GPD students at the end of the curriculum

Conclusions about procedural competence

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
7. At the end of the program the quality of strategy and procedure use will have increased for the weaker RPD students. Therefore the differences in procedural competence between weaker and better RPD students will have become smaller.	8. At the end of the program the results for the weaker and better students will be the same as predicted from this point of view half-way through the second grade. Weaker GPD students will have relatively better scores and the differences between the better and weaker students will be significant larger in the RPD.

Looking at the scores on single-digit problems on the Arithmetic Speed Test, we found a significant effect for type of program for better and weaker RPD and GPD students. It appeared that the better RPD students outperformed the better GPD students on three out of the six problem types. For the weaker students this was even the case for four out of the six problem types For multi-digit problems, again a significant effect for type of program was found. Here we did not find significant differences between better RPD and GPD students. For the weaker students we found that for four out of eight different problem types the weaker RPD students solved more multi-digit problems correctly than the weaker GPD students. This even accounted for the most difficult problem type: subtraction with regrouping with numbers less than 100.

No important differences were found between better and weaker RPD and GPD students in respect of the number of correct answers in the Arithmetic Subject Matter Test and the Arithmetic Scratch Paper Test. The analysis of the type of errors made on both tests did also not reveal significant differences between the two groups of students. In the National Arithmetic Test (CITO LVS E4) too, no significant differences were found between the two groups of students with respect to addition and subtraction up to 100.

In general we can conclude that the procedural knowledge of both the better and weaker RPD students on the speed test was better than that of the better and weaker GPD students. On the other tests the results from both groups of students were comparable. This means that Beishuizen's hypothesis concerning procedural knowledge at the end of the curriculum, should be rejected, while Treffers' hypothesis is confirmed. The results for the weaker RPD students were even better than was predicted: They outperformed the weaker GPD students several times.

Conclusions about strategic use of computation procedures

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
7. At the end of the program the quality of strategy and procedure use will have increased for the weaker RPD students. Compared to half-way through the curriculum the situation will be the same for the weaker GPD students. They will not be amenable to the adoption of new solution strategies or computation procedures, because they will stick to the use of the N10 procedure.	8. At the end of the program the results for the weaker and better students will be the same as predicted from this point of view half-way through the second grade. Weaker RPD students will still be more flexible in using different strategies and procedures, but, compared to the weaker GPD students, this use will be of a lower quality in both a strategic and a procedural sense.

The better and weaker GPD students solved the numerical and context addition problems in the Arithmetic Subject Matter Test in June predominantly by using the N10 procedure. The weaker and better RPD students also used the N10C procedure to solve problems like 55 + 19. For the addition problems that had to be solved in two ways we see a slight difference in procedure use by the better GPD students. The better GPD students seem to be more flexible in using a different procedure, the second time they were asked to solve a problem, than the weaker GPD students. The weaker GPD students continued to use the N10 procedure, even when they were asked to use a different procedure, the second time they had to solve the same problem. This pattern also appears for the subtraction problems: The weaker and better RPD students and the better GPD students are the most flexible in adapting their use of computation procedures to the number characteristics of the problem. Most of the weaker GPD students stuck to using the N10 procedure to solve the problems on the Arithmetic Subject Matter Test in June. For the procedures used on the Arithmetic Scratch Paper Test we also see that for most of the problems the weaker and better RPD students and the better GPD students changed their solution procedure according to the number characteristics of the problem. The weaker students kept using the N10 procedure, also for context problems with differences. These conclusions are supported by the MANOVAS on the frequency of using the N10 procedure and the analyses of the profiles of the solution behavior of the weaker and better students. In general the weaker and better GPD students used the N10 procedure more frequently than the weaker and better RPD students (see Figure 6.5) Furthermore the number of flexible and half-flexible students is the smallest among the weaker GPD students: For the Arithmetic Scratch Paper Test, half of the weaker GPD students used the N10 procedure on a majority of the 21 problems (see also Figure 6.5).

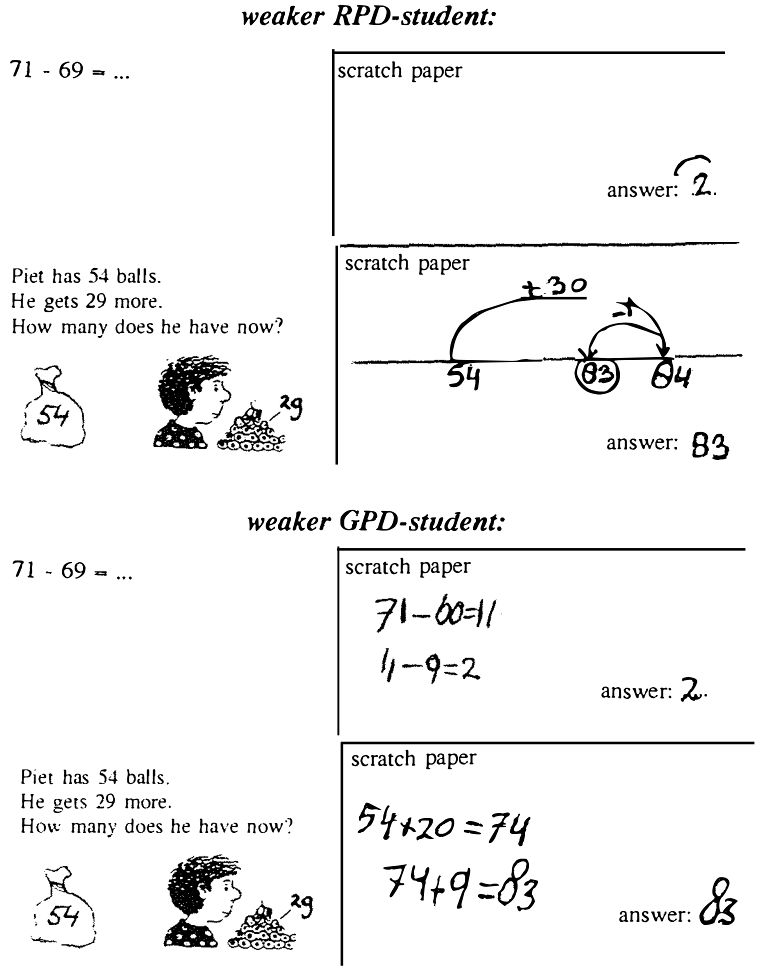


Figure 6.5 A weaker RPD and GPD student solving problems on the Arithmetic Scratch Paper Test in June

We conclude that Beishuizen's hypothesis was partly confirmed. Compared to the weaker GPD students, the weaker RPD students were more flexible in using different solution procedures according to the number characteristics of the problem. However, Beishuizen did not predict that the weaker students would do so well on these problems, sometimes even better than the weaker GPD students. The expected confusion in the weaker RPD students, caused by the introduction of different solution procedures like the Connecting Arc, N10C, A10, and 1010, did not show up in the test results at the end of the curriculum. We only saw temporary compensation mistakes in the students' worksheets (Beishuizen et al., 1996) after the introduction of N10C (see Figure 6.6).

Difference problem "Leiden on Sea" in worksheet:

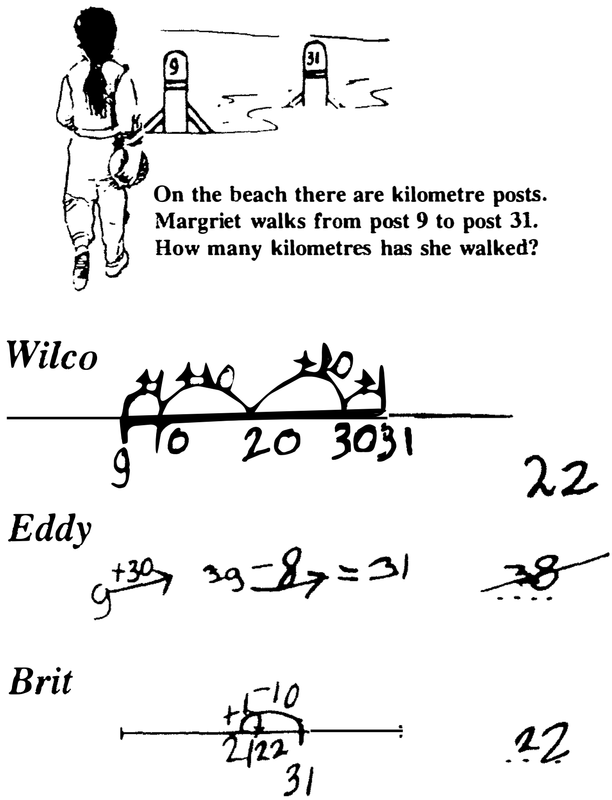


Figure 6.6 Temporary confusion in carrying out the N10C procedure noticed in the worksheets of RPD students

However, by the end of the school year many of these mistakes disappeared, probably with support of the empty number line: There were no differences between weaker RPD and GPD students in the number of procedural errors (although the RPD students used the N10C procedure more frequently). This is all in accordance with the hypothesis formulated by Treffers, which could therefore be confirmed.

6.5 Transfer to addition and subtraction with numbers up to 1000

KLEIN (post hoc questions)
14. Do we see any transfer from what the students learned for addition and subtraction in the number domain of 0-100 to the number domain 0-1000?

Both at the end of the curriculum and at the beginning of the third grade, when the students worked with their original mathematics textbooks again, we asked the students to solve addition and subtraction problems with numbers up to 1000. During the last month of both the Realistic and Gradual Program Design, some attention had been paid to these kinds of problems, so that students would not be confronted for the first time with these problems during the test.

The overall results for the transfer test in June did not show a significant difference between RPD and GPD students. Only the number of correctly solved subtraction problems was higher for the RPD students than for the GPD students. This is probably caused by the solution procedures used by the RPD and GPD students for these problems. Most of the RPD students used the Connecting Arc procedure to solve the subtraction problems 301 - 298 and 404 - 395, where most of the GPD students used the N100 procedure to solve these problems. The N100 procedure takes more steps and is therefore easily prone to mistakes (see Figure 6.7).

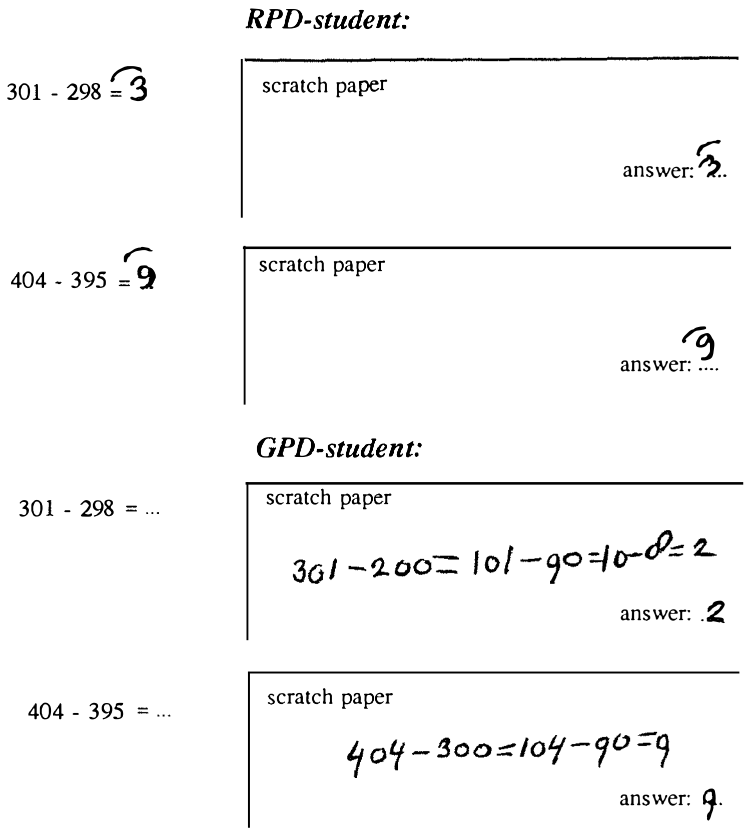


Figure 6.7 A GPD and RPD student solving two problems on the Arithmetic Scratch Paper Test in November

It is remarkable that most of the problems with numbers up to 1000 were already being solved correctly (about 70%) in the second grade. Normally these problems are not introduced until the third grade. We can therefore conclude that especially for the numerical problems, there is a transfer effect to addition and subtraction problems with numbers up to 1000. What is especially interesting is the transfer of the use of the Connecting Arc to problems with larger numbers, where the RPD students do not fall back to less efficient procedures like N100.

In November, at the beginning of the third grade, the Arithmetic Transfer Test was administered again. At this time no significant differences were found between RPD and GPD students. This is probably due to the fact that both groups of students were more familiar by that time with problems with larger numbers. The decline in use of the Connecting Arc procedure by the RPD students may also have contributed to this result. With respect to the context problems we see that the percentage of students that solved the problem correctly had increased from 38% to 50%. However, the *kino* problem (Sundermann and Selter, 1995), in which the students had to calculate how many chairs were not occupied, remained difficult to solve.
We can conclude that there are no differential transfer effects for the two program designs regarding addition and subtraction problems with numbers up to 1000. In the next section we will discuss if this is also true for addition and subtraction problems with numbers less than 100 which were administered at the beginning of the third grade.

6.6 Retention of flexibility in using different solution procedures at the beginning of the third grade

KLEIN (post hoc questions)
15. What is the retention of the strategies and procedures RPD and GPD students have learned for addition and subtraction up to 100, some months after they stopped working with the experimental program?

The Arithmetic Scratch Paper Test, which was administered at the end of the second grade, was also administered at the beginning of the third grade. The test was repeated in order to investigate if the RPD and GPD students were still flexible in using different solution procedures, even when they had stopped working with either one of the program designs.

First we looked at the number of correctly solved problems at the Arithmetic Scratch Paper Test in November. It appeared that the RPD students solved more problems correctly than the GPD students. Especially for the context addition and numerical subtraction problems, the RPD students performed better than the GPD students. With respect to the procedures that were used in the beginning of the third grade, we only noticed some slight differences in the procedures used to solve the problems on the Arithmetic Scratch Paper Test compared to the end of the second grade. The RPD students were still more flexible in adapting their solution procedures to the number characteristics of the problems. This means that these students' flexibility did not vanish after the program had stopped. The solution procedures seem to have been incorporated by the RPD students and to a lesser extent by the GPD students. In the latter case we see more adoption of the 1010 procedure for addition problems. This is illustrated by the refusal of many students to use column-wise arithmetic procedures (which are taught during the third grade) to solve subtraction problems like 301 - 298 or 312 - 189. RPD students continued to prefer to use the mental Connecting Arc procedure to solve the first problem, and both RPD and GPD students were in favor of using the N10C or N10 procedure to solve the second problem. To solve the second problem most of the students preferred to use the empty number line to write down their solution steps, instead of written column-wise arithmetic procedures.

6.7 Motivational processes: RPD versus GPD students

Conclusions about motivational aspects at the domain-specific level

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
Predictions about motivational beliefs	
9. In the RPD the use of self-invented and informal strategies is encouraged. This will make the RPD students feel involved and give them pleasure in arithmetic which will result in a higher motivation towards mathematics than the GPD students.	10. Half-way through the program, there will be no differences between RPD and GPD students with respect to motivation for mathematics in general.
11. At the end of the program the RPD students will have a higher motivation towards arithmetic, but the differences between the RPD and GPD students will be smaller than half-way through the program.	12. At the end of the program there will be no significant differences between the two programs with respect to motivation for mathematics in general.

The data at the domain-specific level half-way through the curriculum revealed that the GPD students showed somewhat more positive affects towards mathematics as a school subject than the RPD students. It also appeared that the GPD students had a higher self-concept of their mathematical ability than RPD students. For value and intended effort towards doing mathematics no significant differences were found between the two groups of students. Only speculations can be offered here. Maybe the RPD students were more challenged during the first half of their curriculum by the introduction of different strategies and the rapid increase of the size of the numbers after the first eight weeks of the program. The introduction of addition and subtraction problems, which require crossing ten and regrouping, also happened faster in the RPD than in the GPD. Although this did not influence their results on the cognitive tests in a negative way, it may have resulted in a lower self-concept of mathematics ability for the RPD students half-way through the curriculum. For some reason this may also have influenced their affects towards mathematics as a school subject.

Concerning the domain-specific hypotheses half-way through the curriculum we can conclude that Treffers' hypothesis was not confirmed. Beishuizen's hypothesis was also not confirmed, but from his point of view the results were even better than expected: The GPD students scored higher than the RPD students on two out of three motivational aspects.

The results on the Mathematics Motivation Questionnaire in June showed that the RPD students scored somewhat higher on motivation towards mathematics. The RPD students were more positive than the GPD students. This means that at the end of the two programs, opposite results were found, compared to the results half-way through the curriculum. We must therefore reject Beishuizen's hypothesis concerning motivational aspects at the domain-specific level. For Treffers the results were even better than expected: The differences between the RPD and GPD students with respect to motivation towards mathematics have not become less, they have reversed in favor of the RPD students.

Conclusions about motivational aspects at the task specific level

TREFFERS (realistic point of view)	BEISHUIZEN (gradual point of view)
Predictions about motivational beliefs	
9. RPD students will have more favorable cognitions and affects towards both number and context problems than GPD students, half-way through the program.	10. Half-way through the program, GPD students will have more favorable cognitions and affects towards numerical problems than RPD students. The opposite will be true with respect to context problems: RPD students will have more favorable cognitions and affects towards these problems than GPD students.
11. At the end of the program the RPD students will still have more favorable cognitions and affects towards numerical and context problems compared to the GPD students. The differences between the two groups of students will be smaller than half-way through the program.	12. At the end of the program there will be no significant differences between the two programs with respect to favorable cognitions and affects towards numerical problems and context problems.

At the task-specific level we measured cognitions and affects before and after solving context and numerical problems. It appeared that, overall, there were no differences between RPD and GPD students in their cognitions and affects at the task-specific level. However, both groups felt more self-confident about solving numerical problems than context problems. The value and intended effort also differed somewhat for the two problem types: Both RPD and GPD students reported attaching more value and intended to investing more effort into context than numerical problems. The interaction "problem type x type of program" appeared to be significant for two motivational aspects before, and one motivational aspect after solving context or numerical problems. Analyses showed that RPD students were more self-confident about solving numerical problems than GPD students. RPD students were also more attracted to numerical problems and felt less threatened after solving numerical problems than GPD students. A possible explanation for these unexpected results can be found in the percentage of correctly solved context and numerical problems half-way through the curriculum. The percentage of correctly solved numerical problems was higher than the percentage of correctly solved context problems for both groups of students. However, it also appeared that the GPD students solved more context problems correctly than the RPD students. This may be an explanation for the fact that the RPD students felt less threatened after solving numerical problems (although this effect was not found for context problems for GPD students). Another explanation for the differences between RPD and GPD students in motivational aspects might be that, in general, numerical problems are found to be easier than context problems. Since much attention is paid to context problems in the RPD, these students may have felt less secure about solving these problems than the GPD students, who may have been more open-minded about their success on these problems. Being less confident about solving context problems may have overruled the positive feelings towards numerical problems, when the RPD students had to indicate their self-concept of mathematical ability. The answers on the OMQ(7-10) at the end of the curriculum sheds light onto the stability of the feelings towards context and numerical problems. The answers given by the better and weaker students also give us more information about the background to these differences between RPD and GPD students. In relation to the first measurement point (January) we are bound to conclude that the Treffers' hypothesis was not confirmed, neither at the domain-specific level nor at the task-specific level. At the task-specific level the opposite to what was expected was even found: RPD students expressed more positive affects after doing numerical problems than context problems. Beishuizen's hypothesis for the task-specific level was confirmed: The GPD students liked numerical problems more than context problems.

For cognitions and affects towards numerical and context problems measured at the end of the curriculum, we found significant differences between RPD and GPD students with respect to self-confidence and, to a lesser extent, for task attractiveness towards both context and numerical problems. It appeared that the RPD students scored higher on these motivational aspects than the GPD students. There also appeared to be differences in feelings of both RPD and GPD students after completing the tasks. After completing the context problems, the students expressed more positive affects than after completing the numerical problems. On the other hand, the students felt less threatened after completing the numerical problems than after completing the context problems. No interaction effects occurred. Recall, that the percentages of correctly solved numerical problems was higher than the percentage of correctly solved context problems.

We can conclude that, beside the difficulty of solving context problems correctly, how students feel about solving context or numerical problems also makes a difference. With respect to the earlier formulated hypotheses our conclusions lead us to reject Beishuizen's hypothesis: we did find differences between RPD and GPD students in motivational aspects at the task-specific level. Therefore, Treffers' hypothesis was confirmed: RPD students were more self-confident and liked both numerical and context problems better than GPD students.

6.8 Motivational processes for the better and weaker RPD and GPD students

Conclusions about motivational aspects at the domain-specific level

KLEIN (post hoc questions)
13. Do weaker and better RPD and GPD students differ in their cognitions and appraisals towards arithmetic as a school subject and more specific towards numerical and context problems?

Half-way through the curriculum there appeared to be a significant difference between better and weaker students regarding affect, self-concept and effort towards mathematics in general. Better students scored somewhat higher on these three aspects than weaker students. These differences were the largest for the perceived self-concept of mathematics ability. There were no differences between RPD and GPD students at the domain-specific level half-way through the second grade. At the end of the second grade, the only differences found were with respect to self-concept of mathematical ability. Both better and weaker RPD students appeared to have a slightly higher self-concept of their mathematical ability than the better and weaker GPD students.

Conclusions about motivational aspects at the task-specific level

At the task-specific level there were no significant differences between RPD and GPD students half-way through the curriculum. Likewise no differences were found between better and weaker students and between numerical versus context problems. The only difference was found for self-confidence about solving numerical problems: Weaker RPD students appeared to be somewhat more self-confident about solving these problems than weaker GPD students. This is in accordance with the scores for the whole group of RPD and GPD students.

We also found that the percentage of correctly solved numerical problems was higher than the percentage of correctly solved context problems and that better students solved more context and numerical problems correctly than weaker students.

At the end of the curriculum, differences were found between RPD and GPD students for self-confidence, task attractiveness and task value/learning intention. It appeared that both the better and weaker RPD students scored higher on these three aspects than the better and weaker GPD students. More detailed analyses showed that the better RPD students scored higher than the better GPD students on task attractiveness of numerical problems, task value and intended effort to solve numerical problems, and positive feelings after solving numerical problems. Compared to the weaker GPD students, the weaker RPD students scored higher on self-confidence about solving numerical and context problems, and intended effort to solve context problems. We found the same percentage of correctly solved numerical and context problems at the end of the program as half-way through the curriculum: Better students solved more problems correctly and a higher percentage of numerical problems was correctly solved than context problems.

The findings for the weaker students at the domain- and task-specific level are in the same direction: Weaker RPD students felt more self-confident about mathematics than weaker GPD students. These findings support the idea that the didactic sequence of the RPD did not harm the weaker RPD students, and that a structured approach towards mathematics, as is done in the GPD, does not necessarily result in a higher self-concept of mathematical ability for the students, even not for the weaker students. We also observed that the numerical problems are regarded as less threatening and more attractive than context problems. This is true for both better and weaker RPD students. The RPD did not succeed in removing these affects, which already existed half-way through the curriculum.

6.9 General discussion: Flexibilization of mental arithmetic strategies on a different knowledge base

In this study two program designs in the domain of addition and subtraction up to 100 were developed and implemented in the second grade of primary education. The Realistic Program Design (RPD) was based on principles of Realistic Mathematics Education (RME). The Gradual Program Design (GPD) had also ideas drawn from RME but follows a psychological conceptualization of stage-wise development. In the previous sections the effects of both program designs on both cognitive and affective variables have been discussed regarding the hypotheses formulated by Treffers, Beishuizen and Klein. In this section we will discuss these findings in a broader sense and reflect on the theoretical framework that was outlined in the first chapter. Reference will also be made to some of the basic features of the two program designs that were mentioned in chapter 2 namely the role of the number line and the role of the teacher.

Mathematics as an activity

One of the reasons for introducing the empty number line as a new didactic model was the expressed need to improve the teaching of basic skills up to 100 in the Dutch primary mathematics curriculum (Treffers & De Moor, 1990). Compared to the outcomes of the national evaluation test (Wijnstra, 1988) the results of both experimental programs, RPD and GPD, demonstrate a higher degree of mastery (about 70% - 80% correct) of the most difficult problem types (subtraction problems which require regrouping) in the program-related tests as well as in the external criterion test (CITO-LVS). Apart from this procedural result, the RPD students used the different mental arithmetic strategies in a flexible way, which was the first objective of this instructional design experiment (Boekaerts & Beishuizen, 1991). In this respect the performance of the weaker RPD students was most convincing: outperforming the weaker GPD students both on procedural speed and flexibility of mental computation. The expected confusion did not turn up - only momentarily in worksheets - in the Realistic condition with multiple strategies, as was predicted in the Beishuizen hypothesis based on psychological theory (Glaser & Bassok, 1989), which arguments are also given by special education experts (Ruijssenaars, 1994; Van Luit & Van der Rijt, 1997).

However, the Gradual point of view that students would develop flexibility later during the last three months of the GPD program, was not confirmed. Some GPD students did (especially the better students), but many students did not and continued with N10 as the one and only solution procedure. The hypothesis put forward by Treffers that flexibility would develop during the curriculum was confirmed. However, flexibility did not occur at the cost of procedural speed. Indeed, despite less emphasis being given to written number exercises in the RPD program, the students (also the weaker ones) caught up speed in combination with variety in strategy use. Our impression is that the early introduction of context problems and different strategies in RPD invited *flexibility*. More specifically the earlier transition from the structured to the empty number line in the RPD program (cf. Table 3.14) challenged the students' *mental activity* to a considerable extent, increasing both their procedural and flexible solution behavior. In addition, the rapid increase of number size and problem difficulty (Table 3.14) made things pretty hard for the RPD students in the beginning, but this may have stimulated a better internalization of the empty number line model than in the GPD condition. Looked at altogether, we conclude from teachers' observations (Bergmans & Leliveld, 1997) as well as from test results that the central Realistic Mathematics principle of *mathematics as an activity* became true to a greater extent in the RPD than in the GPD condition. Passive reading-off behavior, as is often observed when pupils work with arithmetic blocks (Beishuizen, 1993), is hardly possible on the empty number line where mental imagination is elicited before drawing the jumps. We will now turn to a discussion of the modeling function of the empty number line.

The modeling function of the empty number line

The semi-structured number line was introduced early in the experimental programs as a visual representation or *model of* a string of 20 beads (RPD, Table 3.14) or manipulatives in a number track (GPD, Table 3.14). During this first period pupils frequently fell back on the concrete level, but when the longer number line was introduced later, pupils hardly needed the concrete support of the 100-bead-string or 50-number-track any more. The number line up to 20 had become already an internalized mental model which could be easily extended to a longer *model for* the number row up to 100. This observation was an important reason explaining why we could proceed rather rapidly to the larger number operations and more difficult problem types, especially in RPD (Table 3.14). Compared to the original textbook *Rekenen & Wiskunde*, where both the larger number operations and the new model (100-square) are introduced at the same moment, the curriculum sequence used in our experimental number line programs facilitated learning progress much better (without the obstacle of a new 100-square model). Teachers mentioned this argument explicitly (Bergmans & Leliveld, 1997).

In the RPD-program the early introduction of context problems also stimulated the mental *representation* function of the number line model. More time than in the GPD program was spent on whole-class practice in playing games (for instance "Guess my number") to improve *number sense*. This contributed to the use of the empty number line as an appropriate external/internal mental model in particular by the RPD pupils. The outcome that at the end of the program about 40% of the RPD pupils still preferred to draw jumps on an empty line (Table 4.17) above using arrows or just computation steps - compared to only 7% of the GPD pupils - is interpreted as a positive signal of integrated use rather than as a signal of relapse. Also the difference between drawing *one step* by the GPD pupils and split-up CS jumps by the RPD pupils, for problems like 28 + 7 and 41 - 4 in January (Figures 4.2 through 4.5), is interpreted as a better use of the number line's *modeling function* in the latter case. From observations and interviews we draw the conclusion that most one step jottings on the number line are a symptom of discrete (mental) operations without integration: calculating the sum (mentally) before or after drawing the jump and not simultaneously.

On a more general level we come to the conclusion, that the RPD pupils developed a better sense of *cognitive economy* or a functional distribution and cooperation between empty number line support and pupils' own working memory effort. Baroody & Ginsburg (1986) described earlier how such basic psychological principles of avoiding cognitive load may play a role as a driving mechanism behind the choice of computation strategies and procedures. In this respect the weaker RPD students provide interesting evidence, because they used N10C as much - sometimes even more - as the better RPD students. This outcome is contrary to the expectation that such a compensation strategy (45 + 19 via 45 + 20 - 1) would confuse weaker students because of the change in direction of operation.

Even more unexpected was the explanation given by some weaker students in interviews: they argued that by using N10C they could avoid crossing tens. What they meant was that they considered splitting-up units in complements-to-ten (i.e. + 9 via + 5, + 4) to be a greater cognitive load than applying the compensation-rule (i.e. + 9 via + 10, - 1). The students invented a popular name for this N10C strategy. They called it "*SPV*" which means "Spring Verder" (Dutch for: Jump Further). In summary, our interpretation is that due to their functional use of the empty number line, the RPD pupils developed a stronger feeling for cognitive economy and a greater sensitivity for possible short-cut strategies or flexibility in the use of strategies and procedures than the GPD pupils. Lorenz (1997) called this 'just strolling around in an imaginary number space'.

However, the modeling function of the empty number line was not supportive in all cases. The Connecting Arc procedure (i.e. 81 - 79 = ?) was deliberately introduced and practiced on a pure mental level, because in this case the number line model could, in fact, introduce an impediment to seeing immediately that two numbers are close neighbours (Treffers, personal communication). Moreover, we do take notice of the fact that at the end of the program a great number of the RPD students (55%) and almost all GPD students (85%) spontaneously preferred *not* to use the empty number line any more (Table 4.17). Many of them were fed up with the laborious drawings of jumps on the number line and told us they needed this support not any longer, because they could perform all operations mentally. Such statements are considered proof for what Treffers called *progressive mathematization* (Treffers, 1987). We doubt, however, whether this is true for all pupils. We believe that not all of the GPD students have come to use the empty number as a mental model. We think the difference between RPD and GPD students illustrates what probably is happening to many pupils learning to use the empty number line in new editions of realistic textbooks like *Wereld in Getallen* (Van de Molengraaf et al., 1991) and *Pluspunt* (Van Beusekom et al., 1991). Because its introduction is rather sudden and not enough time is spent on the use of the empty number line, we are afraid it fails to become an appropriate mental model a child can rely on when the child encounters a problem which cannot be solved immediately. The GPD pupils' outcomes in our experiment illustrate such limited effects.

Effects for Weaker and Better pupils; differentiation

One of the surprising outcomes of the study were the good results of the weaker RPD students. In January they already showed the same procedural speed and performance as the weaker GPD students, although from both theoretical view points they were expected to lag behind using *non-abbreviated or inefficient procedures* (Treffers' hypothesis) with higher flexibility at the cost of confusion and lower quality of answers (Beishuizen's hypothesis). Temporary confusion, indeed, was visible in weaker RPD students' worksheets, but apparently the empty number line helped them to overcome these learning difficulties. In June they clearly outperformed their weaker GPD fellows both in fluency on many of the two-digit speed tests and in flexibility of strategy use. In fact there was not much difference in the profile of flexibility between the better and weaker RPD students, while the weaker GPD students showed the expected rigidity of use of mainly N10 (Tables 4.37 - 4.39).

Our interpretation is that the weaker RPD students profited, in particular, from the general factors already discussed: *mathematics as an activity* and *the modeling function* of the empty number line. For instance teachers in the RPD classrooms observed a greater pleasure in doing arithmetic problems. Weaker students also started to enjoy presenting each other with *difficult* problems in the classroom like 5 + 72 (Bergmans & Leliveld, 1997). Apart from being a natural and transparent model the empty number line served also a metacognitive function for the weaker students: recording their jumps helped them keep track of the solution steps (which normally is a problem because of a more limited capacity of their working memory; Ruijssenaars, 1994). Further research is needed focusing on weaker students and the empty number line. A more extended analysis of the data we gathered on weaker students will be reported in the future. Beishuizen (1997) analyzed the 'difference problems' in the Arithmetic Scratch Paper Test for better and weaker RPD students. He reported that the empty number line serves also as a natural aid for *differentiation* (Treffers, 1995). At the end of the program in June 60% of the weaker students preferred to draw an empty number line as support for solving such difficult problem types, while this was 40% for the better students. The better use and the greater impact of the empty number line in the RPD condition contributes to the explanation that the weaker RPD students profited more from this model compared to the weaker GPD students.

Motivational aspects

Beside these cognitive explanations also motivational aspects can help explain the differences in strategy use between RPD and GPD students (cf. Boekaerts, in preparation). It appeared that at the end of the curriculum the RPD students liked mathematics more than GPD students. RPD students also had a higher self-concept of their mathematical ability than the GPD students. Boekaerts (1997) found that students who find a mathematics task interesting are willing to expend effort to accomplish this task. In line with this Pintrich & De Groot (1990) found that management of effort plays a crucial role in strategy use. At the task-specific level, we found that half-way through the curriculum all students had a greater learning intention towards context problems and considered context problems as more valuable than numerical problems. We also found that especially the weaker RPD students scored higher on task value and learning intention to solve context problems than weaker GPD students. The weaker RPD students also scored higher on self-confidence to solve numerical and context problems. We think that the use of the empty number line helped the RPD students, and especially the weaker RPD students, to overcome their uncertainty about solving problems, especially context problems. Maybe the empty number line did not have the same effect for GPD students mainly because the empty number line did not become a model on which they had learned to rely. This may have influenced the flexibility in using solution procedures of the weaker GPD students: They would not use another procedure other than their *save* N10 procedure.

Until now we have looked for explanations at the student's level. However, students are members of a small community called the class in which the teacher plays an important role. The teacher also played an important role in our experiment. The different roles the RPD and GPD teachers played, especially during the first half of the school year, may shed some light on the differences we found between the two programs.

Role of the teacher

In chapter 2 we already mentioned the important role of the teacher, which can be seen as an Achilles' heel of RME (cf. Gravemeijer et al., 1993; Hiebert et al., 1996; Selter & Spiegel, 1997). We were afraid that this would also be a vulnerable point in our research project, especially for the RPD because we wanted to create a classroom climate of interactive teaching and discussion about children's solutions right from the beginning of the school year. This is not an easy job for the teacher, since he has to guide the re-invention of different strategies by the students (Freudenthal, 1991; Streefland, 1991a, 1991b). We also tried to establish this atmosphere in the GPD classes but only during the last period of the program. In the first part of the program the teacher had the initiative and showed which procedure should be used to solve the problems. We think that the RPD approach was more successful in creating a climate of *interactive teaching*, than the GPD approach. However, we were not so certain about the success of this kind of teaching, half-way through the curriculum. The motivation data collected in January indicated that the first half of the RPD had not been easy for the students: They were less confident about their mathematical ability than the GPD students and felt more secure about solving numerical problems than the ambiguous context problems. In this period the RPD teachers also expressed their concern about how rapidly the size of the numbers increased, and how rapidly they had to deal with two-digit subtraction problems which required regrouping (in the original textbook *Rekenen & Wiskunde*, these problems are not introduced until the end of the second grade). Teachers were afraid that particularly the weaker students would get confused. These concerns were not so frequently voiced from the GPD teachers. Fortunately we could tell the RPD teachers that their students did very well on the test in January. Based on previous experience we were able to reassure them that things would turn out right by the end of the school year. We think this helped them to continue their way of teaching in which the students take over the initiative in the dialogues and were given the opportunity to visualize their strategies and communicate them to other students (cf. Boekaerts, in preparation; Brown & Palincsar, 1989; Freudenthal, 1991). The introduction of *verbal labels* for strategies and procedures (see Figure 2.11) proved to be very useful in this climate of interactive teaching.

In the Gradual Program Design, the teachers had to establish such a climate in just a few months. Indeed during the first half year of the GPD emphasis was placed on building a firm knowledge base through solving many problems. The shift from using one solution procedure to more attention for informal and different solution procedures that was made during the second half of the year was successful for the stronger GPD students. It came too late for the weaker GPD students to give up their save and successful N10 procedure.

Some remarks and implications for future research

Although our results and conclusions are quite positive, especially concerning the effects of the RPD on flexibility in using different solution procedures (also for weaker RPD students) they are still tentative. It is important to realize that the number of schools was limited and not a representative sample of the total school population in the Netherlands. For instance, the number of immigrant students was limited and the schools were sufficiently motivated to change their mathematics education for at least one year. A second limitation is that our results can not be generalized to students with special needs. Although our weaker students scored below the national mean at the start of the experiment, their results improved considerably in the course of the experiment. Most of the weaker RPD students were quite capable at solving addition and subtraction problems correctly, using different solution procedures in a flexible way. Future research is needed to investigate to what extent these results can also be obtained for children in special education (cf. Ruijssenaars, 1994; Van Luit & Van der Rijt, 1997).

A third remark concerns the way the RPD and GPD were implemented. Gravemeijer (1994) distinguished two paths for implementation of realistic mathematics education: (1) by directly influencing the teachers' views, knowledge, insight and skills or (2) by using direct-teaching to implement realistic mathematics education in which the textbooks are adapted accordingly. Gravemeijer prefers the first path "although it is much more difficult to put into practice and probably not feasible in the short term" (Gravemeijer, 1994, p. 175). In this study we choose the second path: we rewrote the pupils' *textbooks* and described in detail and extensively how the teachers should act during their lessons. We tried to control this by visiting the school every fortnight, talking over the classroom experiences the teachers had and evaluating the test results of the students. Our experience was that during this study the RPD teachers, in particular, were successful in achieving one of the goals of realistic mathematics education, namely making mathematics become an activity in the sense described by Freudenthal (1991). Our results are in line with the recommendation made by Gravemeijer (1994).

Finally the way we implemented the RPD and GPD is probably not always feasible in everyday school practice. However, we think that an innovation in education *does not end* by providing students with new textbooks and teacher with new teacher's handbooks. Attention and time should also be devoted to the implementation of these materials. Information should not only be given through workshops or courses but also through coaching teachers while they are working with the new materials in their classroom (Slavenburg, 1995).

An important lesson we have learned from this experiment is that we should not underestimate the *capacity of our pupils* (cf. Freudenthal, 1991). Providing pupils with a powerful model like the empty number line, establishing an open classroom culture in which students' solutions are taken seriously, and making teachers aware of both cognitive and motivational aspects of learning (cf. Vermeer, 1997), will help every student to become a flexible problem solver.

References

Ames, C., & Ames, R. (1989). *Research on motivation in education: Goals and cognitions.* San Diego: Academic Press.

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*, 260-267.

Ashcraft, M. H. (1985). Is it farfetched that some of us remember our arithmetic facts? *Journal for Research in Mathematics Education, 16*, 99-105.

Baroody, A. (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: Lawrence 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. (1989, September). *Learning addition and subtraction through materialized or schematic representation.* Paper presented at EARLI Conference, Madrid, Spain.

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. (1997). Development of mathematical strategies and procedures up to 100. In M. Beishuizen, K.P.E. Gravemeijer, & E.C.D.M. van Lieshout (Eds.) *The role of contexts and models in the development of mathematical strategies and procedures* (pp.127-155). Utrecht: Freudenthal Institute.

Beishuizen, M., Felix, E. & Beishuizen, J. (1990). A genetic model for tutoring addition and subtraction skills in an adaptive computer program. In J.M. Pieters, K. Breuer, & P.R.J. Simons (Eds.). *Learning Environments - Contributions from Dutch and German Research* (pp. 151-161). Berlin: Springer Verlag.

Beishuizen, M., Klein, T., Bergmans, C., Leliveld, K. & Hoogenberg, E. (1996). Hoe leerlingen zich ontwikkelen op de lege getallenlijn: twee portretten. [How pupils develop on the empty number line: two portraits]. In C. van den Boer and M. Dolk (Eds.). *Modellen, meten en meetkunde - Paradigma's van adaptief onderwijs* (pp. 117-130). Utrecht: Freudenthal Institute.

Beishuizen, M., Treffers, A., & Klein, A.S. (in preparation). *Flexibility of weaker and stronger students on mental arithmetic strategies.*

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.

Beishuizen, M., Wolters, G., & Broers, G. (1991). Mentale rekenprocedures in het getallengebied 20-100 onderzocht met reactietijdmeting en tempotoetsen [Research into mental computation procedures in the number domain 20-100 using response times and speed tests]. *Tijdschrift voor Onderwijsresearch, 16*, 19-38.

Bergmans, C., & Leliveld, K. (1997). Rekenen op de lege getallenlijn in groep 4. Een wereld van verschil [Arithmetic on the empty number line in the second grade. It makes all the difference]. *Willem Bartjens, 16*, 36-38.

Bishop, A.J., Clements, K., Keitel, C., Kilpatrick, J. & Laborde, C. (Eds.) (1996) *International handbook on mathematics education*. Dordrecht, The Netherlands: Kluwer.

Blöte, A.W. (1993, August). *Cognitive strategy use of learning disabled students: Metacognitive and motivational factors*. Paper presented at the Fifth European Conference for Research on Learning and Instruction, Aix-en-Provence, France.

Blöte, A.W., Klein, A.S., & Beishuizen, M. (in preparation). *Flexibility in mental computation up to one hundred. Procedure use in relation to procedure valuing*.

Boekaerts, M. (1987). Situation specific judgments of a learning task versus overall measures of motivational orientation. In E. de Corte, H. Lodewijks, R. Parmentier, & P. Span (Eds.), *Learning and Instruction* (pp. 169-179). Oxford/Leuven: Pergamon Press/Leuven University Press.

Boekaerts, M. (1988). Motivated learning: Bias and appraisals. *International Journal of Educational Research*, 12, 267-280.

Boekaerts, M. (1992). The adaptable learning process: Initiating and maintaining behavioral change. *Journal of Applied Psychology: An International Review, 41*, 377-397.

Boekaerts, M. (1995). The interface between intelligence and personality as determinants of classroom learning. In D.H. Saklofske & M. Zeidner (Eds.), *Handbook of personality and intelligence* (pp. 161-183). New York: Plenum Press.

Boekaerts, M. (1996). Personality and the psychology of learning. *European Journal of Personality, 10*, 377-404.

Boekaerts, M. (1997). Capacity, inclination and sensitivity for mathematics. *Anxiety, Stress, and Coping, 10*, 5-33.

Boekaerts, M. (in preparation). *What is the impact of different types of mathematics problems and different instructional conditions on students' experiental state?*

Boekaerts, M., & Beishuizen, M. (1991). *Flexibilization of arithmetic strategies on a different knowledge base*. Research proposal to The Netherlands National Science Foundation.

Bokhove, J., Van der Schoot, F., & Eggen, G. (1996). *Balans van het rekenonderwijs aan het einde van de basisschool 2* [An account of mathematics education at the end of primary school 2]. Arnhem, The Netherlands: National Institute for Educational Measurement.

Brown, A.L., & Palincsar, A.S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L.B. Resnick (Ed.), *Knowing, learning and* *instruction; Essays in the honor of Robert Glaser* (pp. 393-451). Hillsdale, NJ: Erlbaum.

Buys, K. (1988). Schaduwzijden van het honderdveld [Shady sides of the hundred square]. *Tijdschrift voor nascholing en onderzoek van het reken-wiskundeonderwijs, 7*, 3-10.

Carr, M. (1996). *Motivation in mathematics.* Cresshill, New Jersey: Hampton Press.

Carpenter, Th. P. (1997). Models for reform of mathematics teaching. In M. Beishuizen, K.P.E. Gravemeijer, & E.C.D.M. van Lieshout (Eds.) *The role of contexts and models in the development of mathematical strategies and procedures.* (pp.35-54). Utrecht: Freudenthal Institute.

Cantor, N. (1981). Perceptions of situations. In D. Magnusson (Ed.), *Toward a psychology of situations. An international perspective* (pp. 229-244). Hillsdale, NJ: Erlbaum.

Carpenter, Th.P. & Moser, J.M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education, 15*, 179-202.

Case, R. (1992). *The mind's staircase: States in the development of human intelligence.* Hillsdale, NJ: Erlbaum.

Case, R., & Griffin, S. (1989). Child cognitive development: the role of central conceptual structure in the development of scientific and social thought. In C.A. Haver (Ed.). *Advances in psychology: Developmental psychology.* North-Holland: Elsevier.

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.

Cobb, P., Yackel, E., & Wood, T. (1992). A constructivist alternative to the representational view of mind in mathematics education. *Journal for Research in Mathematics Education, 23*, 2-33.

Crombach, M.J., Voeten, M.J.M., & Boekaerts, M. (1994). A model for explaining individual differences between students in intended effort on curricular tasks. *Tijdschrift voor Onderwijsresearch, 19*, 301-317.

De Lange, J. (1992). Tussen einde en begin [Between the end and the beginning]. In F. Goffree, A. Treffers, & J. de Lange. *Rekenen anno 2002* (pp. 67-93). Utrecht: Freudenthal Institute.

Demetriou, A., Shayer, M., & Efklides, A. (Eds.) (1992). *The neo-Piagetian theories of cognitive development go to school: Implications and applications for education.* London: Routledge.

Felix, E.J.H.M. (1992). *Learner-based verus expert-based tutoring of arithmetic skills.* Doctoral dissertation. Vrije Universiteit, Amsterdam.

Freudenthal, H. (1968). Why to teach mathematics so as to be useful. *Educational Studies in Mathematics, 1*, 3-8.

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. (1990). Conceptual structures for multi-unit numbers: Implications for learning and teaching multidigit addition, subtraction, and place value. *Cognition and Instruction, 7*, 343-403.

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.

Fuson, K.C. & Briars, D.J. (1990). Using a base-ten blocks learning/teaching approach for first- and second-grade place-value and multidigit addition and subtraction. *Journal for Research in Mathematics Education, 21*, 180-206.

Fuson, K.C., Richards, J. & Briars, D.J. (1982). The acquisition and elaboration of the number word sequence. In C. Brainerd (Ed.). *Progress in Cognitive Development Research: Volume 1 Children's Logical and Mathematical Cognition* (pp. 33-92). New-York: Springer Verlag.

Fuson, K.C. & Smith, S.T. (1997). Supporting multiple 2-digit conceptual structures and calculation methods in the classroom: issues of conceptual supports, instructional design, and language. In M. Beishuizen, K.P.E. Gravemeijer, & E.C.D.M. van Lieshout (Eds.) *The role of contexts and models in the development of mathematical strategies and procedures* (pp.157-194). Utrecht: Freudenthal Institute.

Fuson, K.C., Wearne, D., Hiebert, J., Murray, H., Human, P., Olivier, A., Carpenter, T. & Fennema, E. (1997). Children's conceptual structures for multidigit numbers and methods of multidigit addition and subtraction. *Journal for Research in Mathematics Education, 28*, 130-162.

Gagné, R.M. (1977). *The conditions of learning*. Third edition. New York: Holt, Rinehart and Winston.

Gersten, R., & Carnine, D. (1984). Direct Instruction Mathematics: A longitudinal evaluation of low income elementary school students. *The Elementary School Journal, 84*, 395-407.

Gillissen, L. & Klep, J. (1980). *De getallenlijn: tellen, meten, rekenen en denken* [The number line: counting, measurement, arithmetic and thinking]. Tilburg: Zwijsen.

Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology, 40*, 631-666.

Gravemeijer, K.P.E. (1992). Socio-constructivisme en realistisch reken-wiskunde onderwijs (Socio-constructivism and realistic mathematics education). In: *Panama Cursusboek 10*, 50-59.

Gravemeijer, K.P.E. (1994). *Developing Realistic Mathematics Education*. Doctoral dissertation. Utrecht, the Netherlands: Freudenthal Institute.

Gravemeijer, K., & Ruesink, N. (1992). Van een realistische methode naar realistisch rekenonderwijs [From a realistic textbook towards realistic mathematics education]. *School & Begeleiding, 35*, 11-15.

Gravemeijer, K., Van Galen, F., Kraemer, J. M., Meeuwisse, T., & Vermeulen, W. (1983). *Rekenen & wiskunde* [Arithmetic & mathematics]. Baarn, The Netherlands: Bekadidact.

Gravemeijer, K., Van den Heuvel-Panhuizen, M., Van Donselaar, G., Ruesink, N., Streefland, L., Vermeulen, W., Te Woerd, E., & Van der Ploeg, D. (1993). *Methoden in het reken-wiskunde onderwijs, een rijke context voor vergelijkend onderzoek* (Textbook series in mathematics education, a rich context for comparative research). Utrecht: CD-ß Press.

Greeno, J.G. (1992, April). *Generality in the specificities of situations: understanding of concepts viewed as attunement to families of constraints.* Paper presented at the AERA-1992 in San Francisco.

Harskamp, E.G. (1988). *Rekenmethoden op de proef gesteld* (Mathematics textbooks evaluated). Groningen: RION.

Harskamp, E.G. & Suhre, C. (1986). *Vergelijking van rekenmethoden in het basisonderwijs* (Comparison of mathematics textbooks in primary school). Groningen: RION.

Harskamp, E.G. & Suhre, C. (1995). *Hoofdrekenen in het Speciaal Onderwijs - Eindverslag SVO-project 92081.* [Mental Arithmetic in Special Education - Final Report SVO-project 92081.] Groningen: GION.

Hiebert, J. (Ed.) (1986). *Conceptual and Procedural Knowledge: The Case of Mathematics.* Hillsdale, NJ: Lawrence Erlbaum.

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*, 12-21.

Hofstee, W.K.B. (1982). Evaluatie: Een methodologische analyse [Evaluation: A methodological analysis]. *Tijdschrift voor Onderwijsresearch, 7*, 193-202.

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 Dutch Educational Research Association Annual Meeting, Groningen, The Netherlands.

Klein, A.S. (1997). *A realistic versus gradual program design: a complete overview of the test results.* Unpublished manuscript, Leiden University, The Netherlands.

Klein, A.S., & Beishuizen, J.J. (1989). Leren rekenen met een adaptieve of rigide computercoach [Learning arithmetic with a adaptive or rigid computer coach]. In C.M. van Putten & M. Beishuizen (Eds.), *Informatietechnologie en media* (pp. 35-44). Lisse: Swets & Zeitlinger.

Klein, A.S., & Beishuizen, M. (1993, September). *Conflict and adaptation in the development of addition and subtraction strategies up to one hundred.* Paper 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). *Tempotoets rekenen 1995* [Arithmetic Speed Test 1995]. Unpublished test, Leiden University, The Netherlands.

Klein, A.S., & Beishuizen, M. (1995b). *Leerstof gebonden toetsen 1995* [Arithmetic Subject Matter Tests 1995). Unpublished tests, Leiden University, The Netherlands.

Klein, A.S., & Beishuizen, M. (1995c). *Stapjessommen toets 1995* [Arithmetic ScratchPaper Test 1995). Unpublished test, Leiden University, The Netherlands.

Klein, A.S., & Beishuizen, M. (1995d). *Transfertoets rekenen 1995* [Arithmetic Transfer Test 1995]. Unpublished test, Leiden University, The Netherlands.

Klein, A.S., Beishuizen, M. & Treffers, A. (In press). The empty numberline in Dutch second grades: "Realistic" versus "Gradual" program design. *Journal for Research in Mathematics Education.*

Klein, A.S., & Boekaerts, M. (in preparation). *Do different instructional conditions create gender differences in mathmatical problem-solving behavior?*

Klein, T. & Beishuizen, M. (1994a). 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.

Klein, T. & Beishuizen, M. (1994b). Flexibilisering van rekenstrategieën op een verschillende kennisbasis [Flexibilization of arithmetic strategies on a different knowledge base]. *Tijdschrift voor nascholing en onderzoek van het reken-wiskundeonderwijs, 13*, 32-38.

Laros, J.A., & Tellegen, P.J. (1991). *Construction and validation of the SON-R 5 - 17, the Snijders-Oomen non-verbal intelligence test.* Groningen: Wolters-Noordhoff.

Lawler, R.W. (1990). Constructing knowledge from interactions. In L.P. Steffe & T. Wood (Eds.), *Transforming children's mathematics education* (pp. 47-61). Hillsdale, NJ: Erlbaum.

Lazarus, R.S., & Folkman, S. (1984). *Stress, appraisal, and coping.* New York: Springer.

Lewis, & Mayer (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology, 79*, 363-371.

Linssen, I. (1996). *Stuurbaarheid van rekenstrategieën bij optellen en aftrekken tot honderd* [Controllability of arithmetic strategies with adding and subtracting up to 100]. Unpublished master's thesis, University of Nijmegen, The Netherlands.

Lorenz, J.H. (1997). Strategies? Procedures? Or is mental calculation just strolling around in an imaginary number space?

In M. Beishuizen, K.P.E. Gravemeijer, & E.C.D.M. van Lieshout (Eds.) *The role of contexts and models in the development of mathematical strategies and procedures* (pp 199-214). Utrecht: Freudenthal Institute.

McIntosh, A., Reys, B.J. & Reys, R.E. (1992). A proposed framework for examining basic number sense. *For the Learning of Mathematics, 12*, 2-44.

McLeod, D.B. (1992). Research on affect in mathematics education: A reconceptualization. In D.A. Grouws (Ed.), *Handbook of research on mathemtics teaching and learning* (pp. 575-595). New York: Macmillan.

Moerlands, F. (1992). Stappenschema's. Een aanvulling of een vervolg op de lege getallenlijn [Step schemes. An addition to or continuation of the empty number line]. *Willem Bartjens, 12*, 42-46.

Nicholls, J.G. (1983). Conceptions of ability and achievement motivation: A theory and its implications for education. In S.G. Paris, G.M. Olson, & H.W. Stevenson (Eds.), *Learning and motivation in the classroom* (pp. 211-237). Hillsdale, NJ: Erlbaum.

Nicholls, J.G. (1984). Conceptions of ability and achievement motivation. In R. Ames & C. Ames (Eds.), *Research on motivation in education* (Vol. 1, pp. 39-73). San Diego: Academic Press.

Nicholls, J.G., Cobb, P., Wood, T., Yackel. E., & Patashnick, M. (1990). Assessing students' theories of succes in mathematics: Individual and classroom differences. *Journal for Research in Mathemics Education, 21*, 109-122.

Pintrich, P.R., & De Groot, E.V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*, 33- 40.

Plunkett, S. (1979). Decomposition and all that rot. *Mathematics in schools, 8*, 2-5.

Radatz, H. (1987). Vorstellungen, Strategisches Wissen und Algorithmen im Mathematikunterricht. [Representation, strategic knowledge and algorithms in mathematics teaching.] In H.G. Steiner (Ed.), *Interdisziplinäres Kolloquium Heinrich Bauersfeld zu Ehren* (pp. 51-63). [Interdisciplinary colloquium to honor Heinrich Bauersfeld.] Bielefeld: Institut für Didaktik der Mathematik.

Radatz, H. (1993). MARC bearbeitet Aufgaben wie 72 - 59 - Anmerkungen zu Anschauung und Verständnis in Arithmetikunterricht. [MARC's solutions of problems like 72 - 59 - Comments on visualization and understanding in learning arithmetic.] in J.H. Lorenz (Ed.). *Mathematik und Anschauung* (pp. 14-24). [Mathematics and visualization.] Köln: Aulis Verlag.

Radatz, H. & Schipper, W. (1983). *Handbuch für den mathematikunterricht an Grundschulen* [Handbook for mathematics education in primary schools]. Hannover, Germany: Schroedel.

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. (1986). The development of mathematical intuition. In M. Perlmutter (Ed.), *Perspectives on intellectual development* (pp 159-200). 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.

Ruijssenaars, A.J.J.M. (1992). *Rekenproblemen. Theorie, diagnostiek, behandeling.* [Arithmetic learning problems. Theory, diagnostics, treatment.] Rotterdam: Lemniscaat.

Ruijssenaars, A.J.J.M. (1994). Speciaal rekenen - Hoe irrealistisch is een orthopedagogische-didactische benadering? [Special arithmetic - How unrealistic is an orthopedagogical-didactical treatment?] In M. Dolk, H. Van Luit & E. Te Woerd (Eds.). *Speciaal Rekenen* (pp. 29-41). [Special Arithmetic.] Utrecht: Freudenthal Institute.

Schoenfeld, A.H. (1992). Learning to think mathematically: problem solving, metacognition and sense making in mathematics. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 189-215). New York: Macmillan.

Seegers, G., & Boekaerts, M. (1993). Gender-related differences in self-referenced cognitions in relation to mathematics. *Journal for Research in Mathematics Education, 27*, 215-240.

Selter, C. (1994). *Eigenproduktionen im Arithmetikunterricht der Primarstufe* [Children's own productions in primary mathematics teaching]. Wiesbaden, Germany: Deutscher Universitäts-Verlag.

Selter, C., & Spiegel, H. (1997). *Wie Kinder rechnen* [How children do arithmetic]. Leipzig, Germany: Klett Grundschulverlag.

Slavenburg, J.H. (1995). *Educatieve dienstverlening: werkwijzen, effecten en waardering* [Educational service: working-methods, effects and valuing]. Rotterdam: Partners.

Stern, E. (1993). What makes certain arithmetic word problems involving the comparison of sets so hard for children? *Journal of Educational Psychology, 85*, 7-23.

Streefland, L. (1991a). *Fractions in Realistic Mathematics Education: A paradigm for developmental research.* Dordrecht, The Netherlands: KluwerAcademic.

Streefland, L. (Ed.) (1991b). *Realistic Mathematics Education in primary schools.* Utrecht: Freudenthal Institute.

Sundermann, B., & Selter, C. (1995). Halbschriftliches Rechnen auf Eigenen Wegen [Half-written arithmetic strategies follow their own ways]. In G.N. Müller & E.C. Wittmann (Eds.) *Mit Kindern Rechnen* (pp. 165-178). Frankfurt am Main: Der Grundschulverband.

Thompson, I. (1994). Young children's idiosyncratic written algorithms for addition. *Educational Studies in Mathematics*, 26, 323-345.

Thompson, I. (Ed.) (1997). *Teaching and Learning Early Number.* Buckingham, UK: Open University Press.

Torn, M., & Ruyters, M. (1992). *Oplossing-en-inzicht? Evaluatiestudie van een leerlijn voor optellen en aftrekken tot honderd* [Solution-and-insight? An evaluation study of a program design for addition and subtraction up to 100]. Master's thesis, Leiden University.

Treffers, A. (1987). *Three Dimensions. A model of goal and theory description in mathematics education: The Wiskobas Project.* Dordrecht, The Netherlands: Reidel.

Treffers, A. (1991a). Realistic mathematics education in the Netherlands 1980-1990. In L. Streefland (Ed.), *Realistic Mathematics Education in primary schools* (pp. 11-20). Utrecht: Freudenthal Institute.

Treffers, A. (1991b). Didactical background of a mathematics program for primary education. In L. Streefland (Ed.), *Realistic Mathematics Education in primary schools* (pp. 21-56). Utrecht: Freudenthal Institute.

Treffers, A. (1995). Rekenen op maat [Adaptive math teaching]. *Tijdschrift voor orthopedagogiek, 34*, 183-194.

Treffers, A. & De Moor, E. (1990). *Proeve van een Nationaal Programma voor 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: Zwijssen.

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*, 11-14.

Van Beusekom, N. et al. (1991). *Pluspunt.* Den Bosch: Malmberg

Van de Molengraaf, G.W.J. et al. (1981). *De Wereld in Getallen* [The World in Numbers]. Den Bosch: Malmberg.

Van den Berg, W., & Van Eerde, D. (1992). *Geef me de vijf* [Give me five]. Rotterdam: RISBO, Erasmus University.

Van den Brink, F.J. (1974). Wie gaat er mee? Wie niet? [Who is coming along? Who is not?). *Wiskobas-Bulletin*, 3, 250-261.

Van den Heuvel-Panhuizen, M. (1996). *Assessment and Realistic Mathematics Education.* Doctoral dissertation. Utrecht: Freudenthal Institute.

Van den Heuvel-Panhuizen, M., & Gravemeijer, K.P.E. (1991). Numberfacts en de reken-wiskunde methode [Numberfacts and the arithmetic-mathematics textbook]. In P. Tesser & J.H.G.I. Giesbers (Red.), *Schoolorganisatie en Curriculum - Bijdragen ORD 1990* (pp. 245-255). Nijmegen: ITS.

Van der Heijden, M.K. (1993). *Consistentie van Aanpakgedrag - Een Proces-diagnostisch Onderzoek naar Acht Aspecten van Hoofdrekenen.* [Consistency of approach behavior - A process-assessment study into eight aspects of mental arithmetic.] Doctoral dissertation, Leiden University. Lisse: Swets & Zeitlinger.

Van Gelder, L. (1969). *Grondslagen van de rekendidactiek* [Foundations of mathematics education in primary school]. Groningen: Wolters-Noordhoff.

Van Lieshout, E.C. D.M. (1997). What can research on word and context problems tell about effective strategies to solve subtraction problems? In M. Beishuizen, K.P.E. Gravemeijer, & E.C.D.M. van Lieshout (Eds.) *The role of contexts and models in the development of mathematical strategies and procedures.* (pp. 79-110). Utrecht: Freudenthal Institute.

Van Luit, J.E.H. & Van de Rijt, B.A.M. (1997). Effectiveness of the AEM program for teaching children early mathematics. In M. Beishuizen, K.P.E. Gravemeijer, & E.C.D.M. van Lieshout (Eds.) *The role of contexts and models in the development of mathematical strategies and procedures.* (pp. 215-238). Utrecht: Freudenthal Institute.

Van Mulken, F. (1992). *Hoofdrekenen en Strategisch Handelen - Het gevarieerd Gebruik van Twee Grondvormen van Optellen en Aftrekken tot Honderd.* [Mental Arithmetic and Strategic Acting - The Varied Use of Two Basic Types of Addition and Subtraction up to one Hundred.]. Doctoral dissertation, Leiden University.

Vermeer, H.J. (1997). *Sixth-grade students' mathematical problem-solving behavior. Motivational variables and gender differences.* Doctoral dissertation, Leiden University.

Verschaffel, L. (1997). Young childrens's strategy choices for solving elementary arithmetic wor(l)d problems: the role of task and context variables. In M. Beishuizen, K.P.E. Gravemeijer, & E.C.D.M. van Lieshout (Eds.) *The role of contexts and models in the development of mathematical strategies and procedures.* (pp.111-126). Utrecht: Freudenthal Institute.

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.F. Friedrich (Eds.), *Learning and Instruction* (pp. 415-428). Oxford: Pergamon.

Verschaffel, L. & DeCorte, E. (1996). Leren realistisch modeleren van vraagstukken. Een exploratief onderwijsexperiment bij leerlingen van de bovenbouw van de basisschool [Teaching how to interprete and model word problems in a realistic way. An explorative educational experiment with students of the junior department of primary school]. *Pedagogische Studiën, 73*, 322-337.

Voogt, S. (1996). *Rekenen tot honderd. Strategiewaardering en motivatie* [Arithmetic up to hundred. Strategy valuing and motivation]. Master's thesis, Leiden University.

Vuurmans, A.C. (Ed.) (1991). *Rekenen tot honderd* (Addition and subtraction up to 100). 's Hertogenbosch: KPC.

Wijnstra, J.M. (Ed.) (1988). *Balans van het rekenonderwijs in de basisschool: Uitkomsten van de eerste rekenpeiling medio en einde basisonderwijs* [An account of mathematics education in primary school: First results of the Dutch national assessment program in primary education in grades 5 and 8]. Arnhem, The Netherlands: National Institute for Educational Measurement .

Williams, E. & Shuard, H. (1982). *Primary mathematics today* (3rd Ed.). Harlow, Essex, England: Longman.

Willems, F. & Groenewegen, K. (1997). De rekenboog. Een instrument ten behoeve van het rekenen tot 20 [The arithmetic-bow. An instrument for aritmetic up to 20]. *Willem Bartjens*, 16, 26-29.

Wittmann, E.Ch. & Müller, G.N. (1995). *Handbuch Produktiver Rechenübungen - Band 1* [Handbook for Productive Arithmetic Activities - Volume 1]. Stuttgart: Klett Schulbuchverlag.

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.

Appendix A

Table A-1 Mental computation procedures for addition and subtraction up to 100

Acro- nym:	Example Addition 45 + 39; 45 + 8	Example Subtraction 65 - 49; 65 - 8; 51 - 49
N10	45 + 30 = 75; 75 + 9 = 84	65 - 40 = 25; 25 - 9 = 16
uN10	45 + 9 = 54; 54 + 30 = 84	65 - 9 = 56; 56 - 40 = 16
N10C	45 + 40 = 85; 85 - 1 = 84	65 - 50 = 15; 15 + 1 = 16
A10	45 + 5 = 50; 50 + 34 = 84	65 - 5 = 60; 60 - 40 = 20; 20 - 4 = 16 49 + 1 = 50; 50 + 10 = 60; 60 + 5 = 65 answer: 1 + 10 + 5 = 16 (adding-on)
A10C	45 + 5 = 50; 50 + 40 = 90; 90 - 6 = 84	49 + 1 = 50; 50 + 20 = 70; 70 - 5 = 65 answer: 1 + 20 - 5 = 16
10s	40 + 30 = 70; 70 + 5 = 75; 75 + 9 = 84	60 - 40 = 20; 20 + 5 = 25; 25 - 9 = 16
1010	40 + 30 = 70; 5 + 9 = 14; 70 + 14 = 84	60 - 40 = 20; 5 - 9 = 4 (false reversal) 20 + 4 = 24 (false answer)
u1010	5 + 9 = 14; 40 + 30 = 70; 70 + 14 = 84	5 - 9 = 4 (false reversal); 60 - 40 = 20; 20 + 4 = 24 (false answer)
CoW (Column- Wise)	45 39 + 84	65 49 − 16
C(Clever)	44 + 40 = 84	66 - 50 = 16
∩ (Conn. Arc)	–	51 - 49 = 2
CS	45 + 5 = 50; 50 + 3 = 53	65 - 5 = 60; 60 - 3 = 57
SS	45 + 4 = 49; 49 + 4 = 53	65 - 4 = 61; 61 - 4 = 37
1-jump	45 + 39 = 84	65 - 49 = 16
NS (No Steps)[a]	45 + 39 = 84	65 - 49 = 16
UNC (Unclear)	unclearly written steps or drawn jumps	unclearly written steps or drawn jumps

Note. These steps can also be drawn as jumps on the number line or written by means of the arrow scheme.
[a]The difference between 1-jump and no steps is that for no steps only the answer to the problem is given. For 1-jump only one jump is drawn or one step is written down to solve the problem.

Appendix B

Table B-1 Procedural and non-procedural errors for addition and subtraction up to 100

Type of error	Example addition: 45 + 39	Example subtraction: 65 - 49
Non-procedural errors		
Error in number of units (minus or plus 2 at most)	45 + 39 = 85	65 - 49 = 15
Error in number of tens or 10-jump (minus or plus 20 at most)	45 + 39 = 94 or 45 + 30 = 74; 74 + 9 = 83	65 - 49 = 26 or 65 - 40 = 15; 15 - 9 = 6
Error in number of tens and units (minus or plus 22 at most or reversal of numbers)	45 + 39 = 75 45 + 39 = 48	65 - 49 = 25 65 - 49 = 61
Error in splitting of a number	45 + 30 = 75; 75 + 5 =80; 80 + 3 = 83	65 - 40 = 25; 25 - 5 = 20; 20 - 3 = 17
Number error	55 + 39 = 94	65 - 59 = 6
Number perseverence error	45 + 30 = 75; 75 + 9 = 89	65 - 40 = 25; 25 - 9 = 19
Wrong operation error	45 - 39 = 16	65 + 49 = 114
Procedural errors		
Direction error	45 + 40 = 85; 85 + 1 = 86	65 - 50 = 15; 15 - 1 = 14 or 60 - 40 = 20; 20 - 5 = 15; 15 - 9 = 16 or 60 - 40 = 20; 5- 9 = 4; 20 + 4 = 24[a]
Error in number of steps	45 + 30 = 75	65 - 40 = 25
Procedure correct but wrong answer	45 + 5 = 50; 50 + 30 = 80; 80 + 4 = 84; answer 39	49 + 1 = 50; 50 + 10 = 60; 60 + 5 = 65 answer 65
Missing answer	45 + 30 = 75; 75 + 9 = ...	65 - 40 = 25; 25 - 9 = ...
Unclear procedural error	45 + 30 = 47; 47 + 9 = 57	65 - 40 = 51; 51 - 9 = 42
Unclear error	unclearly written steps or drawn jumps	unclearly written steps or drawn jumps

Note. These steps can also be drawn as jumps on the number line or written by means of the arrow scheme.
[a]This type of direction-error is also known as the smaller from larger bug.

Appendix C

Table C-1 Mental computation procedures for addition and subtraction up to 1000

Acronym	Example Addition 225 + 124	Example Subtraction 368 - 234; 301 - 298
N100	225+ 100 = 325; 325 + 20 = 345; 345 + 4 = 349	368 - 200 = 168; 168 - 30 - 138; 138 - 4 = 134
uN100	225 + 4 = 229; 229 + 20 = 249; 249 + 100 = 349	368 - 4 = 364; 364 - 30 = 334; 334 - 200 = 134
N100C	225 + 100 = 325; 325 + 25 = 350; 350 - 1 = 349	368 - 200 = 168; 168 - 40 = 128; 128 + 6 = 134
A100	225 + 75 = 300; 300 + 25 = 325; 325 + 24 = 349	368 - 68 = 300; 300 - 100 = 200; 200 - 66 = 134; or 234 + 66 = 300; 300 + 68 = 368; 66 + 68 = 134
A10	225 + 5 = 230: 230 + 100 = 330; 330 + 19 = 349	368 - 8 = 360; 360 - 200 = 160; 160 - 26 = 134 or 234 + 6 = 240; 240 + 128 = 368; 128 = 6 = 134
100s	200 + 100 = 300; 300 + 25 = 325; 325 + 24 = 349	300 - 200 =100; 100 + 68 = 168; 168 - 34 = 134
100100	200 + 100 = 300; 25 + 24 = 49; 300 + 49 = 349	300 - 200 = 100; 68 - 34 = 34; 100 + 34 = 134
u100100	25 + 24 = 49; 200 + 110 = 300; 300 + 9 = 349	68 - 34 = 34; 300 - 200 = 100; 100 + 34 = 134
CoW (Column- Wise)	225 124 349	368 234 134
C(Clever)	220 + 120 = 340 + 9 = 349	369 - 235 = 134
∩ (Conn. Arc)	-	301- 298 = 3
1-jump	225 + 134 = 349	368 - 234 = 134
NS (No Steps)[a]	225 + 134 = 349	368 - 234 = 134
UNC (Unclear)	unclearly written steps or drawn jumps	unclearly written steps or drawn jumps

Note. These steps can also be drawn as jumps on the number line or written by means of the arrow scheme.

[a] The difference between 1-jump and no steps is that for no steps only the answer to the problem is given. For 1-jump only one jump is drawn or one step is written down to solve the problem.

Summary

In discussions about the renewal of primary mathematics education, mental arithmetic has become more important because it stimulates number sense and the understanding of number relations (McIntosh, Reys, & Reys, 1992). Recent developments in the area of mathematics education in elementary schools also point to a greater emphasis on the role of the student as somebody who actively constructs mathematics (Becker & Selter, 1996). This is in line with the instructional theory of Realistic Mathematics Education (RME) which sees school mathematics as an activity (Freudenthal, 1991). In the Netherlands most of the mathematics textbooks in primary school are based on the principles of RME. However, a first national evaluation study of mathematics education in the Netherlands in primary schools (Wijnstra, 1988) 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. This was an important reason behind the new Specimen of a National Program for Primary Mathematics Teaching (Treffers &De Moor, 1990) and the development and comparison of two experimental programs for teaching mental addition and subtraction up to 100 in the Dutch second grade. The goal of both programs is greater flexibility in mental arithmetic through use of the empty number line as a new mental model, and raising the level of procedural competency for the fore-mentioned type of problems. We were interested in how these two program designs influenced the development of both cognitive and affective processes of the students who worked with one of the experimental programs.

In chapter 1, both a cognitive and an affective perspective towards mathematics education are presented. Earlier models for mental addition and subtraction up to 100 are described in order to elucidate the introduction of the empty number line (Treffers & De Moor, 1990) as a new didactic model. The main mental computation procedures in this domain are also described, as well as the effect of these models on mental computation procedures. The empty number line is incorporated in two program designs that differ in instructional design to enable comparison of two 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 objective, a gradual increase in knowledge through initial emphasis on procedural computation followed by flexible problem solving. Beside the effects of these program designs on cognitive outcome variables, we were also interested in the possible effects on the students' reported experiential states and their motivational beliefs. To interpret our findings, the model of adaptable learning (Boekaerts, 1992, 1995) is described in which cognitive and affective person variables are integrated. We were interested in motivational beliefs at the domain-specific level and task-specific appraisals. The chapter ends with the expected differences in motivational beliefs elicited by the two program designs.

The second chapter describes the background of the two program designs. First some main features of the RPD and GPD are described: the theoretical framework, the way the number line is introduced, the role of mental arithmetic, the role of context problems and the role of the teacher. For each program design, a time schedule and an instructional sequence is given to provide an insight into what the context of the two programs was and how they were arranged. At the end of the chapter, hypotheses are formulated for the outcome for the RPD and GPD condition. For this we used Hofstee's bet-model (1982). According to this model, hypotheses have been formulated by two parties, each of them representing a different point of view. The two parties have to commit themselves to their predictions before the experiment starts. Treffers (one of the proponents of RME designs) formulated predictions about the outcome of the RPD condition, while Beishuizen formulated hypotheses for the GPD. These hypotheses concern the development of procedural and strategic knowledge, results for weaker and better students and development of motivational processes. Finally some post hoc questions were formulated by Klein.

The method of research is described in chapter 3. Subjects were 275 second grade students (7-8 years) selected from 9 primary schools which had experience in working with realistic mathematics textbooks. From this sample 100 students were selected to test the predictions concerning weaker and better students. Teachers and students used experimental teacher's guides and textbooks instead of their regular mathematics textbooks. Measures for non-verbal intelligence, procedural and strategic competence in the domain of addition and subtraction up to 100, transfer to the domain for addition and subtraction up to 1000 and questionnaires for domain-specific motivational beliefs and task-specific cognitions and affects were administered in the group setting. Flexibility in using different computation procedures was measured by both looking at the procedures that were used by the whole group of RPD and GPD students, and the extent to which a student changed his solution behavior across items.

The results are presented in chapters 4 and 5. For both chapters, we only described the results half-way through and at the end of the curriculum because for these moments, hypotheses were formulated by Treffers and Beishuizen. To answer the post hoc questions, some of the results of the tests that were administered in the third grade, after the students had stopped working with the experimental programs, are also described. The cognitive results are described in chapter 4, whilst chapter 5 describes the affective processes.

We found that half-way through the curriculum, compared to the GPD students, the procedural competence of the RPD students in solving numerical problems was of a higher level. This outcome was unexpected since the RPD placed less emphasis on written exercises with numerical problems. Instead, more time was spent on interactive teaching and on oral solution of such problems during whole-class teaching. One outcome that had been predicted was that the RPD students were more flexible than the GPD students in their procedure use. At the end of the curriculum the RPD students still performed better in respect of procedural competence in solving numerical problems. Therefore Treffers' hypothesis regarding procedural competence in addition and subtraction up to 100 was confirmed. His hypothesis concerning strategic knowledge was also confirmed. The RPD students were more flexible in adapting their procedures according to the number characteristics of the problem than the GPD students, also more RPD students could be categorized as flexible or half-flexible than GPD students.

The results for the weaker and better RPD and GPD students point more or less in the same direction. Half-way through the curriculum no differences were found in procedural competence between weaker RPD and GPD students. Regarding strategic knowledge, we found that both the better and weaker RPD students were more flexible in using different solution procedures than the better and weaker GPD students. At the end of the curriculum, the results of the weaker and better RPD students on the speed tests were better than the results of the weaker and better GPD students. The weaker and better RPD students appeared also to be more flexible in using different solution procedures according to the number characteristics of the problem. The expected confusion in the weaker RPD students, did not show up in the test results at the end of the curriculum. In general, we could say that the results of the weaker and better students are in accordance with the hypotheses formulated by Treffers. Beishuizen's hypotheses must therefore be rejected.

With respect to the post hoc questions we found no differential transfer effects for the two program designs regarding addition and subtraction problems with numbers up to 1000. At the beginning of the third grade, after the program had stopped, the RPD students still appeared to be more flexible in adapting their solution procedures according to the number characteristics of the problems. The solution procedures seemed to have been incorporated well by the RPD students and to a lesser extent by the GPD students.

The results regarding the motivational variables are described in chapter 5. At the domain-specific level, it appeared that half-way through the curriculum the GPD students showed more positive affects towards mathematics as a school subject than the RPD students. The GPD students also appeared to have a higher self-concept of their mathematical abilities than the RPD students. At the end of the curriculum opposite results were found. The RPD students had more positive affects towards mathematics than the GPD students. This result was even better than Treffers had predicted and therefore his hypothesis is confirmed and Beishuizen's hypothesis is rejected. At the task-specific level in general, we found no significant differences between RPD and GPD students half-way through the curriculum. When we look at the different tasks, we see that both the RPD and GPD students attached more value and intended to investing more effort into context than numerical problems. Half-way the curriculum, RPD students reported feeling more self-confident about solving numerical than context problems. This finding is not in accordance with Treffers' hypothesis who predicted that there would be no difference in favorable cognitions and affects between numerical and context problems. Beishuizen's hypothesis at the task-specific level is confirmed: The GPD students liked the numerical problems more than context problems. At the end of the curriculum we found that the RPD students scored significantly higher than GPD students on self-confidence and, to a lesser extent, on task-attractiveness towards numerical and context problems. Therefore Beishuizen's hypothesis for the results at the end of the curriculum must be rejected and Treffers' hypothesis is confirmed: RPD students were more self-confident and liked both numerical and context problems better than GPD students.

With respect to the post hoc question about the cognitions and appraisals of the weaker and better students at the domain- and task-specific level we can say the major differences were found at the end of the curriculum. The findings for the weaker students at the domain- and task-specific level are in the same direction: weaker RPD students felt more self-confident about mathematics than weaker GPD students. These findings support the idea that the didactic sequence of the RPD did not harm the weaker RPD students. A structured approach towards mathematics, as in the GPD, does not necessarily result in a higher self-concept of mathematical ability for the students, even not for the weaker students.

In the last chapter the results are discussed and recommendations are given for education and future research. The RPD appeared to be more successful in attaining the objectives of the study than the GPD. The RPD students scored higher on both procedural competence and strategic knowledge in the domain of addition and subtraction up to 100. In this respect the performance of the weaker RPD students was most convincing: outperforming the weaker GPD students both on procedural speed and flexibility of mental computation. This result is probably due to the fact that the central Realistic Mathematics principle of mathematics as an activity was achieved to a greater extent in the RPD than in the GPD condition. Also the way the (empty) number line was used by the RPD students might provide an explanation for these results. It seems that the empty number line as a mental model is not used by all GPD students. These two factors may also be true of the explanation of the good results of the weaker RPD students. The weaker RPD students came to see the empty number line as a natural and transparent model they could rely on when confronted with difficult problems. This is only true to a lesser extent for the weaker GPD students. These conclusions are supported by the results on the motivational questionnaires which were more positive for the RPD than for the GPD students.

Besides the effect of the experimental textbooks on the students, the role of the teacher is also discussed. The teacher plays an important role in implementing the ideas he or she reads in the teacher's guide. We think that the RPD approach was more successful than the GPD approach, in creating a climate of interactive teaching. In the RPD, the teachers had to establish mathematics as an activity from the beginning of the curriculum. The GPD teachers only had the last three months of the curriculum in which to establish a climate of interactive teaching, a time span that was probably too short. The RPD teachers were apprehensive about creating a climate of interactive teaching and placing less emphasis on written exercises, especially during the first half of the school year. The fortnightly visits to the school made by one of the researchers were considered to be very supportive in this period.

Classroom experiences and test results were discussed during these visits, which were also made to GPD teachers. This underlines the importance of how an innovation in education is implemented. The innovation does not end by providing students with new textbooks and teachers with new teacher's guides. In addition to courses and workshops, the teachers should be coached while they are working with new materials in their classroom. Another thing we learned from this experiment is that we should not underestimate the capacity of our students. Providing students with a powerful model like the empty number line, establishing an open classroom culture in which students' solutions are taken seriously, and making teachers aware of both cognitive and motivational aspects of learning, will help every student become a flexible problem solver.

Samenvatting

Flexibilisering van rekenopgaven op een verschillende kennisbasis: de lege getallenlijn binnen een realistische en stadiagewijze leerlijn.

In de discussie over de vernieuwing van reken- wiskundeonderwijs op de basisschool, speelt hoofdrekenen een belangrijkere rol, omdat hierdoor het getalbegrip en het begrip van getalsrelaties gestimuleerd wordt (McIntosh, Reys, & Reys, 1992). Ook de rol van de leerling als iemand die zelf actief wiskunde construeert neemt binnen recente ontwikkelingen op het gebied van het reken- wiskundeonderwijs op de basisschool een steeds belangrijkere plaats in (Becker & Selter, 1996). Deze visie komt overeen met de theorie van het realistisch rekenonderwijs waarbinnen rekenen wordt opgevat als een activiteit (Freudenthal, 1991). De meeste rekenmethoden die in Nederland worden gebruikt zijn gebaseerd op principes van het realistisch rekenonderwijs. De in 1987 gehouden eerste Periodieke Peiling van het Onderwijs (PPON) wees echter uit dat op het gebied van rekenen, bepaalde gebieden onvoldoende beheerst werden (Wijnstra, 1988). Zo bleek bijvoorbeeld slechts 55% van de leerlingen van groep 5 van het basisonderwijs in staat te zijn om de som 64 - 28 goed uit te rekenen. Uit dit onderzoek kwam ook naar voren dat leerlingen over het algemeen weinig flexibel zijn in het gebruiken van verschillende oplossingsstrategieën. Deze twee uitkomsten vormden een belangrijke reden om een nieuwe Proeve van een National Programma voor het Reken-wiskunde- onderwijs op de Basisschool (Treffers & De Moor, 1990) te publiceren, en om twee nieuwe leerlijnen voor het optellen en aftrekken tot 100 te ontwikkelen en met elkaar te vergelijken. Het doel van beide leerlijnen is het bereiken van een hogere mate van flexibiliteit in het gebruik van oplossingsstrategieën en het komen tot een betere beheersing van het hierboven genoemde somtype. Hiervoor wordt in beide leerlijnen de lege getallenlijn gekozen als centraal model. De vraag is welke invloed de twee leerlijnen hebben op de ontwikkeling van zowel cognitieve als affectieve processen, bij de leerlingen die met één van de twee programma's hebben gewerkt.

Hoofdstuk 1 bekijkt het reken- wiskundeonderwijs vanuit een cognitief en een affectief perspectief. Om de introductie van de lege getallenlijn als nieuw didactisch model (Treffers & De Moor, 1990) te verduidelijken worden de voorgaande modellen voor het optellen en aftrekken tot 100 beschreven. Vervolgens beschrijft dit hoofdstuk de belangrijkste oplossingsprocedures voor het rekenen tot 100 alsmede de invloed van voornoemde modellen op deze oplossingsmanieren. De lege getallenlijn is opgenomen in de beide leerlijnen, die qua opbouw van elkaar verschillen om op die manier twee verschillende instructieprincipes met elkaar te kunnen vergelijken. De Proeve-leerlijn stimuleert het flexibel gebruik van verschillende oplossingsmanieren van meet af aan door gebruik te maken van realistische context problemen. De Stadia-leerlijn daarentegen gaat uit van een geleidelijke kennisopbouw en benadrukt eerst de procedurele kant van het rekenen, waarna er ruimte is voor het op een flexibele manier oplossen van problemen. De interesse ging niet alleen uit naar het effect van deze twee leerlijnen op cognitieve maten, maar ook naar het effect op motivationele variabelen. Daarbij wordt een onderscheid gemaakt tussen domein-specifieke en taak-specifieke variabelen. De resultaten worden geïnterpreteerd met behulp van het model van adaptief leren (Boekaerts, 1992, 1995), waarin de samenhang tussen cognitieve en affectieve variabelen is beschreven. Het hoofdstuk eindigt met de verwachte verschillen in motivationele opvattingen, die op grond van het werken met één van de twee programma's zouden kunnen ontstaan.

Hoofdstuk 2 beschrijft de achtergrond van de beide leerlijnen. Het beschrijft eerst de belangrijkste kenmerken van de Proeve- en de Stadia-leerlijn: de theoretische achtergrond, de wijze van introductie van de getallenlijn, de rol die context problemen spelen en de rol van de leerkracht. Om een idee te krijgen hoe de twee programma's eruit zien, wordt er voor elke leerlijn een tijdschema beschreven en de volgorde waarin verschillende onderwerpen aan bod komen. Het einde van het hoofdstuk geeft de formulering van hypothesen. Hierbij wordt gebruik gemaakt van het weddenschapsmodel van Hofstee (1982). Volgens dit model formuleren twee partijen, vanuit verschillende gezichtspunten, verwachtingen omtrent de uitkomst van het experiment. De twee partijen moeten deze verwachtingen vastleggen voordat het experiment begint. Treffers (één van de pleitbezorgers van het realistisch rekenen) formuleerde verwachtingen omtrent het resultaat van de Proeve-leerlijn, terwijl Beishuizen vanuit cognitieve psychologie hetzelfde deed voor het resultaat van de Stadia-leerlijn. Er werden hypothesen geformuleerd omtrent de ontwikkeling van procedurele en strategische kennis, prestaties van de zwakke en goede rekenaars en de ontwikkeling van motivationele processen. Tenslotte formuleerde Klein een aantal post-hoc vragen.

Hoofdstuk 3 beschrijft de methode van onderzoek. Er werden 275 leerlingen uit groep 4 geselecteerd (leeftijd 7-8 jaar) afkomstig van 9 basisscholen die ervaring hadden in het werken met een realistische rekenmethode. Hieruit werden 100 leerlingen geselecteerd om de hypothesen met betrekking tot de goede en zwakke rekenaars te toetsen. De leerlingen en de leerkrachten gebruikten experimentele leerlingboekjes en handleidingen in plaats van hun alledaagse rekenmethode. Toetsen voor non-verbale intelligentie, procedurele en strategische kennis in het domein van optellen en aftrekken tot 100, transfer van dit domein naar het domein van optellen en aftrekken tot 1000, en vragenlijsten voor domein- en taak-specifieke cognities en motivationele opvattingen, werden klassikaal bij alle leerlingen afgenomen. Flexibiliteit in gebruik van verschillende oplossingsprocedures werd zowel op groepsniveau bekeken als individueel. De vraag hierbij was: in welke mate past een leerling zijn of haar oplossingsgedrag aan, aan de aard van de verschillende opgaven.

De hoofdstukken 4 en 5 beschrijven de resultaten van het onderzoek. Beide hoofdstukken vermelden alleen de resultaten, die halverwege en aan het einde van het schooljaar werden behaald. Dit mede omdat voor deze momenten verwachtingen zijn geformuleerd door Treffers en Beishuizen. Om een antwoord te geven op de post-hoc vragen worden ook enige resultaten beschreven van de toetsen die in groep 5 zijn afgenomen, nadat de leerlingen gestopt waren met het werken met één van beide leerlijnen. In hoofdstuk 4 worden de resultaten met betrekking tot de cognitieve variabelen beschreven, in hoofdstuk 5 wordt hetzelfde gedaan voor de affectieve variabelen.

Halverwege het leerjaar bleek dat de Proeve-leerlingen een betere procedurele vaardigheid hadden in het oplossen van kale sommen dan Stadia-leerlingen. Dit was een verassende uitkomst, omdat de Proeve-leerlijn minder aandacht besteedt aan schriftelijk oefenen dan de Stadia-leerlijn. De Proeve-leerlijn besteedt daarentegen meer tijd aan klassikaal interactief mondeling oefenen. Een verwachte uitkomst was dat de Proeve-leerlingen halverwege het leerjaar flexibeler waren in hun proceduregebruik dan Stadia-leerlingen. Aan het einde van het schooljaar presteerden de Proeve-leerlingen nog steeds beter dan de Stadia-leerlingen met betrekking tot de procedurele beheersing van het oplossen van kale sommen. Treffers' hypothese met betrekking tot procedurele beheersing van het optellen en aftrekken tot 100 werd dan ook bevestigd. Dit was ook het geval voor zijn hypothese aangaande de strategische kennis in dit domein. De Proeve-leerlingen waren flexibeler dan de Stadia-leerlingen in het aanpassen van hun oplossingsprocedures aan de aard van de opgave. Ook konden er meer Proeve- dan Stadia-leerlingen gecategoriseerd worden als flexibel of half-flexibel.

De resultaten voor de goede en zwakke rekenaars wijzen min of meer in dezelfde richting. Halverwege het leerjaar werden er met betrekking tot procedurele vaardigheid in het domein van optellen en aftrekken tot 100 geen verschillen gevonden tussen zwakke rekenaars van de Proeve- of de Stadia-leerlijn. Met betrekking tot de strategische kennis in dit domein bleek dat zowel de goede als de rekenzwakke Proeve-leerlingen flexibeler waren in het gebruik van verschillende oplossingsprocedures dan de goede en rekenzwakke Stadia-leerlingen. Aan het einde van het leerjaar waren de prestaties van de goede en rekenzwakke Proeve-leerlingen op de tempo-toetsen beter dan die van de goede en rekenzwakke Stadia-leerlingen. De goede en rekenzwakke Proeve-leerlingen waren ook flexibeler in het aanpassen van hun oplossingsmanier aan de getalskenmerken van de opgave. De verwachte verwarring, die zou optreden bij de rekenzwakke Proeve-leerlingen, werd niet teruggevonden in de toetsresultaten aan het einde van het schooljaar. In het algemeen kan de conclusie gelden dat de prestaties van de goede en de zwakke rekenaars overeenkomen met de verwachtingen, die door Treffers waren geformuleerd. De hypotheses van Beishuizen moeten dus worden verworpen.

Ten aanzien van de post-hoc vragen waren er geen verschillen tussen de beide leerlijnen met betrekking tot transfer naar het optellen en aftrekken tot 1000. Aan het begin van groep 5, nadat het experiment was beëindigd, bleken de Proeve-leerlingen nog steeds flexibeler te zijn dan Stadia-leerlingen in het aanpassen van hun oplossingsprocedures aan de kenmerken van de opgave. De Proeve-leerlingen lijken zich de verschillende oplossingsmanieren meer eigen te hebben gemaakt dan de Stadia-leerlingen.

Hoofdstuk 5 beschrijft de uitkomsten met betrekking tot de motivationele variabelen. Halverwege het schooljaar bleek dat de Stadia-leerlingen op het domein-specifieke niveau meer positieve gevoelens ten aanzien van het vak rekenen hebben dan Proeve-leerlingen. De Stadia-leerlingen bleken een hoger beeld van hun bekwaamheid te hebben dan de Proeve-leerlingen. Aan het einde van het leerjaar was het omgekeerde het geval. De Proeve-leerlingen stonden positiever ten opzichte van het vak rekenen dan de Stadia-leerlingen. Deze uitkomst was zelfs beter dan Treffers had voorspeld en daarom wordt zijn verwachting bevestigd en die van Beishuizen verworpen. Halverwege het leerjaar waren er op het taak-specifieke niveau over het algemeen geen verschillen tussen Proeve- en Stadia-leerlingen. Bij de verschillende type problemen hechten zowel de Proeve- als de Stadia-leerlingen meer waarde aan, en willen beter hun best doen, voor contextproblemen dan kale opgaven. Proeve-leerlingen gaven aan meer zelfvertrouwen te hebben ten aanzien van het oplossen van kale sommen dan het oplossen van context opgaven. Deze uitkomst is niet in overeenstemming met Treffers' verwachting. Hij voorspelde dat er geen verschil zou zijn tussen cognitieve en affectieve verwachtingen ten aanzien van kale en contextopgaven. De door Beishuizen geformuleerde verwachting op het taak-specifieke niveau wordt bevestigd: de Stadia-leerlingen hebben een sterkere voorkeur voor kale dan voor contextopgaven. Aan het einde van het schooljaar bleek dat de Proeve-leerlingen hoger scoorden dan de Stadia-leerlingen op het gebied van zelfvertrouwen en in mindere mate op het gebied van taakattractiviteit ten aanzien van kale en contextopgaven. Daarmee moet Beishuizen's hypothese ten aanzien van de resulaten aan het einde van het schooljaar worden verworpen en Treffers' hypothese worden geaccepteerd. De Proeve-leerlingen hebben meer zelfvertrouwen ten aanzien van het vak rekenen gekregen en ook hebben ze meer zin in het oplossen van zowel kale als contextopgaven dan Stadia-leerlingen.

Ten aanzien van de post-hoc vragen over de cognitieve en affectieve processen op domein- en taakspecifiek niveau kunnen we zeggen dat de grootste verschillen gevonden worden aan het einde van het schooljaar. De resultaten voor de goede en zwakke rekenaars wijzen in dezelfde richting: rekenzwakke Proeve-leerlingen zeggen meer zelfvertrouwen te hebben ten aanzien van het vak rekenen dan rekenzwakke Stadia-leerlingen. Deze uitslag ondersteunt de opvatting dat rekenzwakke Proeve-leerlingen geen nadeel hebben ondervonden van de snelle didactische opbouw van de Proeve-leerlijn. Een meer gestructureerde benadering van het rekenwiskunde onderwijs, zoals vormgegeven in de Stadia-leerlijn, hoeft bij leerlingen niet noodzakelijkerwijs te resulteren in een hoger beeld van bekwaamheid ten aanzien van het vak rekenen, zelfs niet bij de zwakkere rekenaars.

Het laatste hoofdstuk bespreekt de resultaten en geeft aanbevelingen voor het onderwijs en voor verder onderzoek. De Proeve-leerlijn bleek succesvoller te zijn dan de Stadia-leerlijn in het behalen van de voorafgestelde doelen van het onderzoek. De Proeve-leerlingen presteerden op het gebied van optellen en aftrekken tot 100, zowel op procedureel als strategisch niveau, beter dan de Stadia-leerlingen. In dit verband waren de prestaties van de rekenzwakke Proeve-leerlingen het meest overtuigend: zowel hun procedurele als strategische vaardigheid was hoger dan die van de zwakke Stadia-leerlingen. Dit resultaat wordt waarschijnlijk voor een groot deel verklaard door het feit dat het realistisch principe van wiskunde als activiteit meer gestalte heeft gekregen binnen de Proeve- dan binnen de Stadia-leerlijn. Ook de wijze waarop de (lege) getallenlijn werd gebruikt door de Proeve-leerlingen kan een verklaring zijn voor dit resultaat. Het lijkt erop dat niet alle Stadia-leerlingen de lege getallenlijn als mentaal model zijn gaan gebruiken. Deze twee factoren kunnen mogelijk ook een verklaring vormen voor de goede resultaten van de rekenzwakke Proeve-leerlingen. Deze leerlingen zijn de (lege) getallenlijn gaan zien als een natuurlijk en transparant model waarop zij konden vertrouwen wanneer zij met moeilijke problemen werden geconfronteerd. Dit is slechts in mindere mate uitgekomen voor de rekenzwakke Stadia-leerlingen. Deze conclusies worden ondersteund door de antwoorden op de motivatie-vragenlijsten, die meer positief waren voor de Proeve- dan voor de Stadia-leerlingen.

Behalve het effect van de experimentele leerlijnen op de leerlingen wordt ook de rol van de leerkracht besproken. De onderwijsgevende speelt een cruciale rol in het gestalte geven aan de ideeën die in de handleiding worden beschreven. De Proeve-leerlijn blijkt succesvoller te zijn geweest in het creëeren van een interactieve manier van lesgeven dan binnen de Stadia-leerlijn. Binnen de Proeve-leerlijn moesten de leerkrachten van meet af aan rekenen-wiskunde benaderen als een activiteit. De leerkrachten die met de Stadia-leerlijn werkten moesten dit zien te bewerkstelligen gedurende de laatste drie maanden van het schooljaar. Deze periode was daar wellicht te kort voor. De Proeve-leerkrachten waren bereid om hun rekenlessen zo te geven omdat ze daarin ondersteund werden door één van de onderzoekers, die elke twee weken op bezoek kwam om de ervaringen en de toetsresultaten te bespreken (deze bezoeken vonden overigens ook plaats bij de Stadia-leerkrachten). Dit benadrukt nog eens het belang van de wijze waarop een vernieuwing binnen het onderwijs moet worden uitgevoerd en geïmplementeerd. De vernieuwing houdt niet op bij het verstrekken van nieuwe leerlingboekjes aan de leerlingen en nieuwe handleidingen aan de leerkrachten. Naast workshops en cursussen zouden de leerkrachten ook begeleid moeten worden op de werkvloer, terwijl zij met het nieuwe materiaal bezig zijn. Het onderzoek leert bovendien de capaciteiten van de leerlingen niet te onderschatten. Elke leerling kan een flexibele probleemoplosser worden door hem toe te rusten met een krachtig model als de lege getallenlijn, hem/haar onderwijs te geven door leerkrachten, die een open manier van lesgeven creëren waarin eigen oplossingen van leerlingen serieus worden genomen, en leerkrachten oog te leren hebben voor zowel cognitieve als motivationele aspecten van leren.

Curriculum Vitae

Ton Klein werd op 24 augustus 1965 te Harderwijk geboren. Hij behaalde in 1983 zijn VWO- diploma aan het Christelijk College Nassau-Veluwe te Harderwijk. In datzelfde jaar begon hij aan zijn studie Psychologie aan de Vrije Universiteit in Amsterdam. Tijdens zijn studie participeerde hij in het onderzoeksproject "Training van aanpak en oplossing van rekenopgaven volgens een genetisch model". Het doctoraal examen psychonomie (specialisatie cognitieve psychologie), met als uitgebreid bijvak ontwikkelingspsychologie, werd in 1989 behaald aan dezelfde universiteit. Aansluitend was hij enige tijd werkzaam als invalkracht op een kinderdagverblijf. Van juni 1990 tot juni 1991 trad hij als wetenschappelijk onderzoeker in dienst bij het Rotterdams Instituut voor Sociologisch en Bestuurskundig Onderzoek (RISBO) van de Erasmus Universiteit te Rotterdam. Daar werkte hij mee aan een onderzoeksproject naar reken- en wiskundeproblemen bij allochtone leerlingen in het basisonderwijs (REWAL-project). Van juli 1991 tot augustus 1992 werkte hij als wetenschappelijk onderzoeker bij de vakgroep Ontwikkelings- en Onderwijspsychologie van de Rijksuniversiteit Leiden mee aan een onderzoek naar de mogelijkheden tot het gebruik van de computer in het voortgezet onderwijs (Proefstation West-Nederland). Van 1992 tot 1997 was hij als Onderzoeker in Opleiding verbonden aan de vakgroep Onderwijsstudies van dezelfde universiteit. In deze periode werd het onderzoek uitgevoerd dat in dit proefschrift resulteerde. Gedurende de eerste twee jaren van dit onderzoek bleef hij ook voor één dag per week als docent/onderzoeker verbonden aan de vakgroep Ontwikkelings- en Onderwijspsychologie. Vanaf oktober 1996 tot januari 1998 is hij eerst 2,5 dag en later 4 dagen per week gaan werken als schoolbegeleider bij het Centrum Educatieve Dienstverlening te Rotterdam. Vanaf 1998 zal hij in dezelfde functie werkzaam zijn in dienst van het eilandgebied Curaçao, alwaar hij mee zal werken aan de implementatie van realistisch rekenonderwijs op de Curaçaose basisscholen.