# Prototypes of Computer-Assisted Instruction for Arithmetic Word-Problem Solving

*A training study on improving the ability of educable mentally retarded children to solve simple addition and subtraction word problems*

*Monique W.M. Jaspers*

# Prototypes
# of
# Computer-Assisted Instruction
# for
# Arithmetic Word-Problem Solving

## *A training study on improving the ability of educable mentally retarded children to solve simple addition and subtraction word problems*

***Monique W. M. Jaspers***

# Prototypes of Computer-Assisted Instruction for Arithmetic Word-Problem Solving

## *A training study on improving the ability of educable mentally retarded children to solve simple addition and subtraction word problems*

CIP-gegevens Koninklijke Bibliotheek, Den Haag

Jaspers, Monique Wilhelmina Maria

Prototypes of Computer-Assisted Instruction for Arithmetic Word-Problem Solving: A training study on improving the ability of educable mentally retarded children to solve simple addition and subtraction word problems/
Monique Wilhelmina Maria Jaspers
Nijmegen: Department of Special Education, University of Nijmegen.
Thesis Nijmegen- With References- With Dutch Summary
ISBN 90-9003-976-7
Subject headings: arithmetic word-problem solving / computer-assisted instruction / educable mentally retarded children

ISBN 90-9003-976-7

# Prototypes of Computer-Assisted Instruction for Arithmetic Word-Problem Solving

## *A training study on improving the ability of educable mentally retarded children to solve simple addition and subtraction word problems*

Een wetenschappelijke proeve op het gebied van de Sociale Wetenschappen

Proefschrift
ter verkrijging van de graad van doctor
aan de Katholieke Universiteit te Nijmegen,
volgens besluit van het college van decanen
in het openbaar te verdedigen
op maandag 22 april 1991,
des namiddags te 1.30 precies

door

Monique Wilhelmina Maria Jaspers

geboren 1 december 1959
te Amsterdam

Nijmegen 1991

Promotor: Prof. Dr. J. J. Dumont

Co-promotor: Dr. E. C. D. M. Van Lieshout

# Acknowledgements

In developing the ideas expressed in this thesis, I was lucky to have encountered other researchers who refused to believe in them, students who refused to pretend to believe in them and children who refused to behave in accordance with them.
Facing these challenges provided the most satisfying moments of this research.

Ernest van Lieshout made me share his enthusiasm for this topic during the entire time I worked on this project. His continual testing of my ideas led to stimulating discussions, from which I have benefited a lot. Also, his independent proposals on how to grapple with various problems were sometimes a great help. He made me the researcher I am now.

Joep Dumont gave me free rein to execute the project.

Joris Borst Pauwels did a wonderful job in implementing the greater part of the software for the computerized training-procedures and in converting my own poor software products into true software.

Denise Dankers, Saskia van Dongen, Christine Halferkamps, José van Erp, Hans Hinssen, Godelieve van den Boogaard, Mirjam Grüter, Yvonne Reusser and Deana de Zwart deserve thanks for their work as trainers in the schools.

The staff and pupils of the schools in which the field work was carried out cannot be thanked enough for their unfailing support and co-operation: the St. Jozefschool and the Zonnegaard school in Nijmegen, the Wethouder Bakkerschool and the Willibrordusschool in Arnhem, the Vredesschool and the Ewaldenschool in Druten, the Rotonde in Didam and the Koningin Julianaschool in Zaltbommel.

Evert Hoeksma did an excellent job in correcting my English (and removing my Dutchisms).

Egbert was definitely convinced (and I almost) that I would have finished this job equally well without his help. Yet, both his stimulating attitude and ability to put things into perspective gave me the emotional support I needed now and then.

Finally, I owe special thanks to my parents for their implicit faith in me.

Monique Jaspers, January 1991.

Paranymphs: Evert Hoeksma
Rob Jaspers

# Contents

# Introduction

Some children's learning difficulties consistently challenge the skills of special educators as well as psychologists. Often, these children seem to have the ability to achieve, but nonetheless their academic performance as compared to their normally achieving peers is low. They seem to require instructional designs geared to their individual needs. In this context, the microcomputer revolution has exercised the minds of some educators and psychologists. However, the focus of attention of educators and psychologists has not seemed to coincide. Many educators have concentrated almost exclusively on hardware equipment and software already available, whereas psychologists apply themselves to research on learning and try to unravel principles that may eventually be used effectively in (computerized) instruction and remediation. In revealing those principles psychologists have started to use the microcomputer not merely as instruction medium and recording medium for the final instruction programs produced but also as an instrument in the experimental stage. Since the computer is extremely well suited to record the intermediate responses of pupils, detailed information relating to these pupils' problem solving activities can be gathered. This may eventually yield a deeper insight in learning development. Besides, the recording of all intervening products of the ongoing problem-solving process provides insight as to how and which inadequate problem-solving processes could be remediated. As such, the construction of prototypes, which precedes the actual development of courseware, permits the development of theories on the intervening processes which are crucial for adequate problem solving. This thesis is an example of research on prototype development intended to construct arithmetic word-problem solving instruction programs for children with learning difficulties. The reported research principally aims at identifying instructional variables that may improve the arithmetic word-problem solving ability of these children. The main focus will therefore be on fundamental research aiming both at the construction and evaluation of prototypes of computerized instruction and at theory development in the field of arithmetic word-problem solving. In the first chapter, recent research on arithmetic word-problem solving will be described. In particular, both research with the major aim to reveal children's word problem-solving processes and research on integrating these data within the framework of theoretical and computer-simulation models will be discussed. Finally, the practical implications of both the empirical studies on children's problem-solving strategies and the simulation models pertaining to developing prototypes for teaching arithmetic word-problem solving to children with learning difficulties will be discussed.

In chapters 2, 3 and 4, pilot studies will be presented of the training procedures based on these empirical and theoretical insights. These chapters concern both the pilot work with human trainers (chapter 2) and with the computerized prototypes for instructing arithmetic word-problem solving (chapters 3 and 4). The pilot work with human trainers preceded the development and effect studies with the computerized prototypes. This research aimed at revealing potential omissions in the description of the instruction procedure, which could then be repaired in subsequent research with the prototypes.

Eventually, three training prototypes for arithmetic word-problem solving were developed. To examine whether the instruction components incorporated in the training prototypes would have a differential effect on particular problem-solving processes, an experiment was conducted in which these three different prototypes of computerized instruction were evaluated and compared. This experiment is reported in chapter 5.

Chapter 6 describes a study in which the usefulness of a product-oriented approach to disclose the knowledge level and misconceptions of children as regards arithmetic word-problem solving is evaluated. This study serves as a starting point to develop Intelligent Tutorial Systems (ITS) for arithmetic word-problem solving, in which the expertise level and particular difficulties of the child direct the specific remedial path to be followed.

Finally, the conclusions and implications for future research on arithmetic word-problem solving are reviewed in chapter 7.

Chapters 2, 3, 4 and 6 of this thesis are revised versions of articles that have been submitted for publication (chapters 2, 3 and 4 with Dr. E. C. D. M. Van Lieshout as second author). To minimize the redundancy in these chapters, the introductions of these articles were altered. The order in which these articles are presented reflects their chronology.

# Chapter 1

# From Theory on Arithmetic Word-Problem Solving Towards Instructional Design

There is clear evidence that in particular children with learning difficulties experience problems in solving arithmetic word problems. Even when these children are equated with nonretarded children on both mental age and computational ability, their performance in solving word problems is markedly inferior to that of normally achieving children (Bilsky & Judd, 1986; Cruickshank, 1948; Russel & Ginsburg, 1984). Probably, the ability of children with learning difficulties to apply computational skills in solving word problems is more impaired in comparison with regular school children. Although it is still uncertain which factors cause this discrepancy, it has been hypothesized that comprehension factors may account for this contrasting performance (Goodstein, Cawley, Gordon, & Helfgott, 1971; Bilsky & Judd, 1986). In view of these facts, a major issue to be addressed concerns the function of word problems in the mathematical curriculum. If word problems prove extremely difficult to solve for children with learning difficulties, why take the trouble to teach these children arithmetic word-problem solving? This issue will be discussed in the next section.

## The Value of Arithmetic Word Problems

A prominent role of arithmetic word-problem solving pertains to the potential value of arithmetic word problems as representations of real-world situations (Resnick & Ford, 1981, p. 84). An example of a word problem that is often used in school is:

Peter had 8 apples.
Peter lost 2 apples.
How many apples does Peter have left?

Such word problems offer a context in which children can apply formal arithmetical knowledge concerning addition and subtraction. As such, instead of simply executing learned algorithms of adding and subtracting numbers, arithmetic word-problem solving may give meaning to these arithmetical operations. Yet, the transfer value of solving word problems to daily-life situations can be questioned if the word-problem curriculum contains context-impoverished word problems, which in fact have been used very often and of which the word problem cited is a good example. Presumably, enriched story contexts would better serve this function. In this context, Treffers and Goffree (1985) distinguish between traditional-school word problems and context problems. In contrast with traditional word problems, context problems are not formulated according to a stereotyped text frame and the situations described are supposed to be more attractive to children. Besides, the stereotyped text frame of traditional word problems often requires that the child is aware of the implicit assumptions with regard to the situations described that sometimes is contradicted by real-world knowledge. Activating real-world knowledge may therefore be helpful in finding solutions for context problems, while using this same knowledge for solving traditional word problems may sometimes impede the solving process. However, as a starting point for understanding and revealing the nature of the problem-solving processes which are involved during word-problem solving, the use of standard problem types seems preferable to enriched word problems for research purposes.

Another merit of a word-problem curriculum may be the potential both to enhance the informal strategies, which children develop naturally, and to relate these informal insights with formal mathematical knowledge. Recent research on children's preschool strategies demonstrated that children appear to understand the additive or subtractive nature described in word problems intuitively. By using materials or counting strategies, these children spontaneously represent and solve simple word problems. Though these children do experience considerable difficulties in associating their informal solution strategy to a formal number sentence, as can be inferred from the fact that children may fail to write down an appropriate number sentence for a word problem which they already solved by using informal knowledge (Carpenter, Hiebert, & Moser, 1983; Verschaffel, 1984; Lindvall & Ibarra, 1980). However, in most current mathematical programs these informal insights are not used as a starting point in teaching mathematics. Instead the design of many elementary mathematical programs in schools for Special Education primarily aims at teaching basic computational procedures such as adding and subtracting. These instructional designs discourage children to use their informal strategies based on understanding the situations described in word problems. As a consequence of incoherence between the formal mathematical concepts instructed and children's initially developed informal insights, children may adopt less effective methods for solving word problems, such as a superficial so-called "key-word " strategy by which the child determines the arithmetic operation merely on one word presented in the verbal text (e.g. "lost" means "I have to subtract"). By introducing word problems earlier in the curriculum, children may learn to concretely represent the actions and relations described and subsequently to symbolically represent these mathematical relations with a number sentence that reflects the problem structure (Moser & Carpenter, 1982; Bebout, 1990). A curriculum based on an earlier introduction of word problems may aid in connecting formal mathematics to children's informal intuition of arithmetical operations. This could result in a deeper insight in the symbolism of addition and subtraction sentences.

Thus, solving word problems may be important for both understanding and applying formal mathematical concepts which in turn may be useful in real-life situations. Especially children with learning difficulties may benefit from a word problem curriculum, since their notorious mathematical arrearage compared to normally achieving children may emanate from their ignorance of the practical usefulness of mathematical concepts. Word problems may at least offer them the opportunity to understand and apply mathematical knowledge in useful situations they will encounter in the future, such as shopping. In view of the inferior word-problem solving performance of children with learning difficulties, special remediation programs to teach these children arithmetic word-problem solving could be developed. In order to frame such remediation programs, knowledge of "external" task characteristics, such as problem structure, as well as the mental processes and cognitive structures involved in arithmetic word-problem solving is required. In the following sections, important research both on these task characteristics and on arithmetic word problem-solving processes will be reviewed.

## Research on Arithmetic Word-Problem Solving

During the seventies, research on arithmetic word-problem solving was mainly devoted to the analyses of surface characteristics of problem statements as factors influencing problem difficulty (Suppes, Loftus, & Jerman, 1969; Jerman & Van

Rees, 1972; Jerman & Mirman, 1974). Although regression analyses revealed several variables (e.g. problem length and grammatical complexity) that could account for a large proportion of variance in problem difficulty (Loftus & Suppes, 1972), other factors seemed to account for perceived problem difficulty, such as whether the operation required to solve the word problem is suggested by so-called "key words" in the problem text or whether materials are available to solve the problem. For that reason, research in the domain of arithmetic word-problem solving recently has focused on the influence of the semantic problem structure on children's solution processes (Carpenter & Moser, 1982, 1984; Carpenter et al. 1981, 1983; De Corte & Verschaffel, 1981, 1985, 1986, 1987) and on developing explicit models of the internal processes and cognitive structures which come in during arithmetic word-problem solving by building computer-simulation models (Riley, Greeno, & Heller, 1983; Briars & Larkin, 1984; Fletcher, 1985; Dellarosa, 1986). From these studies it becomes clear that understanding the quantitative relations or actions, which characterize the situation described, is most important in finding correct solutions for these problems. Nowadays, word problems are mainly classified by these quantitative relations and quantity-altering actions described in the problem text. The features by which problem types are distinguished will be specified in the next section.

## Classifying arithmetic word-problem types

From an analysis of the semantic structure of word problems by Heller and Greeno (1978), four distinct main categories emerged: change, combine, compare and equalize problems. Since this distinction in categories has been used by most researchers in categorizing word-problem types, the features of each of these main problem types will be described.

*Change* problems are problems that describe a quantitative change due to an action or event, e.g.: "John had 4 marbles. John got 2 more marbles. How many marbles does John have now?" In this problem category addition and subtraction are described as actions that cause an increase or decrease in some quantity. In contrast, both combine and compare problem-categories describe static relations between quantities.

*Combine* problems may involve two distinct quantities which have to be combined in order to determine the total amount of the two quantities or the total quantity is known and the value of one of the parts has to be determined, e.g.: "John has 3 marbles. Mary has 5 marbles. How many marbles do John and Mary have together?"

*Compare* problems refer to a static situation as well. However, this time two quantities are to be compared, e.g.: "Mary has 6 marbles. John has 4 marbles. How many more marbles does Mary have than John?"

A synthesis of the change and compare category produces *equalize* problems, in which a quantity has to be changed so as to be equal to the other quantity, e.g.: "Mary has 6 marbles. John has 4 marbles. How many more marbles does John need to have as many marbles as Mary?"

Within each main category, different problems can be formed by varying the nature of the unknown quantity. In change problems, the start quantity, change quantity or result quantity can be found from the given amount of the other two quantities. Furthermore, the direction of change may be an increase or a decrease. By varying both the nature of the action and the nature of the unknown quantity, six kinds of change problems can be constructed. A similar set of word problems can be created for compare problems for the comparison of sets may involve "more" or "less" and the unknown quantity may be one of the quantities or the amount of difference between two quantities. Although for equalize problems six variations may likewise be created, the description of the unknown is usually restricted to the difference between the two quantities. Only two variations are possible for combine problems, since for this problem type the unknown may be either the combined set or one of the subsets. Yet, while introducing the sets in chronological order seems more natural for change problems and describing the comparison of sets only after introducing the reference set seems rational for compare problems, the order in which the problem statements are introduced in combine problems is not constrained to describing one of the subsets first. Alternatively, the combined quantity may be introduced first without producing an artificial problem text: "Together Mary and John have 8 marbles. Mary has 3 marbles. How many marbles does John have?" Examples of each of these problems are given in Table 1.1.

Although solving all of the cited word problems merely requires adding or subtracting the two given numbers, the problems differ with regard to relative problem difficulty. In general, compare 3 problems are more difficult than either change 1 or combine 1 problems, in all of which adding the two given numbers is

Table 1.1 Examples of Each of the Types of Word Problems

| Action-cued problems | Static-problems |
|---|---|
| Change | Combine |
| 1 John had 4 marbles. John got 3 more marbles. How many marbles does John have now? | 1 John has 3 marbles. Mary has 5 marbles. How many marbles do John and Mary have together? |
| 2 John had 6 marbles. John lost 2 marbles. How many marbles does John have left? | 2a John has 3 marbles. Mary also has some marbles. Together John and Mary have 8 marbles. How many marbles does Mary have? |
| 3 John had 4 marbles. John got some more marbles. Now John has 6 marbles. How many marbles did John get? | 2b Together John and Mary have 8 marbles. John has 3 marbles. How many marbles does Mary have? |
| 4 John had 6 marbles. John lost some marbles. Now John has 4 marbles. How many marbles did John lose? | |
| 5 John had some marbles. John got 2 more marbles. Now John has 6 marbles. How many marbles did John have first? | Compare |
| 6 John had some marbles. John lost 2 marbles. Now John has 4 marbles. How many marbles did John have first? | 1 John has 8 marbles. Mary has 5 marbles. How many more marbles does John have than Mary? |
| | 2 John has 8 marbles. Mary has 5 marbles. How many less marbles does Mary have than John? |
| Equalize | 3 John has 3 marbles. Mary has 4 marbles more than John. How many marbles does Mary have? |
| 1 John has 3 marbles. Mary has 5 marbles. How many marbles does John need to have as many marbles as Mary? | 4 John has 7 marbles. Mary has 3 marbles less than John. How many marbles does Mary have? |
| 2 John has 5 marbles. Mary has 3 marbles. How many marbles does John need to get rid of to have as many marbles as Mary? | 5 John has 8 marbles. John has 5 marbles more than Mary. How many marbles does Mary have? |
| 3 There are 5 birds. There are 3 worms. Suppose the birds race over and each one tries to get a worm. How many birds won't get a worm? (Hudson, 1983) | 6 John has 5 marbles. John has 3 marbles less than Mary. How many marbles does Mary have? |

*Note.* After Riley, Greeno and Heller (1983).

required. Combine and compare problems involving subtraction are in general more difficult than change problems 2 and 4 (Carpenter et al., 1981; Riley et al., 1983;

Ibarra & Lindvall, 1979; Vergnaud, 1981). These findings suggest that solving word problems entails more than just good mathematical ability. Indeed, investigations of the origins of wrong answers to word problems have shown that a majority of incorrect answers are due to misconceptions instead of calculation errors (De Corte & Somers, 1981; De Corte & Verschaffel, 1981). Instead of a lack of elementary mathematical knowledge, an incorrect representation of the specific features within the semantic problem structure seems to produce wrong answers.

In most current models of arithmetic word-problem solving, two major components are stressed, i.e. problem representation and problem solution (Riley et al., 1983; Briars & Larkin, 1984; De Corte, Verschaffel, & De Win, 1985; Kintsch, 1986; Fletcher, 1985; Dellarosa, 1986). Comprehending word problems requires a translation of the word-problem text into a semantic representation, which indicates the choice of the arithmetic or other solution strategy to be used. The quality of this semantic representation to a large extent determines the solution strategy used to solve the problem. Evidence for the influence of the semantic representation formed comes from studies showing that children's wrong solutions to word problems constitute "correct solutions" to incomprehended problems. In analysing errors in recalling problem texts after solution attempts, it was found that children's transformations of difficult problems into simpler ones often resulted in incorrect problem representations and that the incorrect answers they produced often were directly related to the problem representation they had built (Dellarosa, Kintsch, Weimer, & Reusser, 1986; Verschaffel, 1984). This finding strongly indicates that the solution strategy used reflects the kind of semantic representation built. A major issue is how the structure of the problem representation built, can be inferred from the solution strategy used. In the next section, this question will be addressed by discussing the findings of research on the strategies children use to solve word problems together with the basic development of strategies.

## Development of arithmetic word-problem solving ability

Longitudinal studies of children's solution processes for arithmetic word problems have produced consensus as regards the influence of semantic structure on children's use of different strategies to solve these problems (Carpenter et al., 1981, 1983; Carpenter & Moser, 1982, 1984; Hiebert, Carpenter, & Moser, 1982; De Corte & Verschaffel, 1981, 1987; Verschaffel, 1984). In these studies the relation between semantic problem structure and solution strategy was observed by having kindergarten children and, first, second and third graders model the word problems offered. These studies showed that the variety of strategies that children used to solve the problems, represent an attempt to model the semantic structure of the problems. To illustrate, consider the following problems:

John has 8 cookies. John ate 3 cookies.
How many cookies does John have left?

John has 3 cookies. John got some more cookies.
Now John has 8 cookies.
How many cookies did John get?

John has 8 cookies. Ann has 3 cookies.
How many more cookies does John have than Ann?

Although all problems can be solved by subtracting 3 from 8, in solving each problem children use different modeling strategies. To solve the first problem, most kindergarten children and first graders would construct a set of 8 cubes and subsequently take away 3 cubes. Then children would answer this problem by counting the remaining cubes ("separating-from" strategy). In contrast, the second problem would be solved by first constructing a set of 3 cubes, then adding cubes to this set until it contained a total of 8 cubes. By counting the cubes added, the answer would be found ("adding-on" strategy). Finally, the third problem would be solved by a "matching strategy", i.e. after representing both the first and second number by constructing sets of 8 and 3 cubes in a one-to-one correspondence until the smaller set is exhausted, a child would answer the problem by counting the remainder of unmatched cubes. As such, a variety of concrete modeling strategies can be distinguished.

Besides this variety in modeling strategies, both Carpenter and Moser (1984) and De Corte and Verschaffel (1987) distinguished between three levels of abstraction according to the level of internalization of strategies: (a) material strategies based on direct modeling with physical objects or fingers, (b) verbal strategies, based on the use of counting strategies and (c) mental strategies, based on recalled or derived number facts.

Table 1.2 Representational Strategies and Corresponding Verbal Counting and Mental Strategies

---

**Material strategies:**

*separating from:*
The child constructs a set of blocks corresponding to the larger number in the problem, then removes as many blocks as indicated by the smaller number. Finally, the child counts the remaining number of blocks.

*separating to:*
The child constructs a set of blocks corresponding to the larger number in the problem, then removes as many blocks until a number of blocks indicated by the smaller number is left. Finally the child counts the number of blocks that were separated.

*adding to:*
The child constructs a set corresponding to the first number, then constructs a set corresponding to the second number. Finally, the child counts all blocks.

*adding on:*
The child constructs a set corresponding to the smaller number, then adds blocks to this set until there are as many blocks as indicated by the larger number. Finally, the answer is found by counting the number of blocks added.

*matching (direct):*
The child constructs a set corresponding to the smaller number and a set corresponding to the larger number. Subsequently, the child matches them in a one-to-one correspondence until one set is exhausted. The answer is found by counting the number of blocks remaining in the unmatched part of the larger set.

*matching (indirect):*
The child constructs a set corresponding to the first number in the problem, successively the child matches blocks in a one-to-one correspondence until this set contains x blocks "less" or "more" than the first set (x refers to the second relevant number in the problem text). The answer is found by counting the number of blocks in the second set.

**Verbal-counting strategies:**

*counting all starting with first:*
The child starts counting by 1 until the first number is reached and continues this forward count as the second number is enumerated. The last number in the sequence is the answer.

*counting all starting with larger:*
The child starts counting by 1 until the larger of the two numbers is reached and continues counting forwards as the smaller number is enumerated. The last number in the sequence is the answer.

*counting on from first:*
The child starts counting by the first number given and continues this forward count as the second number is enumerated. The last number in the sequence is the answer.

*counting on from larger:*
The child starts counting by the larger number and continues this forward count as the smaller number is enumerated. The last number in the sequence is the answer.

*counting down from:*
The child starts counting backwards by starting with the larger number until as many number words as indicated by the smaller number are counted. The last number in the sequence is the answer.

*counting down to:*
The child starts counting backwards by starting with the larger number until the smaller number is reached. The number of counting words in the sequence is the answer.

*counting up from given:*
The child starts counting by the smaller number and continues counting until the larger number is reached. The number of counting words in the sequence is the answer.

Table 1.2 *(continued)*

**Mental strategies:**

*known fact starting with first:*
The child retrieves an addition number fact, starting with the first number in the problem immediately from long-term memory.

*known fact starting with larger:*
The child retrieves an addition number fact, starting with the larger number in the problem immediately from long-term memory.

*derived fact starting with first:*
The child begins with the first number of the problem and uses other recalled number facts to find the solution.

*derived fact starting with larger:*
The child begins with the larger number of the problem and uses other recalled number facts to find the solution.

*direct subtractive known fact:*
The child retrieves a direct subtractive number fact with the two numbers immediately from long-term memory.

*indirect subtractive known fact:*
The child retrieves an indirect subtractive number fact with the two numbers immediately from long-term memory.

*indirect additive known fact:*
The child retrieves an indirect additive number fact with the two numbers immediately from long-term memory.

*direct subtractive derived fact:*
The child subtracts the smaller number from the larger by using other recalled subtractive number facts.

*indirect subtractive direct fact:*
The child finds the answer by determining what quantity should be subtracted from the larger number to get the smaller number by using other recalled subtractive number facts.

*indirect additive direct fact:*
The child finds the answer by determining to what quantity the smaller number should be added to obtain the larger number by using other recalled additive number facts.

*Note.* After Carpenter and Moser (1984) and De Corte and Verschaffel (1987).

For a full description of the strategies, see Table 1.2.

During subsequent interviews of the longitudinal studies children tended to solve word problems first by using mainly material strategies, then by verbal-counting strategies and finally by applying mental strategies based on number facts. This finding points at a development in the internalization level of strategies. However, despite this development in abstraction level, children at all internalization levels showed a tendency to use strategies in which the semantic structure of the problem was reflected. For example, most children at the verbal strategy level used a so-called "counting-up from given" strategy to solve change 3 problems, in which the transfer set is unknown (see second example cited). Counting-up from given starts with the first given number and continues until the second given number is reached. The answer is determined by the number of counting words in the sequence. This strategy therefore requires some method of keeping track ("double count") of the number of counting steps that represent the added number. The "counting-up from given" strategy is the verbal-counting correlate of the material "adding-on" strategy described earlier. Accordingly, the semantic structure of word problems even influences the strategy employed by children operating at the mental level. For instance, questioning by the interviewer revealed that change 3 problems (second example cited) were most often solved by an "indirect additive known-fact strategy" ($3+x=8$) instead of a "direct subtractive known-fact strategy" ($8-3=x$). In fact, with the exception of the matching strategy, each of the material strategies has a verbal and mental correlate. As such, the material "separating-from" and "separating-to" strategies (see Table 1.2) do have verbal-counting parallels in "counting-down from" and "counting-down to". According to Carpenter and Moser (1982, 1984), change 2 problems are most clearly modeled by "separating-from" and "counting-down from" strategies, whereas "separating-to" and "counting-down to" are best in representing change 4 problems (see Table 1.1 and 1.2). The mental correlate of "separating-from" is the "direct subtractive known-fact strategy" in which the answer is found by subtracting the smaller number from the larger number. Likewise, the mental parallel for "separating to" is the "indirect subtractive known-fact strategy" in which the unknown number is found by determining what quantity has to be subtracted from the larger number in order to get the smaller number.

Although at each internalization level the choice of strategy is mainly determined by semantic structure, the effect of semantic structure on children's choice of strategy becomes less straightforward at each abstraction level. Some verbal-counting strategies are more efficient than material strategies in which all sets are directly modeled. In the material "adding-on" strategy each of the addends is modeled whereas in its verbal "counting-up from given" correlate the child begins counting forward instantly starting with the first addend of the problem. Besides, for some problem types children learn to exchange the "counting-on from first" strategy, in which the counting sequence starts with the first number given in the problem regardless of its cardinality, for the more efficient "counting-on from larger" strategy, in which the child starts counting with the larger of the two addends. By disregarding the order of introduction of sets in the problem text and starting with the larger of the two given numbers, the child reduces the number of "double count" steps and thus the processing load. Children at the mental level also tend to use other number facts to derive solutions for number combinations for which they have no long-term memory trace. For example, children may use their knowledge of 6+6=12 and subsequently subtract 12 by 2 in order to arrive at the solution for 4+6 ("derived-fact starting with larger").

At first sight, this seems to denote that in finding solutions to word problems children become less dependent on semantic structure by the time they use more efficient counting or mental strategies. However, although interchanging solution strategies requires reorganizing the primary problem representation, De Corte and Verschaffel argued (1985) that this rearrangement of sets and set-relations is not equally demanding for all problem types. For instance, interchanging sets that fulfill identical roles in the problem representation seems less profound than switching sets that serve a different function. In line with this argument, they hypothesized that children will be more willing to exchange a "counting-on from first" strategy for a "counting-on from larger" strategy for combine problems, in which the two subsets can be interchanged without affecting the problem representation than for change problems, in which all sets play a distinct part in the problem representation. De Corte and Verschaffel (1987) indeed found that children intended to interchange sets for combine 1 problems in which the second number given was the largest, whereas children were less apt to reverse the roles of sets for change 1 problems in this case. De Corte and Verschaffel argued that the dynamic nature of change 1 problems is so compelling that it triggers solution strategies that model the chronological sequence of events. In contrast, since combine 1 problems describe no implicit action, the order of introduction of the two given subsets seems less constraining in eliciting solution strategies that follow the sequence of introduction of the subsets.

In summary, it can be said that although other factors besides problem structure may influence the solution process of word problems, in choosing an appropriate solution strategy children notably tend to focus on the semantic structure of a word problem. Despite the obvious development in the level of internalization of the strategies, children continue to adhere to the semantic structure in finding solutions for word problems.

### Text-Analysis processes

In addition to studying modeling, verbal and mental strategies, De Corte and Verschaffel (1986) reported an investigation of eye-movement recording for studying text-analysis processes that accompany and contribute to creating problem representations. In this investigation, both gaze durations and sequence of fixations were analysed for different parts of the problem text. The analysis of gaze durations per area showed that high ability children generally spend more time in the question areas and parts of the problem text containing important semantic information (i.e. "more than" or "less than" in compare problems) than low ability children. The sequence of fixations in reading the problem text revealed a relationship between problem difficulty and reading behavior. Difficult problems elicited more rereadings than easy problems. (This finding was replicated in a study in which reading behavior during arithmetic word-problem solving was recorded by monitoring the words that a child selected on a touchscreen (Van Lieshout & Jaspers, 1989).) Furthermore, level of ability corresponded to different patterns of rereading. Whereas high ability children frequently reviewed words and even whole sentences, the rereadings of low ability children consisted almost exclusively in jumping back and forth to the numbers in the problem text. De Corte and Verschaffel concluded (1986) that semantic processing indeed is crucial for skilled arithmetic word-problem solving and that failures in solving word problems are due to superficial analysis of the problems.

Their empirical studies resulted in designing a model of competent word-problem solving. According to their model, while reading the problem text, a competent problem solver constructs an internal representation of the semantic structure of the problem, in which sets and relations between sets are reflected. This representation encompasses an unknown set for which the quantity has to be determined. Starting from this representation, this unknown set is calculated by selecting and executing an appropriate counting strategy. Finally, the answer is given and checked. As will be explained in the next section, the knowledge implemented in computer-simulation models for word-problem solving fits in with this hypothetical model.

## Theoretical Models of Arithmetic Word-Problem Solving

How problem solvers may learn to construct different semantic representations for the different word-problem types is a second matter of interest. In an attempt to specify internal processes and knowledge structures that come in during arithmetic word-problem solving, besides researchers involved in empirical studies on the arithmetic word-problem solving activities of children, other researchers aimed at constructing simulation models. Prevailing models for arithmetic word-problem solving will be outlined in the next sections.

### The arithmetic word-problem solving model of Riley, Greeno and Heller

Riley et al. (1983) distinguish between three kinds of knowledge structures involved in problem solving: (a) problem schemata for understanding the essential components and various semantic relations underlying word problems, (b) action schemata, which refer to knowledge about actions involved in finding the unknown quantity and (c) strategic knowledge for planning solutions to problems. Successful problem solving requires knowledge of problem schemata to represent the particular problem situation in an organized structure consisting of elements and relations between these elements. These structures have the form of semantic networks. Three

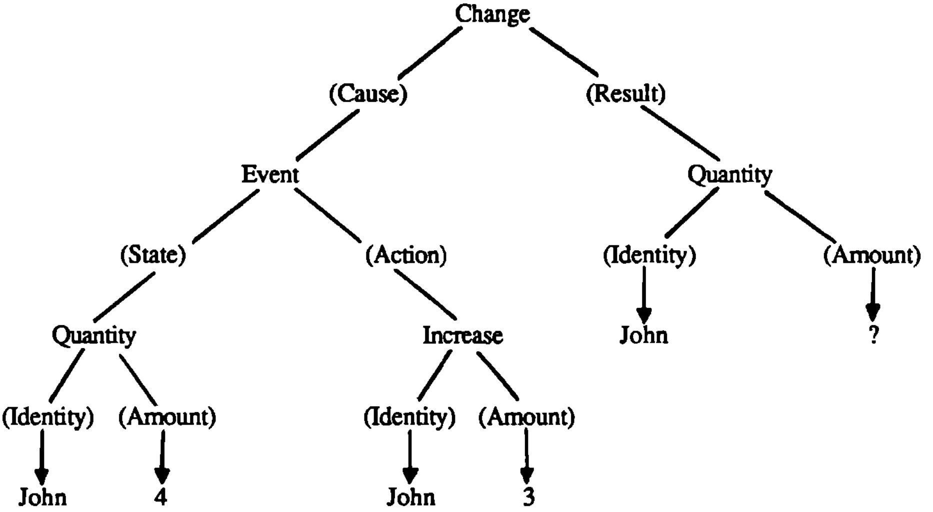


Figure 1.1 Schematized representation of the Riley et al. model for the change 1 problem "John had 4 marbles. John got 3 more marbles. How many marbles does John have now?" After Riley et al. (1983).

main types of problem schemata, so-called superschemata, are proposed for understanding change, combine and compare problems. The process of constructing a problem representation involves mapping the verbal statements onto the problem schema. In outlining the problem schema, the specific quantities described in the text are assigned to slots or elements of the schema structure. For example, Figure 1.1 shows a representation of a change 1 problem.

As can be seen from Figure 1.1, the representation contains three main components; i.e. an initial quantity which represents the start set of John's 4 marbles, some event that causes an increase in the amount of the start set by 3 marbles, and a result set, which in this case is represented as the unknown quantity of which the amount has to be determined. These components are built in accordance to the verbal information received. When sentences containing words such as "gave" "lost" or "won" are encountered, the model infers that the problem is about sets which change in quantity and consequently the change superschema is activated. Thus, the semantic structure described is mapped onto one of the superschemata, change, combine and compare. In doing so, the unknown and known quantities are assigned to different slots given their particular role in the problem text.

Once the problem has been represented, action schemata are used to bridge the gap between the constructed internal representation and the solution strategy employed. Action schemata are organized into different levels of complexity ranging from constructing and counting concrete sets to directly adding and subtracting numbers.

Finally, strategic knowledge is used to choose and carry out a relevant solution plan by working out the solution from the top down. Strategic knowledge includes knowledge of subgoals that are useful in achieving a plan. After the model has selected a plan, it tries to carry out the actions associated with that plan in an attempt to solve the problem. In carrying out the plan, new subgoals replace old subgoals. This process continues until the problem is solved.

Riley et al. distinguish three levels of expertise. At the first level the model merely understands set relations by means of a simple schema for representing the sets described externally with blocks. At the second level set-subset relations can also be established internally. At the third level the model can use its superschemata in a top down manner to build a mental representation of the entire problem before actually solving it. According to this model, children's difficulties in solving some problems are accounted for primarily by differences in the complexity of the superschemata available. For a complete discussion of the model, see Riley et al. (1983).

## The arithmetic word-problem solving model of Briars and Larkin

Whereas the Riley et al. model is a schema-based model, the model CHIPS (Concrete Human-like Inferential Problem Solver) of Briars and Larkin (1984) is a propositional model. Instead of mapping verbal statements directly onto an already available schema, it sets up relatively simple propositions which correspond to the statements of the problem text. Each of these propositions includes cues that cause the model to perform actions. Change problems involving actions can provide special cues for adding and separating chips from a source supply of chips. As such, CHIPS interprets phrases as "gave him" "won" etc. as a cue to make a schema for moving CHIPS from the source into an already existing start set and phrases as "lost" and "gave away" as a cue to create a schema for moving chips out of an already existing start set. This so-called "single-role counters" knowledge corresponds to the first level of expertise of the Riley et al. model. In Figure 1.2 the complete output of CHIPS is presented for the change 1 problem "John had 4 marbles. John got 3 more marbles. How many marbles does John have now?"

In addition to "single-role counters" knowledge, CHIPS uses knowledge concerning "double-role counters" that enables CHIPS to keep track conceptually of the double-role of chips which belong to two different sets. The second knowledge level of the Riley et al. model corresponds to this "double-role counters" knowledge. Also, CHIPS makes use of "rerepresentation" knowledge by which CHIPS can recognize that actions like combining and separating can be reversed in time ("time-reversal" knowledge) or that subsets can be interchanged to produce equivalent situations ("subset-equivalence" knowledge). This knowledge broadly resembles the third knowledge level of the Riley et al. model.

Combine and compare problems involve static relations between sets. In order to use a concrete representation to solve these problems, the situation described in these problems needs to be translated into actions that CHIPS can execute with counters. This is done by interpreting words as "together" "more than" and "less than" as an implicit action cue to move counters into or out of a set or as an implicit cue to match counters one-by-one in order to determine the answer set. Although both the model of Riley et al. and CHIPS can solve the same word-problem types,

| | |
|---|---|
| (problem statement) | John |
| (problem statement) | had |
| (model's inference) | There is a set with cardinality |
| (problem statement) | 4 |
| (production rule activated) | Move chips until a set of 4 chips is created |
| | 1 chip has been moved |
| | 2 chips have been moved |
| | 3 chips have been moved |
| | 4 chips have been moved |
| | enough chips have been moved |
| (present knowledge) | Now 4 chips in set |
| (problem statement) | marbles |
| (present knowledge) | set of marbles |
| (problem statement) | John |
| (problem statement) | got |
| (inference) | increase |
| (problem statement) | 3 |
| (production rule activated) | Add chips until 3 chips are moved |
| | 1 chip has been moved |
| | 2 chips have been moved |
| | 3 chips have been moved |
| | enough chips have been moved |
| (problem statement) | more |
| (present knowledge) | more marbles |
| (problem statement) | marbles |
| (present knowledge) | set of marbles |
| (problem statement) | How many |
| (problem statement) | marbles |
| (production rule activated) | count set of marbles |
| | 1, 2, 3, 4, 5, 6, 7 |
| (present knowledge) | set counted 7 |

Figure 1.2 Schematized representation and output of the CHIPS' model for the change 1 problem "John had 4 marbles. John got 3 more marbles. How many marbles does John have now?" using single-role counters. After Briars and Larkin (1984).

there is a striking difference. While the Riley et al. model represents the situation described by using its internal knowledge of superschemata, CHIPS does not possess superschemata but instead directly represents sets with counters after processing text propositions. Despite these differences in internal knowledge structures, both the model of Briars and Larkin (1984) and Riley et al. (1983) hypothesize three similar levels of competence. At each competence level, both models solve similar problem types and produce equivalent responses.

## The arithmetic word-problem solving model of Kintsch, Greeno and Dellarosa

The simulation ARITHPRO (Dellarosa, 1986) is based on a model of children's problem-solving processes by Kintsch and Greeno (1985) and simulates the performance of third grade children only. Thus, ARITHPRO as opposed to both the model of Riley et al. and CHIPS does not simulate different expertise levels and consequently the presumed development of skill in solving word problems. In accordance with CHIPS, ARITHPRO represents a propositional model of arithmetic word-problem solving. However, in ARITHPRO text-comprehension processes are more emphasized than in CHIPS. ARITHPRO solves word problems by an interaction of text-comprehension processes and arithmetic solution strategies. In solving word problems, ARITHPRO first comprehends the story by building proposition frames which represent the story's text base. This text base is a mental representation of the semantic content of a text. Numeric information in the text is used to build representations of sets, called set frames. The relations between sets are nested in superschemata, which are larger set frames that have whole sets as their components. There are three types of superschemata, which are comparable to the superschemata of Riley et al.: a transfer superschema to represent change problems, a superset schema to represent combine problems and a compare superschema to represent compare problems. For example, the transfer superschema identifies a start set, a set of objects transferred into or out of the start set and a set representing the result of the transfer. In accordance with the Riley et al. model, sets are likewise assigned to slots in these superschemata. Figure 1.3 illustrates the complete sets created during processing of the first two sentences of a transfer problem type (i.e. a change 1 problem). For example, if the text base contains the information that "Tom gave John 3 marbles more" and a set belonging to John already exists, this set is labeled "start set" and subsequently a set with cardinality 3 is created and labeled "transfer set".

Thus, in accordance with the model of Riley et al. ARITHPRO maps the semantic structure described onto a superschema and subsequently selects a solution strategy to solve the problem.

---

```
[TransferInset
 (STARTSET
     (SET 1 (QUANTITY : 4)
              (OBJECTS : MARBLE)
               (ROLE : PRIORSET)
               (SPECIFICATION
                    (OWNER : X)
                    (TIME : PAST))))

(TRANSFERSET
(SET2 (QUANTITY : 3)
      (OBJECTS : MARBLE)
      (ROLE : TRANSFERSET)
      (SPECIFICATION
                    (OWNER : X)
                    (TIME : PAST, AFTER SET1)))) ]
```

---

Figure 1.3 The set structures constructed by ARITHPRO after processing the first two sentences of the change 1 problem "John had 4 marbles. Tom gave John 3 more marbles. How many marbles does John have now?" After Dellarosa (1986).

## The psychological validity of the models

One of the main pleas for developing computer-simulation models bears upon the possibility to test these models as viable models of children's problem-solving processes. By comparing the output of these programs with the empirical data on the problem-solving processes of children, hypotheses as regards several aspects of the arithmetic word-problem solving process of children may be tested. Since thusfar no data concerning the goodness of fit between the simulation model ARITHPRO and empirical data of children's solving processes are available, the discussion of the psychological validity of the simulation models will be restricted to the model of Riley et al. and CHIPS.

With regard to the levels of skill of the models there is a relatively good fit between the models' predictions and the different patterns of performance of children at different grade levels. To illustrate, CHIPS (Briars & Larkin, 1984) could account for 88 percent of the variance of the data coming from eleven different research studies in which children of different grade levels were represented. Besides, the success rates observed in the experiments were very similar to those predicted by CHIPS (Briars & Larkin, 1984). The proportion of response patterns of children from grade one to three consistent with the Riley et al. model ranged from 95 to 100 percent for change problems, from 83 to 100 percent for combine problems and from 65 to 100 percent for compare problems.

Yet, several mismatches exist between the hypothesized levels of skill of the simulation models and the observed performance levels of children. For example, both models of Riley et al. and CHIPS are able to solve, change 1, 2 and 4 problems at the first expertise level, whereas change 3 problems are solved correctly only at the second level of expertise. Although these predictions are confirmed by most studies, Verschaffel (1984) found that a definite proportion of children, who were able to solve change 2 problems correctly, failed on change 1 problems. Also, Verschaffel (1984) demonstrated that the difficulty of change 5 and 6 problems is equal to change 3 problems, for which the models hypothesize a lower expertise level.

Finally, not all solution errors committed by children in solving the various word-problem types are accounted for by the models. For instance, subjects in the study of Verschaffel (1984) often gave the smallest given number as incorrect answer to change 6 problems. Though this error is also produced by CHIPS using "double-role" counters at the second level of expertise, the Riley et al. model does not produce an answer at all at the first and second level.

In summary, though both the models of Riley et al. and Briars and Larkin can account for a large proportion of empirical data concerning problem difficulty, solution strategies and errors, some predictions by the models are not confirmed by these data. For a more complete discussion of the testing of the validity of these models, we refer to Riley et al. (1983), Briars and Larkin (1984), and De Corte and Verschaffel (1988).

## Towards Developing Instructional Programs

All in all, considerable progress has been made in revealing and specifying the problem-solving processes and knowledge structures that children must have in order to solve word problems with understanding. A major issue now is what principles could be used to improve the impaired word-problem solving performance of children with learning difficulties. The leap from the research findings and simulation models described thusfar to designing instruction in arithmetic word-problem solving for these children will be presented in this section.

### From theory to instructional design

As was already mentioned in the first section, children with learning difficulties manifest inadequate word-problem solving skills. The difficulties experienced by these children in many academic skills is primarily due to the absence of adequate information processing strategies and a lack of metacognitive skills (Hall, 1980). However, research of remedial teaching of these children has shown that they are able to learn and to apply these kinds of problem-solving skills. This was accomplished by instructing and training an adequate task strategy and by making the children aware of the cognitive operations they needed (Brown, Campione, & Day, 1981; Hall, 1980). Brown et al. pointed out the necessity of a structured approach, by which these children should not only learn how to use the task strategy, but also how to monitor, check and evaluate this strategy. In the light of the fact that children with learning difficulties apparently are susceptible to training, probably poor instruction practice rather than the characteristics of children with learning difficulties can account for their deficient word-problem solving performance. For once these children have entered Special Education, they are unlikely to be confronted with word problems because mathematical instruction in these schools often primarily focuses on teaching basic computational skills such as addition and subtraction. To prevent the presentation of difficult subject matter, these skills are often taught outside a meaningful context. As long as these children have not mastered these basic skills, arithmetic word-problem solving is not instructed. Moreover, once these children are able to solve addition and subtraction sentences, only a few simple word-problem types are offered, mainly those for which the correct solution can be easily inferred on the basis of the key word described.

Consequently, these children may adopt superficial solving procedures based on the surface structure of word problems rather than the underlying semantics.

How then may children with learning difficulties be helped in acquiring the requisite knowledge essential for adequate arithmetic word-problem solving? Research findings on regular school children's solution errors and problem-solving behavior as well as computer models simulating these children's problem-solving processes prove that besides mathematical ability, procedural and conceptual knowledge is required for solving word problems adequately. Procedural knowledge pertains to knowing how to set subgoals and attaining those subgoals in order to eventually solve the problem. Conceptual knowledge refers to understanding the semantics of the problem, i.e. understanding the logical set-relations or actions described by integrating them in some sort of internal representation of the problem. This representation consists of a complex network of the known and unknown number sets and their mutual relations described in the verbal text. In order to build such a representation, the child first has to analyse the problem text carefully. Especially for children with learning difficulties, text-comprehension processes may play an important role. For these children, like unexperienced children, may depend heavily on a text-driven or bottom-up processing of the word-problem text to arrive at an adequate representation of the problem. Unlike more experienced problem solvers, their processing of the text is not conceptually driven in a top-down manner by activating one of the semantic superschemata and mapping the verbal statements onto these network structures.

Thus it would seem that one way to tackle the problem is to develop remedial instruction programs that concentrate on teaching that links the conceptual understanding of word problems to procedures that support it (see for example Resnick and Ford, 1981). The problem is then, how to increase children's understanding of the semantics defining problem structure. The first training procedure developed that pertains to instructing thorough text analysis is described in the next section.

## A training procedure for instructing text analysis

One way to improve children's understanding of the semantics defining problem structure, is to design instruction methods in which attention is paid to text-analysis processes. Since beginners lack mental models that may be activated and helpful in analysing problem structure, thorough text analysis may play an important role in building internal problem representations. As some researchers report, poor word-problem solving may be the result of poor text analysis. Poorly performing children often do not read the whole text and often do not even look at the question sentence in which as a rule the unknown number set is defined (Goodstein et al., 1971; De Corte & Verschaffel, 1986).

To remediate superficial reading behavior, a first training procedure was developed in which children learned to pay special attention to the word-problem text by pointing at specific sets of words. Decisions as to which words had to be pointed at were based on their presumed function in the problem text; i.e. words were selected on the basis that they seemed to reflect set descriptions, set-relations or actions. Additionally, children learned to pay attention to the steps of the task strategy. By using a so-called "planning list", children were instructed to plan the order of execution of the steps of the task strategy. In other words, this training procedure was aimed at remediating both superficial reading habits and weak metacognitive control (Van Lieshout, submitted). In doing so, this training procedure is expected to support the development of cognitive superschemata internally, although these schemata are not explicitly trained. Some first pilot studies with this training procedure yielded positive results, whether performed by a human trainer (Van Lieshout, 1986) or by the computer (Van Lieshout, submitted). In a later version this training procedure nevertheless underwent some minor revisions. These revisions, which proceeded from research on reading behavior of children during word-problem solving, concerned slight redefinitions of the sets of words to which the children had to pay special attention during training. Although investigations of the reading behavior of children during word-problem solving are rare, the study on children's text-analysis processes by De Corte and Verschaffel (1986) provided some information as regards which different text parts are particularly processed during arithmetic word-problem solving. The study of De Corte and Verschaffel revealed some general trends, e.g. long fixation times on numbers and specific words, such as "more", "together" and short fixation times on question areas. Also, fixation time in the question areas varied according to mathematical ability; i.e. high ability children spend more time in the question areas and reviewed words as "got", "less than" more frequently than low ability children. While this kind of research may eventually contribute to our understanding of the processes intervening in constructing problem representations, these data present only few clues for designing instruction programs to improve reading behavior.

The descriptions of the computer-simulation programs also lack details of the text-comprehension processes that occur during word-problem solving. Thus neither did these models deliver straight directions for designing instruction programs to improve poor text analysis. In fact, all simulation models have a rather primitive ability to understand language. For example, the model of Riley et al. (1983) matches complete sentences against stored "templates" for word strings and the model of Dellarosa (1986) takes a propositionalized problem text as input.

Nevertheless, it was decided to extend the sets of words that had to be paid special attention to, with words comprising information on the underlying chronological structure of change problems and the underlying semantic structure of combine problems. For instance, for combine problems the word "together" had to be touched in the revised training procedure, whereas this word was not included in the set of words in the older training procedure. These revisions resulted in a training procedure aiming at teaching children to select information in the text describing set ownership ("Peter") , set entity ("marbles"), set quantity ("8"), time sequence ("now") and the relations or actions among sets ("lost, "got", "more than", "less than", "together" etc.). Despite the primitive language abilities of the simulation models, the words that were defined as crucial in the revised training procedure for remediating poor reading habits are the words that seem (1) to activate problem schemata in these simulation models, (2) crucial in filling the slots of these schemata and (3) to define problem structure.

Although this type of training focuses the child's attention on problem-solving activities essential for constructing proper representations of word problems, it is uncertain whether the children actually will learn to construct such representations. Indeed, in a first computerized training-study children improved in identifying the relevant number sets among both relevant and irrelevant number sets, but not in formulating an adequate number sentence (Van Lieshout, submitted). In that study, however, children only were allowed to formulate canonical number sentences. Since canonical sentences often seem unsuitable for representing the semantic relations described, children may have had difficulties in translating their problem-solving route into a direct mathematical number sentence. Carpenter and Moser (1984) pointed out that young children do not consider "addition" and "subtraction" as abstract operations to be used for solving word problems. Instead these children tend to define "subtraction" and "addition" in the context of the actions and relations described ("losing", "winning", "together" etc.). Since the position of the unknown quantity described in the problem text determines whether "winning" should factually lead to adding and "losing" to subtracting the numbers given, merely using these key words in choosing the operation may result in the wrong answer. As such, a so-called key-word strategy produces correct answers on "direct" problem types (a+/-b=?) but incorrect answers on "indirect" problem types (?+/-b=c or a+/-?=c).

Concrete modeling of the actions or relations described may overcome these difficulties. By modeling indirect word problems with concrete objects, the number of objects of the unknown set may be estimated and determined by the number of the two other sets given. As a consequence, in order to solve "indirect" problem types adequately, children do not necessarily have to transform their definition of "addition" or "subtraction" into a canonical formal number sentence, which sometimes even requires inverting the mathematical operation required.

Thus, merely reading the verbal text thoroughly will not necessarily lead to a proper representation. In fact, the procedure of most investigations cited (see Carpenter et al., 1981, 1983; Carpenter & Moser, 1982, 1984; Hiebert et al., 1982; De Corte & Verschaffel, 1981, 1987; Verschaffel, 1984) involved having children individually solve word problems read out by an experimenter. To prevent memory difficulties the experimenter usually read the problems slowly and repeated them if necessary. Despite the presence and use of the materials presented to solve the problems, some children who participated in these studies nevertheless made errors that also seem to reveal incorrect representations of problem structure. The frequent occurrence of particular types of errors such as "answering with one of the numbers stated in the problem text" has been shown to be manifestations of such incorrect representations (Verschaffel, 1984). Thus some children continue to commit errors, even when materials are available. The second training procedure developed aimed at teaching children to construct adequate external representations for word problems so as to remediate their own incorrect problem representations. This training procedure will be described in the next section.

### A training procedure for instructing representational strategies

As was demonstrated by the research of De Corte and Verschaffel (1981, 1987) and Carpenter and Moser (1982, 1984) children naturally understand problem situations by acting them out physically. Even kindergarten children, without having received any formal instruction in mathematics, prove extremely sucessful in solving the easier word problems if materials are presented. Since these children lack knowledge of the symbolic system, they cannot possibly translate the problem statement in a formal number sentence. Indeed these children find the answer by modeling the actions and relations described. Various modeling strategies can be distinguished as such for the different word-problem types as pointed out previously (see Tables 1.1 and 1.2).

Ibarra and Lindvall (1982) also noted that children can be helped to solve word problems if they are allowed to really "act out" the problem with concrete aids. However, in agreement with the studies of Carpenter and Moser (1982, 1984) and De Corte and Verschaffel (1981, 1987), the study showed that even under these optimal conditions a definite proportion of children was unable to comprehend even the simplest word problem. Ibarra and Lindvall assumed that these children need specific instruction and experience in word-problem solving, i.e. having them acted out for them as well as acting them out themselves, in order to reach solutions through this type of manipulation. In doing so, four major components should be included in modeling the essentials of a word problem: i.e. (a) set identity, (b) set numerosity, (c) operations on sets, and (d) identity of answer set.

If some specific types of concrete representations aid word-problem solving by normally achieving children, teaching similar representations to children with learning difficulties, who seem to need explicit instruction, may result in an improvement of their arithmetic word-problem solving ability. Because competent regular school children build physical representations that reflect the original problem statements, these representations could give meaning to the semantics underlying the problem. Modeling word problems may be particularly helpful for beginners in analysing and solving the problem. The construction of an external representation takes time and therefore may increase the depth of processing, which reduces the chance of superficial text analysis. Second, external representations could replace internal representations and thereby relieve the working memory (Van Essen, Hamaker, & Van Grafhorst, 1989). Third, counting procedures are externalized by external modeling, which aids in counting the answer without error, also by virtue of relieving working memory. Finally, instruction in external modeling could result in the development of similar cognitive schemata in memory.

Moreover, these representations are rather simple to construct and the procedures for building these representations may easily be broken up in successive steps. Teaching children these procedures step by step allows detailed instruction and explicit performance by a child of all the steps of the task strategy. The execution of each step may directly be evaluated and provided with feedback by which the consolidation of incomplete or inadequate problem-solving procedures may be circumvented.

Besides the four components of Ibarra and Lindvall (1982), the material strategies of regular school children acted as models in designing the representational strategies for the various word-problem types. We also resorted to the representational schemata of the computer-simulation model CHIPS.

It was decided to use the CHIPS model instead of the Riley et al. (RGH) or ARITHPRO model because we believe that an important difference arises between the representational schemata of young children and the computer-simulation models of RGH and ARITHPRO. Whereas the computer-simulation models of RGH and ARITHPRO distinguish between knowledge concerning internal schemata and knowledge concerning counting strategies, empirical studies evidence that these same knowledge components seem to be based on external actions and merged for young children. Young children, with little experience in word-problem solving, seemingly do not possess cognitive schemata comparable with these models' internal schemata. External representations may thus replace internal schemata for beginners. With experience, external modeling may result in consolidating internal cognitive schemata for the different word-problem types. Thus, the computer-simulation models of RGH and ARITHPRO seemingly represent a level of expertise that is not comparable with the expertise of young children.

Whereas both RGH and ARITHPRO first represent sets internally and subsequently set out to count a corresponding set of objects, representing problem structure with objects and counting these objects seem to be merged in the CHIPS model. We believe that the CHIPS model is the most adequate description of the psychological processes of young children. For that reason, with regard to the modeling strategies to be instructed we decided to adhere to the representational strategies both encountered by young children and included in the CHIPS model.

So this instruction program trains children with learning difficulties word-problem solving at the lowest level of development, i.e. at the level of direct modeling. Although one could claim that children with learning difficulties who already master simple addition and subtraction sentences, already bypassed this developmental level, there were several reasons to start instruction at this level of skill. First, proficiency in addition and subtraction is no guarantee for adequate word-problem solving. Even regular school children who are skilled in arithmetic often fail to solve the more difficult word-problem types, because they are unfamiliar with such word problems. To produce an answer anyway, these children often resort to superficial solving strategies. Second, manipulating objects may be a better aid in constructing adequate problem representations owing to the direct correspondence between the material actions performed and the problem situation described. Third, the representational strategies instructed may eventually serve as a starting point for the transition to the next levels of skill, i.e. verbal-counting strategies and number-fact strategies. For the very reason that semantic structure exerts a major influence on the choice of the problem-solving strategy at all internalization levels, additional instruction methods may be designed by which children are smoothly moved through the successive stages of development.

### Pilot work with the training procedures

The pilot studies conducted with the training procedures will be discussed in the next chapters. Chapters 2 and 3 concern the pilot work with the training procedure for instruction on representational skills. In chapter 2 a pilot study with human trainers is reported, while a first study with the computerized version is described in chapter 3. Considering the fact that the pilot work with the training procedure for remediating superficial reading behavior has been well documented elsewhere (Van Lieshout, 1986, submitted), the results of these pilot studies are not discussed here. Instead the pilot study reported in chapter 4 involves a computerized training-procedure in which both representational and text-analysis skills were instructed.

# Chapter 2

# Teaching Concrete Modeling to Solve Simple Arithmetic Word Problems

## Introduction

As was explained in the first chapter, the problems that children with learning difficulties experience in solving arithmetic word problems may arise from the incorrect or incomplete problem representations they build from the semantics underlying the word problem. Since thusfar no research concerned the investigation of the concrete-modeling strategies of children with learning difficulties, it is uncertain whether these children ever used problem-solving strategies, such as modeling with fingers or materials, similar to the ones of regular kindergarten, and first and second graders. These children may as well have lost a major part of these strategies, e.g. after first grade entrance. As from that moment formal arithmetic methods are stressed and the informal problem-solving procedures that these children bring along are ignored. If some types of material representation, as regular first and second graders seem to use, induces correct answers to word problems, then it seems worthwile to investigate whether children with learning difficulties could learn to build such material representations and whether this would result in an improved word-problem solving ability. Therefore, a teaching experiment was performed in which educable mentally retarded children were taught to construct external representations that reflect the problem structures of word problems.

The main goal of our investigation was to examine the effect of this training procedure on their word-problem solving performance. Second, this experiment was to serve as a pilot study for future research with computerized training. This was the main reason to formalize the training procedure. Finally, our interest was in the efficiency of each of the representational strategies instructed to produce a performance increase on the different word-problem types used.

## Method

### Design and subjects

Six children, one boy and five girls (age 9-13 years), from two Special Schools for educable mentally retarded children were trained in a multiple-baseline design across subjects (Kratochwill, 1978). All children could read aloud correctly, understood words like "together", "more than" and "less than", could count on from a given number and had the lowest performance of arithmetic word-problem solving of their class. To assess the starting performance level, for each child baseline data were collected. When the baseline was stable for one of the children of a particular school, training was started for that child. During the training phase, a probe session was inserted after each two training sessions. These probe sessions were inserted in order to attain unobtrusive performance measurements, which were impossible to obtain during the training sessions. If the performance of the subject, who was trained at that particular moment, during the probe series was stable again, training was started for the next child of that school and so on. The trained child entered the posttest phase, in which the follow-up performance of the trained child was assessed. In this way, three children from each school were trained in separate designs.

### Materials

In each session ten addition and subtraction word problems were presented. To prevent attaining the correct answer by simply adding or subtracting all numbers given, all problems contained a third number which was irrelevant to the correct solution of the problem. The position of the irrelevant information was varied across problem types. All numbers and the correct answer were smaller than 10. None of the numbers stated in the problem text corresponded to the correct answer. Each trial contained the following problem types: change 2, change 3, change 4, combine 1, combine 2a, combine 2b (in which as compared to combine 2a the order of introduction of subset-superset is reversed), compare 1, 2, 3 and 4. A full description of these basic problem types is given in chapter 1, Table 1.1. The children had to solve the word problems by using blocks of two different colors. There were as many blocks as were needed to build an external representation in which the total of all the numbers mentioned in the problem text was included.

### Procedure

The procedure consisted of a baseline period and a training phase including probe sessions. In addition, for the first four children trained posttest data were gathered. During the baseline, probe and posttest sessions, the child had to solve the problem independently and without blocks. In order to prevent the discouragement of the children in these sessions the trainer refrained from giving feedback concerning the correctness of the answer of the child.

### Instruction

The training phase started with two instruction sessions. During these sessions the trainer showed the child the task strategy by using blocks to represent the problems. The instruction phase was intended to reveal to the child the representational steps the trainer expected during the succeeding training sessions.

### Training

After the child had read the problem aloud, the child had to represent the word problem with blocks. The representational strategies were divided into three main strategies: joining, separating and matching. The joining strategy consisted of adding blocks to the start set until the number of blocks added corresponded to the given number (combine 1) or until the total number of blocks corresponded to the known end set (change 3 and combine 2a). To distinguish these change and combine problems the blocks had to end up in one line for change problems and two separate lines of blocks had to be constructed for combine problems. Since for change 3 the point at which the addition of blocks started would become invisible after the blocks of the unknown set had been joined in one line to the blocks of the start set, red instead of white blocks had to be used to represent this unknown set. The separating strategy consisted of separating blocks from a start set until the number of blocks separated corresponded to the given number (change 2 and combine 2b) or until the number of blocks remaining corresponded with the known end set (change 4). However, in order to prevent difficulties in determining the number of blocks of the answer set, the children were not allowed to actually remove the separated blocks from the table. Instead two separate lines of blocks had to be constructed for change 2 and 4 as well as for combine 2b problems. The matching strategy consisted of adding blocks in a line below the start set, whereby the number of blocks present in both sets had to be "matched" one-by-one. For compare 1 and 2 problems, the number of blocks added had to correspond with the given number. For compare 3 and 4 problems, the number of blocks added had to be determined by placing $x$ blocks "less" or "more" under the start set. For a full description of the representational strategies, we refer to Jaspers and Van Lieshout (1989a, 1989b) and chapter 3, Figure 3.2. All representational strategies were divided in separate, consecutive steps. A step consisted of one of the following actions: (step 1) representing the first known number set with blocks, (step 2) joining blocks (either in the same or another line) to, matching blocks one-by-one with or separating blocks from the first set of blocks, and (step 3) identifying the answer set by pointing at the relevant blocks. During the execution of each step the trainer observed the modeling behavior and gave standardized feedback, which contained a hint for the correct execution of that step. Several types of errors were distinguished and each type was followed by its own feedback. The types concerned errors in the number of blocks, wrong placement of sets with respect to each other, wrong choice of color of the blocks, and finally errors in locating the answer set.

## Results

### Baseline and probe sessions

Figure 2.1 (solid lines) shows the number of problems which were completed correctly for each child in the baseline and probe sessions. Visual inspection of the curves shows a change from baseline to training phase in all subjects. The mean percentage of correct solutions rose from 12 in the baseline phase to 49 in the probes of the training phase. During the posttest phase of the first four subjects, this percentage increased to 67. In the posttest phase subject 2 even showed a ceiling effect in number of problems correctly solved.

Statistical support of the marked increase from baseline to training phase can be found in a time-series analysis method proposed by Tryon (1982). This method first

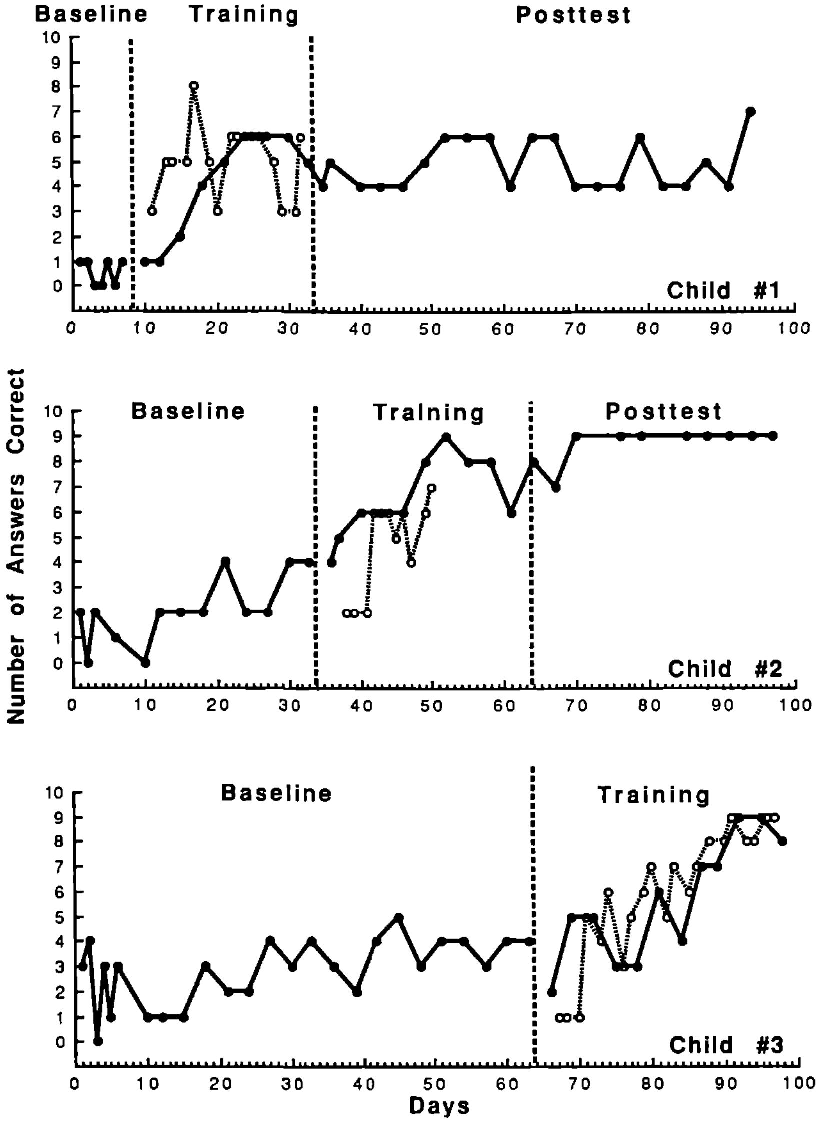


*Figure 2.1* Number of correctly completed problems in each session for each experimental condition and for each of the six children. Solid lines= performance during probes in baseline, training and posttest phase, dashed lines= performance during the training sessions.

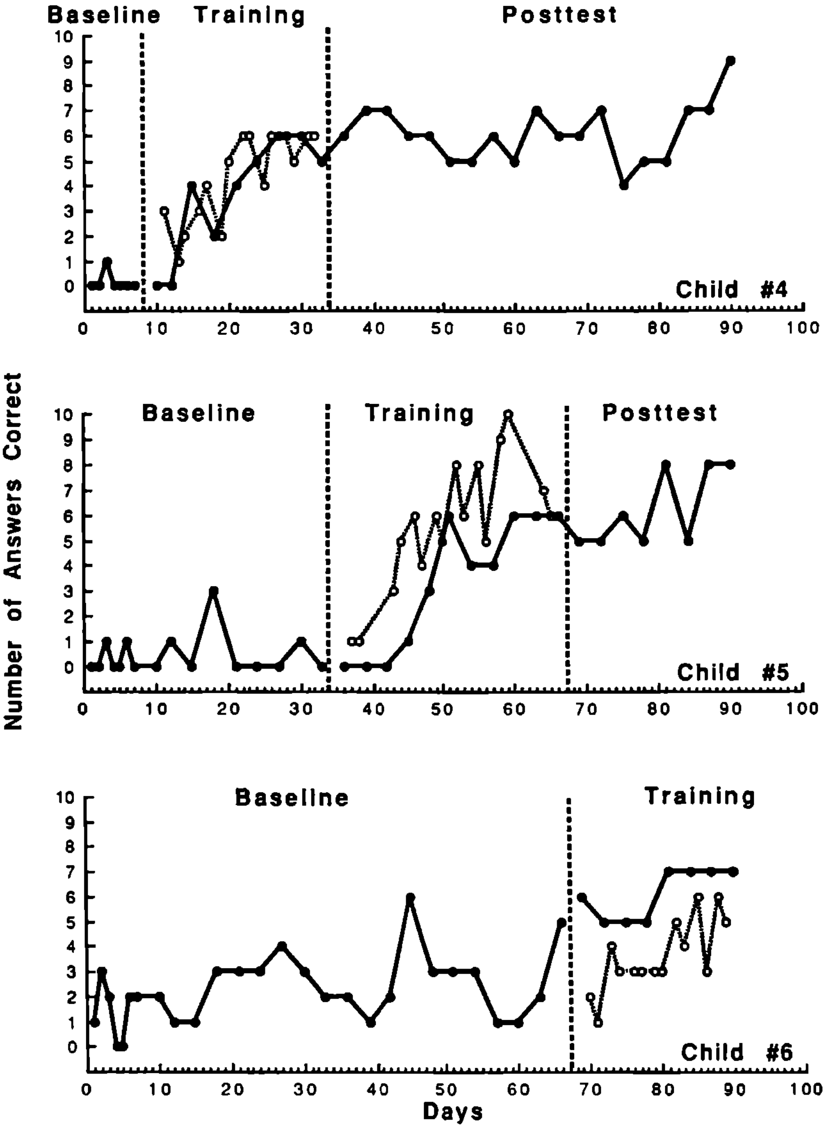


*Figure 2.1 (continued)*

tests whether a trend occurs in the baseline data. No significant trends were found in the baselines of subject 1, 3, 4 and 5; subject 1: $z$=0.99, $n$=6, n.s.; subject 3: $z$=1.04, $n$=25, n.s.; subject 4: $z$=0.52, $n$=7, n.s.; subject 5: $z$=1.32, $n$=16, n.s. In accordance with the visual inspection the subsequent change in performance during the probes in the training phase became significant for all four subjects, subject 1: $z$=3.87, $n$=15, $p$<.01; subject 3: $z$=4.37, $n$=37, $p$<.01; subject 4: $z$=3.60, $n$=16, $p$<.01; subject 5: $z$=4.51, $n$=27, $p$<.01.

However, the baselines of subject 2 and 6 contained a trend, subject 2: $z$=1.94, $n$=13, $p$<.05; subject 6: $z$=2.00, $n$=27, $p$<.05. In this case, Tryon proposes an alternative method to test for trend. First, difference scores of each baseline session and corresponding probe session (i.e. score of the first probe session minus the score of the first baseline session etc.) have to be computed. In the second step these difference scores are examined for trend. As such, the training phase of subject 2 demonstrated a statistically significant trend, subject 2: $z$=2.80, $n$=8, $p$<.01, as opposed to the trend in the training phase of subject 6, which proved insignificant, subject 6: $z$=1.00, $n$=8, n.s. Although Tryon's method conforms nicely to the rationale of the multiple-baseline design, critical comments have been made relating to its validity (Blumberg, 1984). Hence, we performed a second analysis by use of TIDA (Oud, Reelick, & Raaymakers, 1986). TIDA tests whether the polynomial fitted curve of the probes differs significantly from the polynomial fitted curve of the baseline data, which has been extrapolated to the training phase. To this end the largest common number of baseline scores and probe scores (six and eight, respectively) of all subjects just before and after the point of intervention were used. The analysis, which was based on a hierarchical test of a constant and a linear component, demonstrated a significant linear change in performance from baseline to training phase ($F$(1,5)=20.12, $p$<.01).

## Training sessions

The dashed curves in Figure 2.1 represent the performance during the training sessions. During the training sessions, a problem was only scored as correct if all representational steps had been performed without any intervention of the trainer. With exception of subjects 1 and 6, these curves also show a marked performance increase. Thus, despite this stringent scoring method in the training trials, most children showed an increase in an errorless execution of the representational strategies instructed.

The mean percentages of correct execution of the representational strategies for each problem type and each main type (change, combine and compare) in some degree indicate the difficulty level of the various modeling strategies (see Table 2.1).

Table 2.1 Mean Percentage of Number of Problems Correctly Solved During The Training Sessions per Problem Type

| Problem type | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Change | | | | Combine | | | | Compare | | | | |
| 2 | 3 | 4 | Total | 1 | 2a | 2b | Total | 1 | 2 | 3 | 4 | Total |
| 84 | 49 | 85 | 73 | 46 | 34 | 39 | 39 | 52 | 68 | 15 | 37 | 43 |

In general, change problems were the most easy to learn to represent and subsequently to solve, although change 3 problems seemed far more difficult to represent than change 2 and 4 problems. In general, combine problems were more difficult to represent than change problems. In particular children experienced difficulties in representing combine 2a problems. Finally, compare problems also

Table 2.2 Mean Percentage of Number of Problems Correctly Solved During the Baseline, Probe and Posttest Sessions per Problem Type

| | Problem type | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Sessions | Change | | | Combine | | | Compare | | | |
| | 2 | 3 | 4 | 1 | 2a | 2b | 1 | 2 | 3 | 4 |
| Baseline | 34 | 13 | 22 | 40 | 3 | 2 | 7 | 5 | 2 | 5 |
| Training | 71 | 70 | 67 | 71 | 34 | 26 | 26 | 35 | 14 | 37 |
| Posttest | 93 | 87 | 84 | 91 | 69 | 70 | 36 | 56 | 24 | 53 |

seemed more difficult to represent than change problems. For compare problems, the children obviously experienced severe difficulties in representing compare 3 and 4 problems.

However, the mean percentages of correct solutions in the baseline, probe and posttest sessions demonstrated a performance increase for each of these "difficult" problem types too (see Table 2.2). Thus, the overall performance increase in the probes of the training phase is not merely explained by a performance increase on the more easy problem types.

## Discussion

The training procedure clearly improved the children's ability to solve the arithmetic word problems presented. This was evidenced by both the statistical analysis of the subjects' scores individually, which showed a significant change in trend at the onset of the training for five of the six subjects, and the statistical analysis of the set of scores of all subjects together, which demonstrated a significant change in linear trend in performance from the baseline to the training phase. This is in accordance with the results of two other experiments in which the subjects were instructed to model simple arithmetic word problems (Ibarra & Lindvall, 1982; Van Lieshout, 1986). Although in both these teaching experiments the subjects had to make a drawing as representation, the subjects stem from different populations. The subjects of Ibarra and Lindvall (1982) came from a population without learning difficulties, whereas the subjects of Van Lieshout (1986) were learning disabled and the subjects of this experiment were educable mentally retarded. Besides, the degree of formalizing of both the instructed procedure and the feedback seemed far more strict in our studies than in the study of Ibarra and Lindvall (1982).

Likewise the results of this experiment are supported by the results of another study, in which educable mentally retarded children were also instructed to model simple word problems with blocks (Van Lieshout & Jaspers, 1990). In that experiment, the subjects merely had blocks of one color available and had to use a bar to mark off the answer set instead of using blocks of a different color. A more important difference is that the subjects of the present study were not allowed to use materials during the probe sessions, whereas the subjects of the study by Van

Lieshout and Jaspers (1990) did have materials at their disposal in both training and probe sessions. As a consequence, the training effect of this latter study may be regarded as merely the result of an improved ability to externally represent word problems, without improvement of the internal representations. Since materials in the probe sessions of the present experiment were absent, the significant performance increase may be regarded as the result of an improved ability to internally represent word problems. On the one hand, the increase in number of problems correct in the training sessions indicates that the subjects successfully learned to exercise the external-modeling strategies instructed. On the other hand, the performance increase in the probe sessions points at the development of mental representations. Thus, constructing physical representations, which reflect the semantic structure of the different word-problem types, may aid in constructing internal problem representations or so-called problem schemata (see Riley et al., 1983).

Yet some problem types proved more difficult to represent than other types. For example, children had difficulties in representing change 3 problems. The observations during the training sessions showed that this was the result of the obligatory use of red blocks in the second step. Children often used white instead of red blocks to represent the second number set. Since the function of the red blocks, i.e. to mark off the unknown answer set, only became clear in the last step, in which the answer set had to be determined, children presumably did not learn "to look ahead" and to use red blocks to avoid difficulties in locating the answer set. Second, combine 2a problems seemed rather difficult to model. Actually, these problems are comparable to change 3 problems insofar as blocks have to be added until the total number of blocks equals the second number described. Although our subjects could count on from a given number, they apparently did not use this ability to determine the number of the unknown set. Finally, compare 3 and 4 problems proved far more difficult than compare 1 and 2 problems. What could be the reason for this enormous discrepancy in representing and solving compare 3 and 4 problems correctly as opposed to compare 1 and 2 problems? A closer look at the problem texts, and corresponding representational steps, seems to reveal the difference in difficulty level. To represent compare 1 and 2 problems adequately, merely requires identifying and modeling the relevant number sets in the problem text. (An example of a compare 1 problem is: "Peter has 6 marbles. Ann has 8 marbles. How many more marbles does Ann have than Peter?"). Finally, the answer set (i.e. the difference in number of blocks between the first and second set) can easily be located. In contrast, representing the second relevant set in compare 3 and 4 problems demands the notion that the second relevant number given should not be directly modeled ( e.g. "Peter has 6 marbles. Ann has 2 marbles more than Peter. How many marbles does Ann have?"). Instead children have to infer from the problem text that the "more than" or "less than" relationship between the first relevant and second relevant number has to be modeled to represent the second set. In fact, for compare 3 and 4 problems in particular our subjects experienced difficulties in representing this second set.

Although the results of all teaching experiments thusfar appear to be an auspicious start in the development of successful training procedures for children with learning difficulties, one should be cautious in generalizing the conclusions concerning the effectiveness of the training procedures to other children with learning difficulties. First, the subject sample of all experiments was very small and therefore these subjects may not be representative for children with learning difficulties in general. By chance, subjects may have been selected, who are especially susceptible to the training procedure. Second, only a subset of a larger set of problem types were presented. Therefore, it is possible that the subjects merely learned to represent the word-problem types for which they received training, but did not learn to generalize these strategies to new problem types. So a final conclusion has to be postponed until transfer effects on new problem types have been studied with larger group designs.

Thusfar, two different training procedures for arithmetic word-problem solving have been developed, which both have proven to be effective; one, in which children are trained to analyse the problem text carefully, and one in which children learn to represent word problems properly. Since the effectiveness of both training procedures was demonstrated when performed by trainers, one could question whether similar Computer-assisted Instruction (CAI) programs for these procedures need to be developed. A first argument in favor of CAI development applies to the individualization of instruction, which is often recommended but is generally very time-consuming for a teacher. The computer in particular offers the possibility to instruct pupils on an individual basis without immense time investment of the teacher. Second, an important merit of CAI programs may be the motivation increase of the pupils working with such programs. As opposed to a human teacher, the computer records but does not judge the student's performance in terms of "poorly achieving", which may reduce fear of failure. Yet, the possibility to record specific kinds of mistakes and competence on some tasks allows for individualized progress through the course. Especially CAI programs in which the problem-solving route has been subdivided into small separate steps, which have to be performed in succession, permits both guidance of the problem-solving route, and the revelation of misconceptions that would otherwise remain unnoticed. In our research the development of CAI first serves as expedient to manifest and to record the problem-solving processes, misconceptions and knowledge deficits of children with learning difficulties, and subsequently to determine along which lines instruction to these children can be optimalized.

# Chapter 3

# A CAI Program for Instructing Modeling of Arithmetic Word Problems

## Introduction

Since the training procedure constructed to help children with learning difficulties to represent word problems with materials turned out to be effective (see chapter 2), a computerized version of this training procedure seemed feasible.

There were several reasons to use the computer as training device. First, to perform the training procedure smoothly, a human trainer has to make himself or herself acquainted with and keep abreast of all kinds of word-problem types distinguished, all representational strategies defined, all possible error types and feedback contents. As a consequence, performing the training procedure can become a rather difficult task for a human trainer.

Second, monitoring, evaluating and correcting the problem-solving behavior of the child ask for an individual training situation, which often does not fit in very well with classroom settings (Van Lieshout, submitted). A computer-device seems more suited to this job than a human trainer.

As in the study with human trainers, the representational strategies that were instructed in the computerized training-program consisted of several steps that had to be performed in a strict order. Instead of moving blocks on the table, the children had to construct visual configurations on a touchscreen monitor with the aid of squares depicted. The computer rendered assistance only when a child reached an impasse or committed an error, whereupon the child independently continued the problem, until the child reached a next impasse or made a next error. As such, the computer helped the child solving the problem until finally the solution of the problem was reached. The help offered by the computer became more specific when errors occurred repeatedly within a specific problem-solving step. The main idea was that experience would serve the child in acquiring the prescribed problem-solving steps. Consequently, the need for assistance would decrease, until the child eventually performed all steps without the computer offering help.

The main purpose of this investigation was to evaluate the efficacy of this computerized training-procedure. As such, the transfer effect of the training procedure on the performances in a situation without modeling facilities was studied. It was assumed that in this situation children would be handicapped in solving the various word problems, as long as they had not cognitively incorporated the problem schemata. Since it was assumed that a positive change in level of performance would not necessarily become apparent immediately after the point of intervention, it was expected that at least an ascending trend in performance level would be found in the training phase of each subject. To determine whether the children had actually developed cognitive schemata, the performances in a situation without visual aids was studied. Second, our interest was in the kind of errors committed by the children in performing the representational steps during the training sessions. Would these errors resemble the errors of inexperienced normally achieving children and would the children increasingly adhere to the concrete modeling strategies instructed?

## Method

### Design and subjects

From an elementary school for educable mentally retarded children, subjects were selected who could read correctly, could count on from a given number and scored more than 75% correct in a comprehension test (CITO comprehension test). From this group the five children with the lowest performance in arithmetic word-problem solving were chosen. These children, one boy and four girls, were trained in a modified multiple-baseline design across subjects (Kratochwill, 1978). The ages of the subjects ranged from 9 years and 1 month to 11 years and 4 months (mean age 10 years and 5 months). For each of the children baseline data were collected. The children were subjected to the training procedure in random order, one after the other. Every time a subject had completed ten training sessions, the next child was trained. The baseline assessment was continued for each subject who had not yet been trained. A child was probed on his or her performance after each two training sessions. After training, the children were probed on their follow-up performance.

## Materials

In each session, ten single addition and subtraction word problems were presented. All problems contained a sentence in which a number set was described which was irrelevant to the solution of the problem. This irrelevant number set was included to prevent children from attaining the correct answer by simply adding or subtracting the given numbers without understanding the relationship between the two sets. Indeed, less competent problem solvers tend to add all the numbers in the problem text (Goodstein et al., 1971). Hence, the probability of correctly solving the problem by chance was decreased by inserting a third irrelevant number in the problem statement. All numbers contained in the problem text and the correct answer were under 10. None of the numbers stated corresponded with the correct answer. Moreover, neither subtraction nor addition, nor a combination of both on the two or three numbers stated, would result in the correct answer, except for the correct operation on the two relevant numbers of the problem.

The word problems differed with respect to the semantic category to which they belonged. Using the problem type names and numbers of Heller and Greeno (1978), the types presented in each session were: change 2, 3, 4; compare 1, 2, 3, 4; combine 1, 2a and combine 2b in which the introduction of the known superset compared to combine 2a was reversed. Figure 3.2 gives an example of a problem text for each of these types.

In order to make each training session equally difficult, the position of the irrelevant number sentence was varied systematically. In the problem types change 4, combine 2b and compare 1, the first sentence was the sentence describing the irrelevant number set. In the problem types change 3, combine 2a and compare 3, the irrelevant number sentence appeared second in the problem text. Finally, the third sentence was the irrelevant number sentence in the remaining problem types.

The computer program had been written in UCSD-pascal 1.3 for an Apple II E microcomputer with 128 Kb, connected to a Philips VP 120 Touchscreen Monitor.

## Procedure

The procedure consisted of a baseline period, followed by two instruction sessions, a training period including probe sessions and finally a retention period. The follow-up performance of subject 5 was not assessed because summer vacation terminated the experiment. The performance assessment in the baseline period, retention period and the probe sessions during the training phase took place under the same conditions. During these sessions, the word problems were also presented on the touchscreen. However, as opposed to the training sessions, the child had to solve the problems without the computer offering help. When a child had solved the problem, the computer only informed the child whether the solution was correct.

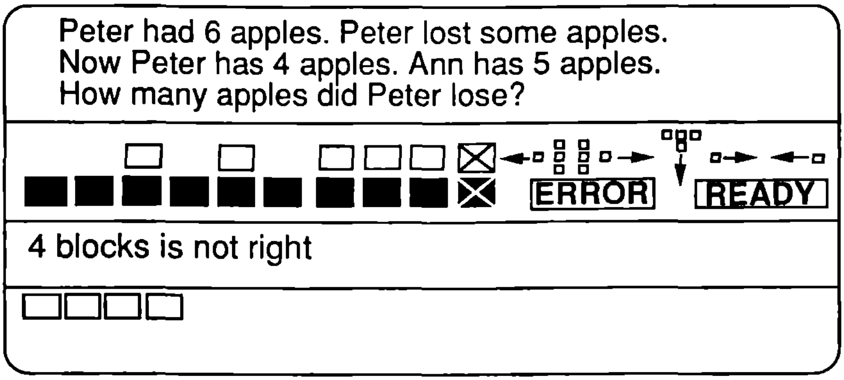
## Instruction

At the start of the training phase there were two instruction sessions. These were the only two sessions in the experiment during which a human trainer was actually involved. The trainer started by verbalizing and showing the activities that the child was expected to carry out during the training phase. In session two, the trainer also showed the child what happened if a mistake was made. Meanwhile, the child was allowed to help represent the problems by touching softkeys on the screen. Finally, the child had to solve the last problem independently.

## Training

When each problem was first presented, only the problem text and a READY-key were displayed on the screen. First, the child was encouraged to read the problem text carefully. After reading the problem text, the child was supposed to construct the requested pictorial configuration for the particular problem type. The construction of this configuration was divided into separate, consecutive steps. In the first step the child had to represent the first relevant set mentioned in the problem text with the appropriate number of squares. The contents of the second step depended on the problem type at hand and consisted of adding squares to, matching squares one-by-one with or separating squares from the first set. The final step was identifying the answer set. The child had to point at the relevant squares, i.e. those belonging to the answer set. For the construction of this pictorial configuration the child was offered a supply of icons from which squares could be moved to a "work-sheet" area. This "icon-supply" consisted of two rows of nine visual squares (one row of black and one row of white squares), the READY-key, an ERROR-key and some other function-keys to manipulate the visual squares on the "work-sheet".

Figure 3.1 gives an example of a screen display after the first (incorrect) attempt of the child to represent the first set on the "work-sheet".



Peter had 6 apples. Peter lost some apples.
Now Peter has 4 apples. Ann has 5 apples.
How many apples did Peter lose?

4 blocks is not right

*Figure 3.1* Example of a screen display after the first (incorrect) attempt of the child to represent the first set on the work-sheet.

At the start of each step, the "icon-supply" was provisioned with white and black squares. The crossed function-keys could be used to place visual squares on the "work-sheet" one-by-one back into the "icon-supply". The arrow-keys were to be used to indicate the position of the second set with respect to the first set. From left to right, the three arrows with four small squares could be used to add squares immediately behind the first set, to add squares to the far right of the first set and to place squares in a one-by-one correspondence beneath the squares of the first set, respectively. The arrows with one small square could be used to separate squares from the first set or to replace squares back into their initial position in the first set, respectively. All actions that a child had performed within a particular step could be cancelled by touching the ERROR-key. Finally, a child had to finish each step by touching the READY-key, after which the computer analysed the child's response and delivered feedback.

For each problem type used Figure 3.2 shows examples of problem sentences and linked representational steps.

Generally, the concrete representations of change problems had to visualize the increment or decrement in number. The representation of combine problems had to show the part-whole character. In the representation of compare problems, the

| Problem type | Example | Representational Steps |
|---|---|---|
| Change | | |
| 2 | Mary had 6 apples. | □ □ □ □ □ □ |
| | She ate 2 apples. | □ □ □ □     □ □ |
| | Paul had 3 apples. | — |
| | How many apples does Mary have left? | ▨ ▨ ▨ ▨     □ □ |
| 3 | Mary had 2 apples. | □ □ |
| | Paul had 3 apples. | — |
| | Mary got some more apples. | — |
| | Now Mary has 6 apples. | □ □ ■ ■ ■ ■ |
| | How many apples did Mary get? | □ □ ▨ ▨ ▨ ▨ |
| 4 | Paul had 3 apples. | — |
| | Mary had 6 apples. | □ □ □ □ □ □ |
| | She ate some of them. | — |
| | Now Mary has 2 apples. | □ □   □ □ □ □ |
| | How many apples did Mary eat? | □ □   ▨ ▨ ▨ ▨ |
| Combine | | |
| 1 | Mary has 6 apples. | □ □ □ □ □ □ |
| | Pete has 2 apples. | □ □ □ □ □ □   □ □ |
| | Paul has 3 apples. | — |
| | How many apples do Mary and Peter have together? | ▨ ▨ ▨ ▨ ▨ ▨   ▨ ▨ |
| 2a | Mary has 2 apples. | □ □ |
| | Paul has 3 apples. | — |
| | Peter also has some apples. | — |
| | Together Mary and Peter have 6 apples. | □ □   □ □ □ □ |
| | How many apples does Peter have? | □ □   ▨ ▨ ▨ ▨ |

Note. □ = white squares had to be used
■ = black squares had to be used
▨ = squares that had to be pointed at to locate the answerset

| Problem type | Example | Representational Steps |
|---|---|---|
| Combine | | |
| 2b | Paul has 3 apples. | — |
| | Together Mary and Paul have 6 apples. | □□□□□□ |
| | Mary has 2 apples. | □□□□ □□ |
| | How many apples does Peter have? | ■■■■ □□ |
| Compare | | |
| 1 | Paul has 3 apples. | — |
| | Peter has 2 appels. | □□ |
| | Mary has 6 apples. | □□<br>□□□□□□ |
| | How many more apples does Mary have than Peter? | □□<br>□□■■■■ |
| 2 | Peter has 2 apples. | □□ |
| | Mary has 6 apples. | □□<br>□□□□□□ |
| | Paul has 3 apples. | — |
| | How many less apples does Peter have than Mary? | □□<br>□□■■■■ |
| 3 | Peter has 2 apples. | □□ |
| | Paul has 3 apples. | — |
| | Mary has 6 apples more than Peter. | □□<br>□□□□□□□□ |
| | How many apples does Mary have? | □□<br>■■■■■■■■ |
| 4 | Peter has 6 apples. | □□□□□□ |
| | Mary has 2 apples less than Peter. | □□□□□□<br>□□□□ |
| | Paul has 3 apples. | — |
| | How many apples does Mary have? | □□□□□□<br>■■■■ |

*Figure 3.2* Examples of problem types used and corresponding representational steps. If a sentence did not require representing the set described, this is indicated by "__".

manifestation of the act of comparing the sets was required. More particularly, to represent change 2, change 4 and combine 2b problems, squares had to be separated in the second step from the start set resulting in two lines of squares next to each other. For change 2 and combine 2b problems the set at the left was the remaining answer set, whereas the answer set for change 4 problems was the set at the right. Combine 1 and 2a problems had to be represented by constructing the first known set and by subsequently adding squares on in a separate line in the second step. Both lines of squares together represented the answer set. To represent change 3 problems, squares had to be added on in the second set to the first set until the second relevant number of the known end set was reached. In this case the squares of the unknown set were joined in one line to the squares of the start set, so the point at which the addition started would become invisible. Therefore, white instead of black squares had to be used for the representation of the added set to enable the child to identify the answer set afterwards. Finally, for the representation of compare problems, squares had to be placed in two separate lines below each other, matching each square of the smaller set to a square of the bigger set. This is called a matching strategy. This representation was taught for all compare problems used (compare 1, 2, 3 and 4).

Several types of errors were distinguished and each type was followed by its own feedback. These types concerned errors in the number of squares, wrong placement of sets with respect to each other (e.g., joining instead of separating or laying squares in one line instead of below each other), wrong choice of color of the squares (e.g., black instead of white squares) and finally pointing at a number of squares that did not constitute the answer set. The feedback consisted of remarks, such as: "Eight blocks is not correct." When the child made a second error, more specific feedback based on error type followed, for example: "Look closely at the text again" or "How many blocks do you need?" In addition the part of the problem text that described the relevant set was highlighted. If the child failed on this occasion as well, the computer displayed the requested pictorial configuration of squares for the particular step on the screen. The computer program was structured to such a degree, that progression through the trial was impossible, unless all preceding steps had been completed correctly. Passing through the modeling steps of the trial resulted in a visual configuration in which the answer set could be located easily.

### Scoring

For both the probe sessions and training sessions, the number of problems correctly solved was scored. However, merely scoring the final solution in the training sessions would result in misleading conclusions with regard to the independence of the children in applying the requested problem-solving strategies. For the requested configuration for each strategy step was automatically displayed after failing on two successive occasions within the specific step. As such, erring frequently in the second step resulted in the computer completing the final configuration of squares, in which the answer set could be located rather easily. As a consequence, the chance of giving a correct answer was increased whereas in the preceding strategy steps the child could have committed many representational errors. Therefore a word problem was only scored as correct in the training phase if all the steps to be executed were accurately performed, without any correcting feedback from the computer.

## Results

### Baseline, probe and retention sessions

Figure 3.3 (curves with solid lines) shows an increase in performance from baseline to training phase in three of the five subjects which is most convincing for the first two subjects trained. On the other hand, the mean performances of the sessions of the baseline and training phase indicated the highest performance increase from baseline to the probe sessions in subject 1 and subject 3. For the first subject the mean percentage of correct solutions rose from 37 to 67 and for the third subject from 42 to 72 correct. These percentages were 37 and 48 for the second subject, for the fourth subject 45 and 55 and for the fifth subject 36 and 38. The average performance increase from baseline to training phase amounted to 16%. The percentage of correct solutions during the posttest sessions was 91 for subject 1, 77 for subject 2, 89 for subject 3 and 60 for subject 4. For subject 5 no data were available. Thus, three subjects also show a marked increase from training phase to follow-up phase.

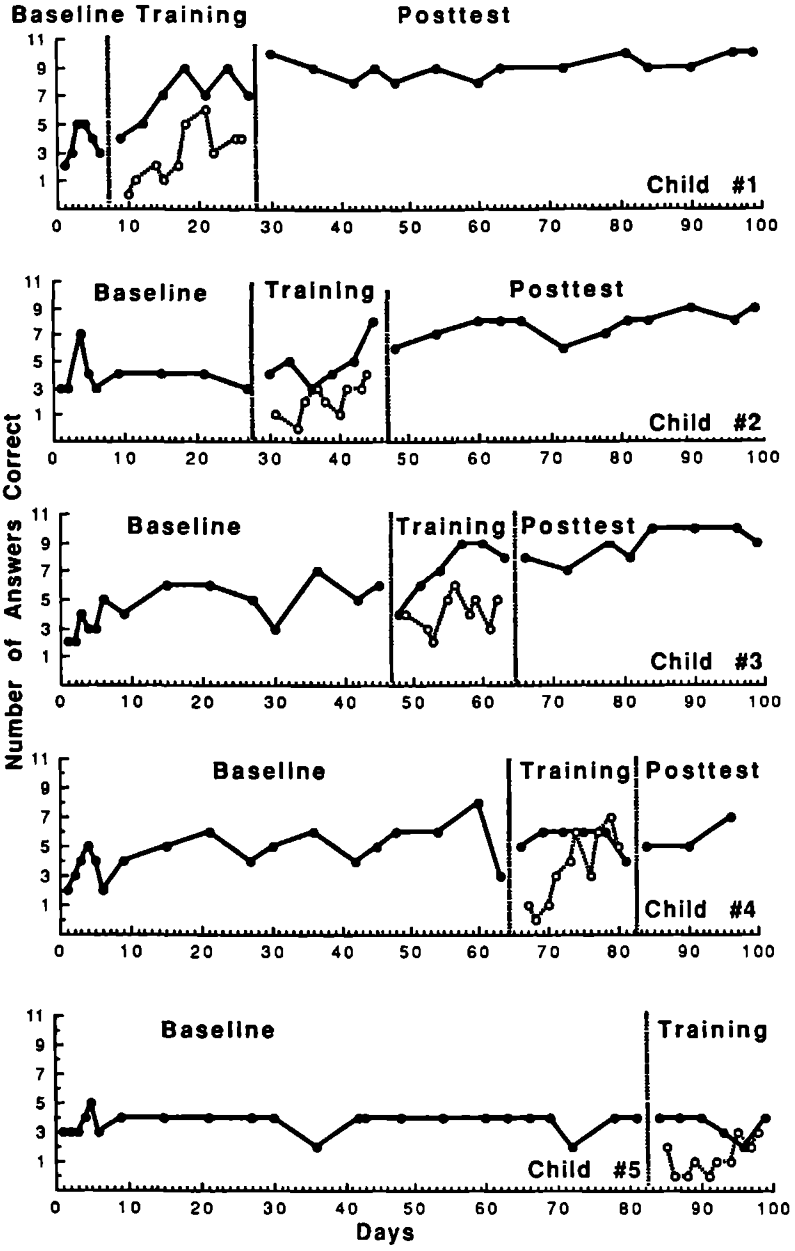


*Figure 3.3* Number of correctly completed problems in each session for each experimental condition and for each of the five children. Solid lines= performance during probes in the baseline, training and retention phase, dashed lines= performance during the training sessions.

Visual inspection of the data presented does not provide clear-cut evidence for the success of the training procedure. Besides, different conclusions proceed from visual inspection of the graphs and inspection of the mean performance. Therefore, a statistical analysis was performed that takes account of changes in both level and trend of the data. The Revusky's Rn analysis (Wolery & Billingsky, 1982) provides a test of significance for cases in which multiple-baseline data are collected across several subjects. In the analysis the score for the subject for whom treatment was introduced and for each subject who had not previously received the treatment are ranked. For both the changes in level and trend the Revusky's Rn is calculated by summing the ranks obtained by subjects for whom treatment was introduced on any given occasion.

The Sum of Ranks for the level changes based on mean performance of the baseline and training phase was significant, Sum of Ranks=5.5, $n$=5, $p<.05$. Likewise, the Sum of Ranks for the changes in slope during intervention was significant, Sum of Ranks=5.0, $n$=5, $p<.05$. In conclusion, the analysis indicated that both level and slope changes were significant, indicating changes in both level and trend in performance level after the training procedure had been put into effect.

To examine whether the significant performance increase from baseline to training phase could be attributed to a performance increase on all problem types or merely resulted from a performance increase on a subset of problem types, for each problem type a monotone-trend analysis was performed on the probe data across all subjects of the training phase (based on Kendall's t, Ferguson, 1971).

This analysis showed significant performance increases for combine 2b and compare 1 problem types (see Table 3.1). It should be noted that problem types change 2, 3, and 4 and combine 1 already showed a ceiling effect or near ceiling effect in the baseline phase. Therefore, it was impossible to find significant z-scores for these problem types.

## Training sessions

Since the main purpose of this study concerned the evaluation of the prestructured teaching of meaningful problem representations, the individual problem-solving steps were checked and scored in their correctness in each training session. The curves with dashed lines in Figure 3.3 display the overall results of each individual training phase. As opposed to the probe sessions in the training phase, in which only a numerical answer was required and scored, a child had to perform all strategy steps without errors for the problem to be scored as correct during the training sessions.

Table 3.1 Mean Proportion of Number of Problems Correctly Solved per Problem Type Across Both Sessions and Subjects and Results of the Trend Analyses of Probe and Training Sessions per Problem Type

| | Sessions | | | | | |
|---|---|---|---|---|---|---|
| | Probe | | | Training | | |
| Problem type | *M* | *z* | *p* | *M* | *z* | *p* |
| Change | | | | | | |
| 2 | 0.9 | 0 | n.s. | 0.6 | 2.57 | <.01 |
| 3 | 0.9 | 1.44 | n.s. | 0.2 | 2.13 | <.05 |
| 4 | 0.8 | 1.03 | n.s. | 0.3 | 2.25 | <.05 |
| Combine | | | | | | |
| 1 | 0.9 | 0.91 | n.s. | 0.2 | 2.84 | <.01 |
| 2a | 0.5 | 1.58 | n.s. | 0.1 | 1.13 | n.s. |
| 2b | 0.6 | 2.26 | <.05 | 0.3 | 2.18 | <.05 |
| Compare | | | | | | |
| 1 | 0.4 | 1.79 | <.05 | 0.3 | 2.06 | <.05 |
| 2 | 0.4 | 1.44 | n.s. | 0.6 | 1.97 | <.05 |
| 3 | 0.2 | 0.23 | n.s. | 0.1 | 1.43 | n.s. |
| 4 | 0.4 | 1.44 | n.s. | 0.1 | 1.03 | n.s. |

Monotone-trend analysis of the training sessions showed a significant overall-performance increase ($z = 3.27$, $p < .01$). Similar trend analyses of the training sessions for each problem type showed significant performance increases for change 2, change 3, change 4, combine 1, combine 2b, compare 1, and compare 2 problem types (see Table 3.1).

### Representational errors in strategy steps

Table 3.2 shows both the error percentages and the total number of errors for each error type within each strategy step. The total percentage of errors in all steps was set to 100, although the total amount of errors committed relative to the potential amount of errors was merely 17%. The total error percentage of the second step exceeds the total error percentages of both the first and third step (step two versus step three, $t(4)=4.86$, $p<.01$). This result does not seem very surprising as the distinction between the various semantic problem types was only reflected in the manipulation of squares in step 2, in which different decisions had to be made concerning number, kind of manipulation (adding versus separating squares) color and placement of squares (below, next to or added to the first set).

Table 3.2 Error Percentages for Each Error Type and Each Strategy Step

| | Step | | | |
|---|---|---|---|---|
| Error type | 1 | 2 | 3 | Total |
| Wrong number of squares | 18.4 | 28.9 | 21.8 | 69.1 |
| Wrong placement of squares | - | 13.0 | - | 13.0 |
| Other error types | 0.4 | 15.1 | 2.4 | 17.9 |
| Total | 18.8 | 57.0 | 24.2 | 100 |

*Note.* - = does not apply.

*Strategy step one*

Subsequently, the error percentages for each error type were scored separately for each problem type within the various strategy steps. As can be seen in Table 3.3, the most frequent error type in step 1 concerned the wrong number of squares (97.8%). This error type is manifold for change 4, combine 2a, combine 2b, and compare 1 (see Table 3.3). Inspection of the subcategories of this error type reveals that for change 4 problems all number errors proceed from moving a number of squares equal to the irrelevant number described in the problem text. This same error is encountered for combine 2a, combine 2b and compare 1. The problem types mentioned, with exception of combine 2a, are precisely those problem types that contain the irrelevant information in the first sentence of the problem text from which the subject presumably starts representing the problem in the first step. In comparison with the other problem types, those with the irrelevant number set in the first sentence on average revealed five times as many irrelevant number errors ($M$=8.6% versus $M$=1.8%). Especially subject 5, who on average made twice as many representational errors as the other subjects, took a number of squares equal to the irrelevant number and made himself responsible for 54% of the errors made within this subcategory. For the other four subjects, this error became less frequent as the training sessions progressed (Kendall t, $z = 2.78$, $p < .01$). Yet for combine 2a, combine 2b and compare 1 problems subjects more often took a number of squares from the supply equal to the second relevant number stated in the problem text.

Table 3.3 Error Percentages for Each Problem Type and Each Error Type Within Step One

| | Problem type | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Change | | | Combine | | | Compare | | | | |
| Error type | 2 | 3 | 4 | 1 | 2a | 2b | 1 | 2 | 3 | 4 | Total |
| Wrong number of squares: | | | | | | | | | | | |
| Irrelevant number | 0.0 | 0.0 | 11.5 | 0.7 | 5.4 | 5.4 | 8.8 | 0.7 | 4.1 | 1.4 | 38 |
| Number 2 | 0.7 | 0.0 | 0.0 | 2.1 | 11.5 | 10.1 | 10.8 | 2.1 | 4.1 | 6.1 | 47.5 |
| Other | 1.4 | 0.0 | 0.0 | 2.6 | 1.4 | 5.4 | 0.0 | 0.0 | 1.4 | 0.0 | 12.3 |
| Subtotal | 2.1 | 0.0 | 11.5 | 5.4 | 18.3 | 20.9 | 19.6 | 0.7 | 9.6 | 7.5 | 97.8 |
| Other error types | 0.0 | 0.0 | 0.0 | 0.7 | 0.7 | 0.0 | 0.0 | 0.7 | 0.0 | 0.0 | 2.1 |
| Total | 2.1 | 0.0 | 11.5 | 6.1 | 19 | 20.9 | 19.6 | 3.5 | 9.6 | 7.5 | 100 |

### *Strategy step two*

Table 3.4 shows the error percentage for each problem type and each error type within the second step. As in the first step, the most prominent error type in the second step was an incorrect number of squares. This error type was most frequently encountered for change 3, combine 2a, compare 3 and compare 4. Inspection of the subcategories of this error type reveals that some of these wrong number errors consist of errors equal to the number of the irrelevant set. In this case too, the problem types indicated are the only ones that contain the irrelevant information in the sentence where the subject probably arrives at the start of the second step.

However, just as in step one, in step two, the most frequent error for combine 2a, compare 3 and 4 also appeared to be moving an amount of squares corresponding with the second relevant number described in the problem text. These three problem

Table 3.4 Error Percentages for Each Problem Type and Each Error Type Within Step Two

| | Problem type | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Change | | | Combine | | | Compare | | | | |
| Error type | 2 | 3 | 4 | 1 | 2a | 2b | 1 | 2 | 3 | 4 | Total |
| Wrong number of squares: | | | | | | | | | | | |
| Irrelevant number | 1.1 | 4.1 | 0.9 | 0.0 | 2.0 | 0.5 | 0.0 | 0.0 | 2.0 | 0.2 | 10.8 |
| Number 1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 1.6 | 0.9 | 0.0 | 0.2 | 3.0 |
| Number 2 | 0.0 | 1.8 | 2.5 | c | 3.6 | 0.9 | c | c | 11.0 | 10.3 | 29.6 |
| Other | 1.1 | 0.9 | 0.0 | 0.7 | 1.1 | 1.1 | 0.0 | 0.5 | 1.8 | 0.2 | 7.9 |
| Subtotal | 2.2 | 6.8 | 3.4 | 0.7 | 6.7 | 2.7 | 1.6 | 1.4 | 14.8 | 10.9 | 51.3 |
| Wrong placement of squares | - | 5.4 | - | 7 | 5.2 | - | 3.4 | 1.6 | 0.0 | 0.2 | 22.7 |
| Other error types | 2.5 | 6.3 | 0.7 | 0.9 | 3.4 | 2.5 | 0.2 | 0.2 | 3.8 | 5.6 | 26.0 |
| Total | 2.7 | 18.5 | 4.1 | 8.6 | 15.3 | 5.2 | 5.2 | 3.2 | 18.6 | 16.7 | 100 |

*Note.* - = does not apply, c = correct execution.

types belong to a subset of five problem types (change 3, change 4, combine 2b, compare 3 and compare 4), in which the second relevant number never had to be directly modeled. Actually, this subset on average revealed eight times as many second relevant number errors as the remaining subset of problem types ($M$=5.8 versus $M$=0.7 respectively). This error type became less frequent as the training sessions progressed (Kendall t, $z$=2.83, $p$<.01). In comparison with the other problem types, the compare problem types on average revealed twice as many errors in the number of squares moved ($M$=67.7% versus $M$=33.3%).

The second prominent error type was the wrong placement of the second set in relation to the first set. In particular problem types change 3, combine 1 and 2a and compare 1 evoked this kind of error. The raw data for change 3 revealed that subjects added squares in step 2 in such a way so as to result in two separate lines of squares (3.6% out of 5.4%). Most wrong placement errors for combine 1 and 2a problems came from placing the squares that represented the second set in a separate line below the first set (5.6% out of 7% and 4.3% out of 5.2% respectively).

### *Strategy step three*

Finally, in step 3 the answer set had to be located by pointing at the relevant squares in the completed representation. This resulted in many errors in the number of squares touched for the problem types compare 3 and 4 (see Table 3.5). As in both steps 1 and 2, in step 3 for these problem types children also pointed at a number of squares equal to the second relevant number (25.3% and 14% respectively).

Table 3.5 Error Percentages for Each Problem Type and Each Error Type Within Step Three

| | Problem type | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Change | | | Combine | | | Compare | | | | |
| Error type | 2 | 3 | 4 | 1 | 2a | 2b | 1 | 2 | 3 | 4 | Total |
| Wrong number of squares: | | | | | | | | | | | |
| Number 1 | 0.0 | 0.0 | 0.0 | 2.4 | 1.7 | 1.7 | 2.4 | 2.9 | 2.9 | 0.6 | 14.6 |
| Number 2 | 4.7 | 2.4 | 7.0 | 4.1 | 4.1 | 5.9 | 6.5 | 5.3 | 25.3 | 14.0 | 79.3 |
| Other: | 0.0 | 0.0 | 0.6 | 0.0 | 1.7 | 1.2 | 1.2 | 1.2 | 0.0 | 0.0 | 5.9 |
| Total | 4.7 | 2.4 | 7.6 | 6.5 | 7.5 | 8.8 | 10.1 | 9.4 | 28.2 | 14.6 | 100 |

## Discussion

The computerized training-procedure increased the children's ability to solve the various word problems, as both a significant change in level and trend at the point of intervention have demonstrated. In spite of this overall result, some subjects did not or only slightly benefit from this training procedure. Moreover, the average performance increase in this computerized training-experiment (16%) compared to the average performance increase of two comparable pilot studies without computers (Van Lieshout & Jaspers, 1989, 55% and Jaspers & Van Lieshout, 1989, 38%, respectively) seems to reveal a difference in training effect at the expense of the computerized training-procedure.

With regard to literature on CAI, this result seems rather surprising, for the surplus value of CAI as opposed to conventional instruction procedures could manifest itself in the reduced time needed to achieve the same performance increase (Conners, Caruso & Detterman, 1986; Hagler & Knowlton, 1987). Therefore CAI procedures seem more rather than less effective compared to human training. Despite the similarity in the instruction procedures for both the pilot studies and this computerized study, there were some differences.

First, subjects in the pilot studies were trained until their performance, reflected in the probe sessions of the training phase, stabilized. In the present study, the number of training sessions, i.e. ten, was fixed, as the statistical analysis by Revusky's Rn required. However, some children in the pilot studies completed more than ten training sessions before their performance stabilized. Therefore some subjects in the present study might have benefited from the procedure pursued in the pilot studies much more than they profited by the fixed amount of ten training sessions.

Second, the sample of subjects was rather small in both the pilot studies and the present experiment. So it could be hypothesized that by chance the subjects in the present experiment stem from a different population than the subjects in the pilot studies, who were rather susceptible to the training procedure. In fact, the overall baseline performance in the pilot studies was much lower than in this experiment. This was due to some problem types that already showed a ceiling effect or near ceiling effect in the baseline phase of this experiment (i.e. four out of ten problem types). Therefore, a significant performance increase on this subset of problems possibly could not be found.

Finally, a last explanation could emanate from the feedback procedure. In the pilot studies the feedback was provided by human trainers, whereas the feedback in the computerized procedure was depicted on the screen. Feedback delivered in writing seems to guarantee the required attention of the child to a lesser degree than spoken feedback.

In spite of this disappointing transfer effect of the computerized training-procedure, at least the subjects proved to develop adequate problem representations for some problem types. This can be inferred from the significant overall performance increase in the probe sessions. Although it may be considered that children already possessed adequate problem representations before training started and as such even had to adjust their own spontaneous strategies, there is no strong evidence for discontinuing proper strategies in this study. First, the performance in the baseline sessions and opening probe sessions should have been higher than actually assessed.

Second, most types of errors made did not seem to be explained by assuming that adequate strategies were corrected. To illustrate this, evidence of alternative explanations for the most prominent error types will be given. For example, the position of the sentence that described the irrelevant number set in the problem text seemed to determine at which moment in the problem-solving route subjects would use this irrelevant number in constructing a representation. This error type seems to reflect a rather superficial analysis of the problem text, as with all problem types this number never had to be used. However, coming to know which given numbers are crucial for reaching the correct solution, first requires thorough reading of the whole text. In contrast with the earlier computerized training-procedure, in which heavy emphasis was placed upon reading and selecting crucial words in the problem text, in the present training procedure accurate reading and actively selecting relevant text parts was not instructed. Furthermore, in the earlier training procedure analysing the question sentence preceded searching the words that described the relevant sets, whereas in the present training procedure the problem text was reread and represented in chronological order. As a consequence, the subjects in this experiment may have experienced difficulties in distinguishing the relevant sets from the irrelevant set since this distinction only becomes clear after reading the question sentence. Particularly subject 5, who proved less competent in representing and solving the word problems, frequently made this kind of error. On the other hand, the remaining subjects used this irrelevant number less and less frequently as the training sessions progressed. Thus, these subjects became more and more competent in determining the relevant sets.

In contrast with error types which seem to reflect a certain lack of knowledge this may not be obvious for some other error types. For example, both combine 2a and compare 1 problems elicited second number errors in the first step. It could be hypothesized that this mistake proceeds from either a misinterpretation of the problem text or from a (correct) separating strategy as an alternative for the strategy instructed. Take for example combine 2a (Figure 3.2). On the one hand, misinterpreting this problem text could result in the conviction that the total set described is a separate set that Mary and Peter own together. As such, the subsets described are not regarded as part of this total set. Consequently, subjects believe that Peter possesses 6 marbles (instead of 4), and that no calculation is required to arrive at the answer.

A similar misconception could account for this same mistake in compare 1 problems (Figure 3.2). Again, misinterpretation of the question sentence may result in the conviction that nothing has to be calculated. Since, as Verschaffel (1984) suggests, this sentence could be interpreted as "Does Mary have more apples than Peter?" and "How many apples does Mary have?" subjects directly try to answer these questions by moving 6 squares instead of the required amount (2 squares, representing Peter's apples).

On the other hand, subjects may have tried to represent both problem types by a separating strategy; i.e. representing the larger number in the problem text (6 for both examples) and subsequently separating squares (2 for both examples) from this set. If this strategy would have been correctly performed, counting the remaining set would have resulted in the correct answer. However, this strategy was not allowed for combine 2a and compare 1 problems and so was evaluated as "wrong" in the first strategy step. As a consequence, the completion of this strategy and potential correct answer could not be evaluated.

Still, although for these problem types a separating strategy may be regarded as expert behavior, research findings on representational strategies of normally performing beginners contradict the employment of such a strategy. In fact, even normally achieving school children, who solve these problem types correctly with materials, rarely seem to use a separating strategy for these problem types. Instead, in representing word problems, they generally seem to analyse the problem texts in chronological order. Besides, this same error type was often committed in the second strategy step with compare 3 and 4 problems for which a proper strategy as alternative cannot be presumed. For compare 3 and 4 problems, this same error type could merely originate from a similar misconception as posited for compare 1 problems. Examples of compare 3 and 4 problems are given in Figure 3.2. Both sentences "Mary has 6 apples more than Peter" and "Mary has 2 apples less than Peter" may be interpreted as "Mary has 6 and more than Peter", respectively as "Mary has 2 and less than Peter". Accordingly 6 squares and 2 squares are moved respectively. This is exactly what happened with these problem types. The subjects seemed rather persistent in their misinterpretation of compare 3 and 4 problems as on average in nine cases out of ten in locating the answer set they pointed at the difference set (e.g. in the examples stated, at 6 squares in compare 3 and at 2 squares in compare 4 problems). Thus, even when these problems are completely represented by squares either by the child or by the computer, this representation did not aid in locating the answer set. In contrast with combine 2a and compare 1 problems, a separating strategy as an alternative cannot be hypothesized, as for compare 3 problems the two relevant numbers always ought to be added and for compare 4 problems this should lead to representing the first relevant number instead of the second number. In general, this error type became less frequent as the training sessions progressed, pointing at more adequate problem representations.

The second prominent error type in step 2 was the wrong placement of set 2 with respect to set 1. In particular, change 3, compare 1, combine 1 and 2a evoked this error. For change 3 problems, subjects added squares resulting in two separate lines of squares. As such, change 3 problems cannot be distinguished from combine problems which represent a distinct semantic problem type. Nevertheless, change 3 problems in particular raise difficulties in locating the answer set if the added squares end up in one line with the start set. Therefore, subjects may not have compensated for this difficulty by using a different colored set as required but instead may have decided to mark off the answer set by keeping the added set at a distance from the start set. In fact, this explanation accounts for 67% of the errors committed for change 3 problems.

Most wrong placement errors for combine 1 and 2a problems came from placing the squares representing the second set in a separate line below the first set. In fact, this was the representation instructed for compare problems. Since these compare problems as a subset in particular revealed an enormous amount of errors in the number of squares moved, it can be inferred that the function of the matching strategy was not clear. As a consequence, subjects may not have distinguished between representations in which the sets are detached by a horizontal or vertical distance. Hence, the assumption that these placement errors are manifestations of adequate modeling strategies, in which the main problem types change, combine and compare are no longer distinguished. However, empirical findings counter this assumption. Actually, even competent children still seem to resort to representational strategies in which change, combine and compare problems are discriminated. Since this distinction between these main types appears to be functional, it was decided to instruct representational strategies in which these main problem types are reflected. Nevertheless, the subjects of this study experienced great difficulties in representing compare problems adequately.

What could be the reason for compare problems being so difficult to represent in this experiment? Since in compare problems no actions are described, either explicitly or implicitly, children could have had difficulties in direct modeling by joining or separating squares. Thus, minor changes in wording compare problems so as to form "action-cued" compare problems should result in richer representations and consequently in a performance increase. This in fact was the result of a study of Hudson (1983), in which even very young children could solve action-cued compare problems (so-called "equalize problems") like "There are 9 birds. There are 7 worms. Suppose the birds all race over and each one tries to get a worm! Will every bird get a worm? How many birds won't get a worm?" The children solved the problem by matching two separate sets one by one and counting the remainder. Likewise, De Corte et al. (1985) found that rewording compare 1 problems as "action-cued" compare problems facilitated the solution processes of regular first and second graders. De Corte et al. argued that the reformulation of compare 1 problems made the problem situation much easier to grasp by avoiding the difficult expression "more than" and, in agreement with Hudson, by more obviously suggesting the use of a matching strategy in the verbal text.

A final question to be answered is whether children with learning difficulties are comparable with some novice problem-solvers in regard to their knowledge and misconceptions concerning word-problem solving. Although the present study involved a rather small sample of subjects, the children with learning difficulties in this study seemingly showed the same kind of errors frequently encountered with poorly performing regular school children.

In particular, for combine 2a and compare 1, 3 and 4 problems, subjects in this study often committed second number errors. This error type seems to correspond with the most common error of poorly performing regular school children, who frequently "solve" these same problem types by answering with the second relevant number in the problem text. Likewise, subjects often represented the irrelevant number stated in the problem text and this error may indicate a tendency to start a so-called "add all the numbers" strategy, which is also frequently encountered with poorly performing regular school children (Verschaffel, 1984). However, since the children's representational errors were evaluated at intervals and, if necessary, corrected by the computer, the subsequent completion of the strategy started by the child could not possibly be verified and interpreted. Thus, although it can be hypothesized that children with learning difficulties are in some respects comparable with poorly performing regular school children and as such presumably only show a developmental lag in their knowledge of arithmetic word-problem solving, this presumption first should be verified by studying the problem-solving strategies of these children without interfering. If these investigations indeed should demonstrate a developmental lag, this lag could be compensated for by teaching these children to build external problem representations, which may finally encourage the development and incorporation of problem schemata in memory.

Besides, the development of computerized procedures for recording word-problem solving procedures in particular may clarify the prior knowledge and misconceptions that underlie errors produced. In future, this could result in the development of CAI, that first diagnoses the misconceptions and prior knowledge of a child and consequently adapts instruction on an individual level. First and foremost, this requires research on diagnosing these answers in such a way as to ascertain the kind of misconception from which the wrong answer stems. Recently, we started research on this question.

# Chapter 4

# A CAI Program for Instructing Text Analysis and Modeling of Arithmetic Word Problems

## Introduction

Although a start has been made in the development of Intelligent Computer-Assisted Instruction (ICAI) in many domains, complete elaborated ICAI systems are scarce. A lack of knowledge concerning learning behavior of novices, their believed misconceptions, their diagnosis, adequate modeling of pupils' behavior and efficiency of adequate teaching strategies seems to underlie the experienced difficulties in constructing complete ICAI systems. However, as was demonstrated in the earlier chapters, less traditional CAI may be developed by making use of both empirical based notions of particular problem-solving processes and computer models simulating these processes. Two prototypic computerized training-programs for word-problem solving were thus developed. The first computerized training-procedure (Van Lieshout, 1986, submitted) focused on reading the problem text carefully and remediating weak metacognitive control, whereas the second computerized training-procedure (chapter 3) concerned representing the semantics of word problems concretely. Although these computerized instruction-programs are still far from being intelligent, these programs do encompass teaching components which seemed to be crucial for adequate word-problem solving and which turned out to be effective in improving the word-problem solving ability of children with learning difficulties. Therefore, it was decided to develop a third training procedure, in which the instruction components of the computerized training-procedures, which were earlier developed, were incorporated. Thus, this third computerized training-procedure teaches children to discover the semantic problem features presented in the verbal text and teaches them to concretely represent the sets, set-relations and actions described by executing the task-strategy steps, described in a "planning list", in a specified order.

## Design of the computerized training-program

In this section, the architecture of the computerized system for arithmetic word-problem solving is described. Since the significance of linguistic information processing as well as schemata activation for representing problem structure is stressed in both empirical and computer-simulation studies, the computerized training-system is directed towards analysing and representing problem structure. As can be inferred from the computer-simulation studies, problem-text analysis and schema activation seems to be a cyclic process rather than a hierarchical process. The psychological validity of such a cyclic process is empirically supported by the fact that normally performing children, after a first reading of the problem text, tend to reread text parts in which sets, set-relations ("together", "more than") or actions ("lost", "got") are presented (De Corte & Verschaffel, 1986; Van Lieshout & Jaspers, 1989). These findings point to a complex solution process in which the construction of an appropriate problem representation is realized by processing and reprocessing the verbal text. Moreover, beginning problem solvers seem to represent the sets described in word problems on the basis of a sentence-by-sentence approach, i.e. they model the sets in the same order as these sets are presented in the problem text. Consequently in our computerized training-program children learn to solve word problems by alternately processing semantic information in the text and representing the sets described with visual squares. The children are thus taught to identify the first set described (in order with its presentation in the problem text) and subsequently to represent this set. Then, the second set has to be located in the verbal text and represented with visual squares. Both identifying sets in the problem text and representing these sets can be accomplished by touching words in the problem text c.q. visual squares in an initial supply of squares. When a child touches words in the problem text, which is depicted on a touch-sensitive screen, these words are outlined. In this way, the child can keep track of the words touched. After a child has analysed a particular sentence, in which a set is described, the child's response is verified by the computer and the child is given feedback on the precision of his or her response. For a more elaborate account of the contents and examples of screen displays, see Jaspers and Van Lieshout (1989c).

## Objectives of the study

The main question of the study performed with this computerized training-procedure concerned the efficacy of this new computerized training-procedure. It was hypothesized that this new computerized training-procedure would also be effective, as all instruction components, which proved effective in improving the word-problem solving performance of children with learning difficulties in previous studies, were included.

Therefore we first examined the efficacy of this computer program. As such, to circumvent the burden for the child of several repeated testing procedures to establish both direct and transfer effects, only the transfer effect of the tutorial program was studied in conditions without text-analysis and modeling facilities. It was assumed that in this situation children would be handicapped in solving the various word problems as long as they had not cognitively incorporated the problem schemata induced by the training procedure. Second, we were interested in the specific errors made by the children in executing the various strategy steps in order to contrast these errors with errors made by the children in the studies of De Corte and Verschaffel (1981, 1987) and Carpenter and Moser (1982, 1984) as well as with the errors performed by CHIPS (Briars & Larkin, 1984). This could aid in defining a psychological model of the word-problem solving processes of children with learning difficulties and in testing the CHIPS model as a viable model of arithmetic word-problem solving by children with learning difficulties.

# Method

## Design and subjects

Five educable mentally retarded children, two boys and three girls, were trained in random order in a modified multiple-baseline design across subjects (Kratochwill, 1978). The ages of the subjects ranged from 8 years and 7 months to 12 years and 7 months old (mean age 10 years and 3 months). The subjects understood concepts as "more than", "less than" and "altogether" and had more than 80% correct on a technical-reading test (CITO Technical-Reading 2, 48 items) and less than 20%

correct on a word-problem test, consisting of the problem types offered in all sessions in this experiment (Word-Problem Test, 10 items).

## Materials

In each session 10 simple word problems were presented. To decrease the probability of correctly solving the problems casually by applying simple tricks, an irrelevant sentence that contained a third, but irrelevant, number was inserted in the problem statement. The following standard problem types were offered: change 2, 3 and 4, combine 1, 2a and 2b and compare 1, 2, 3, and 4 (see Chapter 3, Figure 3.2). The position of the irrelevant information was varied systematically over the problem types. As such, for problem types change 4, combine 2b and compare 1 the first sentence contained the irrelevant number. For problem types change 3, combine 2a and compare 3 the second sentence contained the irrelevant number. Finally, the irrelevant number was included in the third sentence in the remaining problem types (change 2, combine 1, compare 2 and compare 4).

Computational errors were kept at a minimum by restricting the size of the numbers and sums to 9. None of the numbers stated in the problem text corresponded with the correct answer. Moreover, merely executing the mathematical operation required on the two relevant numbers in the problem produced the correct answer.

### Hardware and software

The computer program ran on an Apple II GS microcomputer with 512 Kb connected with a Philips Touchscreen Monitor VP 120. The computer program had been written in TML Pascal, using the Quickdraw Tool from the Toolbox to animate the softkeys and visual objects and using the 320 mode to project text with the aid of letters that are bigger than the standard typeface.

## Procedure

The procedure consisted of a baseline period, followed by two instruction sessions, a training phase including training trials which were preceded by probe trials and finally a retention period. During the baseline period, the probe trials of the training phase and retention period, the child had to solve the word problems presented by entering a numerical answer on the touchscreen. This answer was followed by computer feedback, which only informed the child whether the solution was correct. During the training trials the child had to analyse the problem text and to represent the relevant sets with visual squares in a cyclic process. For each single problem type, a set of crucial words as well as a particular allocation of visual squares per relevant sentence was defined. The computer feedback in the training trials consisted of "hints" that offered help when the child erred in performing the various strategy steps. In each strategy step, three opportunities were offered to the child for correcting errors. The third time a strategy step was incorrectly executed, the correct execution was provided by the computer.

In the two instruction sessions both the course of actions required and the feedback contents that followed specific mistakes were shown to the child on the touchscreen by a human trainer. After instructing the child, the human trainer refrained from intervening during the succeeding training trials. Each training trial was preceded by a probe trial in which the child first had to enter a numerical answer and successively had to insert this numerical answer in the text. During the training trials the child had to perform the instructed course of actions for all problem types independently. The child had to announce which action was to be performed next by pointing at step labels which were depicted on the screen. These step labels briefly described the actions of the task strategy, which were (step 1) read the problem text, (step 2) select the crucial words in the question sentence, (step 3) select the crucial words in the first relevant sentence, (step 4) represent the set of the first relevant sentence with visual squares, (step 5) select the crucial words in the second relevant sentence, (step 6) represent the set of the second relevant sentence with visual squares, (step 7) locate the answer set by pointing at the squares that form the answer set and finally, (step 8) enter a numerical answer. These actions were to be accomplished by pointing at words in the problem text, by pointing at visual squares depicted in a supply of squares and by pointing at one of the numbers displayed on the screen. For strategy step 2, 3, and 5, in which crucial words in a particular sentence had to be touched, children were prompted to touch words in the relevant sentence only. In doing so, several errors were distinguished. For example, the child could have touched words in sentences that described a relevant set or set-relation, that had to be touched in a previous or subsequent problem-solving stage. These words were then outlined in the color orange. When a child touched words in the sentence describing an irrelevant set, these words were outlined in the color red.

Finally, if the child touched words that described the relevant set which had to be located at that particular moment in the problem-solving route, these words were outlined in the color green. Additionally, the child was informed of the meaning of the different colors ("The red words are wrong. The orange .....") and hinted at the correct sentence that described the set relevant for that moment. When a child proved to be able to select words in the relevant sentence independently but could not yet select the crucial words in this sentence, the computer hinted that only these words should be selected. After a particular set had been identified by the child, the child had to model this set with visual squares. The way in which sets presented in the problem text could be modeled was similar to the method described in chapter 3. Each execution error was followed by feedback, which was specific for the error made. When a child proved to master a particular problem type by correct scoring in three successive probe trials, training for this particular problem type was stopped until this series of correct answers was again interrupted by an incorrect answer. For a full report of the computerized training-program, we refer to Jaspers and Van Lieshout (1989c).

## Results

### Probe sessions

Figure 4.1 shows a performance increase from baseline to training phase for subjects 3, 4 and 5. For the third subject the mean performance of correct solutions rose from 4.4 in the baseline to 6.9 in the training phase. These percentages were 0.6 and 2.9 for subject 4 and 2.5 and 5.4 for subject 5.

The data were statistically analysed by the computer program TIDA (Oud, Reelick, & Raaymakers, 1986). TIDA fits two polynomial curves, one through the baseline and one through the probe data. Subsequently, TIDA tests whether the fitted curve of the probe data significantly diverges from the extrapolated fitted curve of the baseline data. To this end, TIDA first computes the mean performance across all subjects for each baseline and probe score and uses only the largest common number of baseline and probe scores just before and after the point of intervention.These data are depicted in Figure 4.2. The analysis revealed neither a significant constant term nor a significant linear component ($F(1,4)=3.66$, n.s. and $F(1,4)<1$, n.s.).

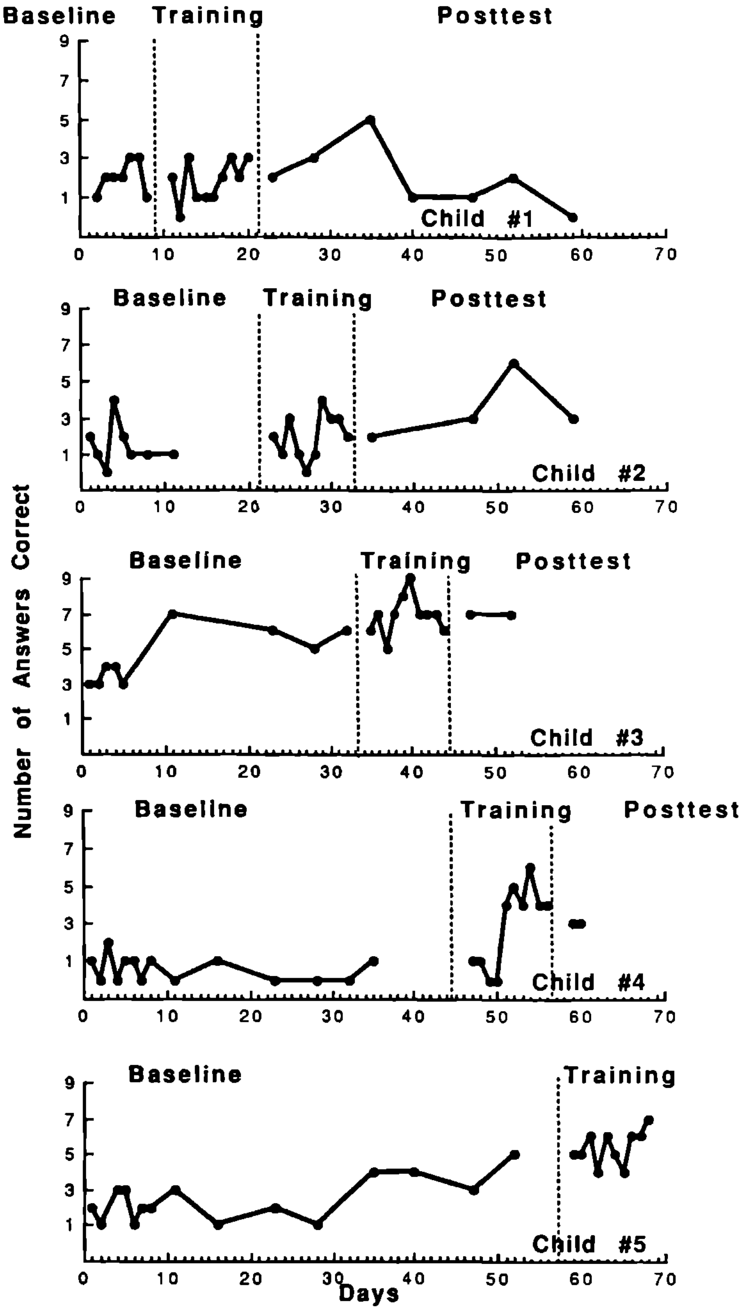


*Figure 4.1* Number of correctly solved problems in each session for each experimental condition for each of the five children.

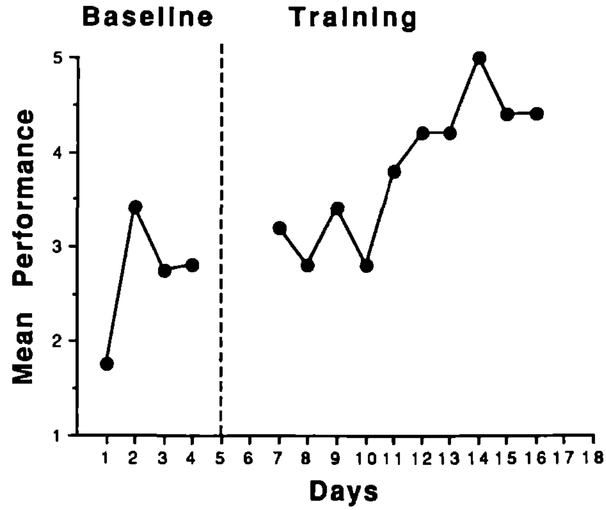


*Figure 4.2* Mean number of correctly solved problems across all subjects in the largest common number of baseline and probe scores.

However, visual inspection of the individual curves (see Figure 4.1) indicates a trend in the baseline phase of subject 3. Actually, statistical analysis of the baseline data of subject 3 revealed a significant trend (Tryon, 1982, $z$=1.77, $p$<.05). It was found that subject 3 had been ill at the start of the experiment, which could have resulted in the low starting performance of this subject in contrast to her performance following her illness. Therefore it was decided to remove the starting baseline sessions of all subjects in a second analysis by TIDA. In contrast with the first analysis, this analysis showed a significant linear component ($F$(1,4)=9.37, $p$<.05) indicating a significant change from the baseline to the training phase.

## Training trials

Several analyses were performed on the data of the training trials. First, to explore the pattern of changes in performing each strategy step, monotone-trend analyses were performed across all error types within each strategy step. For all strategy steps these monotone-trend analyses showed a significant decrease in the number of errors as a function of trials, except for strategy step 4 (see Table 4.1). However, it should be noted that strategy step 4 already showed a floor-effect and that strategy step 8 showed a near floor-effect in number of errors made at the start of the training trials. Therefore it is possible that a significant z-score could not be found for strategy step 4. Despite the near floor-effect of strategy step 8, a negative trend was found for this step, i.e. children became somewhat less proficient in entering the right numerical answer on the screen. Although the mean number of errors seems high for some strategy steps, the total number of errors made relative to the potential number of errors was 32%.

Table 4.1 Mean Number of Errors Committed in Each Strategy Step Averaged Across All Training Trials and Results of the Trend Analyses

| Strategy step | *M* | *z* | *p* |
|---|---|---|---|
| 2. Select cruc. words quest. sent. | 9.1 | 3.03 | <.01 |
| 3. Select cruc. words 1st rel. sent. | 9.9 | 2.29 | <.05 |
| 4. Represent 1st rel. set | 2.9 | <1 | n.s. |
| 5. Select cruc. words 2nd sent. | 15.0 | 2.00 | <.05 |
| 6. Represent 2nd rel. set | 17.0 | 1.94 | <.05 |
| 7. Locate answer set | 9.9 | 2.15 | <.05 |
| 8. Enter numerical answer | 0.5 | 2.13 | <.05 |

*Note.* Strategy step one (read the problem text) was not analysed since no errors were committed in this step.

Second, monotone-trend analyses were performed on specific error types within each step to examine which particular error types showed a marked decline. The results will be reported in two clusters of steps: text-analysis steps and representational steps.

### Text-Analysis steps

The most prominent error type for strategy steps 2, 3, and 5 was touching words in other sentences than the relevant sentence (see Table 4.2). For strategy steps 2 and 3 the decrease in the number of errors made is partly attributable to a significant decline ($z$=2.00, $p$<.05 and $z$=2.44, $p$<.01 respectively) in errors that stem from selecting words outside this relevant sentence (see Table 4.3). Moreover, strategy steps 2, 3, and 5 showed a significant increase in touching only the words that formed the set of crucial words defined (see Table 4.3). It should be noted that

Table 4.2 Mean Percentage of Errors Committed for Different Error Types in the Text-Analysis Steps

| | Strategy step | | | |
|---|---|---|---|---|
| Error type | 2 | 3 | 5 | Total |
| Outside relevant sentence | 19 | 35 | 36 | 90 |
| Other error types[a] | 3 | 2 | 5 | 10 |
| Total | 22 | 37 | 41 | 100 |

[a] (a) Words touched both inside and outside relevant sentence and (b) Words touched inside relevant sentence, but not all crucial words defined.

the significant decline in words touched outside the relevant sentence does not automatically result in a significant increase in the words touched that belonged to the set of crucial words, since this word set merely represented a subset of all words in the relevant sentence.

Table 4.3 Trend Analyses of the Text-Analysis Steps of the Most Prominent Error Types

| | Words touched outside relevant sentence only (decrease) | | Set of crucial words touched within relevant sentence (increase) | |
|---|---|---|---|---|
| Strategy step | *z* | *p* | *z* | *p* |
| Touch crucial words in: | | | | |
| Question sentence | 2.00 | <.05 | 2.80 | <.01 |
| First relevant sentence | 2.44 | <.01 | 2.44 | <.01 |
| Second relevant sentence | 1.26 | n.s. | 1.70 | <.05 |

### Representational steps

The most prominent error type for strategy steps 4, 6, and 7 was moving or pointing at a wrong number of visual squares, i.e. a number of squares moved that was not in accordance with the required amount (see Table 4.4). For strategy step 4,

Table 4.4 Mean Percentage of Errors Committed for Different Error Types in the Representational Steps

| | Strategy step | | | |
|---|---|---|---|---|
| Error type | 4 | 6 | 7 | Total |
| Wrong no. of objects | 11 | 39 | 41 | 91 |
| Other error types[a] | 1 | 8 | 0 | 9 |
| Total | 12 | 47 | 41 | 100 |

[a] (a) Wrong manipulation (adding objects instead of separating objects) and (b) Wrong allocation of objects and (c) Wrong color of objects.

in which the first set had to be represented, this wrong number of squares mainly corresponded with the irrelevant number stated in the problem text across all problem types except for compare 2 and 4 problems. However, this error type represented a mere 4% of the total amount of errors across all representational steps. Besides, this error type became less frequent as the training trials progressed ($z=1.82$, $p<.05$). For compare 2 and 4 problems, children often moved a number of squares that corresponded with the second relevant number of the problem text. For this category the decline in the number of errors made was not significant ($z=0.84$, n.s.).

In strategy step 6, in which the second set had to be represented, for each problem type the most important finding was that subjects made all kinds of errors (see Table 4.5). Still, 73% of the errors was due to making a set of squares that was equal to the second relevant number given for problem types in which this number never had to be directly modeled. Actually, subjects were rather persistent in committing this type of error since for this error category the decrease in number of errors was not significant ($z=1.20$, n.s.).

In strategy step 7, in which the answer set had to be located, across all error types children often pointed at a number of squares that corresponded with the second relevant number of the problem text (see Table 4.6). This error type was manifold for combine 1, combine 2b, compare 1 and 2 problems. On the other hand, for compare 1 and 2 problems subjects also often committed "first number errors",

Table 4.5 Mean Error Percentages for Each Problem Type and Error Type Within Strategy Step Six (Represent the Second Set)

| | Problem type | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Change | | | Combine | | | Compare | | | | Total |
| Error type | 2 | 3 | 4 | 1 | 2a | 2b | 1 | 2 | 3 | 4 | |
| Number of objects equal to: | | | | | | | | | | | |
| Irrelevant number | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 0 | 3 | 3 | 15 |
| Number 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Number 2 | 3 | 9 | 6 | c | 10 | 5 | c | c | 28 | 12 | 73 |
| Subtotal | 5 | 11 | 7 | 1 | 11 | 6 | 2 | 0 | 31 | 15 | 89 |
| Other errors | 0 | 1 | 2 | 0 | 1 | 1 | 1 | 1 | 2 | 2 | 11 |
| Total | 5 | 12 | 9 | 1 | 12 | 7 | 3 | 1 | 33 | 17 | 100 |

*Note.* c = correct execution.

Table 4.6 Mean Error Percentages for Each Problem Type and Error Type Within Strategy Sep Seven (Locate the Answer Set)

| | Problem type | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Change | | | Combine | | | Compare | | | | Total |
| Error type | 2 | 3 | 4 | 1 | 2a | 2b | 1 | 2 | 3 | 4 | |
| Number of objects equal to: | | | | | | | | | | | |
| Number 1 | 0 | 3 | 0 | 6 | 4 | 2 | 11 | 11 | 3 | 3 | 43 |
| Number 2 | 5 | 1 | 2 | 8 | 1 | 8 | 11 | 10 | 2 | 1 | 50 |
| Subtotal | 5 | 4 | 2 | 14 | 5 | 10 | 22 | 21 | 5 | 4 | 93 |
| Other errors | 0 | 0 | 0 | 2 | 0 | 2 | 1 | 2 | 0 | 0 | 7 |
| Total | 5 | 4 | 2 | 16 | 5 | 12 | 23 | 23 | 5 | 4 | 100 |

i.e. pointed at a number of squares that corresponded with the first relevant number of the problem text. Both "first number errors" and "second number errors" declined in number as the training trials progressed ($z=2.11$, $p<.01$ and $z=2.62$, $p<.01$, respectively).

## Discussion

Visual comparison of the data from the baseline sessions, the probe trials of the training phase and the posttest showed that the computer-assisted training procedure was only slightly effective in teaching the educable mentally retarded children of this study to solve word problems. The change in linear trend from baseline to training phase proved to be significant only after removing the starting baseline data in the analysis. So, what could be the main reason for the fact that this computerized training-procedure did not show the expected effectiveness? For this training procedure encompasses instruction components that proved to be essential components of the word-problem solving process in both the empirical studies cited and in the computer-simulation models. Moreover, these instruction elements separately brought about a significant performance increase in earlier training experiments (Van Lieshout, submitted, chapter 3). The principal reason could be that the incorporation of so many strategy steps, which accompany the text-analysis and representational process, obscured the child's understanding of the problem-solving strategy instructed. Since the computerized training-procedure demands that a lot of distinct actions be performed, the child may no longer have a clear view of the relationship between the different steps and the ultimate goal of the strategy. Additionally, it could be hypothesized that the fixed number of training trials was not sufficient to establish a robust training effect. In order to obtain a higher performance increase some subjects may have needed more than ten training trials. If more than ten training trials had been offered, or otherwise, if the training would have been less complex, children might have acquired more insight in the task strategy. To meet these objections, a revised computerized training-procedure has been developed, in which the number of actions to be performed is reduced, but in which training of text analysis and concrete modeling with squares is preserved. Yet, a third reason could be that as a result of the trend in the baseline of subject 1 and 3, which was even significant for subject 3 and almost reached significance for subject 1, a more rigorous training effect could not be established. This practice effect might have occurred incidentally and might have been absent if other subjects had been selected.

Despite the difficulties that the children experienced in performing the various strategy steps, most steps revealed a significant decrease across all prominent error types as the training phase progressed. This general decline in errors indicates the onset of the children's adherence to the instructed task strategy.

The educable mentally retarded children of this study stem from a different population than the children of the studies of De Corte and Verschaffel and Carpenter and Moser. As such, the types of errors committed by our subjects may not be comparable to the errors committed by regular school children. Accordingly, CHIPS could be an accurate model to simulate the problem-solving processes of normally achieving novices, whereas this same model needs adjustment and enlargement of its knowledge structures to account for the psychological processes of children with learning difficulties. Still, although the subjects of our study presumably showed a larger range of different error types than the subjects of the empirical studies cited, some of these prominent error types are also committed by poorly performing regular school children. For example, after reading the problem for the first time, those children reviewed non-numerical facts, such as crucial words, less frequently than the high ability children (De Corte, & Verschaffel, 1986). In comparison, the subjects of this study were not able to select crucial words in the relevant sentences at the start of the training phase but learned to do so during training, which can be inferred from the significant increase in solely touching words that form the set of crucial words defined. Likewise, in encoding word problems successfully as belonging to one of the main problem types change, combine or compare, CHIPS processes crucial words such as "gave" , "together", "more than" or "less than" and utilizes this information in building adequate representations by making inferences. Without this knowledge, CHIPS produces wrong answers.

Additionally, some error types committed by our subjects in the representational steps seem to be comparable to errors committed by poorly performing school children. As such, a prominent error type encountered in the representational steps was moving or pointing at a number of squares that corresponded with the second relevant number of the problem text for problem types in which this number never had to be directly modeled. For example, this error type was often committed for change 3, combine 2a, and compare 3 and 4 problems in strategy step 6 (represent the second set) and for combine 1, combine 2b, compare 1 and 2 problems in strategy step 7 (locate the answer set). Likewise, Verschaffel (1984) encountered a lot of "second number" errors for these problem types in his study concerning the problem-solving processes of young regular school children. CHIPS also produces "second number" answers for these problem types when CHIPS does not have the knowledge required to solve these problems.

Protocol analysis revealed that some of the subjects of Verschaffel interpreted the change 3 problem "Ann had 5 marbles. Peter gave Ann some more marbles. Now Ann has 8 marbles. How many marbles did Peter give Ann?" as "Ann had 5 marbles. Peter gave Ann 8 marbles. How many marbles did Peter give Ann?" and accordingly represented the second relevant number instead of supplementing the first relevant number with objects until this second number was reached. Finally, these children answered with the second number given. In contrast, although CHIPS comes up with the "second number" as the final answer for change 3 when certain knowledge is lacking, in response to processing the second and third sentence of change 3 problems, CHIPS accurately adds squares to the first set until the total set equals the second number. However, CHIPS does not keep track of the number of blocks added and counts the total set to calculate the (wrong) answer. Thus, the subjects of our study are comparable to the poorly performing children of the study of Verschaffel but apparently for change 3 problems CHIPS needs to be revised in order to produce representational errors similar to those encountered with children.

Likewise, for combine 2a problems, ("Mary has 2 apples. Peter also has some apples. Together Mary and Peter have 6 apples. How many apples does Peter have?") the representational errors committed by our subjects conform to the representational errors of the poorly performing subjects of Verschaffel. Both the educable mentally retarded children and the poorly performing regular school children again represented the second relevant number ("6") instead of adding objects until the total set equals this number. Since the only known number set that is associated with Peter is the total set described ("6") children possibly interpret "Together Mary and Peter have 6 apples" as "Mary has 6 apples and Peter has 6 apples" and consequently add 6 objects to represent the set of Peter (Verschaffel, 1984). Accordingly, when CHIPS lacks "set-superset" language knowledge, CHIPS interprets the final set as a set belonging to Mary (and represents this set) and the second set described as a set belonging to Peter and Mary and represents this set. Finally, CHIPS answers this combine 2a problem incorrectly with "6", counting the only set it has associated with Peter. In the case of this problem type our results agree both with Verschaffel and CHIPS.

Finally, Verschaffel argued that for compare problems "second number" answers stem from misconceptions concerning the interpretation of the sentence in the problem text describing the "more" or "less" relation of a set with regard to another set. So, the number in the second sentence described (italicized in the next example) in compare 4 problems like: "Ann has 8 apples. *Peter has 5 apples less than Ann.* How many apples does Peter have?" may be regarded as the size of the answer set since children interpret this second sentence as "Peter has 5 apples" and "Peter has less apples than Ann" (Verschaffel, 1984). Accordingly, the "second number" errors that the subjects committed in our study may proceed from a similar misinterpretation of this second sentence describing the (unknown) set of Peter. Since, as a consequence of this misconception, no calculation is required in determining the number of the set of Peter, children may have represented the number described in the problem text ("5") directly. Indeed, for all compare problem types children seemed rather persistent in their misconception since, after the computer had displayed the requested number of squares representing the set of Peter, in determining the answer set they still pointed at a number of squares that corresponded with the second number described. For the example of the compare 4 problem this response even resulted in pointing at squares in the two different sets displayed, because the number of squares representing the set of Peter (three) was insufficient. Accordingly, CHIPS responds to compare 1, 2, 3 and 4 problems with the second number given if CHIPS does not have the ability to understand so-called "comparison" language. In case of the example given, CHIPS too uses each number in the problem to create a set and subsequently responds with the number of squares of the second set, the only squares it has associated with Peter.

Combine 1 and 2b problems also elicited "second number" errors in representational step 7 (locate the answer set). Although combine 1 problems in general are easy to solve, some children in the study of Verschaffel came up with two numbers as answer, the first relevant number and the second relevant number. Verschaffel believes that the children interpret the word "together" as " each " and interpret a sentence as "How many apples do Peter and Mary have together?" as "How many apples does Peter have and how many apples does Mary have?" Thus, the answer to this question is given by mentioning the two relevant numbers. However, if the subjects of our study had misunderstood combine 1 problems in conformance to the poorly performing children of Verschaffel, they would have pointed at the objects belonging to Peter <u>and</u> to the objects belonging to Mary. Eventually, this would casually have resulted in the correct answer in strategy step 7.

Instead, the subjects that made errors in our study at one time primarily pointed at the objects belonging to Mary, at another to the objects belonging to Peter.

To solve combine 1 problems CHIPS must recognize the static situation described to translate it into a dynamic situation, i.e. CHIPS must recognize that the objects of two sets are moved together to form a superset. When CHIPS is lacking such knowledge, it still creates two sets, but would not move both sets into a new set and count the resulting set. Unfortunately, it is unknown what answer CHIPS would give, since CHIPS always solves combine 1 problems correctly and this simulation was never run.

Combine 2b problems were often solved by pointing at the objects of the second relevant number set. Overall, combine 2b problems are more difficult to solve than combine 1 problems. Since Verschaffel did not use combine 2b problems and Carpenter & Moser (1982) did not report on qualitative data concerning the kind of wrong answers produced it is unknown whether our subjects are comparable with normally achieving school children. Nevertheless, the "second number" errors produced for combine 2b problems cannot be explained by the same misinterpretation as posited for combine 2a problems, since the order of introduction of set and superset is reversed in combine 2b problems (e.g. "Together Mary and Peter have 6 apples. Mary has 2 apples. How many apples does Peter have ?"). Likewise, the educable mentally retarded children of an earlier study (Jaspers & Van Lieshout, 1990) manifested the same kind of error for combine 2b problems, i.e. they most frequently answered wrongly with the "second relevant number" (in the example with "2"). However, without "set-superset" language CHIPS produces "first relevant number" errors ("6" in the example) as answer to these problem types since as with combine 2a problems CHIPS counts the only set it associated with Peter. As such, the errors produced by CHIPS do not fit the wrong answers produced by the educable mentally retarded children of our studies.

Nevertheless, large scale research in which no intervention in the solving route of the child is undertaken seems to be required to investigate the word-problem solving processes of children with learning difficulties. This could eventually answer the question whether children with learning difficulties are comparable with poorly performing school children in their knowledge of arithmetic word-problem solving. Recently, a study in which the final answers of children to word problems with unique number combinations are diagnosed, has been started. In the future we also plan to diagnose the errors of these children committed during word-problem solving. Furthermore, we plan to rebuild a CHIPS-like simulation program in such a way as to simulate the word-problem solving processes of children with learning difficulties. To accomplish this, simulation experiments will be performed in which the kind of errors produced by the simulation model are examined after both removing certain knowledge components and adding certain misconceptions that children with learning difficulties presumably hold. Our final goal is to develop computerized instruction-systems, which first diagnose the misconceptions and lacking knowledge of children with learning difficulties and second remediate pupils on an individual basis. This study is only a very modest starting point in this construction process.

# Chapter 5

# The Efficacy of Prototypes of CAI for Arithmetic Word-Problem Solving

## Introduction

Theoretical models of word-problem solving stress the importance of constructing adequate problem representations. In order to construct such a representation a child first has to identify the relevant sets and relations or actions described in the problem text. This requires a thorough analysis of the problem text. In doing so, external modeling of the sets and actions or relations described may aid in analysing and solving the problem, for external representations allow the permanent reinspection and recounting of the number sets already represented. As such, external representations may substitute internal representations and thus relieve working memory. Besides, modeling word problems concretely may give a deeper insight in the sets and set-relations or actions described and thus may set an example and as such be a useful start in constructing corresponding internal problem schemata.

In the previous chapters various pilot studies were reported in which the effectiveness of two computerized training-procedures for word-problem solving was demonstrated. In the first computerized training-procedure, children learned to represent word problems with concrete aids. In the second computerized training-procedure, children were instructed both to analyse the problem text carefully and subsequently to represent the sets described by concrete aids.

However, these pilot studies merely served as preliminary small-scale investigations of the usefulness of the computerized training-versions as tools in fulfilling our main goal, i.e., to acquire a deeper understanding and knowledge of, first, by which particular training components the word-problem solving deficits of children with learning difficulties may be improved and, second, which problem-solving processes may be influenced by each training component. Besides, up to now the computerized training-procedures were studied on their effectiveness in separate experiments. Therefore statements as regards the differential effectiveness of these training procedures would not be valid. Moreover, the contributions of the Text-

Analysis component on children's achievements was not studied in isolation in the pilot studies. As a consequence, to what extent this instruction component contributed to the overall training effect cannot be judged. Finally, since the sample in each pilot study was rather small, it is questionable whether the subjects were representative for the total population of educable mentally retarded children.

To overcome these weaknesses of the preceding research a training experiment with larger groups was conducted, in which both computerized Text-Analysis instruction and computerized External-Modeling instruction were studied on their efficacy. In order to examine the share of the Text-Analysis as well as the External-Modeling component in children's achievements separately, the presence of the Text-Analysis and External-Modeling component was varied systematically. Another important research question was whether Text-Analysis and External-Modeling instruction would influence specific problem-solving processes differently.

## Modifications in the computerized training-procedures

Both the computerized training-procedure for improving text analysis (see chapter 1, Van Lieshout, submitted) and the computerized training-procedure for improving text analysis as well as concrete modeling (see chapter 4, Jaspers & Van Lieshout, submitted b) encompassed a metacognitive instruction component; children were taught to pay attention to the successive steps of the task strategy by pointing at step labels which briefly described the content of each step. Besides the absence of adequate information-processing strategies, a lack of metacognitive skills has become one of the main explanations of the poor academic performance of children with learning difficulties (Hall, 1980; Sternberg, 1981; Slife, Weiss & Bell, 1985).

Although it is generally assumed that presence of adequate metacognitive skills aids in guidance of the problem-solving process and as such is crucial in finding solutions for problems, there were several reasons to abandon the instruction of metacognitive skills in this experiment. First, one of our major interests was to study the individual contribution of each instruction component on achievement scores and process measures. Consequently, the presence of each instruction component had to be varied systematically across the training procedures. To accomplish this goal, the supplement of a metacognitive instruction component would have required the development of additional computerized training-procedures. Besides, this increase in the number of training procedures would have demanded a larger sample size than that which we had at our disposal.

Second, incorporation of this metacognitive component in the Text-Analysis and External-Modeling training would have required modification of the contents of the step labels of each training. This would have confounded the assessment of the contribution of the metacognitive component and the other instruction components to the overall training effect.

Finally, the results of three pilot studies, in all of which this metacognitive instruction component was incorporated (Van Lieshout, submitted; Jaspers & Van Lieshout, submitted b; Dankers, 1990) showed conflicting results. In both the study of Van Lieshout (submitted) and the study of Jaspers and Van Lieshout (submitted b) the percentage of incorrect choice responses in the planning list, representing the metacognitive component, decreased during the training sessions. In contrast, the subjects in the study of Dankers (1990) did not improve in choosing the correct step label during the training sessions. Therefore, it is uncertain whether the planning list with the different step labels fulfilled the presumed metacognitive function, i.e. to make children aware of the successive actions of the task strategy to be performed.

Another important finding of the pilot studies described in chapters 3 and 4 was that children experienced some difficulties in learning to use different representation strategies for semantic different word-problem types. The main reason could be that the difference in the semantic problem types was reflected in the actions instructed rather than in the final representation on the screen after each step had been performed. Change problem types, in which visual blocks had to be separated from a start set could not longer be distinguished as such from combine problem types, in which two separate sets had to be represented on the screen (see Figure 3.2). Also, the function of the use of a different colored set to represent the (unknown) number of the added set in change 3 problem types ("Peter had 5 marbles. Ann gave Peter some marbles. Now Peter has 8 marbles. How many marbles did Ann give to Peter?") often remained unclear (i.e. to mark off the answer set).

To compensate these difficulties, several changes were made in the representational strategies instructed. The result of these changes was that all arrow-keys were removed from the computer screen. To locate the required position of the second set, the child had to touch specific areas on the "work-sheet" instead of arrow-keys. Besides, the representation of the second set was divided into two separate steps: a child first had to locate the required place of the second set and second had to represent this set with visual squares. The changes made in the contents of the training procedure will be further described in the section "Training". As a result, in nearly all cases, the main problem types change, combine and compare could be distinguished by the final representation on the screen.

A final modification in the computerized training-procedures concerned the feedback contents that followed children's mistakes in representing the sets with squares on the computer screen. The feedback that followed errors in representing sets in the computerized-training studies of chapters 3 and 4, informed the child e.g. that the number of squares moved was wrong but did not inform the child why this number was incorrect (see for example chapter 3, Figure 3.1, the feedback reads "4 blocks is not right"). Hence, children could have experienced difficulties in relating the feedback with their mistake.

To prevent such difficulties, the feedback contents of the computerized-training programs of this study were altered so as to describe clearly why, e.g., the number of squares moved was not right at that particular moment of the problem-solving route by referring back to the problem text. To give an example, when the child had to model the set described in the sentence: "Mary had 6 apples" and only had moved 4 squares to the work-sheet, the feedback would read "Did Mary have 4 apples? Move more blocks until Mary has 6 apples!" It was expected that these changes in feedback would give children better insight in their representational errors.

## Hypotheses

Thus, the main focus of this training experiment was to evaluate the efficacy of the computerized training-procedures for word-problem solving. Hence the most important question of this study was whether children with learning difficulties would improve their ability to solve word problems as a result of one of the training procedures.

*Hypothesis 1:*
Children who receive Text-Analysis or External-Modeling instruction will improve their ability to solve word problems correctly.

A second important question concerned the differential contribution of instructing Text Analysis and External Modeling on both the performance and problem-solving processes of educable mentally retarded children. To that end, the effect of Text-Analysis instruction and the effect of instruction in External Modeling were examined both in isolation and in combination. To see whether these instruction components would influence the word-problem solving process differently, several achievement and process measures were constructed.

In order to examine whether children learned to generalize the strategies instructed to situations without Modeling and Text-Analysis instruction and to untrained problem types, two paper and pencil word-problem tests were constructed. The first paper and pencil test consisted of word-problem types, which were also used during training ("Trained Word-Problem Test"). In line with the earlier mentioned empirical and theoretical insights with regard to the importance of constructing problem representations, it was assumed that Text-Analysis instruction would not explicitly direct the child's attention to the construction of a representation of the problem. Text-Analysis instruction would rather let the child pay attention to the information necessary for building such a representation. As such, Text-Analysis instruction may aid in constructing problem representations internally, but these problem representations were not externalized and thus not instructed explicitly. In contrast, instruction in External Modeling actually directs the child's attention towards constructing proper problem representations for the semantic different problem types. Since constructing a proper problem representation is considered the most important phase in the word-problem solving process, it was hypothesized that the children who receive instruction in External Modeling would perform better on the paper and pencil test with trained problem types than children who merely receive instruction in Text Analysis.

*Hypothesis 2:*

Children who receive instruction in External Modeling will improve their ability to solve the trained word-problem types correctly more than children who receive Text-Analysis instruction.

The second paper and pencil test, however, consisted of new word-problem types for which the children had not received training ("Untrained Word-Problem Test"). This test was administered to explore the transfer value of the training procedures to untrained word problems. Since Text-Analysis instruction aims at constructing adequate internal representations of the word problems by determining the unknown set in the question sentence and subsequently the relevant sets in the sentences preceding the question sentence, Text-Analysis training mainly involves the thorough inspection of the problem text. In contrast, External-Modeling instruction aims at developing proper representations by constructing different external representations for each word-problem type. As such, Text-Analysis instruction is less specific with regard to the word-problem types used than External-Modeling instruction and thus the Text-Analysis instruction procedure may be used as a general strategy to solve both trained and untrained word problems. It was therefore hypothesized that children who receive Text-Analysis instruction would perform better on the paper and pencil test with new problem types.

*Hypothesis 3:*
Children who receive Text-Analysis instruction will improve their ability to solve untrained word-problem types more than children who receive instruction in External Modeling.

Besides the question whether the Text-Analysis and External-Modeling instruction would have a differential effect on the performance tests, another major research goal was to examine the influence of both these instruction procedures on children's problem-solving processes. Since the task strategy of the training procedures differed as regards the problem-solving route instructed, it was assumed that the training procedures would influence particular problem-solving processes differently. More specifically, Text-Analysis instruction mainly focused on a thorough text analysis, whereas External-Modeling instruction principally centered on modeling sets with visual cubes. To investigate the implied differential effects of both training procedures on both the text-analysis processes and the modeling behavior of the children, two process measures were constructed. The hypotheses 4 and 5 concern predictions with regard to the reading behavior of the children, whereas hypotheses 6 through 9 refer to predictions as regards the modeling behavior of the children without intervention.

In order to observe differences in text-analysis processes, the reading behavior of the children was observed by a technique which has the same purpose as eye-movement recording. The test, which we performed with this technique will be referred to as "Reading-Behavior Test".

Although the main goal of both instruction in Text Analysis and External Modeling is to aid the construction of an adequate problem representation, this is accomplished in different ways. The computerized training-program for Text Analysis aims at constructing an adequate internal problem representation by teaching children to identify the relevant sets and set-relations or actions described in the problem text. This identification of known sets and set-relations is preceded by determining the unknown set described in the question sentence. Thus, the construction of this internal representation starts with identifying the unknown set. Subsequently the known quantities are identified ("Backward Analysis"). In contrast, the computerized training-procedure for External Modeling is directed towards constructing a proper problem representation by actually modeling the sets and set-relations or actions described in the problem text with visual squares. The construction of this external representation proceeds chronologically with the order of introduction of each relevant set in the problem text, i.e. starts with the first relevant sentence and ends with locating the answer set by the time the child arrives at the question sentence ("Forward Analysis").

Inasmuch as the effect of the Text-Analysis component on reading behavior could manifest itself on various aspects of looking behavior during reading, several variables were defined. In their investigation of eye movements during word-problem solving De Corte and Verschaffel (1986) found that the solution process does not occur as a linear sequence of sharply distinguished stages, namely a text-analysis, representational and a computational stage but rather that these aspects seem to alternate and interact. Therefore in our study, a distinction was made between the various phases in the reading process; the first phase concerned the initial reading of the problem text (Initial Reading; IR-phase). The second phase was defined as "reading the question sentence" (Question Reading; QR-phase), in which as a rule the unknown quantity is defined. The final phase encompassed reading behavior that followed the reading of the question sentence (ReReading phase; RR-phase). Within each phase per sentence and across sentences the following was recorded: the gaze duration, total number of words read, total number of different words read. De Corte and Verschaffel (1986) found that some children regularly answered a problem without even glancing at important parts of the problem, such as the question sentence or crucial words such as "more than" which define problem structure. However, this finding was not replicated in a study with our technique for recording looking behavior in which the reading behavior of educable mentally retarded children was compared to the reading behavior of normally achieving children (Van Lieshout & Jaspers, 1989). In this study, all of the children demonstrated complete reading of the problem text. It was therefore assumed that the effect of the Text-Analysis instruction would not manifest itself in a more complete reading of the problem text by children who receive Text-Analysis instruction. Instead, it was hypothesized that children who receive Text-Analysis instruction would, after a first reading of the text, analyse the word-problem texts differently than children who receive instruction in External Modeling. The training procedures differed with regard to the order in which the problem text had to be analysed. The children who receive instruction in Text Analysis learn to analyse the problem text by starting with the unknown quantity described in the question sentence (Backward Analysis). Consequently, after the first reading of the problem text (IR-phase), these children would start analysing the problem text backward, i.e. starting from the question sentence (QR-phase). Thus, it was hypothesized, that instruction in Text Analysis would lead to more looking activities in the QR-phase.

This effect would manifest itself on various aspects of looking behavior in the QR-phase, such as gaze durations, number of words touched, number of different words touched.

*Hypothesis 4:*
Children who are trained in Text Analysis will demonstrate more looking activities in the question sentence (QR-phase) than children who receive instruction in External Modeling.

Most word problems used in the Reading-Behavior Test contained one or more sentences that were irrelevant to the solution of the problem. These irrelevant sentences described an actor possessing some set of objects of a certain quantity. Since hypothesis 4 states that subjects who receive Text-Analysis instruction would analyse the problem text starting with the unknown quantity in the question sentence, it was further assumed that these children would distinguish between the relevant number sets and the irrelevant set(s) sooner than the other children. For the first word in each sentence described the actor and thus merely touching this word was sufficient to identify the irrelevant sentence among the other sentences. Therefore, expert reading behavior would be reflected in less words touched in this irrelevant sentence and consequently in a shorter RR-phase. Children who receive training in External Modeling by chronologically representing the sets described would improve less in determining the irrelevant set in the problem text. These children would analyse the word-problem text in accordance with the order of text analysis that was implicitly instructed during training, i.e. from left to right. Hence, children who receive External-Modeling training would demonstrate more rereadings after the initial reading of the word-problem text; i.e. in the RR-phase in order to determine the relevance of the description of the sets encountered during the initial reading of the problem text.

*Hypothesis 5:*
Children who are trained in External Modeling will demonstrate more rereadings than children who receive instruction in Text Analysis, i.e. more looking activities in the RR-phase in word problems containing irrelevant set descriptions.

Again, it was assumed that this effect would manifest itself on various aspects of looking behavior in the RR-phase, such as gaze durations, number of words touched, number of different words touched.

To observe differences in modeling behavior, a word-problem task in which materials were made available to externally model the word problems presented, was constructed. It was hypothesized that children who received instruction in modeling word problems properly would perform better on this word-problem task. Although the computerized training-procedure for concrete modeling ultimately aimed at activating cognitive problem schemata, which in the end should replace the external representations, it was assumed that by removing the materials some children would to some degree be handicapped in solving certain word-problem types because they may not yet have similar cognitive schemata available. Thus a task was constructed in which materials were offered to concretely model the word problems presented.

*Hypothesis 6:*
Children who receive instruction in External Modeling will improve in number of word problems correctly solved on a word-problem test in which concrete materials are presented to model the word problems.

To examine the effect of the External-Modeling component on the representational strategies of the children, several variables were defined. In defining these variables, we resorted to both empirical data with regard to the concrete modeling patterns of young children (De Corte & Verschaffel, 1981, 1987; Carpenter & Moser, 1982, 1984; Riley, Greeno & Heller, 1983) and our own pilot work (see chapters 2, 3 and 4). The most important finding of this research was that the semantic structure of word problems determines the representational strategy selected by the child. For example, children seem to use different locations on the table to represent different word-problem types and children seem to really add blocks to or separate blocks from an existing set in order to represent an increase or decrease described. It was hypothesized that children who receive instruction in the External Modeling of word problems, would show representational strategies similar to those instructed during training for word-problem types on the Free Representational Test, which were also trained.

*Hypothesis 7:*
Children who receive instruction in External Modeling will show representational strategies which are similar to the representational strategies instructed for trained word-problem types.

In line with the assumption that children who receive Text-Analysis instruction would improve in determining the irrelevant set in the problem text more than children who receive External-Modeling instruction, it was hypothesized that children who receive Text-Analysis instruction would represent this irrelevant number less often than children who receive External-Modeling instruction. Thus, children who receive Text-Analysis instruction would skip the representation of this irrelevant number and start directly with representing one of the relevant numbers of the problem. For these children learned to identify the unknown and known quantities in a non-chronological order, the introduction of relevant sets would not necessarily coincide with representing the relevant sets of the problem in a similar order.

*Hypothesis 8:*
Children who receive instruction in Text Analysis will show representational strategies in which the irrelevant number is skipped and in which the answer given is correct more often than children who receive External-Modeling instruction.

Beyond the assumption that children who receive External-Modeling instruction would not so much improve in identifying the irrelevant set as children who receive instruction in Text Analysis, it was assumed that the order of introduction of the relevant and irrelevant sets in the problem text would coincide with the order in which the children who receive External-Modeling training would analyse and solve the problem. The order of analysing and solving the word problems would match the sequence in which the child would represent the relevant and irrelevant sets. Thus, children who receive External-Modeling instruction will not only show more representational strategies in which the irrelevant number set of the word problem is represented, but will also represent this irrelevant number set in sequence with its occurrence in the problem text more often than children who receive Text-Analysis instruction. Since materials are available to represent the word problems, it is assumed that the children who receive External-Modeling instruction, after they finish modeling the sets described, will inspect the sets represented on the table. Consequently, they will identify the irrelevant set as redundant in finding the solution. Thus, children who receive External-Modeling instruction will solve the word problems correctly, although they represent all sets described.
Despite representing this irrelevant number set, children who receive External-Modeling instruction will nevertheless produce the correct answer.

*Hypothesis 9:*
Children who receive instruction in External Modeling will show more representational strategies in which the irrelevant number of the problem is represented in sequence with its occurrence in the text but in which the answer produced is correct than children who receive Text-Analysis instruction.

Finally, a so-called "Word-Problem-Classification Test" was administered to study whether children may be differentiated on their ability to classify word-problem types on the semantic structure of the problem. As Morales, Shute and Pellegrino (1985) demonstrated, more experienced problem solvers show evidence of problem differentiation according to problem schemata, whereas less experienced problem solvers show a less systematic sorting pattern more in line with the surface structure. In line with this finding, it was assumed that "expert" skill would manifest itself in being able to perceive the underlying semantic structure of the word problems, whereas "beginners" would be more distracted by surface features of the problem. Both Text-Analysis and External-Modeling instruction direct the child's attention towards constructing proper problem representations by making the child aware of the semantic structure of the problem. Unlike the External-Modeling training in which the semantic structure of the word-problem types is reflected in the concrete manipulations instructed, in the Text-Analysis training word-problem types are merely distinguished by the set of words that have to be touched. Hence, it was hypothesized that children who were trained in representing the word-problem types externally would improve in classifying word-problem types according to their semantic structure.

*Hypothesis 10:*
Children who receive External-Modeling training will improve in classifying word problems according to semantic problem type.

## Method

### Subjects

Eighty-four educable mentally retarded subjects from seven different Dutch schools for Special Education participated in the experiment. The ages of the subjects ranged from 8 years and 6 months old to 16 years and 8 months old, with a mean age of 12 years and 4 months old. From each school, 12 pupils were selected who had the lowest performance in word-problem solving but who could read sufficiently (more than 80% correct on a Dutch Technical-Reading test, Cito Technical-Reading 2, 48 items), who understood crucial words as "together", "more than", and "less than" (more than 80% correct on a Dutch Comprehension Test, 6 items). The actual performance of these subjects on a word-problem test, in which word-problem types similar to those presented during training were included ranged from 42%-100% incorrect with a mean percentage incorrect of 84 (Trained Word-Problem Test, 43 items). Since these children were tested on different IQ-tests at very divergent time points, it was decided to refrain from using these IQ-scores in characterizing the children on intellectual ability and in further data analyses.

## Design

The influence of both Text-Analysis and External-Modeling instruction on the performance and the problem-solving process was investigated in isolation as well as in combination. In a 2 x 2 factorial randomized-block design the presence or absence

| Pretest | Training period | Posttest | Follow-up |
|---|---|---|---|
| Trained Word-Problem Test | Text-Analysis training | Trained Word-Problem Test | |
| Untrained Word-Problem Test | External-Modeling training | Untrained Word-Problem Test | Trained Word-Problem Test |
| Reading-Behavior Test | Text-Analysis and External-Modeling training | Reading-Behavior Test | |
| Free-Representation Test | Control condition | Free-Representation Test | Untrained Word-Problem Test |
| Word-Problem Classification Test | | Word-Problem Classification Test | |

*Figure 5.1* Overview of the research design.

of the Text-Analysis and External-Modeling component was systematically varied. This resulted in four experimental conditions:

1. TA: Text-Analysis training
2. EM: External-Modeling training
3. TA-EM: both Text-Analysis and External-Modeling training
4. C: Control, neither Text-Analysis nor External-Modeling training

The selected children were trained in a pretest-posttest control group design with follow-up tests. These pretest, posttest and follow-up tests are of particular interest for testing the main hypotheses. In Figure 5.1 an outline of the design is presented. In order to obtain information regarding the performances of the children during the training period, a probe trial was inserted before each training trial.

Although children in the C-condition did not actually receive instruction in word-problem solving, in an equal amount of training sessions these children engaged in practising on the computer the same number and types of word problems as presented in the other experimental conditions.

## Procedure

The procedure will be discussed in two main sections. In the first section the procedure with regard to the pretest, posttest and follow-up period and the materials used during these test periods will be discussed. The second section describes the procedure followed during the training period.

### Pretest, posttest and follow-up

The first part of the experiment consisted of a classroom administration of a word-problem test in 22 schools for children with learning difficulties. This word-problem test consisted of problem types which were also to be used in training and served both as selection test to determine whether a child was suited to participate in the experiment and as pretest to assess the starting performance of the subjects. In addition, the children were tested on their reading ability by a Dutch Technical-Reading Test as well as on their comprehension of concepts as "more than", "less than" and "together" by a Dutch Comprehension Test. After the administration of these three tests, 84 subjects attending seven different schools for Special Education were selected to participate in the experiment. Since only seven computer configurations were available, the seven schools with the pupils who had the lowest performance on the word-problem test were chosen as experimental schools. All testing and training took place at the school the child was attending. Seven orthopedagogists, one male and six female, served both as testers and trainers in the experiment. First, the selected children of each school as a group did the paper and pencil pretest with untrained problem types. The remaining pretests were administered individually in a prefixed order by the testers in five different sessions on five successive days. For practical purposes, each tester was attached to a different school and acted also as trainer during the training period. In order to counter school and trainer bias, all four treatment conditions were run on each school by each trainer, who was replaced by a trainer of a different school in the posttest period to ascertain ignorance of the trainers as to which students had received which treatment.

The experiment took place from April 1989 till October 1989. The training period started around May and stopped after about three months. The children were tested on their follow-up performance in October, after the summer holidays.

At each experimental school the 12 pupils were assigned to matched quartets resulting in three groups. From each of these groups every child was randomly assigned to one of the four experimental groups. To accomplish this, the subjects were ordered according to their pretest performance of the Trained Word-Problem Test (TWPT) and split in groups of four subjects, who had the same rank or were closest in rank. Subsequently, each child of a matched quartet was randomly assigned to one of the experimental groups. Then, the distribution of girls and boys among the experimental groups was checked. An unequal distribution of the sexes among the four experimental groups was controlled by interchanging a boy from one experimental group with a girl from another experimental group, who had the same TWPT rank or were closest in rank on the TWPT. Also, within each experimental group age differences were checked by computing the total sum of ranks according to age. Again, differences were compensated for by interchanging subjects of the same sex and with the same TWPT rank or closest TWPT rank. Finally, each experimental group was randomly assigned to one of the four treatment conditions.

### Materials used during pretest, posttest and follow-up

In this section, the tests which were used to assess the pretest, posttest and follow-up performance of the subjects will be discussed. A distinction will be made in performance tests and process measures. From a practical point of view during the follow-up only the performance tests were administered.

### *Performance tests*

In the pretest, posttest and follow-up period, two performance tests were administered: a test containing trained word problems and a test containing untrained word problems.

*Trained Word-Problem Test* (TWPT). The Trained Word-Problem Test consisted of word problems, which were also used during subsequent training. This test was constructed in order to assess the transfer value of the training procedures to conditions without computerized instruction. The following 10 word-problem types were included: change 3, 4 and 5, combine 1 and 2b, compare 1, 2, 3, 4 and 6. An example of a change 5 problem is "Peter had some marbles. Ann gave Peter 3 marbles. Now Peter has 8 marbles. How many marbles did Peter have first?" An example of a compare 6 problem is "Peter has 4 marbles. Peter has 3 marbles less than Ann. How many marbles does Ann have?" For a description of the other problem types, we refer to Figure 3.2. The test was divided into four subtests. Within each subtest the order of presentation of the items was randomly fixed. The first subtest consisted of all 10 standard word-problem types. In the other three subtests, all word problems contained an extra sentence, in which a third number set was described, which was irrelevant to the correct solution of the problem. The function of this third irrelevant number in the word-problem texts has been explained earlier (see chapters 1, 2 and 3). However, the description of this irrelevant set differed as regards content with the irrelevant set descriptions used in the earlier pilot studies. In these studies, the linguistic form of the sentence describing the irrelevant set included a set-owner and a set of objects with the same class membership as the objects of the relevant sets (e.g. "Peter has 4 apples. *Paul has 3 apples*. Mary has 6 apples. How many apples do Mary and Peter have together?"). In this study the description of the irrelevant set was thus formulated as to enlighten on its redundancy even more (e.g. "Peter has 3 apples. *Ann catches 9 flies*. Mary has 1 apple less than Peter. How many apples does Mary have?"). The position of this irrelevant number sentence was systematically varied within each problem type. This resulted in four different word-problem versions for the change problem-types and in three different versions for the combine and compare problem-types. Thus the word-problem test consisted of 43 items. The order of the three subtests that contained word problems with irrelevant information was interchanged systematically, resulting in six different versions of the TWPT.

All numbers used and the correct answer were under 10. Number combinations were constructed that allowed the revelation of the mathematical operation used (addition or subtraction or a combination of addition and subtraction on the two or three numbers given).

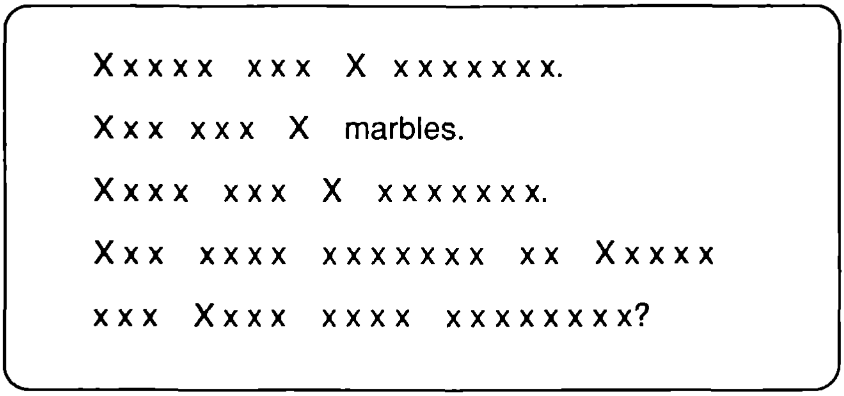
*Untrained Word-Problem Test* (UWPT). This test consisted of six subtests and was constructed to assess the transfer value of the training procedure to new problem types, i.e. to problem types for which the children received no instruction. Again, within each subtest the order of presentation of the items was randomly fixed. The first subtest concerned 6 word-problem types without irrelevant information similar to the change 3, 4 and 5, combine 2b and compare 3 and 4 which were trained. However, the order of introduction of known and unknown quantities had been reversed (an example of a reversed change 3 problem: "Peter just got some marbles. Now Peter has 8 marbles. Peter first had 2 marbles. How many marbles did Peter get?"). The second subtest comprised 20 word problems with irrelevant information which were also similar to the problem types trained. But the irrelevant number sentence was thus formulated as to trigger mistakes concerning the semantic problem type. For instance, merely an analysis of the first two sentences of the problem "Together Peter and Ann have 9 marbles. Peter has 5 marbles. Paul has 3 marbles. How many marbles does Peter have more than Paul ?" could result in the answer "4", since the child expects the question sentence to be "How many marbles does Ann have?" In the third subtest, eight "complex" word problems were included for which the correct answer had to be computed by executing two mathematical operations (subtraction/addition). The fourth subtest consisted of six items representing the remaining standard semantic problem types change 6, combine 2 and compare 5 for which the children did not receive training. In the fifth subtest, 6 action-cued compare problems were used (see Discussion chapter 2). Finally, in the sixth subtest, four syllogisms were included, in which children had to solve "more than" and "less than" problems for which no computation was needed. Thus the Untrained Word-Problem Test consisted of 50 items. For examples of a word-problem type of each subtest, see Appendix B. For each subtest two randomized versions were constructed. The order of the two versions of each subtest was varied. All numbers used and the correct answer were under 10. Again, number combinations were constructed that allowed the revelation of the mathematical operation used (addition or subtraction or a combination of addition and subtraction on the two or three numbers given).

*Process measures*

Both in the pretest and posttest, children were tested on the following process measures.

*Reading-Behavior test* (RBT). In two separate computerized test-sessions both in the pretest and the posttest, children had to analyse and solve 13 word problems. This test was constructed to examine the reading behavior of the children during word-problem solving. The recording technique used resembles the fixation contingent stimulus-presentation technique (McConkey & Rayner, 1975), in which sentences are presented onto a monitor screen with each character masked (e.g. replaced by the character "x"). In the technique of McConkey and Rayner only the window around the fixation point of the subject shows readable text. When the subject fixates another part of the text, the new part is made visible and the text in the former place of fixation is masked again. In the technique we used, the eye fixation is replaced by pointing to the word that one wishes to read on a touchscreen. The only word that is visible at any moment is the word touched by the reader; all other words are masked. A shift of the finger or pencil on the touchscreen results in masking the word that was previously visible. As a result, only one word is readable at a time and this facilitates the recording of the ongoing text-analysis process. An example of a screen display is given in Figure 5.2. Although fixation on a word allows information extraction from a broader area than one word, readers seem to fixate on almost every word, except small words which are relatively unimportant for the comprehension of texts (Just & Carpenter, 1980). In fact, to fully comprehend the text, readers have to fixate on nearly every word since the area from which they are able to extract lexical information appears to be no larger than approximately 10 characters from the point of fixation. Yet readers are able to perceive word-length information from a broader area (12 to 15 characters from the fixation point) parafoveally (McConkey & Rayner, 1975). Therefore, in order to mimic the natural reading state as accurately as possible, the word-length information was retained in the technique used by preserving the interword space (for a full description of the technique, see Van Lieshout, in press).

In each session 13 masked word-problem texts were presented in succession. Every character of the original text had been replaced by the character "x". Capitals and numbers had been printed in upper case ("X"), whereas all other characters had been printed in lower case ("x"). A word was readable as long as the subject touched the word (see Figure 5.2). When the subject had finished reading, the tester entered the numerical answer given by the child. The word-problem types used were classified into five categories: (a) three solvable word problems without irrelevant number information, (b) four solvable word-problem types with irrelevant information, (c) four unsolvable word-problem types with irrelevant information, (d) a "complex" word problem and (e) a word problem for which no computation was required to arrive at the correct answer. The word problems of category c were made unsolvable



Xxxxx xxx X xxxxxxx.
Xxx xxx X marbles.
Xxxx xxx X xxxxxxx.
Xxx xxxx xxxxxxx xx Xxxxx
xxx Xxxx xxxx xxxxxxxx?

*Figure 5.2* Example of a screen display of the Reading-Behavior Test. The word "marbles" is touched. The masked sentence reads "Peter has 4 marbles. Ann has 9 marbles. John has 3 marbles. How many marbles do Peter and John have together?"

by interchanging the number sizes of the relevant sets (e.g. "Peter has 3 apples. *Mary has 2 apples more than Peter.* Ann has 9 apples. How many more apples does Peter have than Ann?"). For examples of word-problem types used, see Appendix C. Category a contained one word problem for which the children received training; the children were unfamiliar with the remaining word problems. For each word problem, four versions were constructed. Each version of a word problem was randomly assigned to one of four tests. For each child, the order in which these tests were administered in the pretest or posttest was randomized.

*Free-Representation Test* (FRT). To explore the representational strategies which children used spontaneously after instruction, a so-called "Free-Representation Test" was constructed. There were two sessions in both in the pretest and in the posttest. In each session the children had to represent and solve 13 word-problem types: three change 1 and change 4 problems, three combine 1 problems, one (unsolvable) compare 1 problem, one compare 5 problem and two compare 3 problems. The compare 1 problem was made unsolvable by interchanging the number sizes of the relevant sets. For examples of word problems used, see Appendix D. The children had red and white squares available for representing the word problems. The word problems were presented one by one on cards by the tester. The order in which the word problems were presented was randomized. First, the child had to read the word-problem text aloud. When a child erred in reading the text properly, the tester hinted the child to start reading again. Following a mistake in the second reading of the text, the tester read the problem aloud. Next, the child started modeling the word problem with squares. In order to reveal which set the child wanted to model, the child first had to read the sentence aloud in which this set was described and subsequently represented this set. During modeling the tester refrained from intervening and merely observed the configuration constructed by the child. After the child had performed a step, the tester recorded the number of the sentence read by the child and drew the configuration, which was present on the table at that moment, on a score form. Then, the child read the next sentence aloud (which was not necessarily the second sentence) and subsequently modeled the set described with squares. When the child finished modeling the word problem, the child read the question sentence aloud and located the answer set by pointing at the corresponding squares on the table. The number of modeling steps allowed was determined by the problem type. As such, a distinction was made in problem types describing two number sets and describing three number sets, for which the maximum number of modeling steps was fixed at two and three steps respectively. Additionally, for problem types in which an unknown number set was described in a sentence preceding the question sentence ("Peter lost some marbles"), the maximum number of modeling steps was increased with two steps. This was decided in order to permit the child to revise the number of squares of this set, since the child might have guessed the number of this set in a preceding step. Thus, the maximum number of modeling steps allowed ranged from two to five steps. The tester only interfered with the child during modeling when the child exceeded the maximum number of modeling steps permitted. In doing so, for each word-problem type the tester had flowcharts available in which the feedback regarding the interventions that were permitted was listed. The drawings of the testers, which represented the modeling steps of the child, were reviewed by three judges according to a classification system, which was framed in line with empirical findings reported by De Corte and Verschaffel (1981, 1987), Verschaffel (1984) Carpenter and Moser (1982, 1984) and our own pilot work (see chapters 2, 3 and 4). The classification system encompassed mutually exclusive categories according to which the successive order of representation of the sets by the child, the number of squares in each set, the kind of manipulation the child had performed (i.e. whether a child had added or removed squares), the placement of sets on the table with respect to each other, the number of representational steps performed, the number and location of the answer set the child identified, were scored. The consensus among the three judges in classifying the modeling patterns according to this classification system was high; Cohen's Kappa amounted to 0.87.

*Word-Problem Classification Test* (WPCT). The Word-Problem Classification Test was constructed to examine the ability of the children to classify word problems according to semantic structure (change, combine and compare problem-types). Seventeen standard word-problem types (six change, five combine and six compare problems) were printed on separate cards. The tester shuffled the cards in front of the child and instructed the child to put the cards with word problems which belonged together in one stack. When the child ordered the cards in more than three stacks, the procedure was repeated and the child was instructed to make three stacks. Both the first and second ordering were scored by the tester. In order to differentiate classifications resulting from an analysis of the "deep structure" of the problem from analyses based on superficial features of the word problem, several superficial problem-text features were systematically varied across the semantic structure of the word problems, which was the first category distinguished. The superficial variables were: (a) name of actor; "Ann", "Peter", "John", "Ann and Peter", "Ann and John" or "Peter and John", (b) name of objects; "marbles", "dolls" or "cars", (c) number pair used; (1,3), (5,2) or (7,3), (d) number of sentences; three or four, (e) first actor mentioned; "Ann", "Peter" or "John", (f) key word presented; "to lose", "to win", "together", "more than" and "less than", (g) number of persons described; one or two, (h) addition or subtraction word-problem type. With regard to the last category, it should be noted that addition was defined as direct addition of the two numbers mentioned and subtraction as direct subtraction of the two numbers mentioned. With respect to each of these superficial features, a different series of ordering and required number of stacks had been defined for the perfect ordering according to each of these features. In scoring the stacks a distinction was made between main features and subfeatures. For instance, a main feature was "first actor mentioned" whereas a subfeature of this main feature was "Ann". Since children might order the problems according to several problem features, each stack constructed by the child was scored per subfeature. For each subfeature the required number of word problems which ought to occur in one stack was predefined. Per subfeature the actual number of word problems encountered in a stack was divided by the number of word problems for this subfeature which was required in case of perfect ordering. For example, in the perfect ordering according to the feature "first actor mentioned", the stack with "Ann" should contain five particular word problems, in which the actor described is "Ann". If the stack merely included two of these word problems, the following score was calculated: actual number of word problems recorded/required number of word problems; 2/5=0.4. For the main feature "first actor mentioned" this procedure was repeated for "Peter" and "John".This yielded a so-called "association score" per subfeature. Then, for each main feature per stack the association scores of all subfeatures were added. So, for the main feature "first actor mentioned" the association scores for "Ann", "Peter" and "John" were added. This total score was divided by the number of subfeatures of a main feature encountered within a stack. For instance, when an ordering on all subfeatures "Ann", "Peter" and "John" was encountered in one stack, the total score for "first actor mentioned" was divided by 3. As such, for each main feature, a score per stack was calculated. These scores were added per main feature across all stacks. Thus, in case of four stacks, for each main feature four stack scores were calculated and added. In order to correct for the number of stacks found, the resulting score was multiplied with the quotient "number of stacks constructed/ number of required stacks" if this quotient was smaller than one, otherwise the reversed quotient was used. Finally, since the number of subfeatures differed per main feature, this score was divided by the predefined number of subfeatures per main feature. As such, an "association score" per main feature ranging from 0 to 1 was computed. The total calculation algorithm is explained by means of an example in Appendix E, in which also the word-problem types used are given.

### Training

The procedure during training consisted of two individual instruction sessions and a training phase including ten individual training sessions. Each presentation of a word problem was preceded by a probe trial. In the two instruction sessions, in the presence of the child the trainer both showed and verbalized all actions of the task strategy for half of the word-problem types to be used during training. Besides, the trainer read the feedback contents aloud that occurred on the screen. In the second instruction session, the child was allowed to perform the actions together with the trainer. Meanwhile the trainer stimulated the child in verbalizing the actions performed. For the control condition the task strategy merely consisted of practising the problem types presented on the screen. During the probe trials the child received no instruction and had merely to answer 10 different word-problem types by entering a numerical answer on the touchscreen. The computer informed the child whether the answer given was correct or incorrect and, if necessary, provided the correct answer. In these probe trials for each word-problem type the uninterrupted series of correct answers was scored. Training stopped for a problem type as soon as a child had answered three problems of a particular word-problem type correctly in at least three successive probe trials. When a child erred in a subsequent probe trial, training for this problem type started again. During training the child solved the word problems independently. The children in the control condition were only instructed to compute the answer. The children in each of the treatment conditions had to perform the required course of actions before they were allowed to answer the problem. Each action was provided with computer feedback. In performing each step, the child was offered three chances to correct mistakes. Erring on the last attempt caused the computer to execute the step. Since the course of actions of each computerized training-procedure has already been discussed in previous chapters, a further description of the details is not repeated here. In essence, the children were taught to analyse the word-problem text by pointing at crucial words reflecting problem structure in the Text-Analysis training. The External-Modeling training concentrated on teaching children to represent word-problem structure by visual squares. Both analysing the word-problem text and modeling the sets and set-relations or actions described was instructed in the combined Text-Analysis and External-Modeling training. In Appendix A flowcharts are presented in which the course of action of each of these training methods is visualized.

For a full report of the representational strategies instructed per word-problem type, we refer to Figure 3.2. However, in order to allow a more explicit distinction in semantic problem types, some changes were made in the representational strategies. The representation of the second set was divided into two separate steps: a child first had to locate the required place of the second set by touching a specific area on the screen, which became marked before he or she could represent the second set with visual squares. When a child touched the area near the start set, besides this area turning grey, a vertical dashed line occurred on the screen immediately behind the start set. The child had to touch this location to represent change problem-types describing an increase. So for change 3 problem-types this vertical mark aided in locating the answer set. For change problem-types describing a decrease, a so-called "throw-away" box, which was depicted on the screen, had to be touched in order to make it active. To represent the second set of combine problem-types, which start with the description of one of the subsets ("Ann has 5 marbles. Peter has 3 marbles. How many marbles do Ann and Peter have together?" or "Ann has 5 marbles. Peter also has some marbles. Together Ann and Peter have 8 marbles. How many marbles does Peter have?") the area to the far right of the first set had to be touched, which resulted in marking this area grey. To represent the second set of combine problems, which start with the description of the whole set ("Together Ann and Peter have 8 marbles. Ann has 3 marbles. How many marbles does Peter have?"), the "throw-away" box had to be touched. Finally, the child had to touch the area beneath the first set to represent compare problem-types. The result of these revisions in the representational strategies was that the main problem types, change, combine and compare, to which a certain word problem belonged could be distinguished by inspecting the final representation on the screen. The only exception was the combine 2b problem-type for which the final representation could not be distinguished from change problems describing a decrease.

### Materials used during training

Ten standard word-problem types were presented in the training phase: change 3, 4 and 5, combine 1 and 2b, and compare 1, 2, 3, 4 and 6. For examples of these problem types, see chapter 1, Table 1.1. Both in the first two probe and two training sessions, these word problems contained no irrelevant information. In the remaining probe and training sessions, these word-problem types contained a third irrelevant number sentence. The position of the irrelevant number sentence in the problem text was systematically varied within each problem type. For change 3, 4 and 5 problems, the first, second, third or fourth sentence described the irrelevant number set whereas the position of the irrelevant sentence for the remaining problem types was first, second or third. In order to make each probe and training session equally difficult, the position of this irrelevant number sentence was systematically varied across problem types. This resulted in 8 pools of 10 word-problem types, in which the order of presentation of the word problems was randomized. For each child, the order of presentation of each of these pools was also randomized. In the next three probe and training sessions all 10 word-problem types were offered. The remaining five probe sessions also consisted of 10 word-problem types. However, the number of word-problem types offered during the remaining 5 training sessions varied according to the performance of the child on each problem type on the previous 3 probe trials.

### Apparatus

The computer programs had been written in TML Pascal, version 1.00b, using the Quickdraw tool from the Toolbox to animate the soft keys and using 320 mode to project text with the aid of letters that are bigger than the standard letter type. The computer programs require an Apple II GS microcomputer with 512 Kb and a Philips VP 120 Touchscreen monitor. The Philips VP 120 consists of a color t.v. monitor with a touch panel which fits over the television screen. Touching this screen causes an interruption of the infrared light beams and results in X- and Y-coordinates which are transmitted to the computer. Subsequently, the computer determines which area on the screen was touched. The hardware interface between the Philips VP 120 and the Apple computer is provided by the printer port. An assembler routine had been written for the software interface.

## Results

The results will be discussed in three main sections. In the first section, the results with regard to the covariance analyses performed to test the hypotheses stated will be presented. Second, the trend analyses concerning the pretest, posttest and follow-up performance of the subjects will be reported. In the final section, the results of the training period will be given.

### Pretest and posttest

The main hypotheses were all tested via multivariate analyses of variance with Text Analysis (2 levels) and External Modeling (2 levels) as between-subjects factors. In order to test these hypotheses by analyses of covariance, the assumption of the homogeneity of the within group regression-line slopes was verified. If the homogeneity of regression hypothesis was not violated, an analysis of covariance was performed with the pretest performances as covariates for each dependent variable. In case of violations of the assumptions of covariance, i.e. when a significant factor by covariate interaction term was found, a covariance analysis was performed in which separate regression-line slopes were fitted for each group.

To investigate the effect of the training procedures on the performance of the children, the number of correctly solved problems across the performance measures of the posttest was studied. In order to test whether Text-Analysis and External-Modeling training had a differential effect on trained and untrained word problems, a distinction was made in trained and untrained problem types. A 2 x 2 x 2 multivariate analysis of covariance with Text Analysis (2 levels) and External Modeling (2 levels) as independent between-subjects factors and with Trained (2 levels) as within-subjects factor yielded neither significant main effects for Text Analysis ($F(1,79)=1.62$, n.s.) and External Modeling ($F(1, 79)<1$, n.s.), nor a significant Text-Analysis x External-Modeling interaction effect ($F(1,79)<1$, n.s.). These results indicate that there was no difference in performance for the different training conditions. For each dependent variable, the mean number of correctly solved word problems for each condition are shown in Table 5.1. It should be noted that these means were corrected for number of items per subtest and for the influence of the covariates (i.e. pretest performance on trained and untrained problem types). Although neither a significant second order interaction effect for Text Analysis x Trained nor for External Modeling x Trained was found, there was a significant main effect for the factor Trained ($F(1,80)=98.53$, $p<.001$). Trained problem types were significantly more often solved correctly than untrained problem types ($M=0.448$ versus $M=0.315$). Thus, against our predictions, first, the main effects Text Analysis and External Modeling were not significant (hypothesis 1). Second, the second order interaction effects External Modeling x Trained and Text Analysis x Trained were not significant (hypotheses 2 and 3).

Table 5.1 Adjusted Means of the Proportion of Word Problems Correctly Solved by Each Experimental Condition, C=Control condition, TA=Text Analysis, EM=External Modeling, TA/EM=Both External Modeling and Text Analysis

| | Training condition | | | |
|---|---|---|---|---|
| Dependent variable | C | TA | EM | TA/EM |
| Trained problem types | 0.418 | 0.473 | 0.427 | 0.473 |
| Untrained problem types | 0.310 | 0.325 | 0.301 | 0.322 |

In order to test whether children who received Text-Analysis instruction would take longer to analyse the question sentence than children who received External-Modeling training, the QR-phase was studied. Since it was assumed that a higher activity of looking behavior by children in the Text-Analysis condition would manifest itself on different process measures, several process measures were used in the analysis. Differences in number of words touched, number of different words touched and gaze durations in the QR-phase were thus studied. These variables were all corrected for number of words per question sentence per word problem.

A 2 x 2 multivariate analysis of covariance with number of words touched, number of different words touched and gaze durations in the QR-phase as dependent variables and Text Analysis (2 levels) and External Modeling (2 levels) as independent between-subjects factors yielded neither significant main effects nor a significant interaction effect (Text analysis: $F(3,75)=2.20$, $p=0.095$; External Modeling: $F(3,75)<1$, n.s. and Text analysis/External Modeling: $F(3,75)<1$, n.s.). Table 5.2 shows the adjusted means per dependent variable in the QR-phase for each experimental condition. The means were adjusted for the number of words touched, the number of different words touched and the gaze durations in the QR-phase of the pretest. From this table it becomes clear that, except for the number of words touched, both the TA and TA/EM conditions have higher, albeit insignificant, scores on number of different words touched and gaze durations. Univariate analysis demonstrated that for the number of different words touched, there was a trend for the Text-Analysis condition in the supposed direction ($F(1,77)=2.85$, $p=0.095$).

Table 5.2 Adjusted Means of the Number of Words Touched, Number of Different Words Touched and Gaze Durations in the QR-Phase per Each Experimental Condition, C=Control condition, TA=Text Analysis, EM=External Modeling, TA/EM=Both External Modeling and Text Analysis

| | Training condition | | | |
|---|---|---|---|---|
| Dependent variable | C | TA | EM | TA/EM |
| Number of words touched | 30.69 | 30.60 | 30.64 | 30.29 |
| Number of different words touched | 22.82 | 23.18 | 22.96 | 24.00 |
| Gaze durations | 703.51 | 725.52 | 700.42 | 755.52 |

Although the results showed no significant differences between the various training conditions with regard to number of different words touched in the QR-phase, the data did indicate a trend in the hypothesized direction (hypothesis 4). Moreover, since in hypothesis 4 the direction of change was predicted, it may be argued that this hypothesis allows one-sided testing, which would result in a p-value of 0.048. Hypothesis 4 would thus be accepted.

In order to examine whether children who received Text-Analysis instruction showed a more general trend to analyse the problem text longer, i.e. not merely in the QR-phase but also in the IR-phase, a second 2 x 2 multivariate analysis of covariance was performed with number of words touched, number of different words touched and gaze durations in the QR-phase that were all corrected by subtracting the value of the IR-phase from their corresponding value in the QR-phase. Since this analysis yielded no significant between-subjects effects (Text Analysis: *F*(3,75)<1, n.s.; External Modeling: *F*(3,75)=1.31, n.s. and Text Analysis/External Modeling: *F*(3,75)<1, n.s.) it can be said that children who received Text-analysis instruction demonstrated a trend of more looking activities in both the IR- and QR-phase instead of only in the QR-phase.

To examine whether children who were trained in External Modeling demonstrated more rereadings than children who received instruction in Text Analysis, the RR-phase was analysed. Again, it was assumed that more rereadings in the RR-phase would show up on different process measures. Hence, differences in number of words touched, number of different words touched and gaze durations in the RR-phase were studied. These variables were all corrected for number of words per sentence per word problem.

A 2 x 2 multivariate analyses of covariance with Text Analysis (2 levels) and External Modeling (2 levels) as independent between-subjects factors and number of words touched, number of different words touched and gaze durations in the RR-phase as dependent variables demonstrated no significant between-subjects effects (Text Analysis: *F*(3,75)<1, n.s.; External Modeling: *F*(3,75)=1.98, n.s. and Text Analysis/External Modeling: *F*(3,75)<1, n.s.). In Table 5.3 the adjusted means per dependent variable in the RR-phase for each experimental condition are given. Each

Table 5.3 Adjusted Means of the Number of Words Touched, Number of Different Words Touched and Gaze Durations in the RR-Phase per Each Experimental Condition, C=Control condition, TA=Text Analysis, EM=External Modeling, TA/EM=Both External Modeling and Text Analysis

| | Training condition | | | |
|---|---|---|---|---|
| Dependent variable | C | TA | EM | TA/EM |
| Number of words touched | 12.73 | 13.87 | 11.79 | 12.26 |
| Number of different words touched | 9.35 | 10.06 | 8.97 | 9.44 |
| Gaze durations | 314.06 | 384.01 | 342.44 | 335.64 |

of these means was adjusted for number of words touched, number of different words touched and gaze durations in the RR-phase of the pretest. From this table it can be seen that for each dependent variable the mean scores for the EM and TA/EM conditions are lower (albeit insignificant) than the mean scores for the other conditions. Thus, these results imply that there was no difference between the training conditions as regards number of words touched, number of different words touched and gaze durations in the RR-phase (hypothesis 5). Instead there was an insignificant trend in the data opposed to the direction hypothesized. As regards the results of the analyses of both hypothesis 4 and 5 it can be said that children who received Text-analysis instruction demonstrated a tendency of more looking activities in all phases, i.e. the IR-, QR- and RR-phase.

To see whether there is a difference in performance on the Free-Representation Test for the training conditions, an analysis of covariance with Text Analysis (2 levels) and External Modeling (2 levels) as between-subjects factors and Trained problem types (2 levels) as within-subjects factor was performed with number of problems correctly solved as dependent variable. This analysis yielded a significant main effect for External Modeling ($F(1,79)=8.12$, $p<.01$). The adjusted means of the proportion of the number correctly solved word problems per dependent variable are shown in Table 5.4. It should be noted that these means are corrected for number of word problems per subtest and for the influence of the covariates (i.e. number of trained and untrained word problems correctly solved in the pretest).

Table 5.4 Adjusted Means of the Proportion of Number of Problems Correctly Solved in the Free-Representation Test per Dependent Variable and Experimental Condition

| | Training condition | | | |
|---|---|---|---|---|
| Dependent variable | C | TA | EM | TA/EM |
| Trained problem types | 0.424 | 0.426 | 0.526 | 0.540 |
| Untrained problem types | 0.382 | 0.315 | 0.464 | 0.422 |

From Table 5.4 it becomes clear that both the External-Modeling and the Text-Analysis/External-Modeling conditions on average have a higher score correct (*M*=0.488) than the other conditions (*M*=0.387) (hypothesis 6). Further, a significant within-subjects main effect for Trained was found (*F*(1,80)=15.93, *p*<.001). As Table 5.4 shows, trained problem types were on average solved correctly significantly more often (*M*=0.479) than untrained problem types (*M*=0.396).

In addition, the representational strategies the children spontaneously used in the Free-Representational Test were studied. It was assumed that children who received External-Modeling instruction would manifest similar representational strategies for trained word-problem types as were instructed during training. Hence, for each trained word-problem type of the Free-Representational Test a so-called "exact representational score" was constructed. These word-problem types were change 4, combine 1 and compare 3. For each problem a different representational strategy had been instructed, i.e. a "separating to" strategy for change 4, an "adding to" for combine 1 and a "match" strategy for compare 3. Three versions of both the change 4 and combine 1 type were offered, whereas two versions were presented of the compare 3 problem type (see Appendix D). For each trained word-problem type per representational step the manipulation with the squares (whether squares were added or separated), the exact number of squares moved in each step and the location of each set on the table were compared with the instructed manipulation, number of squares moved and location for that word-problem type during training. Besides, the final number of squares which the child pointed at in the answer step was evaluated on both number and location on the table. For each word-problem type the "exact representational score" represented the number of times that the overall representational strategy (i.e. the execution of all representational steps) completely resembled the instructed representational strategy.

A multivariate 2 (Text Analysis) x 2 (External Modeling) analysis of covariance with the "exact representational score" of the pretest as covariate revealed a significant factor by covariate interaction term for Text analysis: (*F*(1,77)=4.38, *p*<.05). Therefore a covariance analysis in which for each group separate regression-line slopes were fitted, was performed. This analysis yielded a significant main effect

Table 5.5 Adjusted Means of the Proportion of "Exact Representational Score" in the Free-Representation Test per Experimental Condition

| | Training condition | | | |
|---|---|---|---|---|
| Dependent variable | C | TA | EM | TA/EM |
| "Exact representational score" | 0.429 | 0.286 | 1.238 | 0.810 |

for External Modeling ($F(1,76)=5.77$, $p<.05$). In Table 5.5 the adjusted means of the proportion of the "exact representational score" are given for each experimental condition. Table 5.5 shows that the External-Modeling conditions on the average have a higher "exact representational score" ($M=1.024$) than the other conditions ($M=0.358$) (hypothesis 7).

For problem types with irrelevant information the modeling strategies the children used in the Free-Representation Test were also evaluated on the presence of the irrelevant number in the external representation. These problem types were change 1 and 4 and combine 1. For each of these problem types containing irrelevant information that was correctly answered, a so-called "chronological irrelevant-set representation" score was constructed by counting the frequency with which the irrelevant number set was represented in accordance with its appearance in the problem text. So, if the second sentence described the irrelevant number set, the second representational step was examined on the occurrence of the irrelevant set. In addition, for each correctly answered problem type with irrelevant information, a so-called "no irrelevant-set representation" score was constructed by counting the frequency of modeling strategies in which this irrelevant number set had not at all been represented. Hence, all representational steps performed by the child were examined on the occurrence of the irrelevant set in order to calculate this "no irrelevant-set representation score". Since our main interest concerned differences in strategy scores between the External Modeling-only and Text Analysis-only condition, only the strategy scores of both of these experimental conditions were used in the analyses.

Figures 5.3a and 5.3b show the curves of the mean strategy scores for both "chronological irrelevant-set representation" (Figure 5.3a) and "no irrelevant-set representation" (Figure 5.3b) per training condition over time. As can be seen from these Figures, the mean "irrelevant-set representation" score increases over time for the External-Modeling condition, whereas this "irrelevant-set representation" score decreases for the Text-Analysis condition. In contrast, the mean "no irrelevant-set representation" for the Text-Analysis condition increases over time, whereas this mean decreases for the External-Modeling condition.

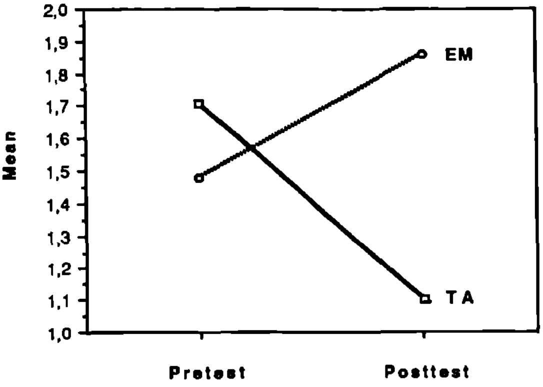


*Figure 5.3a* Mean "chronological irrelevant-set representation" score.

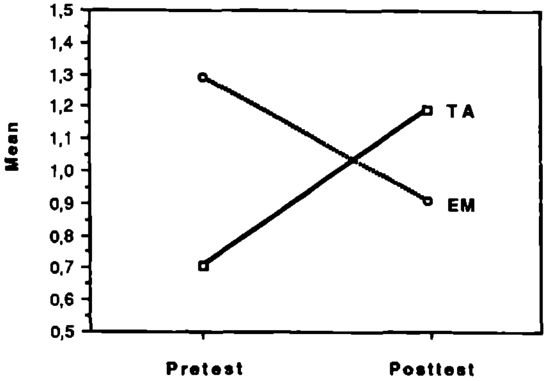


*Figure 5.3b* Mean "no irrelevant-set representation" score.

A first analysis of covariance with Group (2 levels) as between-subjects factor and Strategy (2 levels) as within-subjects factor demonstrated a significant second-order interaction effect for Group x Strategy (*F*(1,39)=7.00, *p*<.01). In Table 5.6 the adjusted means per strategy score are given for both the External Modeling-only and Text Analysis-only condition.

To test these differences in strategy scores between the External Modeling-only condition and Text Analysis-only condition, an analysis of the simple main effect of the second-order interaction effect Group x Strategy was performed. Both the differences in the mean strategy score for "chronological irrelevant-set representation" and "no irrelevant-set representation" between the Text-Analysis condition and External-Modeling condition reached significance (*F*(1,39=9.55, *p*<.01 and *F*(1,39)=5.13, *p*<.05, respectively). These results are in accordance with our predictions (see hypothesis 8 and 9).

Table 5.6 Adjusted Means of the Proportion of "Chronological Irrelevant-Set Representation" Score and "No Irrelevant-Set Representation" Score for the External Modeling-only and the Text Analysis-only Condition

| | Training condition | |
|---|---|---|
| Dependent variable | TA | EM |
| "Chronological-irrelevant set representation" | 1.021 | 1.932 |
| "No irrelevant-set representation" | 1.408 | 0.688 |

Finally, the relation between type of training and the classification of word problems according to superficial and deep-structure features was studied by multivariate analyses of covariance with Text Analysis (2 levels) and External Modeling (2 levels) as between-subjects factors and Main Feature score as within-subjects factor (9 levels) on the Word-Problem Classification Test. The Main Features which were distinguished in the analyses were (a) semantic problem structure (b) name of actor, (c) name of object, (d) number pair used, (e) number of sentences, (f) first actor mentioned, (g) key word presented, (h) number of persons described, (i) addition or subtraction word-problem type. The method of computation of these Main Feature scores has already been explained in the Method (see also Appendix E). In order to study differences between the training conditions in ordering of the word-problem types according to one of the Main Features, for each Main Feature scores ranging from 0 to 1 were calculated, multiplied by 100 and used in the analyses.

Starting with the data of the first attempt, a first 2 (Text Analysis) x 2 (External Modeling) x 9 (Main Feature) analysis of covariance merely yielded a significant main effect for Main Feature ($F(8,73)=40.89, p<.001$) and a significant second-order effect for External Modeling x Main Feature ($F(8,73)=2.16, p<.05$). Table 5.7 shows the adjusted means per Main Feature.

Table 5.7 Adjusted Means of the Ordering per Main Feature on the Word-Problem Classification Test

| | Training condition | | | |
|---|---|---|---|---|
| Main Feature | C | TA | EM | TA/EM |
| Semantic problem structure | 29.33 | 31.00 | 30.91 | 31.81 |
| Name of actor | 19.29 | 15.86 | 11.95 | 14.48 |
| Name of object | 60.48 | 65.62 | 75.05 | 60.33 |
| Number pair used | 42.95 | 42.43 | 35.19 | 38.62 |
| Number of sentences | 28.24 | 30.33 | 33.43 | 32.76 |
| First actor mentioned | 39.38 | 47.10 | 51.10 | 42.24 |
| Key word presented | 20.10 | 18.04 | 14.91 | 17.05 |
| Number of persons described | 27.76 | 30.19 | 33.24 | 33.71 |
| Addition or subtraction problem | 28.57 | 30.00 | 33.38 | 32.00 |

Univariate analysis demonstrated that across all conditions, children significantly more frequently ordered the word problems according to semantic type (*M*=31.01) (*F*(1,80)=182.34, *p*<.001) than according to the key word mentioned in the text (*M*=17.52) or according to name of actor (*M*=15.40; *F*(1,80)=226.24, *p*<.001). Yet, children across all conditions more frequently ordered the word problems according to the object mentioned in the problem text (*M*=65.37, *F*(1,80)=83.67, *p*<.001), according to the first actor mentioned (*M*=44.96, *F*(1,80)=83.58, *p*<.001) or according to the number pair used (*M*=39.89, *F*(1,80)=15.20, *p*<.001) than according to semantic type.

Analysis of the simple main effect of the second order interaction effect, External Modeling x Main Feature, demonstrated that children who received External-Modeling instruction ordered the word problems more according to number of persons (*M*=33.48 versus *M*=28.98; *F*(1,79)=12.19, *p*<.01), addition or subtraction (*M*=32.69 versus *M*=29.29; *F*(1,79)=6.86, *p*<.05), number of sentences (*M*=33.10 versus *M*=29.29; *F*(1,79)=9.85, *p*<.05) than the children in the other conditions.

It was decided to refrain from analysing the second attempt, since only 20% of the subjects constructed less or more than three stacks on the first attempt and thus were offered a second opportunity to order the word-problem types.

### Pretest, posttest and follow-up

In order to examine the significance of any curvilinear trend in the change of performance of the subjects with respect to time, trend analyses were performed on the pretest, posttest and follow-up performances. Since equal time-periods passed between the pretests and posttests and between the posttests and follow-up tests, polynomial curves with equal spacing were fitted to the performances of the pre-, posttests and follow-up tests. For all subsequent analyses of trend, Text Analysis (2 levels) and External Modeling (2 levels) were included as between-subjects factors and Time (3 levels) as within-subjects factor. Since the assumptions for compound symmetry were not tenable, the results of the multivariate analyses will be presented. A first trend analysis with Text Analysis (2 levels) and External Modeling (2 levels) as between-subjects factors and Time (3 levels) and Trained (2 levels) as within-subjects factors yielded significant main effects for Time (*F*(2,67)=53.64, *p*<.001) and Trained (*F*(1,68)=38.20, *p*<.001). Besides, a significant second-order effect for

Time x Trained was found ($F(2,67)=47.83$, $p<.001$). Subsequent analysis of the simple main effect of this second-order interaction effect demonstrated that on the pretest, children did not perform significantly better on trained than on untrained problem types ($F(1,68)<1$, n.s.), whereas on the posttest and follow-up test, children did perform significantly better on trained than on untrained problem types ($F(1,68)=31.11, p<.001$ and $F(1,68)=72.82, p<.001$ respectively).

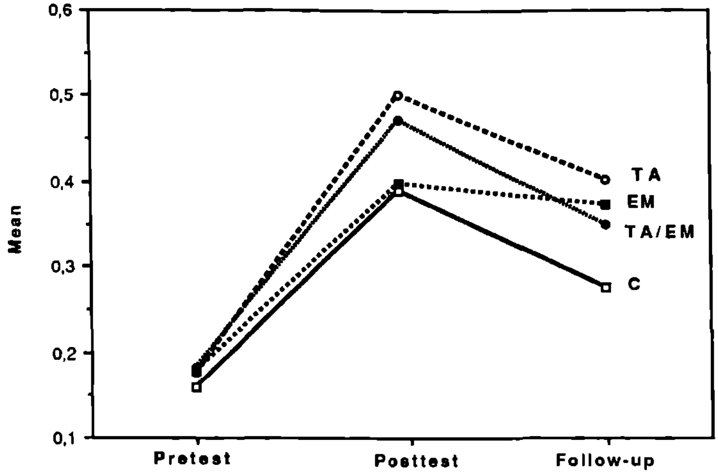


*Figure 5.4a* Mean proportion of number of correctly solved trained problems.

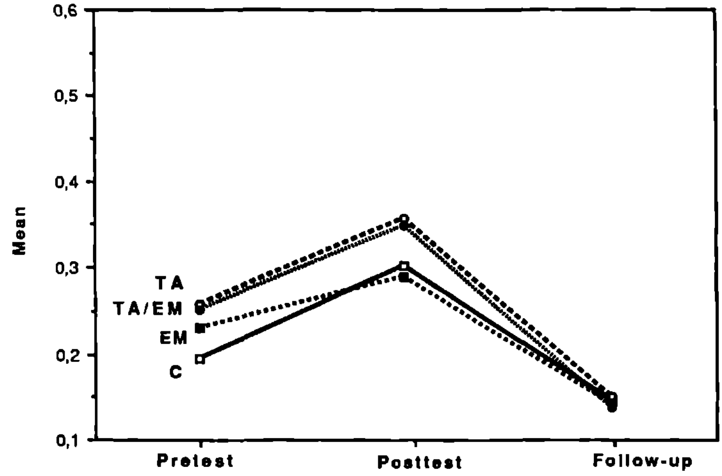


*Figure 5.4b* Mean proportion of number of correctly solved untrained problems.

Univariate analysis of the simple main effect of the second-order interaction effect Time x Trained demonstrated that both for trained and untrained problem types the linear ($F(1,68)=33.47$, $p<.001$ and $F(1,68)=30.71$, $p<.001$) and the quadratic component ($F(1,68)=84.12$, $p<.001$ and $F(1,68)=91.09$, $p<.001$) were significant. Figures 5.4a and 5.4b show that for all conditions, the performance on trained and untrained word-problem types (Figures 5.4a and 5.4b respectively) first increased from pretest to posttest and subsequently decreased from posttest to follow-up test. Yet, the mean performance on trained problem types on the follow-up test is still higher than on the pretest. In contrast, the performance on untrained word-problem types on the follow-up test degenerates beneath the performance level on the pretest.

In general, the performances of the follow-up test declined with regard to the posttest performances. Since most analyses of the posttest had demonstrated no significant results and the performance on the follow-up tests as compared to the posttest performance even decreases, it was decided to refrain from separately analysing the follow-up test.

## Training period

To examine the transfer effect of the different training conditions on the performances during training, multivariate analyses of variance of the probes with repeated measures on the factor Time were performed. Since the first two training and probe sessions contained no word problems with irrelevant information and thus were probably less difficult than the remaining eight sessions, these first two sessions were excluded from the analyses. First, the performances over time of the main problem types change, combine and compare were analysed in a 2 (Text analysis) x 2 (External Modeling) x 3 (Main Type) x 8 (Time) analysis of variance. This analysis yielded significant main effects for Time ($F(7, 74)=12.33$, $p<.001$) and Type ($F(2,79)=54.68$, $p<.001$). Besides, a significant second-order interaction effect was found for External Modeling x Main Type ($F(2,79)=4.11$, $p<.05$). Figure 5.5a, b and c present the mean number of problems correctly solved over time for the main problem types change, combine and compare respectively. As can be seen in these Figures, for each main problem type subjects across all conditions increased in number of problems correctly solved.

Univariate analysis demonstrated ($F(1,80)=101.66$, $p<.001$) that change problems were solved correctly significantly more often ($M=12.43$) than combine

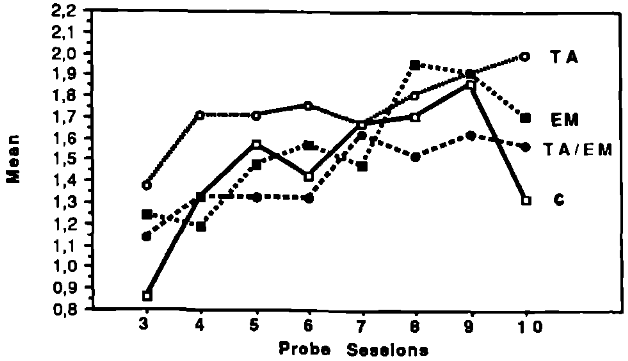


*Figure 5.5a* Mean number of correctly solved change problems.

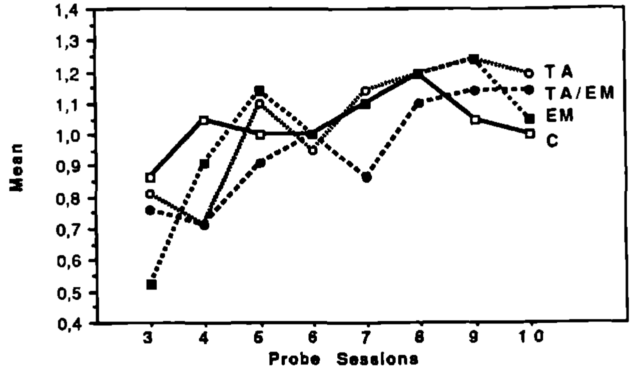


*Figure 5.5b* Mean number of correctly solved combine problems.

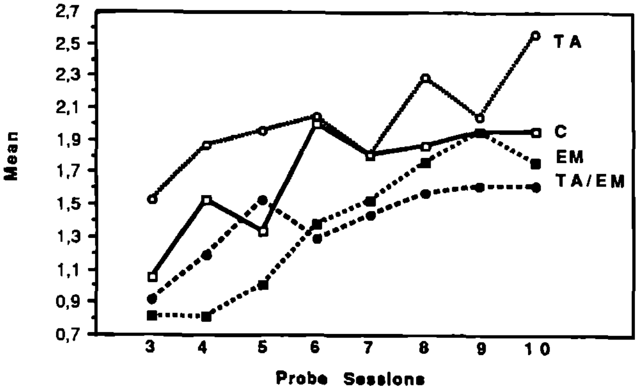


*Figure 5.5c* Mean number of correctly solved compare problems.

problems (*M*=8.08), whereas change problems were not solved correctly significantly more often than compare problems (*M*=12.93; *F*(1,80)<1, n.s.). Analysis of the simple main effect of the second-order interaction effect External Modeling x Type demonstrated that children who received External-Modeling instruction solved compare problems correctly significantly less often (*F*(1,80)=4.22, $p<.05$; *M*=22.14) than the children without External-Modeling instruction (*M*=29.57) whereas children who received Text-Analysis instruction did not differ in performance on compare problems (*M*=27.25) from the children who did not receive Text-Analysis instruction (*M*=24.46) (*F*(1,80)<1, n.s.).

Finally, for each of the training conditions the ability of the children to perform the task strategy was studied. During training, the children in all conditions with the exception of the subjects in the control condition had to perform several problem-solving steps in succession before they were allowed to answer the word problem. In order to examine whether children learned to perform the overall task strategy without errors, a problem was only scored as correct if all steps of the task strategy had been performed correctly at the first attempt. Three multivariate analyses with repeated measures on Time (8 levels) and Main Type (3 levels) were conducted, in which the performance of the Text-Analysis, the External-Modeling and the Text-Analysis/External-Modeling condition were separately analysed. The first analysis, in which the performance of the Text-Analysis condition was examined, yielded neither significant main effects nor significant interaction effects.

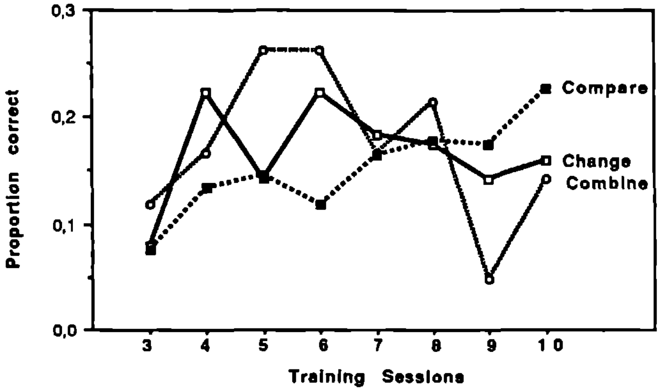


*Figure 5.6a* Mean proportion of number of correctly executed task strategies for the Text-Analysis condition.

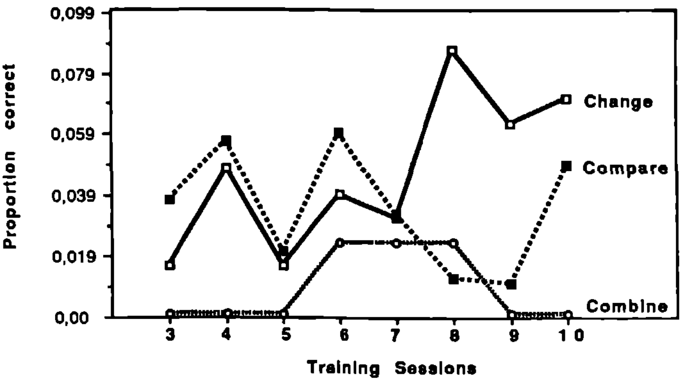


*Figure 5.6b* Mean proportion of number of correctly executed task strategies for the External-Modeling condition.

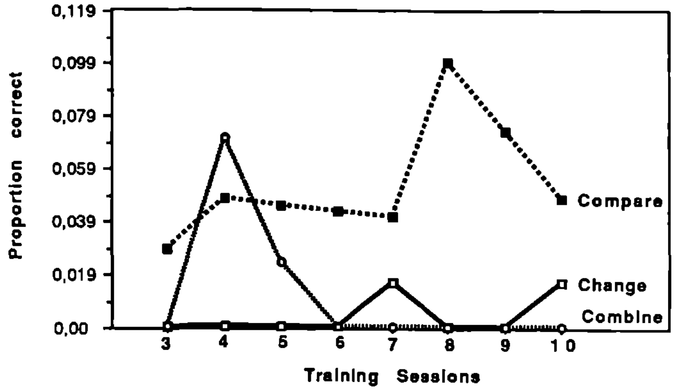


*Figure 5.6c* Mean proportion of number of correctly executed task strategies for the Text-Analysis/External-Modeling condition.

However, both analyses, in which the performance was examined of the External-Modeling condition and the performance of the Text-Analysis/External-Modeling condition respectively, demonstrated significant main effects for Main Type ($F(2,19)=3.72, p<.05$ and $F(2,19)=4.13, p<.05$ respectively).

For each of these conditions Figures 5.6a, b and c present the proportion of correctly executed task strategies for the main problem types change, combine and compare.

## Discussion

One of the main purposes of the present experiment was to study whether the training effects of instructing word-problem solving found in earlier pilot studies (Van Lieshout, submitted; Jaspers & Van Lieshout, 1989a, 1989b; Van Lieshout & Jaspers, 1990; see chapters 2, 3 and 4) could be replicated in a larger sample of educable mentally retarded children. It was expected that as a result of the received training, educable mentally retarded children would improve their ability to solve simple addition and subtraction word problems. More specifically, it was hypothesized that the training procedures would have a differential effect on the performance on trained versus untrained word-problem types. In essence, instruction in External Modeling would induce a performance increase on trained word-problem types inasmuch as External Modeling directs the child's attention towards constructing efficient problem representations for the word problems offered. Yet, the transfer value of these representations to untrained problem types was questioned. At the same time, Text-Analysis training on the other hand was believed to be less specific as regards the word-problem types presented and therefore may be used as general strategy to solve all kinds of word problems. Consequently, Text-Analysis instruction would lead to an improvement in performance on untrained word-problem types.

Contrary to these expectations, the results showed that in general the training procedures were not more effective in inducing a performance increase than the experimental condition in which children merely practised the word problems on the computer. Besides, the training procedures did not induce divergent effects on word problems for which children received training or on new word problems. On the posttest, trained word problems appeared easier to solve than untrained word problems. Still, this finding may be attributed to practising the word problems as well as to particular differences between the word-problem types used and not used during training. In general, the trend analyses of the pretest, posttest and follow-up demonstrated that the subjects in all conditions improved their word-problem solving performance as a result of practise. Likewise, the results of the probe trials showed an overall performance increase. Despite these relatively weak results, the children in the External-Modeling conditions proved better word-problem solvers than the other children when materials were available to model the word problems. This finding at least indicates the efficacy of External-Modeling training in inducing adequate modeling strategies by which children could solve the word problems correctly. Still, the results of the tests without materials were rather weak as opposed to the results of the previous pilot studies.

There are a number of factors that may have been responsible for the relative weakness of the results of the present experiment as opposed to the results of the preceding pilot studies.

In the first place, as a consequence of the employment of the computer the feedback was delivered in visualized form in contrast to the pilot studies with human trainers in which the feedback was presented orally. As was already discussed in chapter 3, oral feedback seems to draw and held the child's attention to a higher degree than visualized feedback. Children may have decided to refrain from reading the feedback delivered on the screen in the long run for two reasons. First, a well-known finding is that children with learning difficulties respond very impulsively when confronted with word problems (Goodstein, Cawley, Gordon, & Helfgott, 1971; Campione & Brown, 1977) and often do not "read" word problems and as such disregard important information statements (Goodstein et al., 1971; Goodstein, Bessant, Thibodeau, Vitello, & Vlahakos, 1972). To pursue this finding, the educable mentally retarded children of this experiment may likewise have "read" the feedback delivered on the screen insufficiently.

Second, although the children in the present experiment were screened on their technical-reading ability, some feedback contents may have been rather wordy and cumbersome. Besides, in order to use feedback contents in correcting mistakes efficiently, children had to refer back to the problem text and in the External-Modeling conditions also to the representation of squares present on the screen.

Although the unsatisfactory results of this study as opposed to the pilot studies with human trainers may be explained by the visualized form of the feedback, this argument will no longer serve in clarifying the positive results of the pilot work with the computerized training-prototypes in which the feedback contents were also visualized on the screen. It could be argued that besides this factor, other factors may account for the present findings and that the accumulation of these factors may illuminate the contrasting results of the pilot work and the present study with the computerized training-programs.

One of these factors is the subject of another explanation. This second explanation for the findings may come from the elucidative effect which the feedback in the control condition may have induced. Although in the probe trials the computer merely gave the correct numerical answer when a child answered the problem incorrectly, for the children in the control condition this computer feedback may have hinted at the mathematical operation required to solve the particular word-problem type and consequently at the correct solution to the presentation of a similar word-problem type that immediately followed in the subsequent training trial. Despite the fact that children in the training conditions received identical feedback in the probe trials, the feedback may not have been so enlightening as to hint to these children the operation required to solve a similar word problem in the training trial since this feedback could not be used until the child arrived at the last step of the task strategy, i.e. the answer step. Therefore, children in the training conditions may not have benefited from this feedback as much as the children in the control condition.

Since in the baseline and probe trials of the pilot studies of chapters 2 through 4 both the human trainers and the computerized training-programs did not deliver the correct answer to the child after a mistake, the training effects of these pilot studies cannot result from this presumed elucidative effect of the feedback. Therefore this argument may be conceivable in explaining the differential results of the pilot studies and this study.

Third, in line with the abovementioned argument, the effect that emanates from merely practising the word problems could have overruled potential training effects. Since in most Special Schools word problems are introduced after formal instruction in addition and subtraction, children may have been selected who had little experience in word-problem solving. A consequence of this may have been that the effect of solely practising the word problems rendered potential training effects marginal. However, one could question why a similar effect of practising did not counterbalance the training effects of the pilot studies. For the baseline assessment of the subjects in the pilot studies which preceded training could have caused ascending trends, which may have exceeded the training effects established. Though it should be noted that in several pilot studies the rate of baseline sessions was less frequent than the rate of the succeeding probe and training sessions. Therefore a potential effect of practising would have been accelerated in the probe sessions and would thus have coincided with the training effects. In light of these facts, the training effects established in the pilot studies for which the rate of the baseline sessions was less frequent than the rate of the probe sessions should be reconsidered (see chapters 3 and 4).

In the fourth place, before the actual training period started, children in each experimental condition received two instruction sessions, in which the trainer in the probe trials used his fingers to calculate the numerical answer. In demonstrating the way in which fingers can be used in counting, some children may have used and profited of this procedure in the subsequent probe trials. In fact, the trainers reported to have observed children making use of finger-counting in determining the answer. Nevertheless, children in all conditions should have taken advantage of this instruction and this effect should thus have been huge enough to overcome the expected differential training effects.

Likewise this finger-counting was demonstrated by the trainers in the instruction sessions of the pilot studies with computerized training-procedures (see chapters 3 and 4). Yet, in these studies this finger-counting instruction may have strengthened instead of leveled the established training effects, since the baseline phase, that served as control phase for each child, preceded the instruction sessions. In contrast, the trainers in the pilot study without computerized instruction merely showed the task strategy that was to be performed during the training sessions and thus did not demonstrate the usefulness of finger-counting to solve the word problems of the probe sessions. As a consequence, this effect of finger-counting may have been present in the computerized training-studies but not in the study without computerized instruction.

Together, all the factors cited may have obscured conceivable training effects to a larger or smaller extent. It is, however, questionable whether training effects would have been established by merely extending the training period in order to exceed the effects of all factors cited. For as long as children do not read the feedback contents depicted on the screen, training effects may still not be demonstrated. Since the feedback given in the training conditions as compared to the feedback given in the control condition was far more wordy and complex, this factor seems the most plausible in explaining the findings of this study. On inquiry, five out of seven trainers reported that children often did not read the feedback depicted on the screen or stopped reading the feedback after the first training sessions. Although such an effect may also have been present in the computerized pilot-studies for which training effects could nonetheless be demonstrated, the negative influence of the written form of feedback contents should not be underrated. For it is very likely that an accumulation of the factors described may have produced noneffects of which the effect of written feedback may be the most important. In fact, the idea of the potential noneffect of written feedback already arose in explaining the results of the computerized pilot-study for External Modeling (see Discussion of chapter 3).

If this fact could largely account for our failure to establish training effects, this would imply that rather than changes in the content of the task strategies of our training procedures, feedback is required that makes more clear which particular mistake the child made and that this feedback should be presented orally.

While the desired training effects could not be established, some interesting findings point at alterations of problem-solving processes. Illustrative are the findings of the performances on the Free-Representation Test, in which children were allowed to use blocks to represent the word problems. In solving the word problems correctly by manipulating blocks, the children who received training in External Modeling outperformed the other children. Thus, whereas children in the External-Modeling conditions did not solve trained word-problem types correctly more often than the other children in situations without modeling facilities, they proved to be better than the other children in solving word problems when materials were present. Furthermore these children showed modeling strategies similar to the modeling strategies instructed during training.

An explanation for these findings may be that the subjects in the External-Modeling conditions may have learned to represent and subsequently solve word problems correctly with objects available, but may not have learned to transfer and incorporate these problem schemata in a mental representation that could also be used in situations without modeling facilities. This may explain why children in the External-Modeling conditions did not perform better on trained word-problem types on the paper and pencil test than the other children.

The results of the Word-Problem Classification Test once more seem to confirm this idea. Since children in the External-Modeling conditions did not perform better in classifying word problems according to semantic word-problem type, these children presumably did not have cognitive problem schemata available that they could use in classifying the semantically different word-problem types.

If the children in the External-Modeling conditions used the instructed representational strategies to model and solve the word problems of the Free-Representation Test, then why did the children in the External-Modeling conditions not learn to execute these representational strategies during training? For they did not show a vast increase in number of correctly executed task strategies in the training period. To answer this question, one should look at the difference in word problems used in training sessions three through eight and the word problems of the Free-Representation Test for which exact strategy scores were computed. The word problems in the training sessions all contained irrelevant information, whereas the word problems of the Free-Representation Test for which the exact strategy scores were analysed did not contain irrelevant information. Therefore in solving the problems during training, children in the External-Modeling conditions may have tried to represent both relevant and irrelevant sets, which may explain the low mean number of correctly executed task strategies during training. But since the word problems of the Free-Representation Test for which exact strategy scores were computed did not contain irrelevant information, the children in the External-Modeling conditions could practise the representational strategies instructed more or less automatically. Evidence for this assumption is that when the word problems in the Free-Representation Test did contain irrelevant information, children in the External-Modeling conditions actually modelled both relevant and irrelevant sets. Despite this finding, these children nevertheless solved the word problems correctly, which points at an eventual awareness of the irrelevance of the third set described in finding the correct solution. Thus, our speculation is that these children automatically practise the procedure of representing the word problems step by step as soon as materials are present, and although they did not learn that representing the irrelevant number set is a redundant step in solving word problems correctly, they were aware of the irrelevance of this number set. In view of the lower frequency of irrelevant set representations, the children who received Text-Analysis instruction presumably learned that representing the irrelevant set is not a crucial step in solving word problems. In fact, these children more frequently skipped representing this irrelevant set in the Free-Representation Test. It thus seems that children who received Text-Analysis instruction improved their ability to distinguish between relevant and irrelevant sets as was reflected in their representational strategies. Supporting evidence for this finding could nevertheless not be found in less rereadings in the RR-phase by these children on the Reading-Behavior Test.

Although the predicted differential effects of instruction in Text Analysis and External Modeling on the reading behavior of the children failed to occur, some interesting trends were found. More specifically, children who received Text-Analysis instruction showed a disposition to read and reread more words and more different words in all reading phases than children in the other conditions. It is obvious that these results do not warrant firm conclusions about the thoroughness or deepness of the text analysis by these children. However, since children in the Text-Analysis conditions did not merely tend to reread more words but also tended to reread more different words, a very tentative conclusion may be that children who received Text-Analysis instruction analysed the word problems more thoroughly than the other children. The prediction that Text-Analysis instruction would improve the ability to distinguish the relevant sets from the irrelevant set and consequently would result in less rereadings in the RR-phase could not be confirmed. On account of the fact that the children who received Text-Analysis instruction showed a tendency to reread more words and more different words in the RR-phase than the other children, they probably had not become so expert as to be aware of the sophisticated strategy of skipping irrelevant information in the problem text. In contrast, their level of expertise may be reflected in the tendency to reread more different words than the other children. Even when children already knew which sentence contained irrelevant information, they may have decided to check their assumptions by reading the rest of the irrelevant sentence. De Corte and Verschaffel (1986) already remarked that incomplete readings can reflect both superficial solution strategies and so-called "economical" strategies, which indicate expertise in word-problem solving. Yet by using "economical" strategies, even an expert problem solver may be mistaken in the schema activated, since activating the adequate schema sometimes requires reading up to the question sentence. Inspection of the raw data showed that none of our subjects stopped reading before they had read the whole word-problem text. In fact, all children demonstrated rereadings of text-parts. Our subjects thus at least proved that they did not merely look for the numbers or the "key word" in the problem text but it seems that they really tried to analyse the word problem.

The conclusion to be drawn from this experiment is that the results provide no directions for desirable modifications in the instructional design of the training procedures. However, it seems compelling to investigate the possibility to present feedback contents in oral instead of in written form. The present user interface should therefore be reconsidered in future research. Perhaps, our hardware could be extended to allow the use of speech in presenting feedback. Besides, if the children's technical-reading ability can account for skipping the feedback, also presenting the word-problem text orally would be the next logical step. While presenting the word-problem text at the start of each trial seems necessary and inevitable, this would probably not suffice for the child to execute the complete task strategy subsequently. In performing each step of the procedure, a child would probably refer back to the problem text in order to determine which words define problem structure or which number set described should be represented. Therefore, at any moment during problem solving the child should be offered the facility to evoke particular parts of the problem text orally. In doing so, the child ought to indicate which part of the text should be verbalized by, for example, pointing at the relevant words. The question is then whether training effects indeed can be established by both presenting the word-problem text and feedback contents orally. Inasmuch as sufficient technical-reading ability would no longer be required to enable a child to follow the computerized instruction, an additional advantage of these modifications may be the extension of the population of educable mentally retarded children for which the computerized training-procedures may be suitable.

Aside from potential training effects that may be established with this revised feedback procedure, a main question to be answered is whether training procedures should be computerized at all. In fact, given our positive experiences with human trainers in the past, the use of teachers to remediate the poor word-problem solving performance by the old training procedures may be reconsidered. On the one hand, as opposed to a computer device, teachers are able to understand speech and thus to interpret the verbal utterances of a child. In addition, other nonverbal behavior by the child, such as "frowning", "sighing" or "hesitating" may be noted by a teacher. On the other hand, teachers may deliver feedback contents orally and this is far simpler than presenting feedback orally by a computer. Since oral feedback releases children from reading and verbally encoding the written information, they may understand feedback contents better and pay more attention. Also, a child may express to the teacher the wish to reread parts of the problem text aloud. All in all, it seems that teachers may compensate for the presumed difficulties of the children of this study in reading and comprehending feedback contents.

However, as an instructional device, the computer offers certain unique technical aspects that may compensate the disadvantages of computer usage cited.

After all, because of its brute force, the computer permits the implementation of sophisticated decision rules according to which the nature of errors during word-problem solving may be evaluated. This allows for timely feedback deliverance conditional on the particular error committed. This continuous monitoring of the performance and of the errors made may also be utilized to modify instruction according to the child's changing abilities. Although teachers may be trained so that they are able to instruct children in similar ways as the computer, they will probably be slower in taking decisions to which specific error occurred and what particular feedback should follow. As the training procedure would become more complex, e.g. as the number of problem-solving steps required to solve the problem increases, the training task would likewise become more complex and prone to wrong decisions by the trainer. In view of the scarcity of human resources, most classroom settings are not suited for teachers practising such complex and time-consuming training procedures. Thus, the computer allows the implementation of instructional designs that would be impossible or difficult to implement with human trainers. For example, an important characteristic of computer-based instruction is the possibility of individualization, which is often quoted as being particularly effective for children in special education. Up to now, in most computerized training-programs individualized practice, i.e. offering different kinds of instruction based on the specific needs of the child, is not really attained. In the next chapter a study is reported that serves as a starting point in the development of instructional designs geared to the specific knowledge deficits and errors of children with learning difficulties.

# Chapter 6

# Diagnosing Wrong Answers to Arithmetic Word Problems

## Introduction

In chapters 3 through 5 experiments with respect to the effectiveness of several prototypes for arithmetic word-problem solving were reported. In designing these prototypes, both theoretically and empirically based notions regarding adequate word-problem solving were used. The last chapter demonstrated that the computerized prototypes were not more effective in improving word-problem solving than the computerized condition in which children merely practised the word problems offered. But besides our main interest in the effectiveness of the prototypes, our research also serves a theoretical purpose. Our second interest is in the problem-solving processes exposed by the children. In line with both this practical and theoretical interest, this chapter describes an experiment in which the usefulness of a product-oriented approach to reveal the knowledge level and misconceptions of children with regard to arithmetic word-problem solving was studied. This research served a triple goal.

First, in previous research both the classification of children according to competence level and the classification of problem types by difficulty level on a product-oriented approach has been complicated by the fact that the answers produced often may not have unequivocally reflected the strategy employed and thus the knowledge deficits of the child. As a consequence, the reported competence levels and difficulty levels of the various word-problem types may not be valid. In order to get round these difficulties, in this study word-problem types were selected for which the problem-solving strategy employed could be revealed unambiguously. This was accomplished by first listing all word-problem solving strategies of young children reported in the literature, and second by checking for each word-problem type whether each of these strategies would produce a unique answer in combination with unique three-number triplets.

Second, although the disclosure of knowledge and misconceptions is generally supported by a process-oriented approach, a product-oriented approach would be less time-consuming and much easier and simpler to apply than a process-oriented approach.

Third, despite the weak results of the latest training study, the main future goal of this research is (still) to develop computerized instruction-programs, in which divergent remediation paths in the program are followed which depend on the expertise level and specific errors of the child. In light of the latest results, this will require immense research work and revisions of our prototypes. The results of this study may eventually be used in the development of more refined computerized prototypes that take into account the child's knowledge level. In the latest prototypes, the child's progress through the training program was merely conditional on his or her previous performance on each particular problem type. The specific errors made by the child were not diagnosed nor used as a starting point in remediating the child. The development of more intelligent computerized instruction-programs, in which instruction links up with the expertise level of a child, first requires diagnosing the present knowledge, knowledge deficits and misconceptions of each individual child. This study just represents the beginning of constructing a simple method to infer the expertise level and misconceptions from the answer given.

## Difficulties in Diagnosing Wrong Answers

Over the past decades, research in the word-problem domain has been dominated by two different tendencies. First, empirical studies on the word-problem solving processes of young children resulted in the formulation of process models. Various investigators have been trying to reveal the origins of mistakes and to determine the difficulty level of the various word-problem types (De Corte & Verschaffel, 1981, 1987; De Corte, Verschaffel, & De Win, 1985; Carpenter & Moser, 1982, 1984, 1985; Riley, Greeno, & Heller, 1983). Second, other research aimed at developing computer models to simulate the problem-solving processes of these children (Briars & Larkin, 1984; Riley et al., 1983; Dellarosa, 1986).

Although many of the computer-simulation models' predictions conform to the results of the empirical studies cited, the models cannot account for all data. The broad scope of the word-problem domain may explain this mismatch between errors produced by these models and by the children. As opposed to errors on arithmetic problems, for which diagnosis seems possible due to the narrowness of the domain

(Brown & Van Lehn, 1982), errors on word problems may not only stem from calculation errors but also from logical errors or wrong text inferences. Brown and Van Lehn (1982) first started to describe a taxonomy of errors produced in arithmetic, which arise from knowledge deficits and misconceptions typical for a pupil at a certain developmental level. As yet, a similar taxonomy of mistakes made in arithmetic word-problem solving is not available.

Solving word problems requires both arithmetic knowledge and insight in the underlying semantic structure of the problem. Hence the importance of processing semantic information is reflected in most arithmetic word-problem solving models. The semantic analysis should precede the choice of the arithmetic operation to be performed. For according to the nature of the actions and relations described, one should either add or subtract (addition and subtraction are the required operations in most previous research). In doing so, an important distinction arises between so-called "direct" and "indirect" problem types. For "direct" problems the mathematical operation can be easily determined by a superficial text analysis, whereas solving "indirect" problems correctly requires a more thorough text analysis. The so-called "key word" presented in the problem text triggers the correct operation for "direct" problems. For instance, in the "direct" problem "Peter has 2 apples. Ann has 3 apples. How many apples do Peter and Ann have together?", the key word "together" may hint at the correct arithmetical operation, i.e. addition. In contrast, for "indirect" problems the key word presented may suggest addition while actually subtraction is needed. In the "indirect" problem "Together Peter and Ann have 5 apples. Peter has 2 apples. How many apples does Ann have?" the child may be misled by this similar key word "together" and may consequently add the numbers, whereas subtraction is required. In fact, most errors in choosing the correct mathematical operation are committed in "indirect" problems (Verschaffel, 1984; Russel & Ginsburg, 1984). Other error types, that are often cited, are: (1) answering with one of the numbers given, (2) "impossible" answers, (3) "0" answers and (4) no answers.

A potential drawback of this type of research is that correct answers may reflect various strategies employed and may even be based on imperfect knowledge. For instance, a pupil who lacks the knowledge to solve a problem like "Peter has 3 marbles. Ann has 2 marbles more than Peter. How many marbles does Ann have?" could thoughtlessly decide to add all the numbers given, which by chance yields the correct answer. Likewise, superficial knowledge may trigger a "key-word" strategy ("since a 'more than' relation is described I add the given numbers"), which for the example cited also happens to produce the correct answer. Consequently, problem types that are classified as "difficult" may indeed be solved correctly by incompetent problem solvers, whereas more competent problem solvers may err on these problems. For example, change 6 problems, in which the two relevant numbers should be added, seem easier to solve than change 5 problems, in which subtraction of the relevant numbers is required. However, this difference in difficulty level may merely result from the preference of children to add rather than to subtract. This fact complicates both the classification of children by competence level and problem types by difficulty level. In addition, given the answer, the problem-solving strategy that the child employed cannot always be determined unambiguously (see the example cited). As a result, a deeper understanding of the misconceptions and knowledge deficits of the children is hindered.

Generally, a process-oriented approach is used to reveal these knowledge deficits and misconceptions. Indeed, wrong answers may stem from different strategies and may be attributed to various misconceptions. The nature of the underlying misconception may particularly be exposed by a process-oriented approach. But the question is whether a product-oriented approach can also disclose typical mistakes in the problem-solving process, since the actual problem-solving process is not easily accessible. If a product-oriented approach indeed allows the disclosure of the expertise level and misconceptions of pupils from the answer given, this will result in a much simpler and much easier method for diagnosing the knowledge level of a child. To this end, the strategies used in word-problem solving reported in the literature were listed. Subsequently, word-problem types were constructed for which the strategy employed could be traced on the basis of the numerical answer given. Although this procedure allows a more accurate assessment of the knowledge deficits of a child, it does not ensure an unequivocal ascertainment of the strategy employed, since the inventory cited may not cover all strategies used in word-problem solving. For example, "wild guesses" may result in any answer that would also be obtained by use of a more sophisticated strategy. One should thus be cautious in interpreting the data. This study is merely a first attempt to determine both the knowledge deficits and the misconceptions of pupils by a product-oriented approach.

# Method

## Subjects

The subjects were 66 pupils attending a Dutch school for educable mentally retarded children. Two pretests were administered in four classes representing the four highest grade levels of this school (i.e. fourth, fifth, sixth and seventh level): (1) a technical reading test (Cito Technical Reading II, 48 items) and (2) an arithmetic test with addition and subtraction items in both canonical and noncanonical form ( 35 items, single-digit problems with the correct answer smaller than 17).

Subjects were selected who could read sufficiently (more than 75% correct on Technical Reading II) and solved more than 80% correct of the arithmetic test. For this reason, three children in grade four were excluded from participation in the experiment. The final selection consisted of 19 subjects in grade four (Mean age=10.4, *sd*=1.11), 16 subjects in both grade five (Mean age=10.8, *sd*=0.98) and six (Mean age=11.9, *sd*=1.26) and 15 subjects in grade seven (Mean age=13.5, *sd*=0.61). Although the subjects of grade four and five hardly differ in age, according to the school, these grade levels represent different performance levels in academic skills.

## Materials

Ten tests were administered of 7 word problems each. Word-problem types were constructed for which the strategy used could unequivocally be determined given the numerical answer of the child. In order to distinguish a superficial so-called "add all numbers" strategy (see Goodstein, Cawley, Gordon, & Helfgott, 1971) from a thoughtful addition strategy, in each problem type a third set was described that was irrelevant for the correct solution of the problem. The position of this irrelevant set was varied within each problem type. Further, ten three-number triplets were constructed that allowed the disclosure of the mathematical operation performed unequivocally (addition or subtraction or a combination of both on two or three given numbers). The following triplets were used: (1,3,9), (2,3,11), (1,6,9), (1,4,11), (1,3,10), (2,3,9), (3,4,7), (3,4,9), (1,3,11).

In order to select proper word-problem types, an inventory was made of all problem-solving strategies cited in the literature for the 14 main problem types (see

Heller & Greeno, 1978). For each of these 14 problem types it was verified whether each strategy described resulted in a unique answer, given one of the three-number triplets. Problem types were excluded in which the "key word" triggers the correct arithmetic operation and thus produces the correct answer. This procedure resulted in the selection of the following problem types: change 3, 5 and 6, combine 2 and compare 1, 5 and 6 problem types. Table 6.1 gives an example for each of these problem types.

Table 6.1 Examples of Problem types Used in the Study

| | |
|---|---|
| Change | |
| 3 | Mary had 3 dolls. Peter gave Mary some dolls more. *Ann had 2 dolls*. Now Mary has 9 dolls. How many dolls did Peter give to Mary? |
| 5 | Mary had some dolls. *Ann had 2 dolls*. Mary got 3 dolls more. Now Mary has 9 dolls. How many dolls did Mary have first? |
| 6 | Mary had some dolls. Mary lost 3 dolls. Now Mary has 4 dolls left. *Ann had 9 dolls*. How many dolls did Mary have first? |
| Combine | |
| 2 | Mary has 3 dolls. *Ann has 2 dolls*. Peter has some dolls too. Together Mary and Peter have 9 dolls. How many dolls does Peter have? |
| Compare | |
| 1 | Mary has 3 dolls. Peter has 9 dolls. *Ann has 2 dolls*. How many more dolls does Peter have than Mary? |
| 5 | *Ann has 2 dolls*. Mary has 9 dolls. Mary has 3 dolls more than Peter. How many dolls does Peter have? |
| 6 | Mary has 4 dolls. Mary has 3 dolls less than Peter. *Ann has 9 dolls*. How many dolls does Peter have? |

*Note*. The irrelevant number sentences are italicized.

Take for instance the change 3 problem from this table "Mary had 3 dolls. Peter gave Mary some dolls more. *Ann had 2 dolls*. Now Mary has 9 dolls. How many dolls did Peter give to Mary?" The number-triplet used is (2,3,9). If a child answers this problem with one of the numbers stated in the problem text, this would yield 2, 3 or 9 as answer. If a child adds all the numbers, this would produce 2+3+9=14 as answer. Subtracting all the numbers would result in 9-2-3=4. A child who decides to add the two relevant numbers described on the basis of the key words "gave more" would give 12 as answer. The correct answer is 6, i.e. the two relevant numbers have to be subtracted. A zero-answer and no-answer would, of course, be easy to distinguish from the other answer classes. Possibly, a child who could not distinguish the irrelevant number from the relevant numbers in the verbal text, would perhaps either subtract the irrelevant number from one of the relevant numbers, producing 1 or 7, or add the irrelevant number and one of the relevant numbers, yielding 5 or 11 as an answer. Each strategy would thus provide a unique answer. The answer-classes distinguished are listed in the section "Scoring".

Ten variants were constructed for each problem type, in which the number triplets and position of the irrelevant number set in the problem text were systematically varied. Finally, the 70-item test was divided into ten 7-item tests. One of the ten variants of each problem type was randomly assigned to each 7-item test.

## Procedure

The ten 7-item tests were randomly administered to the selected children on 10 separate schooldays. The child was instructed to solve all items. During testing no time pressure was exerted.

## Scoring

The answers of the children were scored and categorized into one of the following classes:

-1. answers with one of the numbers given ("one number")
-2. all given numbers added ("add all")
-3. two numbers subtracted from largest number ("subtract all")
-4. "key word" used in determining the operation ("key word")
-5. correct answer ("correct")
-6. "0" answer ("zero")
-7. no answer ("none")

-8. irrelevant number added to one of the relevant numbers ("irr. no. added")
-9. irrelevant number subtracted from one of the relevant numbers ("irr. no. subtracted")

## Results

All nine answer classes were used by the children of all grade levels. A multivariate test of variance with Grade Level (4 levels) as between-subjects factor and Problem Type (7 levels) and Answer Class (9 levels) as within-subjects factors and number of responses as dependent variable showed that there was a significant main effect for Answer Class ($F(1,8)=405.97$, $p<.001$) and that there were significant second-order interaction effects for Grade Level x Answer Class ($F(1,24)=2.206$, $p<.01$) and for Problem Type x Answer Class ($F(1,48)=7.77$, $p<.001$). The third-order interaction effect Grade Level x Problem Type x Answer Class also reached significance ($F(1,144)=1.581$, $p<.031$). Children in the different grade levels, differed in use of the nine answer classes for the various problem types. Figure 6.1a-g shows the distribution of answers for each problem type among the different answer classes per grade level.

Various analyses of variance were performed for each problem type. When analysed separately it appeared that children of the various grade levels significantly differed in use of answer classes 1, 5 , 8 and 9 for change 3 problem-types ($F(1,3)=5.41$, $p<.01$; $F(1,3)=10.07$, $p<.001$; $F(1,3)=4.22$, $p<.01$ and $F(1,3)=4.19$, $p<.01$). Subsequent analyses demonstrated that grade four children answered change 3 problems significantly more often with one of the numbers given ($t=2.46$, $p<.05$) and added the irrelevant number significantly more often than the children in grade 5 ($t=2.80$, $p<.01$). Grade five children subtracted the irrelevant number significantly more often than grade six children ($t=3.09$, $p<.003$). Grade four children answered change 3 problems less frequently correct than grade five children ($t=2.29$, $p<.05$), who again answered these problems correctly significantly less frequent than grade six children ($t=2.43$, $p<.05$) (see Figure 6.1a).

The separate analyses for the remaining problem types (change 5 and 6, compare 1, 5 and 6 and combine 2) showed that children in the four grade levels significantly differed in answering the problems with one of the numbers given (change 5: $F(1,3)=6.05$, $p<.001$; change 6: $F(1,3)=5.47$, $p<.01$; compare 1:

$F(1,3)=8.94$, $p<.001$; compare 5: $F(1,3)=3.83$, $p<.05$; compare 6: $F(1,3)=8.10$, $p<.001$) (see Figure 6.1b-6.1g). Children in grade four answered these problem types significantly more often with one of the numbers given than the other children (change 5 : $t=3.09$, $p<.01$; change 6: $t=2.74$, $p<.01$; compare 1: $t=3.78$, $p<.001$; compare 5: $t=2.53$, $p<.01$; compare 6: $t=3.35$, $p<.001$ and combine 2: $t=2.89$, $p<.01$) (see Figure 6.1b-6.1g).

Besides, for problem types change 5 ($F(1,3)=3.76$, $p<.05$), compare 1 ($F(1,3)=3.19$, $p<.05$) and compare 6 ($F(1,3)=3.10$, $p<.01$) children in the four grade levels differed significantly in use of answer class 8, adding the irrelevant number (see Figure 6.1b, 6.1e and 6.1g). For problem types change 5 and compare 1, grade four children added the irrelevant number significantly more often than grade five children ($t=2.44$, $p<.05$ and $t=1.96$, $p<.05$), whereas for problem type compare 6 children in grade five added the irrelevant number significantly more often than grade six children ($t=2.46$, $p<.05$) (see Figure 6.1b, 6.1e and 6.1g).

For problem types change 5 and 6, compare 1, 5 and 6 and combine 2 problems children in the various grade levels significantly differed in the number of correct answers (change 5: $F(1,3)=8.70$, $p<.001$; change 6: $F(1,3)=3.91$, $p<.05$; compare 1: $F(1,3)=8.94$, $p<.001$; compare 5: $F(1,3)=11.86$, $p<.001$; compare 6: $F(1,3)=6.07$, $p<.001$; combine 2: $F(1,3)=14.06$, $p<.001$) (see Figure 6.1b-6.1g).

Grade four children answered problem types change 5, compare 1 and 5 and combine 2 correctly significantly less often than children in grade five ($t=2.15$, $p<.05$; $t=3.50$, $p<.001$; $t=3.04$, $p<.01$ and $t=3.62$, $p<.001$ respectively), who in turn answered problem types change 5 correctly ($t=2.33$, $p<.05$), compare 5 and 6 ($t=2.56$, $p<.01$ and $t=2.74$, $p<.01$) and combine 2 ($t=2.25$, $p<.05$) significantly less frequently than grade six children (see Figure 6.1b, 6.1d, 6.1e, 6.1f and 6.1g).

Finally, for compare 5 and 6 problem-types children from distinct grade levels significantly differed in use of answer class 4, "key word used" ($F(1,3)=3.46$, $p<.05$ and $F(1,3)=5.31$, $p<.01$). Grade four children used a "key-word" strategy significantly less often than the other children (compare 5: $t=5.77$, $p<.01$ and compare 6: $t=6.80$, $p<.005$) (see Figure 6.1f and 6.1g). For compare 6 problems children in grade five significantly less often ($t=2.41$, $p<.05$) and children in grade seven significantly more often ($t=2.49$, $p<.05$) used a "key-word" strategy than the other children (see Figure 6.1g).

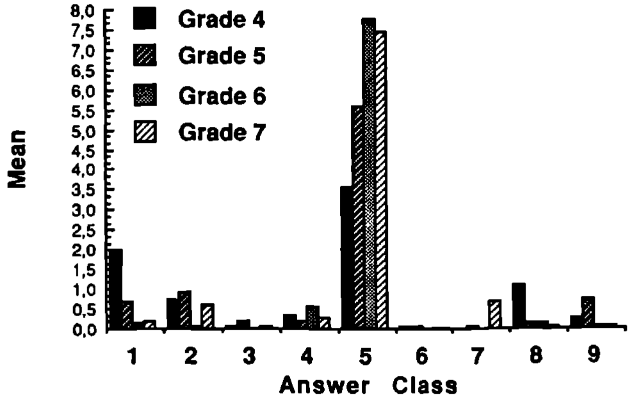


*Figure 6.1a* Mean number of responses for each answer class for the change 3 problem-type.

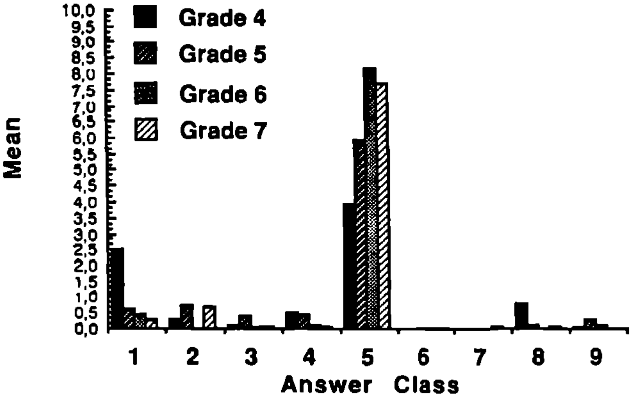


*Figure 6.1b* Mean number of responses for each answer class for the change 5 problem-type.

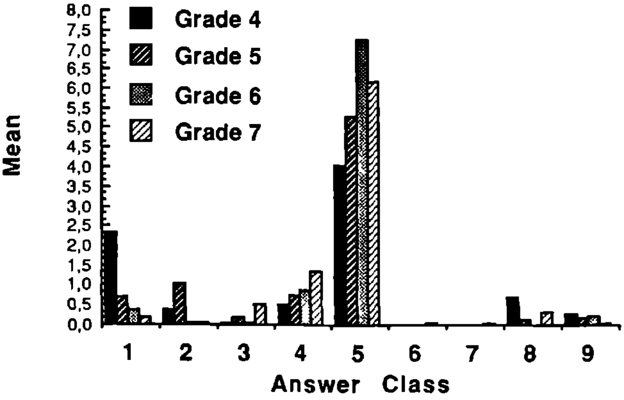


*Figure 6.1c* Mean number of responses for each answer class for the change 6 problem-type.

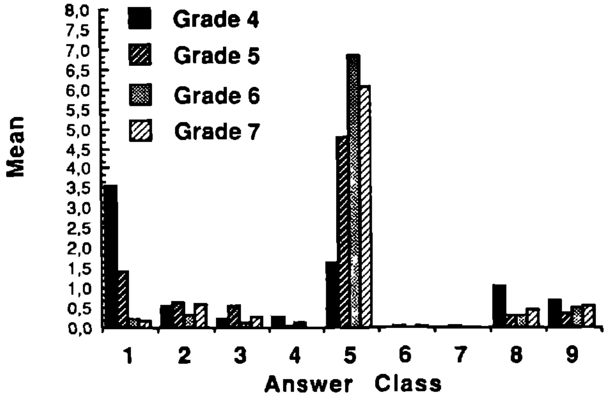


*Figure 6.1d* Mean number of responses for each answer class for the combine 2 problem-type.

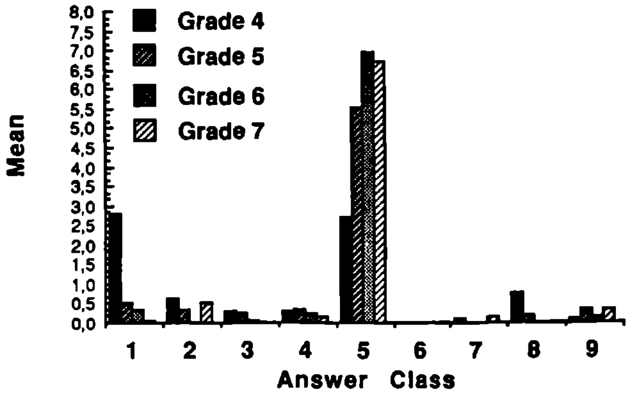


*Figure 6.1e* Mean number of responses for each answer class for the compare 1 problem-type.

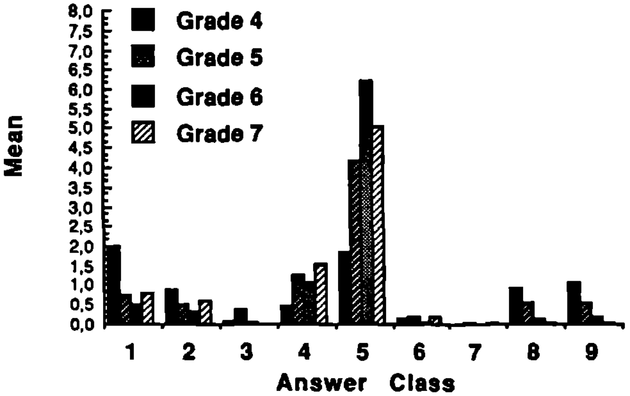


*Figure 6.1f* Mean number of responses for each answer class for the compare 5 problem-type.

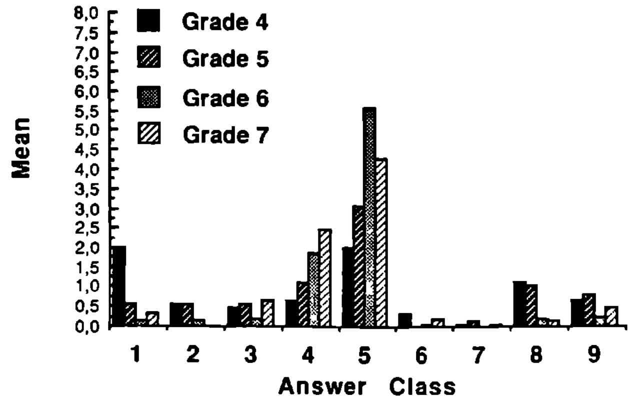


*Figure 6.1g* Mean number of responses for each answer class for the compare 6 problem-type.

## Discussion

The results showed that, when confronted with the various problem types, the children in different grade levels used different answer classes. First, for all problem types children of different grade levels differed in number of problems that were solved correctly. This frequency showed a certain development in the knowledge of arithmetic word-problem solving. For instance, subjects in grade six generally performed better than the other subjects, whereas subjects in grade four performed worst. A remarkable finding was that although children in grade seven represent a higher academic performance level, these subjects never differentiated from grade six with regard to the number of problems correctly solved and use of the other answer classes.

For all problem types, subjects in grade four more frequently exercised answer class 1, i.e. "answering with one of the numbers given", as compared to the other subjects. Visual inspection of the raw data revealed that subjects who exercised this answer class were rather persistent in answering most problem types with one of the numbers given. Obviously, these children do not yet know that solving word problems requires applying some mathematical operation on the numbers given. Verschaffel (1984) also found that novice problem solvers often answer certain word problems with a given number. Verschaffel termed this finding as a lack of "word-problem" knowledge. Beginners are not yet initiated in the "game of word problems" and thus do not yet know what precisely is expected when they first are confronted with word problems. This answer tendency may reflect both "answer guesses" and misconceptions. For since they do not know what to do, children may be lead to guess the answer by the given numbers, or they may believe, owing to their misconception of the word problem, that their "one number" answer is correct. In fact, the problem types that were used in this study may have particularly elicited "one number" answers, since they represent a set of problem types for which "one number" answers are often cited (see Verschaffel, 1984).

Likewise, CHIPS, a computer-simulation model for arithmetic word-problem solving (Briars & Larkin, 1984), produces "one number" answers for these same word-problem types if certain knowledge required to solve a particular problem type is missing. CHIPS solves simple addition and subtraction word problems by processing the problem text word by word and by constructing displays of counters. Since our children did not have concrete objects available, they may have had to resort to their internal representation of the problem. However, since the numbers used in our word problems were rather small, the children may also have used their fingers to directly model the word problems. Although CHIPS understands word problems by acting them out physically, with minor changes CHIPS is also able to solve word problems internally with imagined counters. Since Briars and Larkin (1984) made it plausible that the processes required to build these external and internal representations are similar in nature, we decided not to distinguish between the processes of external and internal modeling in the description of the problem-solving processes of the children in our study.

CHIPS solves the easiest problems by using "single-role counters" knowledge, i.e., by simply building sets and counting these sets. However, more difficult problems are solved by exercising advanced knowledge concerning "double-role counters", "subset-equivalence", "re-representation", " consistent-comparison" or "conflict-comparison". To solve change 3 and combine 2 problems (A+X=B) correctly, knowledge is required of how to find and count sets of objects that are specified as belonging to two sets ("double-role counters" knowledge). Lacking this knowledge, CHIPS has to resort to its "single-role counters" knowledge and consequently answers these problems by counting the set in its final problem representation (the final set it constructed, B). CHIPS thus comes up with one of the numbers given.

To solve change 5 and 6 problems (X+A=B and X-A=B respectively) correctly, so-called "transfer-schema" knowledge is required. This knowledge represents a deeper understanding of the reversibility of actions in time: i.e. adding ("to get") and separating ("to lose") are actions that can be reversed, resulting in the unknown becoming the result set instead of the start set. However, Briars and Larkin suggest that to solve change 5 problems adequately, knowledge concerning "subset-equivalence" may also be used. This knowledge enables the child to reverse the position of the number sets described in the problem text. In this way, a problem like X+A=B may be changed in A+X=B and subsequently solved with the knowledge used to solve the easier change 3 problems. However, a similar problem structure emerges (X-B=A) by using this "subset-equivalence" knowledge for change 6 problems, which is equally difficult to solve as the original problem structure (X-A=B). Without "transfer-schema" or "subset-equivalence" knowledge, CHIPS has to depend on its "double-role" or "single-role" counters knowledge. If CHIPS has only "single-role" counters available, then again, CHIPS counts the objects in the final set and comes up with the answer B. If CHIPS has the ability to use "double-role" counters for these problems, it comes up with the wrong answer A. Since CHIPS does not know how to respond to the initial description of an unknown set, ("some") it merely constructs a set of A objects and adds (change 5) or removes (change 6) objects to obtain a set of B objects. Subsequently, in response to the question sentence ("How many before") by using its "double-role counters" knowledge CHIPS counts the initial set of A objects. Thus, both exercising "single-role" and "double-role" counters knowledge results in answers that represent one of the numbers given in the problem texts.

Finally, solving compare 1 problems correctly, requires "consistent-comparison" knowledge. This enables CHIPS to match sets of counters and interpret "How many more than" as a cue to count the left over. In contrast, to solve compare 5 and 6 problems, CHIPS needs "conflict-comparison" knowledge. "Conflict-comparison" knowledge enables CHIPS to interpret the word "more" as a cue to move counters *out* of a set and to interpret the word "less" as a cue to move counters *into* a set. Without both "consistent-" and "conflict-" comparison knowledge CHIPS resorts to its "single-role counters" knowledge and uses each number in the problem to create a set. CHIPS thus builds a set of A and a set of B objects. But when asked how many more (or less) it simply responds with the number of the second set, B.

In conclusion, it can be hypothesized that some children in grade four, in terms of CHIPS, merely had "single-role counters" knowledge available, which resulted in "one number" answers for all problem types used in this study.

Then, subjects in grade four, who do know that a mathematical operation is demanded to solve word problems, seem unable to distinguish the relevant sets from the irrelevant set, given the high frequency of answer class 8 ("irrelevant number added") for most problem types. Both Cruickshank (1948) and Goodstein et al. (1971) already demonstrated that inserting a third irrelevant-number set in word-problem texts resulted in a decreased performance by children with learning difficulties. As far as we know, CHIPS merely solves word problems without irrelevant-number sentences. Consequently, it is unknown whether CHIPS would produce similar answers for word problems in which a third irrelevant-number sentence is inserted.

In comparison to the subjects in grade four, most subjects in grade five seem to know that word-problem solving requires the practice of an arithmetic operation, given the lower frequency of strategy 1 answers to all problem types. However, in calculating their answer these children frequently used the irrelevant number for compare 6 problems and change 3 problems. Apparently, these children also analyse the problem texts of change 3 and compare 6 rather superficially. It seems that some children in grade five, like some children in grade four, do not analyse the problem thoroughly, but rather perform a mathematical operation on two numbers in the problem, regardless of the relevance to the correct solution.

In view of the higher frequency of class 4 answers ("key-word" answers) of subjects in grade five, six and seven for compare 5 and 6 problems as compared to the children in grade four, at least some subjects seem to be guided more by problem features. But these children's analysis of the problem text is still rather shallow since they merely use the "key word" presented in choosing an arithmetic operation. For "direct" word-problem types, which were not used in this study, this "key-word" strategy would have resulted in the correct answer and thus would have never been revealed as a wrong strategy. As already said, in classifying compare problem-types Briars and Larkin (1984) distinguished between "consistent-comparison" and "conflict-comparison" knowledge. Consistent-comparison knowledge enables a problem solver to answer compare 1 and 2 problems correctly, by interpreting the words "how many more" or "how many less " as a cue to count the left-over of two previously matched sets. Solving compare 3 and 4 problems correctly, also requires "consistent-comparison" knowledge. For compare 3 and 4 problems the phrase "more than" or "less than" involves triggering a move-schema for moving x counters "more than" or x counters "less than" an already existing set. Yet, to solve compare 5 and 6 problems requires "conflict-comparison" knowledge to adequately process the "conflicting" information in the problem. Conflict-comparison knowledge enables the child to interpret the phrase "more/less" ("Peter has A apples. Peter has B more/less than Ann.") as a cue to move objects "out of" or "into" a set respectively. If CHIPS can interpret a "consistent-comparison" word as "more" or "less" as a cue to increment or decrement a set, it may increment or decrement an initial set A by B objects and subsequently count the A-/+B objects as a final answer. Eventually, this would result in "key-word" answers. In view of the high frequency of wrong "key-word" answer classes for compare 5 and 6 problem types, the subjects of grade five, six and seven presumably had "consistent-comparison" knowledge but did not have "conflict-comparison" knowledge available. In contrast, as said before, children of grade four only may have had "single-role counters" knowledge available to solve these problems and therefore may have answered all compare problem-types mainly with one of the numbers given.

These results may allow judgements on the efficacy of a product-oriented approach with regard to the revelation of strategies and thus of misconceptions and knowledge deficits of pupils. By now it is clear that a product-oriented approach based on a thoughtful construction of items not only increases our insight in the ability level of a pupil, but also in the kind of strategies used. A careful selection of problem types and an inventory of reported strategies combined with unique number triplets puts the researcher in position to accurately assess the developmental level of a pupil and to ascertain the kind of (wrong) strategy the child used to solve a particular problem type. However, certain difficulties still arise with a product-oriented approach. First, the problem types used in this study necessarily represent a subset of all main problem types described in the literature, and therefore may not constitute a good sample. Consequently, for the remaining problem types the kind of strategies used cannot be assessed unambiguously.

Second, various misconceptions may result in using one and the same strategy. Thus, attributing wrong answers to a certain strategy does not guarantee the ascertainment of the kind of misconception.

Finally, misconceptions cannot be distinguished from impasses reached during the problem-solving process (see Brown & Van Lehn, 1982) by a product-oriented approach that is merely based on unique answers. But misconceptions may produce rather stable error patterns, whereas impasses may yield less constant error patterns. In fact, as was demonstrated by Hamaker, Van Der Baaren and Brand (1985), the stability of errors proceeding from impasses may be manipulated by varying the time-interval between two similar impasses or by offering an alternative repair that may be used to overcome all sorts of impasses. On the one hand, it seems that problem solvers are not aware of their own held misconceptions and of their wrong choices regarding the employed strategy caused by this misconception. On the other hand, impasses seem to result from a conscious lack of knowledge, which the problem solver tries to overcome by using a "repair" to solve the problem. Misconceptions seem to develop earlier than impasses and seem to represent certain knowledge deficits. Impasses rather seem to represent an intermediate stage between rejecting old knowledge and acquiring new knowledge. Misconceptions and impasses therefore seem to require different instruction methods emphasizing the knowledge stage of the child.

In conclusion, a product-oriented approach in which problem types and number triplets are carefully selected allows a more precise judgement of the knowledge level of pupils and their preferences for certain strategies. Since various misconceptions may underlie one and the same strategy, a product-oriented approach seems limited in revealing all kinds of misconceptions held by pupils. Thus process-oriented research still seems required to verify whether a certain strategy results from one and the same misconception, and not from some makeshift contrivance that pupils accidently use to overcome their impasse.

# Chapter 7

# General Discussion, Implications for Future Research on Instructional Programs

In this final chapter endeavors are undertaken to infer some implications for future research on instructional designs, founded on the results of the studies described in this thesis. In the first section, the reasons for developing prototypes for research on unraveling the processing deficits associated with learning difficulties will be presented. Subsequently, it is discussed whether the decision to use the problem-solving strategies of competent problem solvers as models for the remediation of children with learning difficulties was valid. Then, in the third section, the prototypes developed thusfar are reviewed with regard to the extent to which they match the deficits particular for children with learning difficulties. More specifically, the importance of metacognitive skills for adequate arithmetic word-problem solving is stressed and the degree in which these skills were implicitly taught by the prototypes. Finally, in the last sections directions for instructional research on arithmetic word-problem solving are discussed.

## The Function of Prototype Development as a Research Tool

One of the major aims of this thesis was to study whether the arithmetic word-problem solving performance of children with learning difficulties could be improved by prototypical computerized training-programs, in which instructional elements were included which were presumed to underlie adequate arithmetic word-problem solving. Another important object of this thesis was to study the arithmetic word-problem solving processes of these children during training. More specifically, the interest was in the nature of information processing and the processing deficiencies that could account for the lower performance level of children with learning difficulties.

An issue to be addressed is whether instructional research is the most suitable method for studying the processing deficits of these children. For another way to investigate the differences in problem-solving behavior between these children and regular children, is by using a research approach in which the performance of learning impaired and regular school children is compared. It would seem that both an instructional and comparative approach have their advantages and disadvantages.

At first sight, a comparative approach may seem to suit the study of problem-solving processes better than an instructional approach, in which only children with learning difficulties are trained. Since a comparative approach allows the unraveling of variations in problem-solving processes, the causes of performance differences between the children with learning difficulties and the normally achieving group may be revealed. However, an important reason to conduct instructional research is that by demonstrating the effect of particular interventions on presumed problem-solving processes, an instructional approach may likewise yield statements about the causes of potential performance increases and thus about feasible deficits of children with learning difficulties as opposed to normally achieving children. In order to specify the relative contribution of each instructional variable to the performance gain, such statements require that the individual instructional variables are carefully controlled. This was the main reason that the presence of the instructional variables was systematically varied in the training procedures constructed and evaluated in the experiment described in chapter 5. Besides, it was presumed that the computerized prototypes of this thesis would permit the recording of the relevant intervening responses of a child during arithmetic word-problem solving, and may also give a deeper insight in the learning potentials of a particular child.

A second important reason pertains to research on Aptitude Treatment Interactions (ATI) by conducting training experiments in which several instructional variables are evaluated. Training experiments may reveal differences among subjects in performance gains, i.e. some children may benefit from a particular instruction procedure where others do not. If a training procedure can be shown to capitalize on some children's abilities, the prerequisites for training may be examined. These findings could eventually be used to assign children to optimal instructional designs.

In the third place, successful instructional research on training prototypes may lead to developing training procedures for practical applications sooner than comparative research. Rather complex instruction programs may be developed after the instructional components that proved to be functional for a particular group of children have been unraveled. In short, it was assumed that the instructional approach could aid in both improving arithmetic word-problem solving and in revealing the causes of this potential performance increase.

Despite the fact that the study reported in chapter 5 showed that the performance of the children who received training did not improve more than children who merely practised the word problems, it was demonstrated that children who received different instruction procedures varied in their use of particular problem-solving strategies. For instance, children who received External-Modeling instruction more often than the other children, used representational strategies similar to those instructed during training, i.e. representational strategies in which the semantic set-relations or actions described are stressed. Whereas children who received Text-Analysis instruction showed a tendency to reread more words and more different words in the word-problem texts than the other children. Since the children were randomly assigned to the experimental conditions, it may be assumed that the variations in instructional design induced these differences in strategies used. Against the background of the recommendations for altering the user interface of the instructional designs, more rigorous and differential training effects on performance and strategies used during problem solving may be hoped for.

In conclusion, a comparative approach as well as an instructional approach may be used to reveal the underlying causes of poor performance. An instructional approach may, however, be useful for disclosing these causes and developing various training procedures simultaneously. Children could subsequently be assigned to one of these training procedures according to their deficits.

## The Performance of Regular Children versus Children with Learning Difficulties

In designing the prototypes, reliance was on the processes involved in adequate problem solving. Instruction was focussed on the teaching and reinforcement of the problem-solving strategies manifested by regular school children. Since the efficacy of the computerized training-procedures in improving arithmetic word-problem solving in children with learning difficulties remains in doubt, the question emerges whether the problem-solving strategies of regular school children should have been used as models for the strategies that were taught to the children with learning difficulties. Although comparative research on the differences in performance between older and younger children indicate that poorer performance is correlated with inefficient use or ignorance of strategies, one could wonder whether the divergence between children with learning difficulties and regular school children mirror those differences between older and younger children. As some researchers have pointed out, children with learning difficulties are, like younger children, less efficient in acquiring and producing strategies to tackle a variety of problems (Campione, Brown and Ferrara, 1982; Rabinowitz & Chi, 1987). Besides, it is often argued that children with learning difficulties compared to normally achieving children generally do not demonstrate any unusual patterns of responding (Russel & Ginsburg, 1984; Judd & Bilsky, 1989). Instead, their difficulties usually seem to result from immaturities of academic knowledge. For example, according to Russel and Ginsburg (1984) children with learning difficulties exhibit bugs characteristic of younger children. More specifically, the errors committed by educable mentally retarded children parallelled the errors of normally achieving children, although higher proportions of errors were made by the retarded group (Judd & Bilsky, 1989). Finally, the educable mentally retarded children in our study demonstrated error types on the pretest comparable to the error types committed by the first graders of Verschaffel (1984) (see Appendix F).

In this context, the contrast between educable mentally retarded and nonretarded children mimics the differentiation between older and younger children. It may therefore be argued that the problem-solving strategies initially exhibited by young regular school children can indeed act as models to remediate children with learning difficulties. Thus the implementation of these strategies in our instructional programs for these children seemed warranted. Despite the fact that teaching arithmetic word-problem solving by such an approach seems justified, an alternative approach may have been to give more attention to the specific deficits of children with learning difficulties and attuning the training procedures more specifically to these deficits. That is to say, though the differentiation between educable mentally retarded and nonretarded children may correspond with younger- older children's differences, an important distinction seems to exist. Regular youngsters may exhibit inefficient use or ignorance of strategies but may eventually learn to use such strategies during their development, whereas, according to Campione, Brown and Ferrara (1982), children with learning difficulties may be unsuccessful in attaining and generating these strategies because of a failure to learn incidentally or to profit from regular instruction.

As was already mentioned in chapter 5, the low word-problem solving performance of children with learning difficulties may emanate from a lack of metacognitive skills besides from specific knowledge deficits. In fact, failures in metacognition, strategic processing and self-regulation (Brown & Campione, 1986; Campione, Brown, & Ferrara, 1982; Rabinowitz & Chi, 1987; Borkowski, Johnston, & Reid, 1987; Ryan, Weed, & Short, 1986) are often mentioned as one of the causes of low academic performance of children with learning difficulties. Even after intensive training in the use of a strategy, these children may fail to apply the learned strategies on transfer tasks that are similar to the training task (Gelzheiser, 1984). Nevertheless, the evidence suggests that these children can behave strategically if instructed to do so (Brown, Campione, & Day, 1981; Torgesen, 1977).

Though we never disguised the importance of metacognitive skills for adequate problem solving, these skills were not explicitly trained in this study for the reasons mentioned in the introduction of chapter 5. Yet, some characteristics of the instruction procedures implicitly pertain to training of metacognition, since guidelines for teaching self-instruction were used in designing the training procedures. In the light of the results, the question rises to what extent the subjects were remediated on metacognitive deficits at all by the prototypical training procedures used in this study. To answer this question, implications of applying the technique of cognitive behavior modification for designing instruction to teach children with learning difficulties adequate problem solving, will be discussed in the next section.

## Cognitive Behavior Modification as an Intervention Technique

Cognitive behavior modification has proved a beneficial technique for remediating the inefficient problem-solving behavior of children with learning difficulties (Meichenbaum, 1976). It seems that the potential of the cognitive behavior approach lies in the distinction between process and performance. In explaining and altering behavior, reference is made to the underlying cognitive processes. Essential to this technique is that children are both trained to perform a specific task strategy and to monitor the execution of the task strategy. To emphasize the relation between the actions to be performed and the final task outcome, training procedures should be characterized as follows: (1) children should be involved as active participants in the learning process, (2) the desired response should be identified by a series of discrete steps, (3) modeling of the target strategy should be employed, (4) overt verbalization usually is required at some point, and (5) the goal of training should be a planful, reflective response style (after Ryan et al., 1986). In discussing the extent to which the prototypes of arithmetic word-problem solving developed thusfar address the self-regulation of strategy use, reference will be made to these five features.

With regard to the first feature, it may be claimed that all prototypes for arithmetic word-problem solving can be characterized as an environment, in which children actively engaged in solving word problems by performing each step of the task strategy independently. The response required at each point of the task strategy of the different training procedures consisted of actively pointing at words in the problem text, pointing at specific parts on the screen to indicate the desired placement of squares, moving visual squares on the screen from a supply to a worksheet, answering the problem by entering a numerical answer on some sort of number bar etc. In performing the strategy steps, children were only given computer assistance when they made a mistake in performing a strategy step or reached an impasse. It was argued that these training procedures would posit control within the child and consequently would result in greater feelings of self-management and greater expectations for success. But since the children did not read the feedback contents displayed on the screen, the function of each step within the total task strategy may not have occurred to them. Consequently, children may have lost track of the task strategy instead of the training inducing a feeling of self-control over the process.

Despite the child's autonomy in performing the strategy steps, the training procedures followed a rather strict structure. This was decided on the principle that poor problem solvers evidently profit more from a structured approach than from a procedure in which the course of actions is determined by the children themselves (Brown, Campione, & Day, 1981; Bernaert, 1984). With regard to the second feature mentioned, in all training procedures the children had to perform a sequence of strategy steps in a strict successive order. It was believed that children would thus be helped in attaining the knowledge to use a word-problem solving strategy by making them aware of each step of the task strategy (see Brown, Campione, & Day, 1981). But again, the children's success rates in performing the task strategy adequately did not increase, presumably owing to their reluctance in reading feedback contents. As a result, the children did not acquire the problem-solving strategies instructed.

With regard to both requirements of active-learning participation and division of total response in discrete steps, it may be argued that delivering feedback contents orally, which was already proposed in chapter 5, would aid a better understanding of the function of each step in the total task strategy. Eventually, this could stimulate children to actively engage in performing each step of the task strategy and thus promote feelings of self-control as well as attainment of the task strategy.

Modeling of the target strategy by the trainer preceded the actual training period. During two instruction sessions, the trainer both overtly modeled and verbalized the actions that the child was expected to perform independently in the subsequent training sessions. Besides, in the second instruction session, the word problems were tackled both by the trainer and the child. In carrying out the task strategy together with the trainer, the child was allowed to help perform the actions by touching specific areas on the screen while the trainer encouraged the child to verbalize the actions performed. Finally, the child had to perform the task alone. As outlined by Meichenbaum and Goodman (1971), the child may progress from external regulation to internalized self-regulation of problem-solving behavior by a transition from overt verbalizations by the trainer to covert verbalizations by the child. The rationale for verbalizing is based on the idea that verbalizations can induce self-control over the entire problem-solving process by continuously acting as prompts for the following actions to be performed. Yet, the aim of both modeling and verbalizing the steps to be performed was primarily to explain the task strategy to the child and check for any misunderstandings by the child of the training procedure. In contrast, verbalizations in the self-instruction training aims at promoting the internalization and maintenance of the instructed task strategy by having the child internally regulate the entire problem-solving route. Since the instruction sessions were not explicitly intended to support self-instruction by verbalizations and presumably were too short to teach the children to verbalize the actions they performed, self-control was not promoted by the computerized prototypes. Although explicit instruction in overt verbalizing as well as the prolongation of the instruction phase until the child is at least able to carry out and verbalize the task strategy independently could be considered, it is uncertain whether children will continue to verbalize during training in which they are no longer encouraged to verbalize their actions and these verbalizations are not evaluated.

On the one hand, as regards the need to control these overt verbalizations during training, computerized instruction poses a severe problem because computers are still unable to understand the verbal input of a child. On the other hand, besides promoting the self-regulation of the problem-solving strategy, verbalizations serve another function in the self-control training, which may be fulfilled by the computer interacting with the child. The second function of verbalizations is the opportunity offered to the human trainer to check and regulate the ongoing problem-solving process in detail by giving feedback at any moment during this process. An important merit of the segmentation of the problem-solving route in small successive steps in the computerized instruction-programs is the possibility to monitor and check each intermediate response in the problem-solving route. Since the child's responses consisted of pointing at specific areas on the screen, this allowed the computer to evaluate and correct these responses. In short, although the computer device did not allow for the controlling of the child's verbalizations, the child's pointing at specific screen areas offered the possibility to monitor and check the ongoing problem-solving process. Nevertheless, revisions of the training procedures which instruct the children to verbalize each action they perform may produce stronger training effects. But to monitor and evaluate the overt verbalizations of the child during training would still require the presence of a human trainer, who could intervene with the verbal utterances of the child. As long as computers are unable to interpret speech, this aspect of self-instruction training will be impossible to implement in computerized instruction.

Finally, the goal of training should be a planful, reflective response style. Presumably, this goal pertains to the transfer and generalization of strategies as a most stunning finding is the persevering failure of children with learning difficulties to apply previously learned strategies to new but similar tasks (Campione & Brown, 1977; Campione, Brown, & Ferrara, 1982). The problem of strategy generalization seems to emanate from the child's failure to recognize and understand that a strategy can be applied in other situations as well. As Brown et al. (1981) put it: "Children should be fully informed participants in any training enterprise, i.e. they should be helped to understand why they should be strategic and when it is necessary to do so." Brown et al. (1981) further argue that children should be trained in the self-management of the strategies they must employ and that children with learning difficulties in particular may need explicit training. Self-regulation training complements instruction in carrying out a strategy with information concerning its significance and with training in planning, monitoring and checking. Presumably, with regard to these last requirements the computerized training-procedures thusfar developed demonstrate some shortcomings. Despite the fact that all training procedures aimed at providing information regarding the usefulness of the task strategy by means of the contents of feedback, this feedback proved too complex to achieve this requirement. Moreover, though the trainer acted as a model in planning, monitoring and checking and drew the child into these activities during the instruction sessions, instead of the child, the computer took care of these activities in the training phase. Some earlier training procedures in which children were explicitly trained to plan, monitor and check their problem-solving route by means of a "planning list" proved effective (Van Lieshout, 1986, submitted), whereas the efficacy of another training procedure, in which this metacognitive component was likewise incorporated, was not demonstrated (Dankers, 1990). Unfortunately, the training procedures differed in other aspects as well, such as training and feedback contents, which complicates matters for sound statements of the function of the planning list. Moreover, the step labels of the planning list of these training procedures varied as a consequence of these differences in training contents. Nevertheless, since the first training procedure developed by Van Lieshout in which both Text Analysis and Planning were trained, proved rather effective, it can be assumed that the metacognitive component and not so much the Text-analysis component caused the strong training effect. But the efficacy of this training procedure could also result from combining Text-Analysis instruction with this metacognitive-instruction component. It therefore seems compelling to investigate the possibility of installing this planning list in each of the prototypes and examining the contribution of this instruction component on the total training effect of each prototype by systematically varying its presence and absence.

With regard to all the requirements of self-instruction training mentioned, it may be argued that aside from the potential revisions of the prototypes already discussed, another extension of the computer equipment may be considered. A striking finding in a study on the word-problem solving performance of educable mentally retarded and nonretarded children was that the problem context, which was manipulated by varying the problem cover-story, differentially affected the performance of the educable mentally retarded and nonretarded children (Judd & Bilsky, 1989). Problem contexts describing changes in abstract entities, such as miles run or hours worked, proved far more difficult for educable mentally retarded subjects than changes in concrete quantities, such as dolls given, candy lost etc. Another notable finding was that even for adult students the solving of change 3, 4, 5 and 6 word problems that described unrelated coactors took considerably longer than similar word problems describing related coactors (Reusser, 1989), which points at an increase in difficulty to comprehend the problem situation described. As Reusser put it "...coactors relating to each other in familiar ways are more easily contributing to the forming of a more familiar and more coherent story script than do coactors which are not related at all and therefore are beneficial to the construction of a situation model."

According to Van Dijk and Kintsch (1983) automized problem solving is characterized by the practice of the "generalization operator" which is used to strip the names and objects of their identity and subsequently to specify these entities in terms of set roles and relations in the problem representation. In order to solve a word problem, it is unnecessary to memorize that dolls were given instead of miles run. Yet it is crucial to remember and understand that 5 was the first quantity of dolls specified in the problem text. Judd and Bilsky (1989) argued that the educable mentally retarded children were not using this "generalization operator" mechanism consistently, since the difficulty of the problems should not have been affected by the specific entities described. In view of the fact that children with learning difficulties indeed prove poor word-problem solvers, they probably do not yet have problem schemata internally available that they can activate and fill in by replacing specific names and objects into set labels.

In pursuing the comparison further, the children in this study may also have failed to apply the "generalization operator". Although the word problems used in this study did not describe such abstract entities as "hours worked", a certain level of abstraction was required in order to understand the function of the visual squares on the screen as representations of the entities mentioned in the problem text. This function would probably have become apparent if the children had read the feedback contents in which reference was made to the quantity described in the problem text and the visual squares representing this quantity. But in spite of the failure of the feedback contents to clarify this function, the manipulation of visual squares on a computer screen nevertheless seems far more alienated from the problem situation described, than manipulating objects on a table in the presence of dolls (see for example Verschaffel, 1984). As the research finding of Judd and Bilsky (1989) suggested, especially children with learning difficulties may experience difficulties in ignoring certain contextual features that in fact are irrelevant to finding the solution of the problem.

In view of this finding, extension of the prototypes with animations could be considered. Especially the dynamic nature of change problems could be expressed by making use of animations in which the actions described are actually performed by for instance two actors exchanging candy, dolls etc., or by one actor actually loosing or finding candy etc. Since animations may induce a rather passive attitude with the children, they may eventually be encouraged to mimic the animation with blocks and dolls. Another possibility is to install user-controlled animations by which the child may actively engage in modeling the word problem. In this way, a gradual transition from animations to less concrete representations may take place.

## Future Research on Instructional Designs

In the preceding paragraphs an account is rendered of the development of prototypes as tools to investigate the children's thought processes in detail, and the practical implications proceeding from the research findings with regard to potential revisions of the prototypes are discussed. Since these adjustments may be realized in the short term, future research could examine the efficacy of these revised prototypes in remediating the poor word-problem solving behavior of children with learning difficulties. Next to research on these new prototypes, other research could be directed towards developing more intelligent remedial tutorial systems for arithmetic word-problem solving. The purpose of the next sections is to outline a possible research route from which the construction of a more Intelligent Tutorial System (ITS) for arithmetic word-problem solving may eventually be attained. First, some shortcomings and additional requirements of the current prototypes to remediate children on an individual basis are discussed. Subsequently, directions for empirical research subservient to designing instructional programs that attune to the developmental level of children are discussed. Finally, desirable adjustments to the present computer-simulation models are presented in view of their utility as a starting point in the development of a more intelligent tutorial system for arithmetic word-problem solving.

## Comments on the current prototypes

Despite the fact that after revision the computerized prototypes may induce stronger training effects, these prototypes would still fall short of remediating children on an individually tailored basis. Although at present for each word-problem type the responses of the child on the probe trials preceding the actual training trials determine whether training will be started, the current computerized training-programs merely offer one route along which children are remediated.

Moreover, while the current prototypes allow the recording of detailed information regarding the problem-solving behavior of the child, these prototypes evaluate and correct the intermediate problem-solving responses of a child and thus intervene with the problem-solving process. As a consequence, the informal problem-solving route of the child and the final outcome of such a route cannot be fully disclosed by using these prototypes. Such inquiries seem crucial in linking up instruction with the child's present and absent knowledge. To detect and diagnose children's informal problem-solving strategies would require research in which guidance and intervention of the problem-solving route remain in abeyance.

## Research on informal strategies

In order to detect and diagnose children's misconceptions and incomplete knowledge of arithmetic word-problem solving, two possible routes may be followed.

On the one hand, research could be started which aims at disclosing the informal word-problem solving strategies of children with learning difficulties by examining the strategies they use spontaneously to solve word problems. To reveal particular misconceptions and knowledge deficits of the child, the gathering of information on the children's performance on different tasks may be considered. The children may thus be asked to answer the word problem first and subsequently to model the word problems by using materials such as blocks and a number-line while verbalizing their problem-solving route. Likewise, children may be asked to recall problem texts either before or after solving the word problem (see Dellarosa et al., 1986) and children may be requested to formulate a number sentence (see Verschaffel, 1984). Answering the word problem directly may disclose mental strategies, whereas the modeling of word problems could be useful in revealing material strategies and using the number-line could be helpful in disclosing verbal-counting strategies. Along these lines the developmental level of a child in solving word problems may be uncovered. In addition, verbal protocols and error patterns may provide data concerning children's individual misconceptions and incomplete knowledge. Presumably, the child's particular misconceptions may also be disclosed by the method of contrasuggestion; if the verbal protocol of the child seems to point at certain incomprehensions, word problems could be presented that counter these misconceptions. To investigate whether children with learning difficulties are actually comparable to younger nonretarded children, such research could include both these children and nonretarded children equalized on important domain-specific knowledge such as word-problem solving performance. These groups could be followed on their development in word-problem solving performance by investigating their problem-solving behavior at subsequent time points (see Carpenter & Moser, 1982, 1984; Verschaffel, 1984). Research of this kind has been started recently.

On the other hand, the construction of a computerized diagnostic-instrument could be considered to record children's problem-solving processes. Designing such a computerized diagnostic-instrument has the advantage that it is relatively easy to record several kinds of intermediate responses of a subject during word-problem solving. Another advantage pertains to the possibility to incorporate such a computerized diagnostic-instrument in planned ITS. In designing this diagnostic instrument, use could be made of the existing computerized program for recording reading behavior and the instructional program for teaching representational strategies. The instructional program for teaching representational strategies would have to undergo changes in order to permit the recording of children's informal representational strategies without intervening. However, in order to follow the child's problem-solving process, this process need to be structured in some degree. The actual computerized diagnostic-instrument could allow children to alternate between analysing the problem text and representing the actions or relations described by modeling with objects on the screen. Unfortunately, except for the loss of information concerning the verbalizations of the child, the recording of responses by the diagnostic instrument would be confined to the material strategies of the child. Besides recording material strategies, the recording of verbal-counting strategies and mental strategies would become very difficult or even impossible to implement in the existing computer program for teaching representational skills. It seems therefore advisable to both investigate children's word-problem solving processes with and without computers. Both verbal protocols and error patterns could be gathered during the solution process, which would provide data concerning individual misconceptions and incomplete knowledge that may account for poor word-problem solving performance. As was described in chapter 6, a modest start has been made to reveal the kinds of strategies used by children producing wrong answers. Although the usefulness of the product-oriented approach described as a method to disclose the underlying misconception is limited, such an approach may at least yield firm statements concerning the degree in which the child masters certain problem types. Presumably, such knowledge could also aid and be used in designing ITS.

## From simulation to Intelligent Tutorial System

Recently, three computer-simulation models have been proposed as cognitive models of the knowledge and procedures required for solving elementary word problems (Riley et al., 1983; Briars & Larkin, 1984; Dellarosa, 1986). The architecture of the models has already been described in full in chapter 1 and in summary in chapter 4. Inasmuch as these models can account for existing data of the word-problem solving processes of regular school children, these simulation programs may be used as a starting point in designing ITS for arithmetic word-problem solving. However, some reservations should be made with regard to the direct use of these models as viable models of the word-problem solving performance of children with learning difficulties.

First, as De Corte and Verschaffel (1988) pointed out, at least the simulation models of Riley et al. and Briars and Larkin cannot account for all empirical findings with regard to children's problem-solving processes. In view of the fact that ARITHPRO (Dellarosa, 1986) merely simulates the performance of third graders, all the more this argument holds for ARITHPRO.

Second, on account of the fact that the computer models developed thusfar set out to simulate the performance of regular school children, it is questionable whether the same fit between these models' performances and the performance data of children with learning difficulties will be found. In view of the lower performance of children with learning difficulties as opposed to normally achieving children in general, it may be that in particular the low performance level of the youngest children with learning difficulties is not simulated by the models. To illustrate, some first graders in the study of Verschaffel (1984) were unable to solve even the easy change 1 problem at the start of the school year. Verschaffel argued that these children lack so-called general "word-problem schema" knowledge that pertains to knowledge about the implicit rules assumed in word-problem solving. All simulation models start from the principle that these rules are known to a beginning problem-solver and thus all models solve change 1 problems correctly. Although the subjects of the studies reported in this thesis demonstrated error types comparable to the subjects of Verschaffel (see for example Appendix F), the educable mentally retarded children were far older than the subjects of Verschaffel. However, especially the performances of younger children with learning difficulties may not match the performance of Verschaffel's subjects. Consequently, these children in particular may show a certain lack of this "word-problem schema" knowledge and thus solve even the easy change 1 problem incorrectly.

Third, the simulation models of Riley et al. and Dellarosa postulate the construction of mental problem schemata during development according to which children store and process word-problem text information. In processing this text information, all slots of these schemata are filled with text information according to a strict principle. The simulation models postulating these problem schemata are rather stringent in the kind of word problems which they are able to solve. But for a problem solver who is already able to solve change 1 problems it may pose no problem to similarly solve "Ann had 3 marbles. John gave Ann 2 marbles and Peter gave Ann 4 marbles. How many marbles does Ann have now?" correctly. Thus it seems that human problem solvers may be able to enlarge or change their problem schemata rather easily according to small changes in word-problem type, whereas these schemata are rather static in the simulation models.

Fourth, both simulation models of Riley et al. and Briars and Larkin solve word problems by creating sets, enlarging or reducing sets and counting sets. The simulation model of Dellarosa solves word problems merely by triggering arithmetic counting procedures. In this way, the simulation models only simulate the problem-solving performance of children operating on the material or verbal-counting level. If these simulation programs are to be used as a starting point in designing ITS that links up with the developmental level of the child, the problem solving performance of children operating at the mental level ought to be simulated as well.

In the fifth place, all three simulation models process word-problem texts in a left to right order without returning to previous text parts and construct problem schemata in a similar sequence. It is, however, questionable whether the text-analysis processes and representational strategies of children resemble the models' sequence of problem-text processing and filling slots in the problem schemata.

Finally, although the simulation models describe several developmental stages at each of which certain knowledge is added to existing knowledge structures, a description of how the transition from one developmental stage to the next occurs is lacking. In view of designing ITS, knowledge of these transitions is extremely important because such knowledge may yield precise guidelines for instruction. For instance, the transition from direct modeling to the use of counting strategies seems to involve a critical improvement in procedural skills and understanding. In direct modeling, each of the entities is represented sequentially. This simplifies problem solving because it is only necessary to keep an account of one number set of the problem at a time. In contrast, the relationship between each of the different number sets described is represented in a single counting sequence in the primary verbal-counting strategies. To enable the use of these counting strategies a better understanding of the relation between the different sets described in the problem is required. Besides, since each of the sets is constructed individually, direct modeling merely requires knowledge of how to construct a set of a given size. However, the use of counting strategies requires knowledge of some form of double counting, because it is essential to count forward or backward from a given number and to keep track of the number of steps in the counting sequence in order to know when to stop counting. As long as it remains unclear how more complex knowledge structures emanate from more simple knowledge structures, precise directions for training will remain undiscovered.

All in all, if the abovementioned imperfections of the simulation models are to be removed, a considerable amount of research is still required. As a start, one could think of using the model of Briars and Larkin, since this simulation model as opposed to the other models does not postulate distinct schemata for representing the main problem types, change, combine and compare. It may thus be hypothesized that Briars and Larkin's simulation model is less fixed as regards its knowledge structures and thus may be the most easy to enlarge or change. Besides, it may be argued that starting problem solvers do not have cognitive schemata available such as postulated by Riley et al. and Dellarosa by which they can organize the word-problem information beforehand. Despite the presumed better fit of the simulation model of Briars and Larkin, this model would still need to be revised to account for transition stages in development, mental strategies, crossing text-analysis processes etc. As suggested by Riley, et al. (1983), transition in learning stages may be elucidated by examining the learning principles or heuristics used in prevailing theories of learning and development by which children are assumed to exchange strategies for others. For example, a child may learn to "count on from a given number" and as such avoid "redundant processing of information" by counting the start set once more and thereby noticing the redundancy of counting the similar set all over again. Consequently the child may learn to eliminate this redundancy and may "count on" by beginning with the cardinality of the set already counted (Neches, 1981; Klahr & Wallace, 1976 in Riley et al., 1983, pp. 190-191). On the one hand, the development of more efficient procedures may lay the foundations of more sophisticated knowledge structures. On the other hand, existing knowledge structures may influence what procedures will be acquired. That is, it is unlikely that a child would acquire "count on" until the child at least has the knowledge available for representing quantitative information. Similar learning heuristics such as the one cited may be implemented in the simulation model for arithmetic word-problem solving.

Finally, this revised simulation model could be tested on its sufficiency by clustering and using data sources concerning text-analysis processes, material, verbal-counting and mental strategies and stratified verbal protocols of children with learning difficulties as input for the computer-simulation model. The steps and errors at each level produced by the simulation model would then have to be compared with empirical data of children with learning difficulties. Presumably, this simulation model would have to be extended and revised until it would account for the word-problem solving processes of children with learning difficulties.

Besides making use of the data sources cited for extending and revising the simulation model, one could think of intermediate refinements of the computerized training-prototypes developed thusfar. A first refinement could be to incorporate a diagnostic product-oriented component by which the typical errors of the children are analysed and subsequently remediated (see chapter 6). Another possibility may lie in extending the training procedures with the diagnostic instrument for recording children's reading and representational processes. Since installing this diagnostic instrument would permit decisions regarding the moment at which intervention in the problem-solving process of the child should take place, the remediation could be even more individually tailored than at present. Eventually, integration of this revised simulation model and the computerized diagnostic-instrument could be pursued. This would result in a first prototypic ITS which could subsequently be investigated on its efficacy in comparison with traditional CAI.

## Summary

Especially children with learning difficulties prove poor word-problem solvers. Since word problems offer a context in which arithmetic knowledge can be applied, the solving of these problems may enlighten the practical usefulness of mathematical concepts. Word-problem solving may thus prepare for real-life situations in which arithmetic knowledge has to be used. The research reported in this thesis is concerned with remediating word-problem solving in children with learning difficulties. In doing so, several prototypic instruction programs were developed. The main focus of these instruction programs was on training specific principles required for proficient word-problem solving. To reveal those principles, both empirical studies of the problem-solving processes of normally achieving children and research on developing computer-simulation models for word-problem solving were reviewed in the first Chapter.

Results of these studies lend support to the notion that adequate word-problem solving requires some sort of representation of the problem, whether internally or externally, in order to decide on a solution strategy. First, to build such representation, essential text information has to be identified and processed. Since children with learning difficulties display superficial reading habits, a first training procedure that focuses on thorough text analysis was constructed. In this training procedure children are taught to select the text information that reflects set-descriptions, the underlying semantic set-relations and actions of the problem. The results of the pilot work with this first training procedure, which has been reported elsewhere (Van Lieshout, 1986; Van Lieshout, submitted), proved the efficacy of this "Text-Analysis" training procedure when performed by a human trainer as well as the efficacy of the computerized version.

Although this training procedure let the child select information defining the problem structure and thus attempts to guide the construction of a proper representation, building such representation is not explicitly trained. Since regular school children naturally build physical representations that reflect the original problem statements, another way to remediate word-problem solving in children with learning difficulties is to teach them to use representational strategies similar to those used by regular school children. Hence a second training procedure aiming at teaching representational skills to children with learning difficulties was developed, the "External-Modeling" training (see last sections Chapter 1).

The contents of this training procedure are described in Chapters 2 and 3. The results of the pilot studies with this training procedure, which were obtained in multiple-baseline designs across subjects, demonstrated that this training procedure was effective in improving word-problem solving in educable mentally retarded children, whether performed by a human trainer (Chapter 2) or presented by the computer (Chapter 3).

A computerized training-procedure was also developed that focuses on both remediating poor text analysis and representational skills. The contents of this computerized training-procedure are described in Chapter 4. In fact, the instruction components of both the earlier training methods were incorporated in this training procedure. The results obtained with this training procedure were far less convincing than the results with the older training procedures. The training procedure was only slightly effective in improving the performance of five mentally retarded children on word problems. It was hypothesized that this result should mainly be attributed to the larger number of steps that had to be performed by the children, which presumably obscured the overview of the task strategy instructed.

In light of the limited success of this latest training procedure, a revised computerized prototype was developed in which the number of actions of the task strategy has been reduced, but in which training of both text analysis and external modeling has been preserved. Consequently, both the computerized training-procedure for teaching text analysis and for teaching representational skills were likewise revised.

Chapter 5 constitutes the experimental part of this thesis, and concerns the study of the contribution of the Text-Analysis and External-Modeling component to the overall training effect. In a 2 x 2 factorial randomized-block design, the influence of both Text-Analysis and External-Modeling instruction on the performance and problem-solving behavior on word problems was investigated in isolation as well as in combination, which resulted in four experimental conditions. Eighty-four educable mentally retarded children from seven different schools for Special Education participated in the experiment. Various ability tests and process measures were used to assess the pretest, posttest and follow-up performance of the subjects. To investigate the effect of the Text-Analysis and External-Modeling component on the reading behavior of the subjects, a touch-contingent computerized recording technique was used by which children's text-analysis processes were studied. To investigate the effect of the External-Modeling and Text-Analysis component on the representational strategies employed by the subjects after training, a test was constructed by which a tester could trace the concrete modeling behavior that the child spontaneously exposed.

One of the main hypotheses concerned differences in the number of correctly solved problems. Besides, it was assumed that children who received External-Modeling instruction would outperform children who received Text-Analysis instruction on the trained word problems, whereas children who received Text-Analysis instruction would outperform children who received External-Modeling instruction on untrained word problems. A multivariate covariance analysis revealed no significant training effects on the number of correctly solved word problems, and no differential training effects on trained versus untrained word problems.

Two other central hypotheses concerned differences in the activity of looking behavior with respect to question sentence reading and rereading. It was assumed that children who received Text-Analysis instruction would demonstrate more looking activities in the question sentence than children who received instruction in External-Modeling, whereas the latter would demonstrate more rereadings than the former. The assumption that children who received Text-Analysis instruction would show more looking activities during question reading was not confirmed by a multivariate covariance analysis in which the number of words touched, the number of different words touched and the gaze durations served as dependent variables. Instead, children who received Text-Analysis instruction showed a disposition to read more words and more different words in both the initial reading of the problem and the reading of the question sentence. A similar analysis of the rereadings yielded an insignificant trend in the data opposed to the direction hypothesized. Again, children in the Text-Analysis conditions showed a tendency to reread more words and more different words than the other children.

The hypothesis that children who received External-Modeling instruction would improve in the number of word problems correctly solved on the representation test with concrete materials available was confirmed. Also the hypothesis that these children would demonstrate representational strategies conforming to the modeling strategies instructed during training was confirmed. Finally, as predicted, children who received Text-Analysis training more often skipped the irrelevant number set in representing the word problems with objects. In contrast, as predicted, children who received External-Modeling instruction did not only represent this irrelevant set more often, but also represented this irrelevant set in sequence with its occurrence in the problem text.

A number of factors were cited that might have been responsible for the non-occurrence of training effects on the number of correct answers, of which the failure of the user-interface to stimulate children to read feedback contents was the most important. While training effects could not be established, some interesting findings pointed at altered problem-solving processes. Broadly speaking, children who received Text-Analysis instruction seemed to analyse word-problem texts more thoroughly, whereas children who received External-Modeling instruction were better in solving word problems for which materials to model the actions or relations described were offered.

The main conclusion drawn from this experiment was that extension of the user-interface of all computerized prototypes to allow the use of speech in presenting both feedback contents and problem texts orally, would probably lead to more rigorous training effects in the future.

However, in designing instructional programs the focus could also be on individualized practice by taking into account the particular deficits and errors of the child. In Chapter 6 a first study attempting to infer the expertise level and misconceptions of children with the use of a product-oriented approach is described and evaluated. The approach consists of constructing a typical set of word-problem types and number triplets for which specific combinations enable the revelation of the strategy used by the child. The results indicated that educable mentally retarded children of various grade levels differed in the use of strategies for the various word-problem types. It was argued that these differences reflected different levels of expertise. The main conclusion drawn from this study was that a product-oriented approach in which problem types and number triplets are carefully selected admits a more precise judgement of the expertise level of a child than a random selection of word problems and numbers used. However, such an approach seems limited in revealing the particular misconception fundamental to the specific answer produced, since various misconceptions might underlie the use of one and the same strategy.

Finally, implications for future research on instructional design are discussed in the last Chapter. In this Chapter both direct implications for practical revisions in the prototypic training programs as well as possible directions for developing more intelligent tutorial systems (ITS) for word-problem solving in the future are presented. One of these direct implications pertains to the possibility to install instruction on metacognitive skills in the computerized prototypes more explicitly and to examine the contribution of this instruction component on potential training effects. A more theoretical implication refers to the possibility to study the problem-solving processes of children with learning difficulties compared to those of normally achieving children equalized on word-problem solving ability without intervention. Additionally, the utility of computer-simulation models as starting point in designing ITS is discussed. More specifically, suggestions for changing and extending the simulation model of Briars and Larkin (1984) are given. It is argued that both research lines could converge by using empirical data as input for the simulation model, as well as by comparing the output of the model to the empirical data concerning children's problem-solving processes. It is assumed that the simulation model could be revised and eventually used in designing ITS.

## Samenvatting

Met name kinderen met leerproblemen zijn slechte redactierekenaars. Aangezien redactieopgaven een context bieden waarin deze kinderen rekenkundige kennis kunnen toepassen, kan hen het practische nut van mathematische begrippen duidelijk worden door deze opgaven op te lossen. Zo kan redactierekenen voorbereiden op dagelijkse situaties waarin rekenkennis gebruikt dient te worden. Het onderzoek waarvan dit proefschrift verslag doet, betreft de remediatie van het redactierekenen van kinderen met leerproblemen. Hiertoe werden verschillende prototypische trainingsprogramma's ontwikkeld, waarin de nadruk lag op het instrueren van specifieke kennis die vereist lijkt voor redactierekenen. Ter vaststelling van deze kennis werd in Hoofdstuk 1 een overzicht gegeven van zowel het empirische onderzoek naar de oplosprocessen van reguliere basisschool leerlingen als het onderzoek op het gebied van de ontwikkeling van computersimulatiemodellen voor redactierekenen.

De bevindingen van dit onderzoek ondersteunen de idee dat voor het adequaat oplossen van redactieopgaven een of andere probleemrepresentatie, intern danwel extern, vereist is om de juiste oplossingsstrategie te kunnen kiezen. Ten eerste vereist de constructie van zo'n representatie een adequate analyse van de opgavetekst. Aangezien kinderen met leerproblemen onder andere gekenmerkt worden door oppervlakkige tekstanalyse werd een eerste training ontwikkeld die gericht is op een grondige analyse van de opgaveteksten bij het redactierekenen. Deze training leert kinderen met leerproblemen informatie uit de tekst te selecteren waarin de verzamelingen, relaties en acties tussen verzamelingen beschreven wordt. De werkzaamheid van deze "Tekstanalyse" training werd aangetoond in een vooronderzoek met een menselijke trainer (Van Lieshout, 1986) alsook in een vooronderzoek met de computergestuurde versie (Van Lieshout, submitted).

Hoewel deze training kinderen leert belangrijke tekstinformatie te selecteren, die de onderliggende probleemstructuur definieert en zo de vorming van een juiste interne probleemrepresentatie ondersteunt, wordt het construeren van zo'n representatie niet expliciet getraind. Aangezien reguliere basisschool leerlingen spontaan materiële representaties opbouwen die de oorspronkelijke probleemstructuur weergeven, is het aanleren van deze representatiestrategieën aan kinderen met leerproblemen een andere mogelijkheid om slecht redactierekenen te remediëren. Hiertoe werd een tweede training ontwikkeld om kinderen met leerproblemen het concreet representeren van redactieopgaven te leren, de "Representatie" training (zie laatste paragrafen van Hoofdstuk 1).

De taakstrategie van deze training is beschreven in Hoofdstuk 2 en 3. De resultaten van de vooronderzoeken met deze training, die via een meervoudig-basislijn design over proefpersonen werden verkregen, toonden aan dat de training effectief was in het verbeteren van de redactierekenvaardigheid in moeilijk lerende kinderen, zowel waneer deze verricht werd door een menselijke trainer (Hoofstuk 2) als door de computer (Hoofdstuk 3).

Tenslotte werd een derde computergestuurde training ontwikkeld met het doel zowel oppervlakkige tekstanalyse te remediëren als representatievaardigheden te bevorderen. De taakstrategie van deze training is beschreven in Hoofdstuk 4. In feite bestaat de training uit de instructiecomponenten van de eerder ontwikkelde trainingen. De resultaten die met deze training werden verkregen waren minder overtuigend dan de resultaten met de twee eerder ontwikkelde trainingen. De training bleek matig effectief in het verbeteren van de redactierekenvaardigheid van vijf moeilijk lerende kinderen. Verondersteld werd dat dit resultaat voornamelijk toe te schrijven was aan de toename van het aantal stappen in de taakstrategie die de kinderen moesten uitvoeren voordat zij mochten antwoorden. Door het grote aantal stappen raakten kinderen vermoedelijk het overzicht over de taakstrategie kwijt.

In het licht van het matige success van de laatste training werd een gereviseerde trainingsversie ontwikkeld waarin het aantal uit te voeren stappen werd teruggebracht maar waarin een grondige analyse van de opgavetekst en het representeren van de opgave nog steeds getraind wordt. Als gevolg van deze wijzigingen werden ook de training voor tekstanalyse en de training voor representatievaardigheid dienovereenkomstig aangepast.

Hoofstuk 5 betreft het experimentele deel van dit proefschrift waarin de bijdrage van de Tekstanalyse-component en de Representatie-component op het totale trainingseffect systematisch werd nagegaan. In een 2 x 2 factorieel gerandomiseerd block-design werd de invloed van de Tekstanalyse- en de Representatie-component op de prestatie en op de probleemoplossingsstrategieën, zowel in combinatie als afzonderlijk, nagegaan, hetgeen vier experimentele condities opleverde. Om de voortoets-, natoets- en follow-up-prestaties van de kinderen te bepalen werden verschillende vaardigheidstoetsen en procesmaten gebruikt. Om het effect van de Tekstanalyse-component en de Representatie-component op het leesgedrag van de kinderen na te gaan werd een aanraak-contingente, computergestuurde registratietechniek gebruikt waarmee de tekstanalyse-processen van de kinderen werden bestudeerd. Om het effect van de Tekstanalyse- en Representatie-component op de representatiestrategieën van de kinderen te onderzoeken werd een toets geconstrueerd waarmee een proefleider de externe representatiestrategieën die de kinderen na de training spontaan hanteerden, kon achterhalen.

Een van de voornaamste hypothesen betrof verschillen in het aantal correcte antwoorden. Daarbij werd ondermeer voorspeld dat kinderen die getraind werden in het representeren van redactieopgaven na de training beter zouden presteren op getrainde opgaven dan kinderen die getraind werden in tekstanalyse, terwijl de laatsten beter zouden presteren op ongetrainde opgaven dan de eersten. Via een multivariate covariantie-analyse werden geen significante trainingseffecten op het aantal correcte antwoorden aangetoond en ook geen differentiële trainingseffecten op de getrainde en ongetrainde opgaven.

Twee andere belangrijke hypothesen hadden betrekking op verschillen in leesgedrag met betrekking tot het lezen van de vraag en het herlezen van de opgavetekst. Verondersteld werd dat kinderen die getraind werden in tekstanalyse meer leesactiviteiten ten toon zouden spreiden in de vraagzin dan kinderen die getraind werden in het representeren van de opgave, terwijl de laatsten delen van de opgavetekst vaker zouden herlezen dan de eersten. De aanname dat kinderen die getraind werden in tekstanalyse meer leesactiviteit in de vraagzin zouden manifesteren werd niet bevestigd via een multivariate covariantie-analyse met het aantal woorden aangeraakt, het aantal verschillende woorden aangeraakt en de kijkduur in de vraagzin als afhankelijke variabelen. In tegenstelling hiermee bleken de kinderen die getraind waren in het analyseren van de tekst de neiging te vertonen om meer woorden en meer verschillende woorden te lezen zowel tijdens de eerste lezing van de opgave als tijdens het lezen van de vraag. Een vergelijkbare analyse van de herleesfase gaf een niet-significante trend te zien in de richting tegengesteld als voorspeld werd. Ook nu bleken kinderen in de Tekst-Analyse condities een tendens te vertonen om meer woorden en meer verschillende woorden te herlezen dan de andere kinderen.

De hypothese dat kinderen die getraind werden in het representeren van redactieopgaven beter zouden presteren op de representatietest waarin fiches ter beschikking stonden om de opgaven te representeren werd bevestigd. Ook werd de voorspelling bevestigd dat deze kinderen representatiestrategieën zouden gebruiken die grote overeenkomst vertoonden met de getrainde representatiestrategieën.

Tenslotte werd bevestigd dat kinderen die getraind werden in het analyseren van de opgavetekst het concreet representeren van de irrelevante verzameling zouden overslaan op de representatietest. In tegenstelling hiermee werd aangetoond, zoals voorspeld was, dat kinderen die getraind werden in het representeren van de opgave, deze irrelevante verzameling niet alleen vaker representeerden maar bovendien ook nog in volgorde met de beschrijving van deze verzameling in de opgavetekst.

Verschillende factoren die mogelijk verantwoordelijk zijn voor het niet optreden van trainingseffecten werden toegelicht, waarvan het falen van de user-interface in het aansporen van de kinderen de feedback te lezen de belangrijkste was. Ofschoon trainingseffecten niet vastgesteld konden worden, wezen andere bevindingen op veranderde probleemoplossingsprocessen. In het algemeen leken kinderen die getraind werden in het analyseren van de opgavetekst na training de opgaveteksten daadwerkelijk beter te analyseren, terwijl kinderen die getraind werden in het representeren van de opgave beter presteerden op de representatietest met materiaal.

De belangrijkste conclusie van dit experiment was dat aanpassingen in de user-interface van alle prototypen, zodat spraak gebruikt kan worden om de feedback en de opgavetekst mondeling aan te bieden, wellicht in de toekomst tot sterkere trainingseffecten zal leiden.

Bij het ontwerpen van instructieprogramma's kan de nadruk ook liggen op geïndividualiseerde instructie door de specifieke kennisleemten en fouten van een kind in aanmerking te nemen. In Hoofdstuk 6 wordt een eerste studie besproken en geëvalueerd waarin gepoogd wordt via een product-gerichte aanpak het vaardigheidsnivo en de misvattingen van kinderen te achterhalen. De benadering bestaat uit het samenstellen van een verzameling redactieopgavetypen en getallencombinaties waarmee via specifieke combinaties de strategie te achterhalen is die een kind bij een bepaalde opgave gebruikte. De resultaten gaven aan dat moeilijk lerende kinderen van vier verschillende klassenivo's verschillende strategieën hanteerden voor het oplossen van de verschillende opgavetypen. Geconcludeerd werd dat deze variaties in strategiegebruik verschillende nivo's van expertise weergaven. De voornaamste conclusie van dit onderzoek was dat een product-gerichte benadering waarin probleemtypen en getallencombinaties zorgvuldig geselecteerd zijn, een meer nauwkeurige inschatting toestaat van het kennisnivo van een leerling dan een toevallige selectie van opgaven en getallen. Een product-gerichte benadering lijkt echter te beperkt om de specifieke misvatting te achterhalen die leidde tot het gegeven antwoord, omdat verschillende misvattingen aan één antwoord ten grondslag kunnen liggen.

Tenslotte worden in het laatste hoofdstuk aanbevelingen gedaan voor trainingsonderzoek in de toekomst. In dit hoofdstuk worden zowel directe implicaties voor practische aanpassingen van de prototypische trainingsprogramma's besproken alsook aanbevelingen gedaan voor het ontwikkelen van intelligente tutoriële systemen (ITS) voor redactierekenen in de toekomst. Een van deze aanbevelingen betreft de mogelijkheid in alle prototypen een instructiecomponent te installeren die meer gericht is op expliciete training van metacognitieve vaardigheden en de bijdrage van deze component aan mogelijke trainingseffecten te onderzoeken. Een indirecte implicatie heeft betrekking op de mogelijkheid de oplossingstrategieën van kinderen met leerproblemen te bestuderen en te vergelijken met die van reguliere schoolkinderen van hetzelfde redactierekennivo, zonder interventie. Ook wordt de bruikbaarheid van de computersimulatiemodellen voor redactierekenen als uitgangspunt voor de ontwikkeling van ITS besproken. Met name worden suggesties aangaande verbeteringen en uitbreidingen van het simulatiemodel van Briars en Larkin gegeven. Beargumenteerd wordt dat beide onderzoekslijnen zouden kunnen convergeren door aan de ene kant de empirische bevindingen met betrekking tot de oplosprocessen van kinderen als input voor het simulatiemodel te gebruiken en aan de andere kant de output van het model te vergelijken met deze empirische gegevens. Uiteindelijk zou het simulatiemodel na een dergelijke revisie gebruikt kunnen worden bij het ontwerpen van ITS.

# References

Bebout, H. C. (1990). Children's symbolic representation of addition and subtraction word problems. *Journal for Research in Mathematics Education, 2*, 123-131.

Bernaert, G. F. (1984). Sturing van het onderwijsleerproces in relatie tot cognitieve capaciteit [Guiding the learning process in relation to cognitive capacity]. In A. Dirkzwager, S. D. Fokkema, G. C. Van Der Veer, & J. J. Beishuizen (Eds.), Leren met computers in het onderwijs [Learning with computers in education] (pp.96-123) SVO reeks. Den Haag: SVO.

Bilsky, L. H., & Judd, T. (1986). Sources of difficulty in the solution of verbal arithmetic problems by mildly retarded and nonretarded individuals. *American Journal of Mental Deficiency, 4*, 395-402.

Blumberg, C. J. (1984). Comments on a "simplified time-series analysis for evaluating treatment interventions". *Journal of Applied Behavior Analysis*, 539-542.

Briars, D. J., & Larkin, J. H. (1984). An integrated model of skill in solving elementary word problems. *Cognition and Instruction, 1*, 245-296.

Borkowski, J. G., Johnston, M. B., & Reid, M. K. (1987) Metacognition, motivation and controlled performance. In S. J. Ceci (Ed.), *Handbook of cognitive, social and neuropsychological aspects of learning disabilities, Volume 2* [pp. 147-173]. London: Erlbaum Press.

Brown, A. L., & Campione, J. C. (1986). Psychological theory and the study of learning disabilities. *American Psychologist, 14*, 1059-1068.

Brown, A. L., Campione, J. C., & Day, J. D. (1981). Learning to learn: on training students to learn from texts. *Educational Researcher, 2*, 14-21.

Brown, J. S. , & Van Lehn, K. (1982). Towards a generative theory of "bugs". In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction, a cognitive perspective* (pp. 117-135). Hillsdale NJ: Erlbaum Press.

Campione, J. C., & Brown, A. L. (1977). Memory and metamemory development in educable mentally retarded children. In R. V. Kail, & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp.376-406). Hillsdale, NJ: Lawrence Erlbaum Associates.

Campione, J. C., Brown, A. L., & Ferrara, R. A. (1982). Mental retardation and intelligence. In. R. J. Sternberg (Ed.), *Handbook of human intelligence* [pp. 392-490]. New York; Cambridge University Press.

Carpenter, T. P., Hiebert, J., & Moser, J. M. (1981). Problem structure and first grade children's solution processes for simple addition and subtraction word problems. *Journal for Research in Mathematics Education, 12*, 27-39.

Carpenter, T. P., Hiebert, J., & Moser, J. M. (1983). The effect of instruction on children's solutions of addition and subtraction word problems. *Educational Studies in Mathematics, 14*, 55-72.

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

Carpenter, T. P., & Moser, J. M. (1984). The development of addition and subtraction problem solving skills. In T. P. Carpenter, J. M. Moser, & T. A.

Romberg (Eds.), *Addition and subtraction: a cognitive perspective* (pp. 9-24). Hillsdale, New Jersey: Erlbaum Press.

Carpenter, T. P., & Moser, J. M. (1985). Representation of addition and subtraction word problems. *Journal for Research in Mathematics Education, 19,* 345-357.

Conners, F. A., Caruso, D. R., & Detterman, D. K. (1986). Computer-assisted instruction for the mentally retarded. *International Review of Research in Mental Retardation, 14,* 105-134.

Cruickshank, W. M. (1948). Arithmetic ability of mentally retarded children: I. Ability to differentiate extraneous material from needed arithmetic facts. *Journal of Educational Research, 42,* 161-170.

De Corte, E., & Somers, R. (1981). Het schatten als heuristiek bij het oplossen van redactieopgaven. [ Estimation as heuristic for the solving of arithmetical word problems]. *Pedagogisch Tijdschrift/Forum voor Opvoedkunde, 6,* 351-361.

De Corte, E., & Verschaffel, L. (1981). Children's solution processes in elementary arithmetic problems: analysis and improvement. *Journal of Educational Psychology, 6,* 765-779.

De Corte, E., & Verschaffel, L. (1988). Computer simulations as a tool in research on problem solving in subject-matter domains. *International Journal of Educational Research, 1,* 49-69.

De Corte, E., & Verschaffel, L. (1987). *Do non-semantic factors also influence the solution process of addition and subtraction word problems ?* Paper presented at the Second European Conference for Research on Learning and Instruction, September, Tübingen, Germany.

De Corte, E., & Verschaffel, L. (1986). *Eye-movement data as access to solution processes of elementary addition and subtraction problems.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

De Corte, E., & Verschaffel, L. (1985). Oplossingsstrategieën van eersteklassers bij eenvoudige redactieopgaven over optellen en aftrekken. Pedagogische Studiën, 62,125-138.

De Corte, E., & Verschaffel, L. (1987). The effect of semantic structure on first graders' strategies for solving addition and subtraction word problems. *Journal for Research in Mathematics Education, 7,* 363-381.

De Corte, E., Verschaffel, L., & De Win, L. (1985). Influence of rewording verbal problems on children's problem representations and solutions. *Journal of Educational Psychology, 77,* 460-470.

Dankers, D. (1990). Spelen met woorden en blokken. Tekstanalyse en materiële representatie als basis voor een redactiereken-training [Playing with words and blocks. Text analysis and concrete modeling as basis for a training for word-problem solving]. Unpublished manuscript, Department of Special Education, Nijmegen, The Netherlands.

Dellarosa, D. (1986). A computer simulation of children's arithmetic word problem solving. *Cahier Research Methods, Instruments & Computers, 18,* 147-154.

Dellarosa, D., Kintsch, W., Weimer, R., & Reuser, K. (1986).Understanding and solving arithmetic word problems. Paper presented at the AERA Annual Meeting, San Francisco.

Dumont, J. J. (1982). *Leerstoornissen. Deel 3. controversen en perspectieven.* Rotterdam: Lemniscaat.

Ferguson, G. A. (1971). Statistical analysis in psychology and education (pp.369-373). Tokyo: McGraw-Hill Kogakuska.

Fletcher, C. R. (1985). Understanding and solving arithmetic word problems: a computer simulation. *Behavior Research Methods, Instruments, & Computers, 17,* 565-571.

Fuson, C. F., & Willis, G. B. (1989). Second graders' use of schematic drawings in solving addition and subtraction word problems. *Journal of Educational Psychology, 81,* 514-520.

Gelzheiser, L. M. (1984). Generalization from categorical memory tasks to prose by learning disabled adolescents. *Journal of Educational Psychology, 76,* 1128-1138.]

Goodstein, H. A., Bessant, H., Thobideau. G., Vitello, S., & Vlahakos, I. (1972). The effect of three variables on the verbal problem solving of educable mentally handicapped children. *American Journal of Mental Deficiency, 6,* 703-709.

Goodstein, H. A., Cawley, J. F., Gordon, S., & Helffgott, J. (1971). Verbal problem solving among educable mentally retarded children. *American Journal of Mental Deficiency, 76,* 238-241.

Hagler, P., & Knowlton, J. (1987). Invalid implicit assumption in CBI comparison research. *Journal of Computer-Based Instruction, 14,* 84-88.

Hall, R. J. (1980). An information processing approach to the study of exceptional children. In B.K. Keogh (Ed.), *Advances in Special Education: Vol. 2. Perspectives on Applications* (pp. 9-15). Greenwich, CT: Jai Press.

Hamaker, C., Van Der Baaren, J., & Brand, M. (1985). De reparatietheorie en de verklaring van fouten in het oplossen van eenvoudige redactieopgaven [Repair theory and the explanation of errors in solving simple word problems] in H. G. L. C. Lodewijks & P. R. J. Simons (Eds.), *Zelfstandig leren* (pp.174-182). Bijdragen tot de Onderwijs Research Dagen 1984. Lisse: Swets & Zeitlinger.

Heller, J. I., & Greeno, J. G. (1978). Semantic processing of arithmetic word problem solving. Paper presented at the Midwestern Psychological Association Convention, Chicago, May.

Hiebert, J., Carpenter, T. P., & Moser, J. M. (1982). Cognitive development and children's solutions to verbal arithmetic problems. *Journal for Research in Mathematics Education, 2,* 83-98.

Hudson, T. (1983). Correspondences and numerical differences between disjoint sets. *Child Development, 1,* 84-90.

Ibarra, C. G., & Lindvall, C. M. (1979). *An investigation of factors associated with children's comprehension of simple story problems involving addition and subtraction prior to formal instruction on these operations.* Paper presented at the Annual Meeting of the National Council of Teachers of Mathematics, Boston, April.

Ibarra, C. G., & Lindvall, C. M. (1982). Factors associated with the ability of kindergarten children to solve simple arithmetic story problems. *Journal of Educational Research,* 149-155.

Jaspers, M. W. M., & Van Lieshout, E. C. D. M. (1989a). A training procedure aiming at improving word problem solving for children with mathematical deficiencies. In J. J. Dumont, & H. Nakken (Eds.), *Learning disabilities: Vol. 2: cognitive, social and remedial aspects* (pp. 117-137). Amsterdam: Swets & Zeitlinger.

Jaspers, M. W. M. & Van Lieshout, E. C. D. M. (submitted b). A tutorial CAI for instructing text analysis and modeling of word problems to children with learning deficiencies. Manuscript submitted for publication.

Jaspers, M. W. M., & Van Lieshout E. C. D. M. (1989b). Een trainingsprogramma voor kinderen met leerproblemen gericht op het aanleren van concrete representaties voor redactieopgaven [A training procedure for children with learning problems aiming at teaching modeling strategies for word problems]. *Pedagogische Studiën, 6,* 240-255.

Jaspers, M. W. M., & Van Lieshout, E. C. D. M. (submitted a). Improving word problem solving in mildly mentally retarded children by instructing modeling strategies. Manuscript submitted for publication.

Jaspers, M. W. M., & Van Lieshout, E. C. D. M. (1989c). Training specific modeling strategies for word problem solving in mildly mentally retarded children. Paper presented at the Computers in Psychology conference, Groningen, The Netherlands. Manuscript accepted for publication in "Computers in Psychology".

Jaspers, M. W. M., & Van Lieshout, E. C. D. M .(1990). Visuele modellen leren opbouwen in een computergestuurde training voor redactierekenen [Learning to construct visual models in a computerized training-procedure for word problem solving]. *Tijdschrift voor Orthopedagogiek, 5,* 277-286.

Jerman, M. E., & Mirman, S. (1974). Linguistic and computational variables in problem solving. *Educational Studies in Mathematics, 5,* 17-62.

Jerman, M. E., & Van Rees, R. (1972). Predicting the relative difficulty of verbal arithmetic problems. *Educational Studies in Mathematics, 4,* 306-323.

Judd, T. P., & Bilsky, L. H. (1989). Comprehension and memory in the solution of verbal arithmetic problems by mentally retarded and nonretarded individuals. *Journal of Educational Psychology, 4,* 541-546.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329-354.

Kintsch, W. (1986). Learning from text. *Cognition and Instruction, 3,* 87-108.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 85,* 363-394.

Kratochwill, T. R. (1978). Foundations of time-series research. In T.R. Kratochwill (Ed.). *Single subject research strategies for evaluating change* (pp. 1-100). New York: Academic Press.

Lindvall, C. M., & Ibarra, C. G. (1980). Incorrect procedures used by primary grade pupils in solving open addition and subtraction sentences. *Journal for Research in Mathematics Education, 11,* 50-62.

Loftus, E. F., & Suppes, P. (1972). Structural variables that determine problem-solving difficulty in computer-assisted instruction. *Journal of Educational Psychology, 6,* 531-542.

McConkey, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17,* 578-586.

Meichenbaum, D. (1976).Cognitive factors as determinants of learning disabilities: A cognitive functional approach. In R. Knights & D. Bakker (Eds.), *The neuropsychology of learning disorders: Theoretical approaches* (pp. 423-442). Baltimore: University Park Press.

Meichenbaum, D., & Goodman, J. (1971). Training impulsive children to talk to themselves: a means of developing self-control. *Journal of Abnormal Psychology, 77,* 115-126.

Mojet, J. W. (1984). Toepassingsmogelijkheden van computers in het speciaal onderwijs [Application of computers in special education]. *Pedagogisch Tijdschrift, 9,* 94-104.

Morales, R. V., Shute, V. J., & Pellegrino, J. W. (1985). Developmental differences in understanding and solving simple mathematics word problems. *Cognition and Instruction, 2,* 41-57.

Moser, J. M., & Carpenter, T. P. (1982). *Using the microcomputer to teach problem solving skills: program development and initial pilot study* (Working paper No. 328). Madison: Wisconsin Center for Education Research.

Oud, H., Reelick, F., & Raaymakers, M. (1986). TIDA, KUNST keyword-version. Input-output description. Unpublished manuscript, Department of Special Education, University of Nijmegen, The Netherlands.

Rabinowitz, M, & Chi, M. T. H. (1987). An interactive model of strategic processing. In S. J. Ceci (Ed.), *Handbook of cognitive, social and neuropsychological aspects of learning disabilities, Volume 2* [pp. 83-102]. London: Erlbaum Press.

Resnick, L. B., & Ford, W. W. (1981). *The psychology of mathematics for instruction.* Hillsdale, New Jersey; Erlbaum Associates.

Reusser, K. (1989). Textual and situational factors in solving mathematical word problems. Paper presented at the Third Conference of the European Association for Research on Learning and Instruction, Madrid, September 4-7.

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.

Russel, R. L., & Ginsburg, H. P. (1984). Cognitive analysis of children's mathematic difficulties. *Cognition and Instruction, 2,* 217-244.

Ryan, E., Weed, K. A., & Short, E. J. (1986). Cognitive behavior modification: promoting active, self-regulatory learning styles. In J. K. Torgesen and Wong B. Y. (Eds.) *Psychological and educational perspectives on learning disabilities* [pp. 367-397]. New York: Academic Press

Sleeman, D., & Brown, J. S. (1982). Intelligent tutoring systems. London: Academic Press.

Slife, B. D., Weiss, J., & Bell, T. (1985). Separability of metacognition and cognition: problem solving in learning disabled and regular students. *Journal of Educational Psychology, 4,* 437-445.

Sternberg, R. J. (1981). Cognitive-behavioral approaches to the training of intelligence in the retarded. *The Journal of Special Education, 15,* 165-183.

Suppes, D., Loftus, E. F., & Jerman, M. (1969). Problem solving on a computer-based teletype. *Educational Studies in Mathematics, 2,* 1-15.

Torgesen, J. K. (1977). Memorization processes in reading disabled children. *Journal of Educational Psychology, 69,* 571-578.

Treffers, A., & Goffree, F. (1985).Rational analysis of realistic mathematics education: the wiskobas program. In L. Streefland (Ed.), *Proceedings of the Ninth International Conference for the Psychology of Mathematics Education. Vol. 2. Plenary addresses and invited papers* (pp.97-122). Utrecht, The Netherlands: State University of Utrecht, Subfaculty of Mathematics, Research Group on Mathematics Education and Educational Computer Center.

Tryon, W. W. (1982). A simplified time-series analysis for evaluating treatment interventions. *Journal of Applied Behavior Analysis, 15,* 423-429.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Van Essen, G., Hamaker, C., & Van Grafhorst, H. (1989). Zelf geconstrueerde grafische modellen als strategie voor het oplossen van redactiesommen. [Selfconstructed graphic models as problem solving strategy for word problems]. In P. Span, E. De Corte, & B. Van Hout Wolters (Eds), Onderwijsleerprocessen: Strategieën voor de verwerking van informatie (pp.171-183). Amsterdam: Swets & Zeitlinger.

Van Lieshout, E. C. D. M. (1985). Aandachtsproblemen en hun behandeling [Attention problems and their treatment]. *Tijdschrift voor Orthopedagogiek, 24,* 218-238.

Van Lieshout, E. C. D. M. (submitted). *A computer assisted training procedure to improve arithmetic word problem solving in mildly mentally retarded children.* Manuscript submitted for publication.

Van Lieshout, E. C. D. M. (1986). *Developing a computer-assisted strategy training procedure for children with learning deficiencies to solve addition and subtraction word problems.* Paper presented at the AERA, San Francisco (ERIC Documents Reproduction Service No. ED 271 285).

Van Lieshout, E. C. D. M. (in press). Recording of reading behavior with touch contingent word presentation. Manuscript accepted for publication in "Computers in Psychology".

Van Lieshout, E. C. D. M. (1984). Strategietraining voor het oplossen van visuele vergelijkingstaken ter bevordering van aandacht bij moeilijk lerende kinderen: een oogbewegingsonderzoek. Unpublished doctoral dissertation, University of Nijmegen, The Netherlands.

Van Lieshout, E. C. D. M., & Anbeek, G. H. J. (1988). Using a touchscreen to monitor and to modify task behavior during arithmetic word problem solving. In F. J. Maarse, L. J. M. Mulder, W. P. B. Sjouw, & A. E. Akkerman (Eds.), *Computers in psychology: methods, instrumentation and psychodiagnostics* (pp. 201-205). Lisse: Swets & Zeitlinger.

Van Lieshout, E. C. D. M. , & Jaspers, M. W. M. (1990). A training procedure for children with learning deficiencies to improve their representation of simple arithmetic word problems. In H. Mandl, E. De Corte, N. Bennett, & H. F. Friedrich (Eds.), *Learning and instruction. European research in an international context.* Vol. 2.2, analysis of complex skills and complex knowledge domains (pp. 431-444). Oxford: Pergamon.

Van Lieshout, E. C. D. M., & Jaspers, M. W. M. (1989). Reading and solving arithmetic word problems by poorly performing and normal children. In A.

Rosa (chair), *Learning environments for special educational needs: techniques for improving academic performance*. Symposium conducted at the third conference of the European Association of Research on Learning and Instruction, Madrid, Spain.

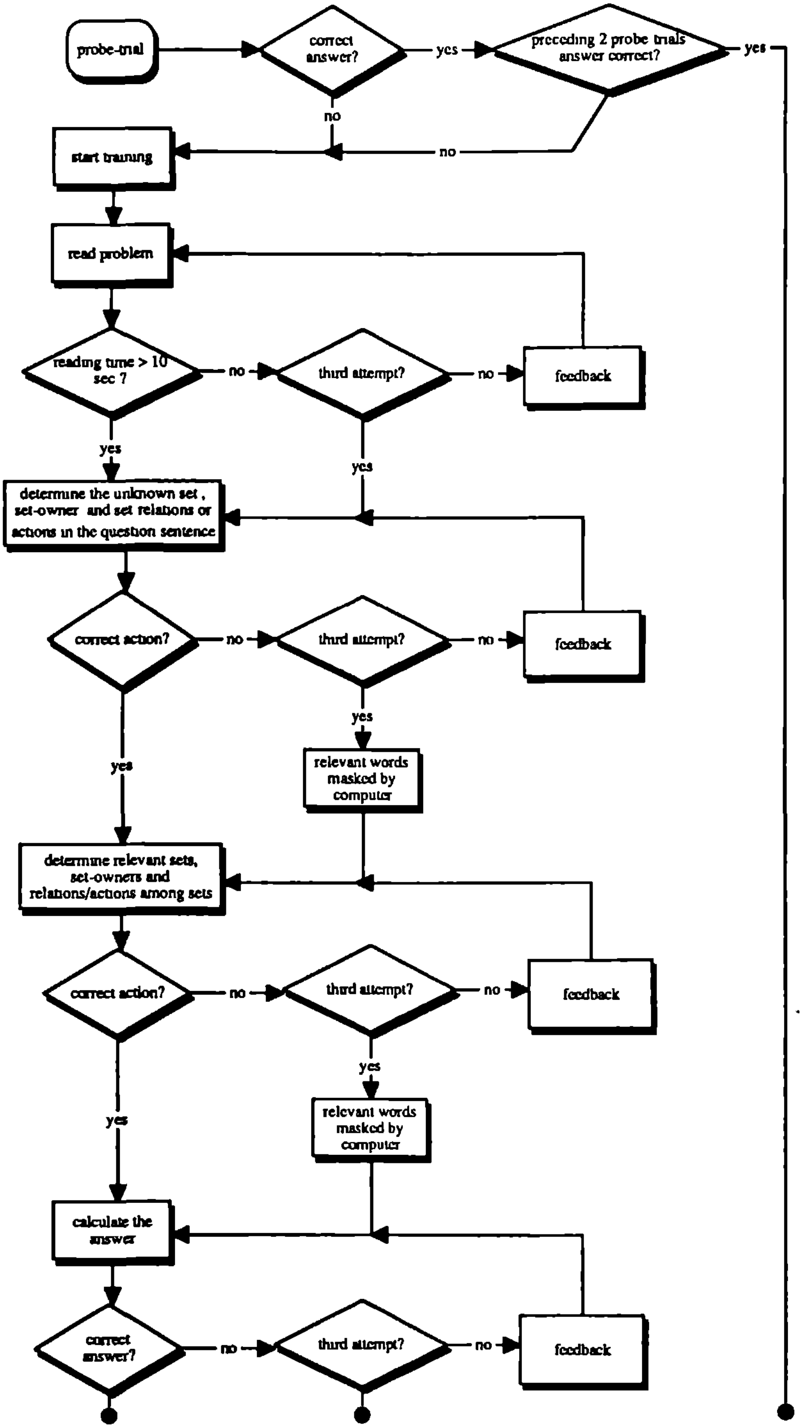
Vergnaud, G. A. (1981). A classification of cognitive tasks and operations of thought involved in addition and subtraction problems. In T. P. Carpenter, J. M. Moser & T. Romberg (Eds.), *Addition and subtraction: Developmental perspective*. (pp.39-59) Hillsdale, New Jersey: Erlbaum Associates.

Verschaffel, L. (1984). *Representatie- en oplossingsprocessen van eersteklassers bij aanvankelijke redactieopgaven over optellen en aftrekken. Een theoretische en methodologische bijdrage op basis van longitudinale, kwalitatief psychologische studie* [First graders' representations and solving processes on elementary addition and subtraction word problems. A theoretical and methodological contribution based on a longitudinal, qualitative psychological investigation]. Unpublished doctoral dissertation, University of Leuven, Belgium.

Wolery, M., & Billingsky, F.F (1982). The application of Revusky's Rn Test to slope and level changes. *Behavioral Assessment, 4,* 93-103.

# Appendices

## Appendix A


probe-trial
correct answer?
yes
preceding 2 probe trials answer correct?
yes
no
no
start training
read problem
reading time > 10 sec ?
no
third attempt?
no
feedback
yes
yes
determine the unknown set, set-owner and set relations or actions in the question sentence
correct action?
no
third attempt?
no
feedback
yes
yes
relevant words masked by computer
determine relevant sets, set-owners and relations/actions among sets
correct action?
no
third attempt?
no
feedback
yes
yes
relevant words masked by computer
calculate the answer
correct answer?
no
third attempt?
no
feedback

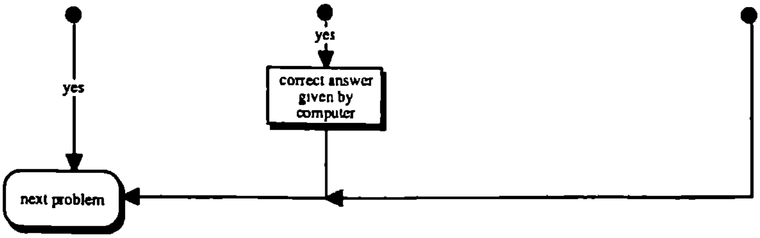
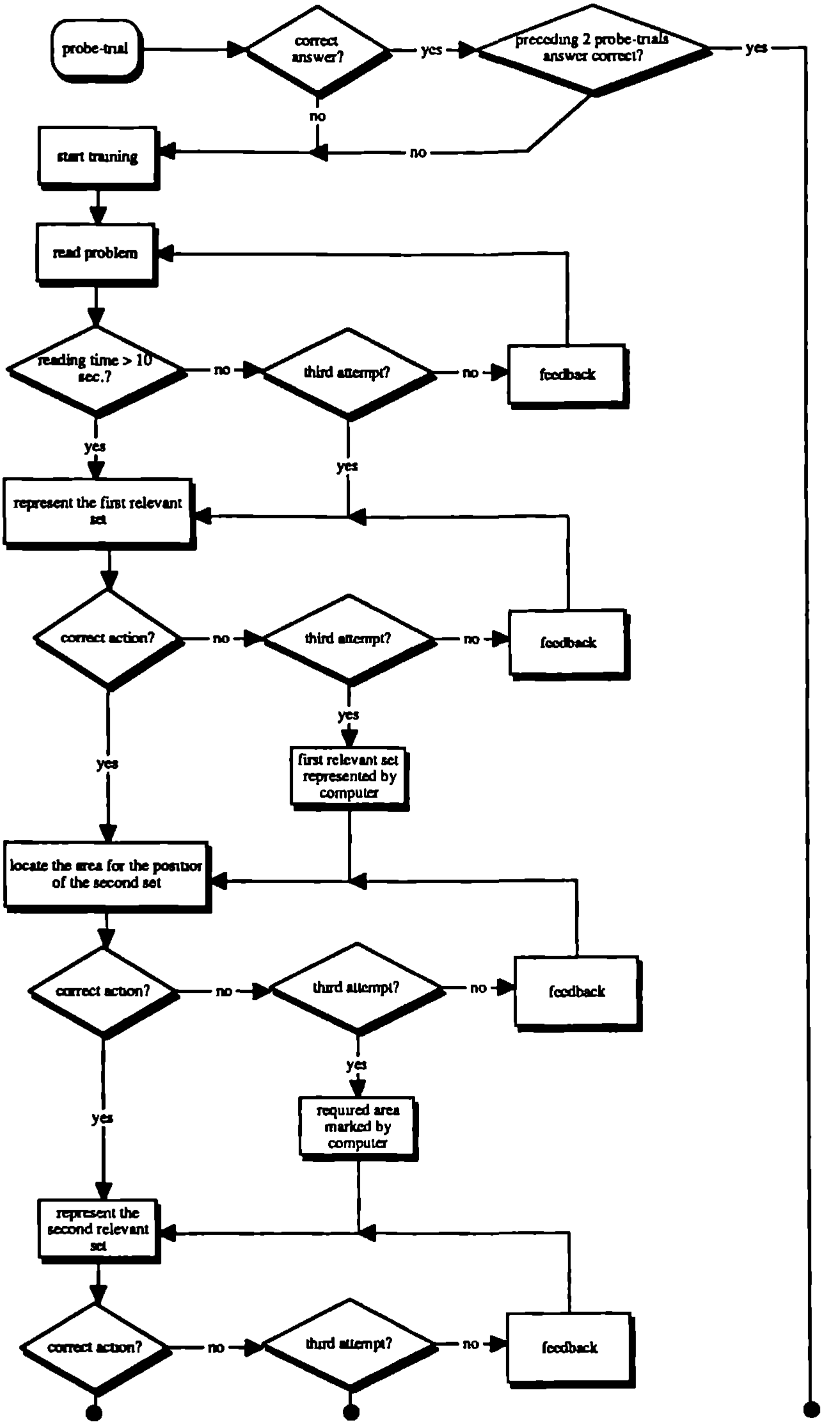


*Figure 1* The task strategy of the Text-Analysis training.


probe-trial
correct answer?
yes
preceding 2 probe-trials answer correct?
yes
no
no
start training
read problem
reading time > 10 sec.?
no
third attempt?
no
feedback
yes
yes
represent the first relevant set
correct action?
no
third attempt?
no
feedback
yes
yes
first relevant set represented by computer
locate the area for the position of the second set
correct action?
no
third attempt?
no
feedback
yes
yes
required area marked by computer
represent the second relevant set
correct action?
no
third attempt?
no
feedback

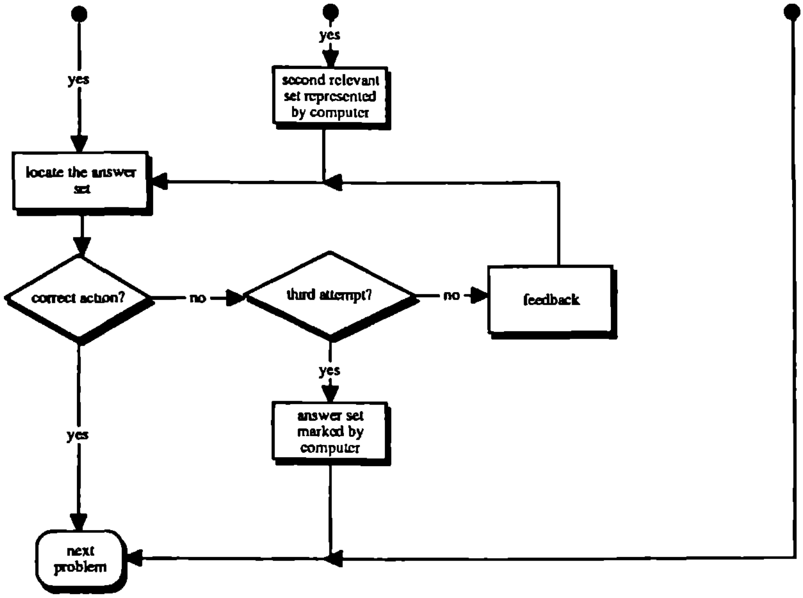
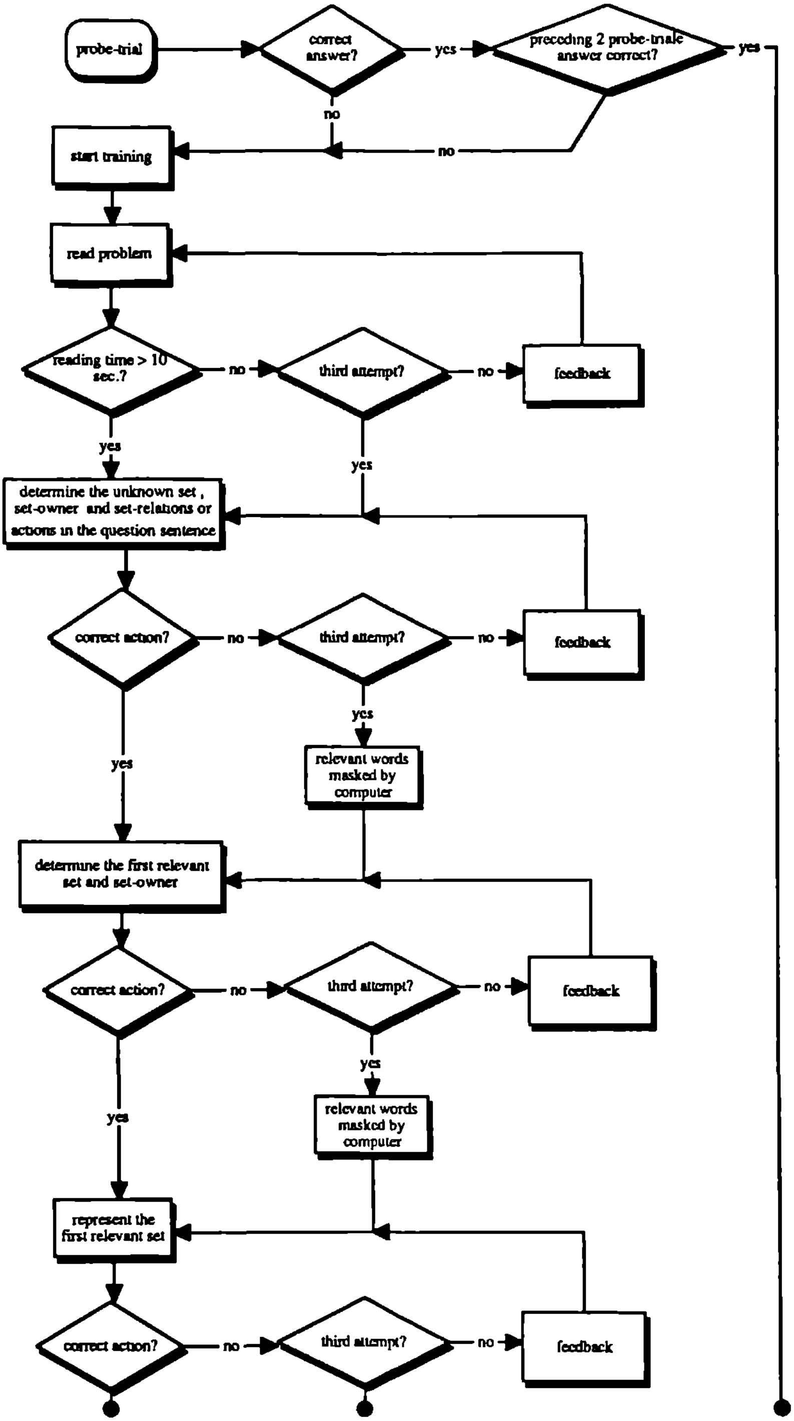


*Figure 2* The task strategy of the External-Modeling training.


probe-trial
correct answer?
yes
preceding 2 probe-trials answer correct?
yes
no
no
start training
read problem
reading time > 10 sec.?
no
third attempt?
no
feedback
yes
yes
determine the unknown set, set-owner and set-relations or actions in the question sentence
correct action?
no
third attempt?
no
feedback
yes
relevant words masked by computer
yes
determine the first relevant set and set-owner
correct action?
no
third attempt?
no
feedback
yes
relevant words masked by computer
yes
represent the first relevant set
correct action?
no
third attempt?
no
feedback

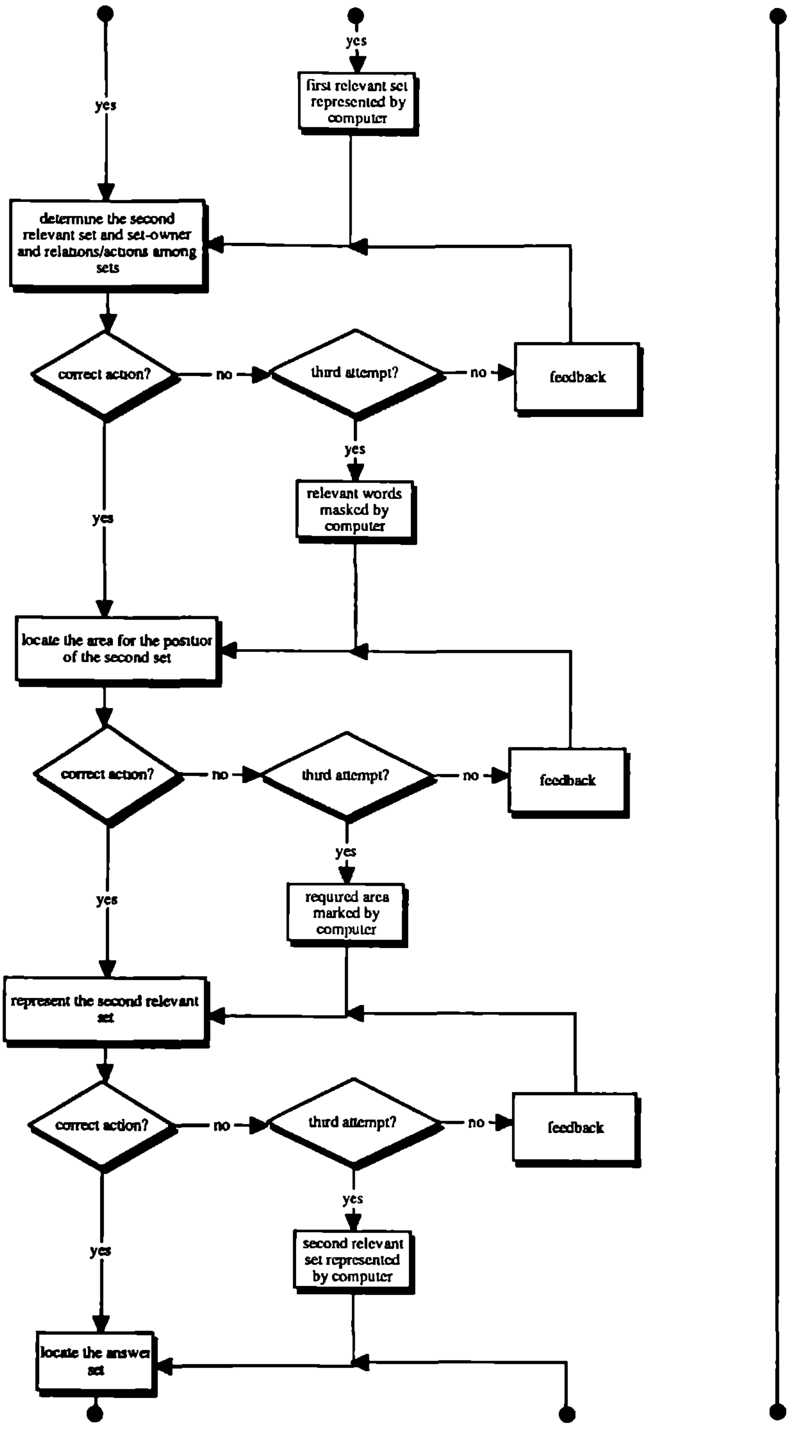


*Figure 3* The task strategy of the Text-Analysis/External-Modeling training.

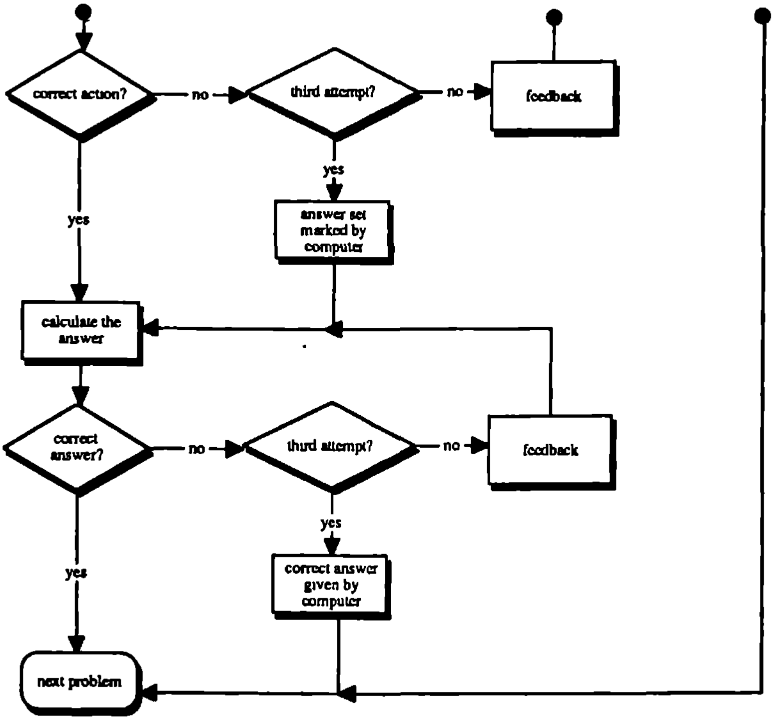


*Figure 3 (continued)*

## Appendix B

Examples of word-problem types used in the Untrained Word-Problem Test. The irrelevant sentences are italicized.

---

**sub-test 1**: reversed order of introduction of known and unknown quantities:

Marc has 2 books more than Peter. Peter has 6 books. How many books does Marc have?

**sub-test 2**: irrelevant information triggering mistakes concerning semantic problem type:

Peter has 4 pencils. *Peter has 3 pencils more than Ann.* Peter got some pencils. Now Peter has 9 pencils. How many pencils did Peter get?

**sub-test 3**: "complex" word problem type:

Together Peter and Ann have 9 dolls. Ann has 4 dolls. Peter got 3 dolls more. How many dolls does Peter have now?

**sub-test 4**: standard untrained type:

Peter had some marbles. Peter lost 4 marbles. Now Peter has 5 marbles. How many marbles did Peter have first?

**sub-test 5**: action-cued compare problem:

There are 9 children. There are 5 candies. How many children won't have a candy?

**sub-test 6**: syllogism:

Sue has got more candy than Peter.
Sue has got less candy than Ann.
Ann has got <u>more</u> / <u>less</u> candy than Peter.

---

## Appendix C

Examples of word-problem types used in the Reading-Behavior Test. The irrelevant sentences are italicized.

---

**Change 2 containing 2 sentences with irrelevant information (Ch22):**
*Peter had 9 books.* John had 6 books. *Peter won 1 book.* John lost 2 books. How many books does John have left ?

**A problem for which no calculation is required (NT3):**
*Ann has 6 books. Peter has 4 books.* John has 3 books. *Mary has 5 books.* How many books does John have ?

**A complex word problem for which two calculations are required (Ch4C1):**
Mary has 5 books. John has 2 books less than Mary. How many books do Mary and John have together ?

**An unsolvable Compare 1 problem with irrelevant information in the second sentence (Cp1-2):**
John has 3 books. *Peter has 2 books more than John.* Ann has 9 books. How many more books does John have than Ann?

**An unsolvable Compare 1 problem with irrelevant information in the third sentence(Cp1-3):**
John has 3 books. Ann has 9 books. *Peter has 2 books more than John.* How many more books does John have than Ann?

**An unsolvable Compare 2 problem with irrelevant information in the second sentence (Cp2-2):**
John has 3 books. *Peter has 2 books more than John.* Ann has 9 books. How many less books does Ann have than John?

**An unsolvable Compare 2 problem with irrelevant information in the third sentence (Cp2-3):**
John has 3 books. Ann has 9 books. *Peter has 2 books more than John.* How many less books does Ann have than John?

**A Combine 1 problem with irrelevant information in the second sentence (C1-2):**
John has 3 books. *Peter has 9 books.* Ann has 4 books. How many books do John and Ann have together?

**A Combine 1 problem with irrelevant information in the third sentence (C1-3):**
John has 3 books. Ann has 4 books. *Peter has 9 books.* How many books do John and Ann have together?

**A complex Combine problem with no irrelevant information (CC):**
John has 3 books. Ann has 4 books. Peter has 9 books. How many books do John, Ann and Peter have together?

**A Compare 1 problem (Cp1):**
John has 6 books. Peter has 2 books. How many more books does John have than Peter?

**An extended Compare 1 problem (Cp1-E):**
John has 6 books. John has more books than Peter. Peter has 2 books. How many more books does John have than Peter?

**A Compare 3 problem with irrelevant information in the second sentence (Cp3-2):**
John has 4 books. *Peter has 9 books more than John.* Ann has 3 books more than John. How many books does Ann have?

**A Compare 3 problem with irrelevant information in the second sentence and relevant information in the question sentence (Cp3-Q):**
John has 4 books. *Peter has 9 books more than John.* How many books does Ann have if Ann has 3 books more than John ?

## Appendix D

Examples of word-problem types used in the Free-Representation Test. The irrelevant sentences are italicized.

---

### Change 1 problems:

Peter gave 7 apples to Ann. Ann already had 2 apples. How many apples does Ann have now?

Peter gave 2 apples to Ann. Ann already had 6 apples. How many apples does Ann have now?

*Peter had 9 apples.* Peter gave 3 apples to Mary. Mary already had 5 apples. How many apples does Mary have now?

### Change 4 problems:

Peter had 9 apples. Peter lost some apples. Now Peter has 2 apples. How many apples did Peter lose?

*Together Ann and Peter have 9 apples.* Peter has 4 apples. Peter lost some apples. Now Peter has 3 apples. How many apples did Peter lose?

Peter had 8 apples. Peter lost some apples. Now Peter has 6 apples. How many apples did Peter lose?

### Combine 1 problems:

Peter has 3 apples. *Paul has 9 apples more than Peter.* Ann has 5 apples. How many apples do Peter and Ann have together?

Peter has 6 apples. Ann has 2 apples. How many apples do Peter and Ann have together?

Peter has 2 apples. Ann has 7 apples. How many apples do Peter and Ann have together?

### Compare 1 problem (unsolvable):

Peter has 5 apples. Ann has 3 apples. How many more apples does Ann have than Peter?

### Compare 3 problem:

Peter has 7 apples. Ann has 2 apples more than Peter. How many apples does Ann have?

Peter has 2 apples. Ann has 6 apples more than Peter. How many apples does Ann have?

### Compare 5 problem:

Peter has 6 apples. Peter has 2 apples more than Ann. How many apples does Ann have?

## Appendix E

The word-problem types used in the Word-Problem Classification Test.

| **Change** | | **problem-number** | **first-actor** |
|---|---|---|---|
| 1 | Peter had 1 marble. Peter won 3 marbles.<br>How many marbles does Peter have now? | (1) | "Peter" |
| 2 | Ann had 5 dolls. Ann lost 2 dolls.<br>How many dolls does Ann have left? | (2) | "Ann" |
| 3 | John had 3 cars. John won some cars.<br>Now John has 7 cars.<br>How many cars did John win? | (3) | "John" |
| 4 | Peter had 7 dolls. Peter lost some dolls.<br>Now Peter has 3 dolls.<br>How many dolls did Peter lose? | (4) | "Peter" |
| 5 | John had some marbles. John won 1 marble.<br>Now John has 3 marbles.<br>How many marbles did John have first? | (5) | "John" |
| 6 | Ann had some cars. Ann lost 2 cars.<br>Now Ann has 5 cars.<br>How many cars did Ann have first? | (6) | "Ann" |
| **Combine** | | | |
| 1 | Peter has 1 car. Ann has 3 cars.<br>How many cars do Peter and Ann have together? | (7) | "Peter" |
| 2a | John has 2 marbles. John and Peter have 5 marbles together.<br>How many marbles does Peter have? | (8) | "John" |
| 2a | Peter has 1 marble. Peter and Ann have 3 marbles together.<br>How many marbles does Ann have? | (9) | "Peter" |
| 2b | Ann and John have 7 dolls together. Ann has 3 dolls.<br>How many dolls does John have? | (10) | "Ann" |
| 2b | Peter and John have 7 cars together. Peter has 3 cars.<br>How many cars does John have? | (11) | "Peter" |
| **Compare** | | | |
| 1 | Ann has 3 dolls. Peter has 1 doll.<br>How many more dolls does Ann have than Peter? | (12) | "Ann" |
| 2 | Peter has 5 marbles. John has 2 marbles.<br>How many less marbles does John have than Peter? | (13) | "Peter" |

Table Appendix E *(continued)*

| | | | |
|---|---|---|---|
| 3 | John has 7 cars. Ann has 3 cars more than John.<br>How many cars does Ann have? | (14) | "John" |
| 4 | John has 3 marbles. Peter has 1 marble less than John.<br>How many marbles does Peter have? | (15) | "John" |
| 5 | Peter has 5 dolls. Peter has 2 dolls more than Ann.<br>How many dolls does Ann have? | (16) | "Peter" |
| 6 | Ann has 7 cars. Ann has 3 cars less than John.<br>How many cars does John have? | (17) | "Ann" |

## Calculation example of the algorithm used in determining the "association-score" for the main feature "first actor mentioned" for the Word-Problem Classification test.

The main feature "first actor mentioned" comprises three subfeatures; i.e. "Ann", "Peter" and "John". Take for example the case in which a child constructed two stacks consisting of the following problem numbers:

| stack 1 | stack 2 |
|---|---|
| 1 | 4 |
| 2 | 5 |
| 3 | 7 |
| 6 | 10 |
| 8 | 12 |
| 9 | 13 |
| 11 | 14 |
| 15 | 16 |
| 17 | |

To determine the "association-score" for the main feature "first actor mentioned", the following would be calculated: First, in case of perfect ordering the subfeature "Ann", would require a total of five word problems in one stack, e.g. the problems numbered 2, 6, 10, 12 and 17. The first stack merely contains the problems numbered 2, 6 and 17. Thus, the number of problems actually encountered is 3, which is divided by the required number of problems 5, 3/5=0.60. This procedure is repeated for "Peter". The actual number of problems encountered in stack one for subfeature "Peter" is 3 (the problems numbered 1, 9 and 11), whereas the required number in case of perfect ordering is 7 (the problems numbered 1, 4, 7 ,9, 11, 13 and 16). Thus, the score for subfeature "Peter" for the first stack is 3/7=0.43. Likewise, the score for "John" would become 3/5=0.60. These scores are first added, resulting in a total score of 1.63 for the first stack. But in

case of perfect ordering merely the problems belonging to one subfeature would be encountered in a stack. Since the first stack in this example comprises problem numbers of each of the subfeatures "Ann", "Peter" and "John", the total score for the first stack is corrected for the number of subfeatures actually found, i.e. 1.63/3=0.54. The procedure cited is repeated for the second stack. For the subfeature "Ann" the score for the second stack becomes 2 (problems numbered 10 and 12) divided by 5 (number of required problems), 2/5=0.40. For "Peter", the second-stack score becomes 4 (the problems numbered 4, 7, 13 and 16) divided by the required number, 7, 4/7=0.57. Finally, for "John" the second-stack score becomes 2 (the problems numbered 5 and 14) divided by the required number, 5, 2/5=0.40. The total score for the main feature "first actor mentioned" of the second stack becomes 0.40+0.57+0.40=1.37. Again, this total score is corrected for the actual number of subfeatures recorded in the stack, 1.37/3=0.46. The first and second stack scores are added, 0.54+0.46=1. When a child would had perfectly ordered the problems according to the main feature "first actor mentioned", three stacks would have occurred, i.e. a stack for problems starting with "Ann", a stack for "Peter" and a stack for "John". Hence, in order to correct for the number of stacks constructed, the total score is subsequently multiplied by
(the number of stacks constructed) / (number of stacks required) since this quotient is smaller than one, i.e. (1) (2/3)=0.67. Finally, since the number of subfeatures differed per main feature, this score was divided by the predefined number of subfeatures per main feature. In this case, 0.67/3=0.22. In case of perfect ordering, the final score would be 1.00. As such, this score indicates the degree of deviation from perfect ordering. This calculation procedure would be repeated for all main features.

## Appendix F

## Explorations

To examine whether the educable mentally retarded subjects in our study demonstrated errors similar to those of normally achieving children, the frequency of the errors committed by our subjects on each word-problem type of the pretest was compared with the error types reported by Verschaffel (1984) and Carpenter and Moser (1982). Since only five of the problem types used in our study were similar to those used by Verschaffel (1984) and Carpenter and Moser (1982), this post-hoc analysis merely concerned similar problem types, i.e. the problem types change 3, combine 1 and 2b, and compare 1 and 3. Although almost all children answered combine 1 correctly (88%), many errors were made on the other problem types. Likewise, Verschaffel (1984) and Carpenter and Moser (1982) reported that first graders generally experience no difficulties in solving combine 1 problem correctly. More specifically, 73% of the first graders of Verschaffel answered the combine 1 problem type correctly at the beginning of the school year. In contrast, combine 2b problems proved rather difficult to solve; 86% of our subjects came up with wrong answers to these problems. The first graders of Verschaffel likewise experienced difficulties with combine 2 problems; 70% of the first graders answered this problem incorrectly at the start of the school year. Although compared to the combine 2 problem used by Verschaffel (1984) the introduction of the superset and subsets was reversed in the combine 2 problem of our study, the most common error was answering with the larger number given (Verschaffel: 30%; our subjects:35%). Change 3 problems likewise were often wrongly answered by the subjects of Verschaffel (53%) as well as by our subjects (77%). Yet, the first graders of Verschaffel mostly demonstrated "the larger given number" errors (30%), whereas this percentage only was 17 for our subjects. Instead our subjects mostly committed "wrong operation" errors, i.e. our children added the two given numbers instead of subtracting the smaller number from the larger (46%). Compare 1 problems were correctly solved by 37% of Verschaffel's first graders, whereas only 18% of our subjects came up with the right answer on the pretest. Both the first graders of Verschaffel (47%) and our subjects (45%) most often came up with the "largest given number" for this problem type. Finally, 52% of our subjects and 76% of the subjects of Verschaffel answered compare 3 problems with "the second given number".

# Curriculum Vitae

Monique Jaspers, born on december 1st 1959 in Amsterdam, received her V.W.O. diploma, Atheneum B in 1978 at the St. Nicolaas Lyceum in Amsterdam. She entered the psychology program at the University of Amsterdam and graduated in 1984 in the Experimental Psychology; her supervisors were Prof. Dr. Frijda and Prof. Dr. Elshout.

From 1984 to 1985 she was employed as a research assistant at the Department of Methodology at the same university. From 1986-1990 she worked at the Department of Special Education of the University of Nijmegen as a junior-researcher on the subject of developing and evaluating computer-assisted instruction programs for children with learning problems. The results of this research are the subject of this thesis.

Presently she is employed as post-doc researcher at the Department of Special Education and engaged in research on designing Intelligent Tutorial Systems.

Stellingen
behorende bij het proefschrift

Prototypes of Computer-Assisted Instruction for Arithmetic Word-Problem Solving

*A training study on improving the ability of educable mentally retarded children to solve simple addition and subtraction word problems*

1. Naarmate een onderzoeker in leerexperimenten meer controleert voor het optreden van experimentele bias neemt de kans op aanname van de onderzoekshypothese af.
(Dit proefschrift)

2. Voor het oplossen van redactieopgaven door moeilijk lerende kinderen geldt dat niet in de eerste plaats het correct uitvoeren van de rekenkundige bewerkingen, maar primair kennis van het juist toepassen van deze bewerkingen een probleem vormt.
(Dit proefschrift)

3. Het feit dat oefening door het herhaald oplossen van redactieopgaven ook bij kinderen in het mlk-onderwijs tot een verhoogde redactierekenvaardigheid leidt, indiceert dat zij ten onrechte als moeilijk lerend worden getypeerd.
(Dit proefschrift)

4. De onderzoeker dient zich bewust te zijn dat er tussen computergestuurde en computerloze instructie meer verschillen kunnen zijn dan alleen de aanwezigheid van de computer.
(Dit proefschrift)

5. De externe validiteit van experimenteel onderzoek wordt beter gewaarborgd door het gebruik van de computer als instructie-medium.

6. In tegenstelling tot hetgeen Kirk en Gallagher (1979) beweren, blijkt dat moeilijk lerende kinderen nieuw aangeleerde oplossingsstrategieën kunnen toepassen op transfertaken, die grote overeenkomst vertonen met de trainingstaak.
(Kirk, S. A., & Gallagher, J. J. (1979). Educating exceptional children. (3rd ed.). Boston: Houghton Mifflin.)
(Dit proefschrift)

7. Ofschoon in theorie het aantal mogelijke Aptitude Treatment Interactions slechts beperkt wordt door de mate waarin leerlingkarakteristieken en instructiemethoden te genereren zijn (Fuchs & Fuchs, 1986), lijkt een inperking van het aantal leerlingkarakteristieken en instructiemethoden op basis van een zorgvuldige theoretische afweging een zinvolle onderzoeksbenadering.
(Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: a meta-analysis. Exceptional Children, 53, 199-208, page 199)

8. De toenemende druk op onderzoekers een voldoende aantal wetenschappelijke publicaties te produceren, gepaard aan het beleid van gerenommeerde tijdschriften overwegend positieve onderzoeksresultaten te publiceren, leidt tot een toename van het aantal schendingen van de wetenschappelijke ethiek.

9. De exponentiële toename van het aantal vrouwelijke promovendi bij de instelling van het AIO-stelsel is, gezien de lage financiële vergoeding, het slechte toekomstperspectief en de daarmee gepaard gaande lage status, goed verklaarbaar.

10. De 19e eeuwse stelling dat wetenschappelijk arbeid door vrouwen tot miskramen en onvruchtbaarheid zou leiden (Winckler, 1898) is gefalsificeerd doordat met de toename van het vrouwelijk wetenschappelijk personeel een drastisch tekort aan crèches is ontstaan.
(Winckler, C. (1898). De vrouw en de studie. Voordrachten en debatten gehouden in den vergaderingen van 3 maart en 10 november 1898 van de "Vereeniging ter behartiging van de belangen der vrouw te Rotterdam." Haarlem, De Erven F. Bohn, pag. 52-53.)

11. Sommigen menen dat het risico van besmetting voor het behandelend personeel door het bestaan van speciale afdelingen voor AIDS-patiënten in algemene ziekenhuizen beter beheersbaar is. Om dezelfde reden zou dan ook het opzetten van aparte afdelingen voor seropositief medisch personeel overwogen moeten worden.

12. Ofschoon de mate van investering het welslagen van zowel zakenrelaties in het bedrijfsleven als relaties tussen partners in de persoonlijke sfeer bepaalt, is in het bedrijfsleven de partij die het meest investeert de machthebbende, terwijl binnen privé-relaties de partij die het minst investeert de macht heeft.