TEACHING AND LEARNING THINKING SKILLS

EDITED BY
J.H.M. HAMERS,
J.E.H. VAN LUIT
AND
B. CSAPÓ

CONTEXTS OF LEARNING

SWETS & ZEITLINGER
PUBLISHERS

TEACHING AND LEARNING THINKING SKILLS

CONTEXTS OF LEARNING
Classrooms, Schools and Society

TEACHING AND LEARNING

THINKING SKILLS

EDITED BY

J.H.M. HAMERS, J.E.H. VAN LUIT
Utrecht University, Department of Education

AND

B. CSAPÓ
Szeged University, Department of Education

SWETS & ZEITLINGER
PUBLISHERS

LISSE ABINGDON EXTON (PA) TOKYO

Library of Congress Cataloging-in-Publication Data

Teaching and learning thinhing skills / edited by J.H.M. Hamers,
J.E.H. van Luit, B. Csapó.
p. cm. -- (Contexts of learning)
Includes bibliographical references and index.
ISBN 9026515456 (hard)
1. Thought and thinking--Study and teaching. 2. Learning, Psychology of. 3. Mathematics--Study and teaching (Elementary) I. Hamers, J. H. M. (Jo H. M.), 1945- . II. Luit, J. E. H. van, 1953- . III. Caspó, Benő. IV. Series.
LB1590.3. T36 1996
370. 15'2--dc21

Printed in The Netherlands by Krips, Meppel

ISBN 90 265 1545 6

Preface

How do people solve complex problems whether in mathematics, in reading comprehension, in text production or in everyday life? Or in brief: how do we think? And how can skills and abilities of these processes be taught, developed or their development be stimulated? For decades, these have been the most intensively researched questions in educational psychology. Despite its long history and the large number of related publications, this field of study is still vital; new psychological theories and results of laboratory studies are searching for their ways to the application in educational practice.

The main goal of this volume is to provide an up-to-date, grounded review of the psychological and educational literature on thinking, problem-solving and the teaching of thinking and to introduce some approaches and concrete projects of the most recent waves of research and development. By carefully selecting papers, which represent different types of studies, an attempt is made to illustrate the vital diversity of the field. Thus, the chapters of this book cover most of the main topics that traditionally form the foundation of the field, but they do not cover, of course, every topic that might conceivably be included. Furthermore, theoretical frames, psychometric, Piagetian, neo-Piagetian, information processing, constructivist and socio-historical approaches are discussed in several chapters. However, these theories are represented with different weights in educational applications. Specific implementations of the main approaches to teaching thinking are presented and examples for development of a large array of thinking skills can be found.

Nearly each of the chapters involves theoretical considerations and presents research findings, although the proportion of these components varies. Those chapters that provide an overview of broader areas and place more emphasis on reviews are balanced with others that present meta-analytic synthesis, results of decade-long research programmes or discuss specific projects. A separate chapter deals with methodological issues in more detail.

The text is unique in its crisscrossing of the major approaches to development and education. Several chapters deal with more than one theoretical framework and attempt to synthesize different approaches. Another unique feature of the book is that it links European and North-American research: literature from both continents are extensively reviewed and European approaches are compared with mainstream works from the United States. The authors represent a number of countries and the cross-fertilizing effects of ideas born in different places and of projects carried out in different cultures and educational contexts can also be observed.

The book can be read as an introduction to the study of thinking and problem solving in several areas. Since both the approaches and the chapters themselves are diverse, we have written and introduction (Chapter 1) and an epilogue (Chapter 14) which may provide help to tie the chapters together.

The Editors

J.H.M. Hamers, Department of Education, Utrecht University, the Netherlands
J.E.H. Van Luit, Department of Education, Utrecht University, the Netherlands
B. Csapó, Department of Education, Szeged University, Hungary

Contents

1

Teaching thinking

J.H.M. Hamers & B. Csapó

Introduction

In the history of education there has never been so much interest in the teaching of thinking and problem solving as there is today. This interest is not new. There have always been educators who see the cultivation of the thinking ability of students as important objectives. In particular, mathematics (Nelissen; Van Luit; Verschaffel, this volume) and science courses (Adey; Csapó, this volume) have traditionally paid much attention to the teaching of problem solving in these specific domains. Dewey (1910/1991) was among the first to focus attention on the enhancement of thinking ability. Largely due to his influence, considerable energy was and is today devoted to the development of thinking.

Other important impulses in the past came from the Würzburger School (Külpe, Selz) and Gestalt psychology (Wertheimer, Duncker, Maier). In 1930 Kohnstamm wrote about Selz: "Originally he was only interested in discovering the problem solving methods for himself but in the course of time he started to inquire into the learnability of these methods" (p. 28). Selz (1935) conducted research into inductive reasoning amongst other topics. He assumed, as a starting point, that the thinking process consists of applying means of ordering, thinking schemes, which determine the course of the thinking process by their character of anticipating the solution. According to Selz, who assumed this to hold, it must be possible to raise the functioning of human intelligence to a higher level by providing the necessary thinking tools. He thought it must be possible to learn problem solving methods. In his experiments 'Versuche zur Hebung des Intelligenzniveaus' he worked according to the principle of the 'kleinstmögliche Hilfe', now known as the heuristic solution method.

According to Resnick (1987), today's need for teaching thinking is created by the rapid changes taking place in society. Knowledge and information are becoming ever more complex and soon may become dated. Children, therefore, have to be equipped with the skills of evaluating choices, and identifying and solving problems using logical reasoning. Thus, it is not enough to have a considerable amount of knowledge at one's disposal (declarative knowledge), but the questions of how to acquire knowledge, and how to apply this knowledge are also important (procedural knowledge). It is also claimed (e.g., Halpern, 1992; Resnick & Klopfer, 1989) that having only a limited command of thinking skills is one of the reasons for falling behind in school. This can be seen in mathematics, reading, and writing, where all sorts of activities come to the fore in which thinking skills play a central role. Examples are the ability to describe and to compare objects, to group objects, to associate one thing with another, to form concepts, and to generalize. Thus, mental processes which are normally associated with the concept of 'thinking' are not limited to some kind of 'higher order' of mental development. On the contrary, thinking processes play a role in a broad range of learning activities in school. This means that these thinking processes should form an integral part of the school curriculum.

This issue of adjustment or innovation of educational aims is an actual topic of study in education. However, the realization of these aims is not simple. The reason is that too many questions are still not, or insufficiently, answered. For instance: What is thinking? Are we able to teach children to think? Which thinking skills can be assessed? Thinking is partly the result of an autonomous process in the development of children. The question that arises from this fact is: What is left to be taught? Which part of the behavioural changes in children can be attributed to spontaneous 'development' and which part to 'learning'? Generally, in education no attention is paid to the explicit stimulation of thinking skills and there is no such thing as a school subject called 'thinking' or 'stimulation of thinking' (Presseisen, 1987). Usually teachers assume that thinking skills develop spontaneously as a by-product of the teaching of regular school subjects. Nowadays, the current view is that this assumption is only partly true (Resnick, 1987). Deprived children, and children with learning difficulties, can benefit from explicit stimulation of thinking, and children who do not belong to such groups can also learn to think more efficiently.

In the United States there have been many proposals to explicitly stimulate thinking and a variety of general and specific programmes.have been developed and described (e.g., Baron & Sternberg, 1987; Costa, 1991a; Idol & Jones, 1991; Jones & Idol, 1990). Variations of general and specific programmes have been designed and collected (e.g., Chipman, Segal, & Glaser, 1985; Costa, 1991b; Nickerson, Perkins, & Smith, 1985). In Europe much attention is also paid to this subject (e.g., Adey & Shayer, 1994; Demetriou, Shayer, & Efklides, 1992). Coles and Robinson (1991), Fisher (1990), McGuinness and Nisbet (1991) and Nisbet and Davies (1990) have published reviews, mainly of British programmes. Hamers and Overtoom (1997) have published an inventory of programmes as well, and extended the field of research by including a greater part of Europe in their inventory.

These programmes and methods appear to be of diverging theoretical orientation: Vygotskyan, neo-Piagetian and along the direction of information processing. Furthermore, the range of themes is wide: programmes for training general reasoning skills, critical thinking, problem solving, memory, comprehensive reading, composition, arithmetic, and secondary school subjects such as science. Terms or concepts are used in a variety of ways, sometimes interchangeably, sometimes synonymously: e.g., thinking, problem solving, reasoning, decision making. Other terms such as analysing, imaging, inferring, inventing, and reflecting are used with greater specificity. It is not our aim to precisely define any of these terms precisely here. In the chapters of this volume, thinking (the most general term) is broadly conceived and it includes much of what is discussed under the other, more specific terms. The chapters in this volume will show that not everyone focuses on the same aspects of the multifaceted activities of thinking.

In this chapter, we will describe some theoretical and practical trends in the research on the stimulation of thinking and pay attention to the impact of these theories on educational research and practice, as demonstrated in several chapters in this volume. In addition, we will describe some of the main issues in the research on the development of thinking and teaching thinking. Finally, we will draw conclusions and give an introduction to the following chapters.

Theories on thinking

Thinking is a broad and relatively abstract concept that is discussed and defined in many variations. Several disciplines consider it as a central concept, and a number of research paradigms examine it using a broad range of approaches and applying a variety of research methods. The list of adjectives used in conjunction with thinking (e.g., from convergent thinking to critical thinking) is virtually endless, and - from the works of Greek philosophers to today's psychologists' publications - many attempts have been made to classify types of thinking or at least to enumerate the relevant forms of thinking (e.g., Chipman et al., 1985; Segal, Chipman, & Glaser, 1985) or to just create a working taxonomy of thinking skills (e.g., Ennis, 1987). However, there are two main fields of research into human thinking that have especially important implications concerning the subject of the present book: (a) the perspective of developmental psychology that identifies types and forms of thinking and describes how the states of these forms are changing over the individuals' life span in a qualitative or quantitative sense; and (b) the educational, cognitive and learning-psychological perspective that deal with the problem of how the development of thinking can best be stimulated by organizing the most influential learning environments for the developing individuals. Both perspectives have lead to the rise of theories emphasizing development (Piaget) or development and learning (Vygotsky, Bruner and neo-Piagetian theories). Those learning theories that aimed at explaining the cognitive processes which take place between the input and output of information, deeply influenced by the conceptual framework of computer science, have evolved towards the information processing approach. The processes involved in perceiving, storing, memorizing and applying information are being studied. Figure 1 shows a global division of the most important theories. In this section we will briefly introduce these theories and some others.

One of the most well-known theories of cognitive development, Piaget's theory, is rooted in the rationalist tradition and has several traits of constructivism as well. According to this theory, the development of thinking in children progresses according to successive, discrete stadia. Thinking in a certain stadium is qualitatively different from the thinking in the previous or the next stadium. Piaget sees development as the emergence of new structures of knowledge or schemas, and as the transformation and refinement of these schemas. The result is equilibration, the attainment of balance between the schemas and the environment. Piaget's classification into stadia of development is based on this principle. The four stadia he distinguished (the senso-motor, the pre-operational, the concrete-operational and the formal-operational) are always passed through in the same order and they are considered to be universal.

In the neo-Piagetian option, the issue of universality of developmental stadia is dropped. Partly as a consequence of learning theory, the possibility of stimulating thinking is being studied, as well as breaking through the stadia and establishing larger individual differences in cognitive schemas. Case (1985) integrated Piaget's theory with information processing theory. By learning or training, children will

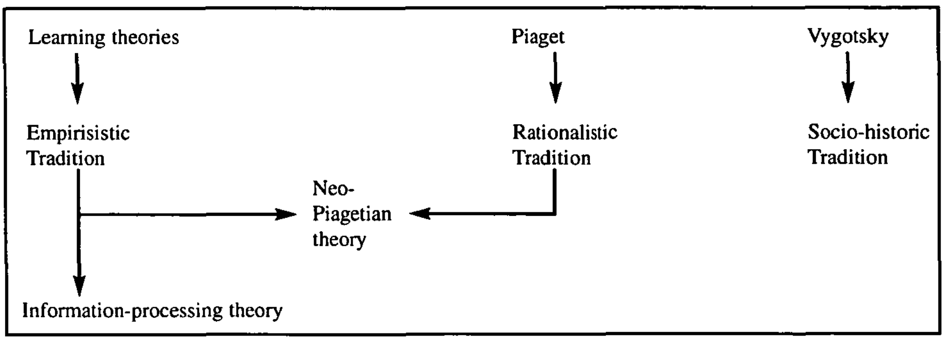


Figure 1 Main theories on thinking

become more skilled in the processing of information. This means that they are increasingly able, and sooner than Piaget assumed, to perform cognitive operations. In this way, the learning environment plays a more balanced part in development. It makes children's development more heterogeneous.

Bruner, Goodnow, and Austin (1956) conducted some of the first research on learning viewed as a product of thinking. What made their work 'A Study of Thinking' distinctive was that it studied learning as an outcome of thinking, using methods that made thinking amenable to objective study. The kind of learning they studied was 'concept attainment'. One can think of concept acquisition as analogous to learning a set of rules for classifying objects, e.g., 'If x has warm blood and fur, then it is a mammal'. This is an important kind of learning because classifying objects is essential to thinking. Bruner and his colleagues thought of the processes of learning as similar to hypothesis testing. A learner uses some strategy to generate possible rules for defining a concept and then tests these hypotheses against actual instances until one is found to survive the tests. In this way, Bruner et al. (1956) found strategies which where well-adapted to the practical difficulties of the situation, such as its complexity. Thus, learning involved thinking (generating and testing hypotheses on the basis of their implications) and it depended on the particular strategy being used (rather than being determined by the environment alone). Bruner's (1990) work is widely accepted, in part, because it made the study of thinking 'tough-minded'.

Vygotsky's socio-historic theory is primarily a learning theory and, from this point of view, applies to learning to think (Case, 1996). Vygotsky values the interaction between parents and children throughout their development. A central concept in his theory is the 'Zone of Proximal Development' (ZPD). This zone refers to the difference in what a child can accomplish on its own and what it might be able to perform with the help of competent others. With such help the child can reach a higher level of development. Language plays an important role in this process. A child's first words are communicative actions that shape its interactions with others. According to Vygotsky, during the first two years of life the development of language and thinking occur along more or less parallel, but relatively separate lines. Around the age of two, a fundamental change takes place in the child's relation between language and thinking. Thinking becomes verbal. Language originates by way of thinking, but subsequently language fosters the further development of thinking.

Characteristic learning processes that occur between the input and output of information processing are determined in learning theory, and on the same lines, in information processing theory. Research is carried out on cognitive processes involved in the perception, storing, memorizing and application of information. Duijker (1977) described the concept of thinking as follows: "Thinking denotes for psychology a coherent complex of specific theoretical problems, dealing with the complexity of the information processing activities (what do they consist of and how are they controlled?) and with the roles these representations of information play (how are they established, what is their nature and structure?). Cognitive psychology mainly occupies itself with problem solving, which means that activity of information processing in which the subject tries to find an answer to a question that is difficult for him" (p. 89). The central concepts in this quotation are: information processing, representations and problem solving. In today's cognitive psychology these concepts are crucial. In information processing theory much attention is paid to incorrect and inefficient thinking. These evaluations of the thinking processes are considered essential, and are useful in improving thinking. In particular, differences in the use of control mechanisms or metacognition and in the speed of information processing account for differences in the development of information processing activities. One of the central questions is how metacognition can be guided or influenced (Boekaerts & Simons, 1993).

Constructivism is another theoretical orientation that nowadays receives growing attention (see Philips, 1997). It has many faces and several sub-branches or sub-theories, so today it can be considered a broad paradigm rather than a specific or consistent theory. Both 'realistic' and 'radical' constructivists are active on a number of fields of instruction, especially in reforming mathematics education (Cobb, 1996). The 'new math' movement traces its origin back to the Piagetian framework while the more recent realistic mathematics education (see also Nelissen, this volume) is influenced by the information processing approach and Vygotsky's social constructivism as well. Although constructivism is not a sophisticated theory yet, it may be a good candidate for integrating several aspects of some competing theories of development (Piaget, Vygotsky and information processing). Constructivists emphasize more strongly that learning is an active, constructive process. Learning is productive, useful, achieves results only if the students are actively involved in the subject matter. The art of learning is to connect new information to existing knowledge. This active connecting process consists of involving all kinds of prior knowledge in the construction of new representations of information. Because each person has individual experiences and different foreknowledge to build on, these new representations are unique. Possibilities are sought to facilitate the active, constructive learning in 'rich learning environments' by involving modern technologies, for instance.

Another important theory is that of psychometry. The psychometric tradition has had a long-term impact on the research of thinking in at least two significant ways: (a) by making psychological traits measurable and in doing so opening the way for quantitative analyses (e.g., factor analysis) and (b) by launching the concept of intelligence (see Carroll, 1993). Galton (1822-1911) is considered to be the founder of this theory of the individual differences in human faculties, while Binet was the first to develop an intelligence test. Binet was already well known for his work on the development of thinking in children who achieved very different levels of attainment. He took it upon himself to develop a tool which could be used to select those children who need special educational help. Binet and Simon's test has been translated and adapted into many languages. In the US this led to the Stanford-Binet

Intelligence Scale, which in its current version is one of the most widely used IQ tests. Later, other influential tests such as the Wechsler Scales followed.

Following on from Binet's work, and from the results of factor analysis, intelligence was described in terms of a collection of latent variables called factors. Individual differences in test achievements, the manifest behaviour, were derived from individual differences in these factors. Typical intelligence theories which can be considered as belonging to this perspective are Spearman's two-factor theories, Thurstone's primary mental abilities theory, Guilford's structure of intellect theory and Cattell-Vernon's hierarchical theory. These factor analytical traditions differ from each other in the number of skills mastered and their inter-relationships. The ones most known today are a few modern variants of intelligence structures (Gustaffson, 1984; Snow, Kyllonen, & Marshalek, 1984). Efklides (this volume) based her research on Gustaffson's intelligence structure.

Binet and Wechsler were both interested in the development of a theory about intelligence. Both had an opinion about intelligence that was instrumental in the choice of tasks and for the cognitive processes which underlie tasks and the development of these processes (Sternberg, 1985). The designers of the factor theories thought they should get a better grip on the cognitive process of the operation which enables an individual to reach an answer during the solving of a problem (Kail & Pellegrino, 1985). This is the reason that in intelligence research a cognitive-psychological approach appears in which analyses of processes have to offer an explanation for the solving of a problem, whether successful or not. Thus the Russian researchers, such as Gal'perin and Zaporozec (in Van Parreren & Van Loon-Vervoorn, 1975) conducted studies of Thurstone's factors 'numeric' and 'space', respectively, from an active learning perspective. De Groot (1965) partly laid the basis for the studies of the novice-expert perspective. These studies investigated how novices differ from experts in their solving of complex problems, for example, in chess problems. Schoenfeld (1985) adapted this idea to educational problems occurring in physics and mathematics. The research used introspection and protocol analysis to gain access to human cognitive processes (see also Elshout, 1988).

Although intelligence and its measurement proved to be a useful concept for several practical purposes, improving its identifiable components (e.g., inductive reasoning, see several chapters in this volume) seems to be a more fruitful and realistic enterprise for education. Sternberg (1985) was one of those theorists who renewed intelligence research by applying the framework of the information processing approach. Sternberg (1985) attempted to answer three questions. The first concerned which elementary cognitive operations (components) were involved in solving a certain type of problem. The second was about the amount of time an individual needed to solve a problem and how accurately the individual worked. The third question concerned the inter-individual differences in the speed and accuracy of the processes. Sternberg (1985) distinguished three types of thinking skills in his component subtheory: (a) executive processes which are used to plan, monitor and evaluate one's own thinking (meta-cognition); (b) performance processes which are used to actually carry out that thinking; and (c) learning processes which are used to learn how to think in the first place. Examples of executive processes include identifying and formulating a question, keeping the situation in mind and organizing one's thoughts. Examples of performance processes include seeing similarities and differences, deducing, and making value judgements. Asking and answering questions of clarification such as 'What do you mean by that?', and listening carefully to other people's ideas, are examples of learning processes. De Koning and Hamers (this volume) give examples of Sternberg's performance processes applied to inductive reasoning (encoding, inference, mapping, application, comparison, response).

The learning and thinking theories discussed here emphasize the development of thinking (Piaget), the development of thinking and education (Bruner; Vygotsky), learning (information processing, constructivism) and measuring individual differences in mental capacities (psychometry). It is primarily the opinions of Piaget, Vygotsky and Bruner which are often compared to each other. In general, we can say that they agreed on the sequence in which thinking developed: from concrete actions via increasing reflection to abstraction. In fact, Vygotsky and Bruner assume much more strongly than Piaget that education has an essential function with respect to a child's development of thinking. The cognitive approach to thinking and intelligence is significant because it can offer an explanation for the way in which achievement in all sorts of tasks is accomplished.

Many curricula for the teaching of thinking are influenced by one or more of these approaches (see Hamers & Overtoom, 1997; Nickerson et al., 1985). The latter proposed five approaches:

(a) In the cognitive operations approach, it is assumed that thinking problems are caused by an insufficient mastering of basic operations like classification and seriation. The training programmes in this approach could be particularly suitable for weaker students who have not yet mastered these operations (emphasis on neo-Piagetian and information processing theory).

(b) In the heuristic approach all kinds of problem solving operations are taught, like problem analyses, planning, representation and verification. The essence of this approach is the task analysis, in which a task is split up into manageable parts or subtasks. After the analysis, attempts are made to improve a person's performance in the subtasks by training in the problem solving strategies and by involving metacognitive skills (emphasis on Vygotskyan and information processing theory).

(c) In the formal thinking approach, the starting point is the neo-Piagetian or Piagetian theory. The programmes aim at effecting the transitions between the different stadia, for example, between the concrete-operational and the formal-operational stadia. A characteristic of this approach is the integration of thinking operations into school subjects like science (emphasis on neo-Piagetian theory).

(d) In thinking as manipulation with language and other symbols, teachers stimulate the use of thinking skills by means of the regular school subjects (emphasis on Vygotskyan theory).

(e) In thinking about thinking (metacognition), it is assumed that a better understanding of the nature of one's own thinking process will improve one's competence in thinking. Students are encouraged to think about thinking in general, and to become more aware of their own thinking processes (emphasis on Vygotskian and information processing theory).

There is some overlap between these five approaches. For instance, in the last approach, heuristics are being used, and thinking through the curriculum content includes various elements of the other approaches. Besides, none of these approaches is superior to any other. All aspects of these approaches are also discussed in chapters in this volume. In 'teaching thinking' the various theories mentioned are brought together; they are seen as compatible and complementary (Sternberg & Berg, 1992). They all contribute in their own specific way to understanding and optimizing learning conditions for the teaching of thinking.

Teaching thinking

Most educators agree on at least one general point, namely that a central aim of education is to take the knowledge that has been acquired by one generation and to create conditions such that this knowledge can be acquired and extended by the next generation. This point does not require any particular view on educational aims and methods. One reason for this diversity is that there is no agreement as to the nature of knowledge itself since there are several different theories of knowledge (Case, 1996). These views have their roots in British empiricism (Watson, Thorndike, Hull), in continental rationalism (Piaget), and in the socio-historic theory (Vygotsky) (see Figure 1). The theories of knowledge deal with the issue of the relationship between man and knowledge. For a long time, philosophical schools such as empiricism and rationalism, for example, formed the foundation for two distinct paradigms in which psychologists formulated their research on the relationship between man and knowledge. This originally led to contrasting views on how people use knowledge, how they reason, and how education can best take advantage of this. De Koning and Hamers (this volume) describe how so-called pragmatic deductive- and inductive reasoning schemes have brought about a synthesis between both paradigms. These context-free schemes ensure the orderly processing and application of knowledge and the organization, reorganization and storing of knowledge.

Views of knowledge (Case, 1996) and knowledge acquisition or learning are inextricably linked. Learning comprises many processes which a person can work through. On the one hand, these learning processes have several common features.

For example, each process will provoke a number of changes in people which are relatively enduring. On the other hand, all types of learning show a great variety of processes. Examples of types of learning include: learning that aims at insight (learning to think), learning facts, memorizing, and learning automatisms. It is the differences which make it difficult to consider or describe learning as an unambiguous concept.

Thinking occurs in a situation when a person is presented with a problem, i.e., a task for which there is no immediate solution. In the most favorable case, the person will allow himself to assess the problem, look at the different aspects, and find a suitable solution by way of insight. The psychology of thinking concerns the issue of how someone acquires that insight. In school we encounter many forms of learning that encourage insight, as in comprehensive reading, arithmetic, and text production in which thought relations must be adapted (e.g., agreement - difference; cause - effect). The essence of this sort of learning processes is primarily the learning of accurate concepts and general rules with which children can tackle new tasks (Van Parreren, 1990), for example, in order to solve a problem involving areas children must be familiar with the concepts of area, right angle, length, width and circumference. The formula for solving this problem provides a rule linking these concepts.

As well as acquiring concepts and general rules, learning which encourages insight also requires mastering the actual methods for solving a problem (Van Parreren, 1990). Research into methods or processes for solving problems has always been an important issue in thinking psychology (Piaget, Wertheimer, Duncker, De Groot, Bruner; for overviews see Dumont, 1966; Frijda & Elshout, 1976). A standard Anglo-Saxon work in this field was written by Newell, Shaw, and Simon (1958). But East European researchers, such as Kuljutkin, Ponomarev and Puskin, also studied solution methods (see Van Parreren & Van Loon- Vervoorn, 1975).

Puskin (1975) distinguished three phases in solution processes. The first phase leads awareness of the problem as such. If all attempts at applying known concepts and rules fail to solve the problem then one is compelled to look for new solution methods. The second phase is closely connected to looking for new solution methods. The actual solution process takes place in this phase. The person must find those operations which, when applied, lead to the desired target situation. This process can be considered the application of transformation methods. In the third phase there is a check on these and the results are integrated into the personal motives structure. In general, three such methods can be distinguished:

(a) The algorithm method: the problem is transformed according to fixed, always valid steps until the solution is reached. The solution is guaranteed. Strictly speaking, this is not about problem solving.

(b) Blind exploration (trial and error): different possible solutions are tried out, without using information about the possibilities themselves (Frijda & Elshout, 1976). This can also be called trials (Podd'jakov, 1979) or guess exploration (De Corte & Verschaffel, 1980).

(c) The heuristic method: in this method sensible ways of operating are selected without having to search every possibility. This method only provides a greater chance of finding a solution, but no guarantee of finding it. In addition, various heuristic principles can be distinguished (Duncker, 1935; Frijda & Elshout, 1976), such as target-means analyses, material and conflict analyses, and splitting the main problem into subproblems.

Learning in all its forms can also be self-taught, one can 'learn to learn' (Van Parreren, 1990). An important starting point here is that learning progresses as the students become more aware of their own activity: learning as a conscious activity or metacognition. In that framework it is important that students orient themselves well to the task, make a plan of how to tackle the task so that they work systematically, analyse mistakes and analyse successful attempts retrospectively, i.e., reflect on the mental steps taken in order to achieve integration of the knowledge. Thus, a goal that the developers of mental stimulation methodologies aim to achieve is that children start thinking more efficiently, partly by acquiring insight into their own mental processes and partly by actively directing these processes. An important means for achieving this aim is reflection, which means stimulating thinking about one's own thoughts or metacognition (Boekaerts & Simons, 1993): "People have knowledge to a greater or lesser degree about their own cognitive system and how it works. The knowledge can concern their own thinking, memory, fantasy, reasoning, etc. and that of other people ..." (pp. 88-89). It is generally accepted that people who have a relatively greater metacognitive knowledge are better able to direct and improve their thinking.

In the process of designing programmes for teaching thinking, the choice of tasks is of great importance. There are many kinds of tasks: some demand primarily motor activity, others demand more mental or thinking activity, as in analogies (e.g., client = doctor : ....), completing series (e.g., 2, 6, 11, 17 ...) and classification (e.g., What does not belong here?: cat, dog, elephant, guinea pig). Cognitive psychologists have tried to describe and analyse the characteristic difficulties and processes of these and other thinking tasks. The most well-known classification of tasks or problem types is that designed by Guilford (1956). He constructed a division of tasks from three starting points: (a) the contents of the thinking task or the nature of the material that has to be worked with; (b) the actions or operations that have to be performed; (c) the result or product of the actions. In his option, a problem of analogy like leg : knee = arm : ... could be characterized as convergent thinking (operation), as semantic (concerning the contents) and as relation (product). Particularly in the field of inductive reasoning many new tasks have been added and investigated (Jacobs & Vandeventer, 1972). Jacobs and Vandeventer taught subjects to solve so-called double classification tasks. These tasks consisted of a 2 x 2 or a 3 x 3 matrix, which presented figures that varied horizontally and vertically and in which the figure in the lower right corner was omitted. The tasks belong to what Guilford (1956) calls the 'cognition of figural relations' in his Structure of Intellect (SI) model. Guilford's

SI model recently has been strongly criticized by several psychometricians (see Carroll, 1993). During the last few decades many new tasks have been developed and examined (e.g., Csapó, this volume; Sternberg, 1985; Vosniadou & Ortony, 1989).

Some relevant issues in teaching thinking

As already stated, thinking is approached theoretically in different ways. Theoretical starting points have implications for the construction of a programme. A logical conclusion is that there is no such thing as one kind of stimulation of thinking. In this section we will describe some widely discussed themes with respect to the teaching of thinking.

Skills versus infusion approach

There is discussion on whether thinking skills apply to all domains of the school curriculum (domain-general) or whether they are specific to the school subjects (domain-specific). And, if the thinking skills are general, are they best taught via discrete programmes or should they be fused into subject areas? This discussion has resulted in two approaches (Maclure & Davies, 1991): (a) the general approach with separate courses for teaching thinking; and (b) the specific approach with integrated courses, which means that the thinking skills are embedded in the school subjects.

In the first approach the basic assumption is that thinking skills can be taught explicitly and independently of the regular school curriculum (the 'skills' or 'across-the-curriculum' approach). In this view there are certain more or less universal thinking skills that can be generalized in the school subjects. A prerequisite for the occurrence of a positive effect on, for instance, reading, writing and arithmetic is that, during the training, a 'bridge' is built between both. These general thinking skills are mostly trained with 'content-poor' tasks (see Figure 2). The question in this case is: In which of the four squares would the figure on the right fit best?

The second approach assumes that thinking skills can best be taught embedded in the school subjects (the 'infusion' or 'within-the-curriculum' approach). Thinking skills are being taught in specific or 'content-rich' domains like reading, writing and science. The following text is an example of this approach.

The zoo

The teacher visited the zoo with the third grade pupils. The children were very pleased. The crocodiles attracted most attention in this wonderful zoo. How big they were! The elephants were funny. They sprayed each other with their long trunks. And those beautiful birds in all kinds of colours ..., etc.

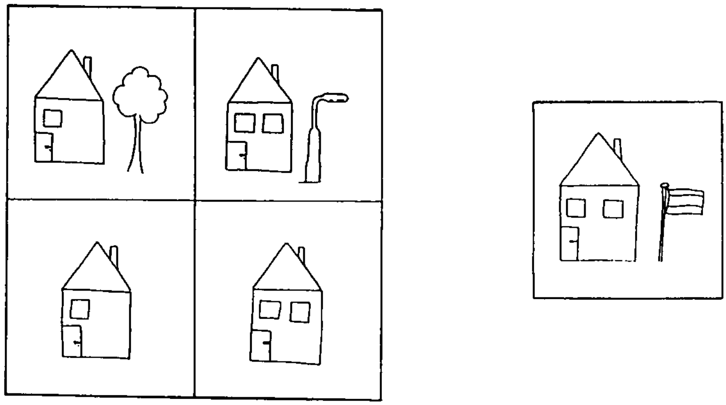


Figure 2 Example of a 'content-poor' item (test item from De Koning, Hamers, & Sijtsma, 1996)

At the end of this reading lesson the teacher will categorize the animals from the story into, for instance, land animals, winged animals and water animals. This text involves the same processes as those required for the item shown in Figure 2.

This integrated mode of operation requires fundamental changes in content and presentation of the subject matter (Csapó, 1990, this volume). Examples of this approach can be found in arithmetic (Nelissen; Verschaffel, this volume), in comprehensive reading (De Koning & Hamers; Oostendorp & Elshout-Mohr, this volume) and in text composition (Chanquoy, this volume). For example, De Koning and Hamers' programme (1995) uses texts which demand thinking operations like classification (grouping on the basis of attributes of objects in the text) and seriation (the formation of a logical sequence on the basis of, for instance, cause-effect relations in the text).

Advocates of the second approach hold the view that teaching thinking programmes should be embedded in the school subjects because a great number of these skills are content-specific and are not easily transferred to other content domains. We are of the opinion that content-relative programmes are preferable to general programmes, unless the students have great difficulty with the domain contents.

Transfer

Transfer of thinking processes can be considered an end-target of learning. The belief in the possibility of transfer can be traced to the institution of school itself and in the assumption that what one learns at school can also be useful in life outside school. Different types of transfer can be distinguished (Salomon & Perkins, 1989; Simons & Verschaffel, 1992). One of the most well-known distinctions is between the so-called near and far transfer. Near transfer is when a skill is learned in a certain context and applied in an almost identical situation. The features of the transfer task are largely similar to the original task, e.g., using the alphabet for looking up a word in a dictionary. Far transfer concerns applying a skill in a completely different context.

One generalization about transfer is that it requires explicit focus in any approach to the teaching of thinking. Whether thinking skills or problem solving strategies are taught in the context of subject areas or in separate courses, students need to be aware of the applicability of what they are learning to contexts other than that in which they are learning it. One problem is that most of the research on thinking has been done in classrooms and psychological laboratories and relatively little is know about the transfer effects to performances in everyday situations. So how reliable are the predictions of demanding tasks used in an academic setting to demanding tasks encountered in everyday life? In this volume the chapters on so-called realistic mathematics (Nelissen; Verschaffel) describe research focused on the kinds of thinking and problem solving taught in schools and outside academic contexts.

Process paradigm

Research (Hamers, De Koning, & Sijtsma, 1998) aimed at studying the interaction between teacher and students during the implementation of programmes to encourage thinking shows that education generally can be characterized as product-oriented. This 'product paradigm' is widely held by teachers and teacher-trainers. Education is primarily thought of as the reproduction of knowledge from particular subject areas. This knowledge is the product of a scientific research process and it consequently has an absolute value, which can best be transferred by an expert in the relevant discipline.

At present a new paradigm has gradually been formulated in which the principal temporary nature of openness of scientific knowledge is emphasized and in which the status and expertise of the teacher is put into perspective. Furthermore - and this is essential -, in this paradigm it is not only the end-result (of the thinking) that is emphasized but also and primarily the preceding process. One of the fundamental premises of this 'process paradigm' is that education must not involve reproduction of knowledge as much as the development of skills and capacities which can be applied to knowledge. In other words, it is more important that a person knows what he can do with certain information or how to acquire it (procedural knowledge) than that he should only have such knowledge, available in his memory (declarative knowledge). In this opinion knowledge functions more as a means than an end in itself.

The shift from factual knowledge to knowledge of procedures of how to acquire knowledge and to organize it assumes a more open, dynamic concept of knowledge. The development or stimulation of thinking assumes teaching-learning situations which primarily run according to the problem-process-solution paradigms. The problems which are offered to children should therefore be organized so that they show a discrepancy between the desired end-situation and the undesired starting situation. And the solution process can be changed by varying the complexity, the scope and the degree of structure of the knowledge.

Concluding remarks

Encouraging thinking has been an important subject since the criticisms of and reinterpretation of Piaget's theory on the development of thinking (e.g., Brown & Desforges, 1979; Donaldson, 1978). These criticisms have pointed to the supposed limits in children's capacity for reasoning and abstract thinking. Thinking can be trained or remedied (Sternberg, 1984) and it has been claimed that children with an apparently limited capacity have a greater potential (Hamers, Sijtsma, & Ruijssenaars, 1993). It has been asked whether that potential can in fact be exploited or achieved with suitable training; opinions are divided about this. We refer to the nature-nurture debate on thinking where the arguments for intelligence as an inherited or acquired feature have been detailed. Intelligence could be defined as a person's 'rough' intellectual power, and thinking as the 'skilled' use of that power. In other words, thinking deals with how people use their intelligence and what they do with it. Thinking could therefore be encouraged 'to a certain level' (the potential). For lack of experience it may underperform. The methodologies for encouraging thinking are used to compensate for the lack of experience or to remedy (or in any case improve) how a person thinks.

The division into general and specific programmes and the preference for one or the other is a subject for discussion. Which of these programmes do we most need in schools? The answer to this question determines to a great extent how thinking will be taught. If thinking is taught in an 'across-the-curriculum' course, objectives for thinking skills and strategies will be the basis of the programme. If thinking is taught in the context of a school subject, content objectives will be the basis. There is considerable debate as to which context is more effective for teaching at-risk students (Resnick, 1987). Proponents of the first approach argue that low-achieving students may experience overload if they have to learn both content and skills simultaneously. For instance, Feuerstein (1980) developed content-free programmes using geometric shapes and pictures. Most other programmes for teaching thinking, however, use a combination of content-free and daily life formats (e.g., Klauer, this volume). Proponents of the second approach argue that programmes should be content-related because a substantial part of skills and strategies is content-specific and these skills and strategies cannot be easily transferred to other areas (Resnick, 1987). We agree with Presseisen's (1987) compromise that content-related programmes, in which skills are learned as a means to learning how to solve problems, are generally preferable, unless students have great difficulty with the content. In which case, an 'across-the-curriculum' course might be the best choice, provided that transfer is built into the programme and that the substance of the programme is well coordinated with the school content courses.

Difficulties encountered in most intervention studies are more complex than is often suggested in the literature. It is widely accepted that the effectiveness of cognitive programmes depends highly upon the degree of transfer from the present task to other tasks. Transfer is extremely difficult to achieve in practice. Many programmes (see e.g., Hamers & Overtoom, 1997) do not provide sufficient guidelines on how to facilitate transfer although many authors acknowledge that transfer is an area which requires further work. Experiences in many studies suggest that curriculum-independent materials (see Figure 2), while having the advantage of being novel to the children, make the 'bridging' task even more arduous. Less abstract and more curriculum-based materials may be required for effecting transfer to other academic areas (assuming that the aim of the intervention is transfer to other school domains).

The teacher plays a crucial role in implementing a programme to encourage thinking skills. The programmes keep provoking new views on instruction. The interactions described between students and teacher must encourage students to be more active participants in the learning process, e.g., by creating new ways of working together and by role changes between teacher and students. It has also been our experience (Hamers et al., 1998) that implementing these programmes demands a thorough reorganization of the way in which the teacher teaches the students. The teacher must master and use a greater variety of didactic strategies (process-directed versus product-directed teaching, encouraging thinking aloud, the dialogue form, guiding versus leading the way, algorithm versus heuristic, mutual learning and teaching, stimulating reflection about own thinking, etc.).

If we want to teach children to think then we are aiming to improve their problem solving ability. We can aim to achieve this by putting children to work systematically and methodically on different types of problems. In this book we take a look at thinking processes and how to encourage them from different points of view. Sometimes there is more emphasis on the development of thinking or on the identification of thinking skills and sometimes on encouraging thinking with the aid of training programmes. The chapters also deal with some special themes such as thinking and mathematics, reading comprehension or text production. In addition, in the different chapters we can recognize combinations of theoretical view points and training goals. The book is completed with a contribution about the methodology of improving thinking and, finally, several general and provisional conclusions are given in Chapter 14.

Introduction to the chapters

Csapó (Chapter 2) discusses cognitive research, which has revealed a broad range of approaches to teaching thinking; several of these can be applied to mainstream school instruction. The chapter reviews the current literature that proposes integrating teaching thinking with subject-matter instruction and then summarizes those theoretical considerations which provide resources for designing teaching materials that foster thinking. Two paradigms, the Piagetian cognitive theory and the information processing approach, are discussed. The chapter then examines how some aspects of these two paradigms can be integrated into a consistent framework for improving thinking through the teaching of subject-matter in regular school classes. Then, based on this framework, a method is presented for analysing teaching materials, designing structured exercises and embedding them into the regular instructional processes. The model presented in this chapter can be applied: (a) to thinking skills that can be identified by their structure, and (b) when teaching materials are given (e.g., as defined by a prescribed curriculum). The process of identifying the suitable content of teaching materials, designing exercises and placing them into the regular instruction is illustrated by examples of: (a) several thinking skills from the domains of deductive, combinatorial and inductive reasoning for which exercises are devised and then embedded in teaching material, and (b) several school subjects, e.g., chemistry, physics and grammar.

Adey (Chapter 3) proposes a programme called Thinking Science, designed for use in schools with students aged 11 - 14 years. He describes the basic Piagetian and Vygotskyan theories and the ways in which these theories are worked out into practical activities. In particular, the ideas of cognitive conflict, metacognition, and bridging are called on in the design of the activities and in the method of teaching. Students who experience Thinking Science activities once every two weeks for two years show significant gains in levels of cognitive development compared with controls and with national norms, and they subsequently show enhanced performance in national tests of science, mathematics, and English. These long-term far transfer results are presented as evidence of the effectiveness of the intervention programme for increasing students' general cognitive processing mechanisms. Conclusions are drawn for the design of programmes for promoting thinking, and also for the methodology of evaluating programmes intended to promote thinking skills.

Scheinin's and Mehtäläinen's (Chapter 4) Formal Aims of Cognitive Education (FACE) is a school intervention project based on a philosophical theory of knowledge. In this project the teaching of skilful thinking was integrated into instruction in most of the subjects in the school curriculum. The project was implemented over a period of three years with students in a junior high school. The evaluation of the effectiveness of the project included both process evaluation of the implementation as well as examination of the effects on students' cognitive abilities, formal cognitive skills, cognitive self-concept, and self-esteem. When interviewed, the teachers claimed that their teaching had changed and that this was influencing the students. Test results show no significant broad transfer effects in cognitive abilities, but clear improvement was found specifically in formal cognitive skills. The change was well beyond the age-typical development of the control group. There were significant positive changes in the cognitive self-concept of the students but no change in their self-esteem.

Efklides (Chapter 5) presents two studies which aimed to investigate the possible acceleration of domain-specific abilities. The theoretical assumptions underlying the two studies were derived from experiential structuralism, a theory of cognitive development, which postulates general and domain-specific abilities, including the quantitative-relational (QR) and the causal-experimental (CE) abilities. The studies involved three sets of tasks: QR tasks, CE tasks, and general intelligence (G) tasks. There were three treatment groups: the first was trained in QR abilities, the second in CE abilities and the third was a control group with no training at all. The results showed that the mechanism of cognitive change involves either both G and domain-specific abilities or only the latter. There was also evidence of transfer from QR to CE abilities but not vice versa. The constraints to change and cognitive acceleration were related to the cognitive level and age of the subjects. Finally, the various methods of training imposed different cognitive demands that influenced their effectiveness. These results imply that the teaching of thinking cannot be reduced to skills' acquisition but is a more complex mechanism which may take different forms depending on the teaching method we use and the ability trained.

Klauer (Chapter 6) presents a theory of inductive reasoning, specifying both processes which enable one to solve inductive problems and the various kinds of problems which are inductive in nature and solvable by the defined procedure. Following this, three hypotheses are derived in order to test the theory. Such tests were possible since training programmes based upon the theory have been developed, programmes which address children and youths of different ages and different ability levels. Moreover, a rather large number of training experiments have previously been run where a training group was contrasted to a non-training control group or to a control group which participated in a non-inductive training. Using meta-analytic methods, the existing body of training experiments is synthesized. Based on the meta-analyses it is concluded (a) that inductive reasoning is amenable to training, (b) that training to reason inductively transfers to tests of fluid intelligence, (c) that the training effects last for at least some months, and (d) that the training transfers to and fosters acquisition of declarative knowledge. Unexpectedly, the effect on learning in school are even higher than that on fluid intelligence. One can assume that the theory of inductive processes has proved to be useful, particular if it is taken as an objective for educational measures.

De Koning and Hamers' (Chapter 7) projects are based on Klauer's theory of inductive reasoning. Firstly, the authors discuss the philosophical schools such as empiricism and rationalism in which research is conducted regarding the relationship between man and knowledge. In addition, attention is paid to how people deal with knowledge, how they reason and how education can best take advantage of this. They also describe how research in each of these schools has led to a new synthesis which can be put into practice by means of so-called pragmatic deductive- and inductive reasoning schemes. These context-free schemes appear to represent the abstract level of man's reasoning. They ensure the orderly processing and application of knowledge and the (re-)organization of stored knowledge. Secondly, the importance of induction for young children is considered. The probabilistic character of induction requires the application of so-called intuitive statistical schemes which monitor the balance between acquired knowledge and the adaptation of knowledge organization on the basis of new knowledge. Adults use inductive pragmatic schemes of the law of large numbers and regression to do this. It is assumed that young children must first form an image of group- and row structures. Thirdly, the authors discuss research regarding the training of children using visual, numeric and verbal inductive reasoning tasks which illustrate the general character of the reasoning schemes. The results show the importance that group- and row structures can have in the relation between man and knowledge.

Nelissen's (Chapter 8) Thinking skills in realistic mathematics focuses on an important approach to mathematics instruction, the so-called realistic approach or realistic school. This approach has brought about far-reaching changes in mathematics instruction. Two factors have had a significant influence on the development of the realistic school: firstly, the mathematicians who developed a different view of mathematics and, secondly, a new conception of how children learn mathematics, which draws both from cognitive psychology and from the cultural-historical tradition. This chapter concentrates on mathematics learning and instruction in primary schools, using three key concepts: construction, interaction, and reflection or metacognition. The chapter then proceeds to explore which cognitive processes are fundamental to solving mathematics problems and, finally, to discuss developments within the field of educational psychology which may be relevant to mathematics instruction. Although the theoretical basis for construction, interaction and reflection is quite solid and there is a high level of agreement about the three concepts, more research is needed at all levels of mathematics instruction to increase our understanding of these cognitive processes and the role they play in mathematics learning and instruction.

Verschaffel (Chapter 9) addresses the issue of teaching and learning how to solve application problems in upper elementary school children. He presents a model of expertise in the solution of mathematical application problems, involving the integration and interactive application of different categories of aptitudes during the distinct phases of the problem solving process. This model is then used to describe and analyse some well-documented research findings about elementary school students' difficulties with modelling and solving application problems, and about the characteristics of the current practice and culture of elementary school mathematics that are (partially) responsible for these difficulties in students. A subsequent review describes three recent experiments in the domain of elementary mathematics education that have been explicitly set up to answer the question of how mathematical modelling and problem solving can be successfully taught to (upper) elementary school children. A critical discussion of some problematic elements in designing these experiments leads to some suggestions for further research and development work.

Van Luit (Chapter 10) argues that more than ten percent of all children in primary education cope with learning difficulties. Most of them need special methods of instruction either in special schools or in remedial classes in regular primary schools, in order to learn adequate problem solving strategies and to practice using them. Many children who are labeled as learning-disabled or as educable mentally retarded do have a lack of knowledge and capacity to understand problem solving strategies in mathematics. Easy strategies for addition can be taught. However, in the domain of multiplication and division it is often too hard for them to learn and select adequate strategies. There are too many diverse strategies and they do not understand why they should use one strategy for one problem and a different, more suitable strategy for another problem, or how to choose which strategy to use. This chapter discusses the effect of a strategy instruction training for teaching multiplication and division problems to 30 children in schools for educable mentally retarded and 30 children in schools for learning-disabled. All these children had severe mathematics disabilities. The effectiveness of this training can be explained in terms of its successful integration into parts of the mathematics curriculum, the thinking about different ways of problem solving and the self-instruction. The results show that teaching these children to think mathematically, based on a programme that includes self-instruction for strategy use, was significantly more effective than teaching them in control groups based on a regular mathematics programme. They developed a capacity to think about the best problem solving strategy for doing a specific mathematics task. Furthermore, the results suggest that most of the experimental children with learning disabilities were far better able to deal with generalization tasks after the experimental training. The results are consistent with previous findings, suggesting the importance of implementing strategy instruction in mathematics training programmes.

Chanquoy (Chapter 11) aims at analysing the variations in connectives used in three textual genres, in relation to the development of writing abilities and thinking processes. To reach this objective, connectives are investigated in texts written by 10 and 13 year olds. Connectives have been considered as appropriate surface indicators of deep writing processes. Eighteen fifth-graders and eighteen eight-graders (French native speakers) successively wrote a narrative, a description and an argumentation concerning similar topics. An analysis of the mean proportion of interclausal connectives was conducted to study the children's ability to use specific connectives according to specific genres. Results showed that both 5th- and 8th-graders used similar proportions of connectives, and that these proportions varied according to textual genres. Connectives were more numerous in argumentations and in descriptions than in narratives. The analysis of linearity markers and textual structuring markers revealed that 5th-graders used more linearity markers than 8th-graders, who used more textual structuring markers. Descriptive texts had more linearity markers than argumentative and narrative texts, respectively. Conversely, argumentative texts exhibited more structuring connectives than narrative and descriptive texts. A descriptive analysis of different categories of connectives indicated a high degree of specialization in the use of the different categories of connectives (and spatial, temporal, non temporal and argumentative connectives) for different types of text. These results were globally in agreement with other works. Connectives were not randomly used, but varied according to the text, and their diversification was accompanied by a relative specialization, and by a reorganization of their functions. Finally, the findings of this experiment should prove particularly interesting for those concerned with instruction. These results highlight the variable and adaptable nature of young writers' composing processes, and also the types of text as an important and influential factor in these processes and in the acquisition of writing skills.

Van Oostendorp and Elshout-Mohr's Chapter 12 on thinking skills in reading and text studying, deals with the complicated activities and the thinking skills which are involved in comprehending the meaning of texts and in processing text information according to the situation and purposes at hand. Relevant thinking skills are identified at various levels. They distinguish and discuss the basic level, the strategic level and the higher order level. Furthermore, they focus on a sample of difficulties that frequently occur in reading and studying informative texts, and on thinking skills that might be helpful in dealing with those difficulties. Three categories of difficulties are discussed. The authors describe how these difficulties have been highlighted from the perspective of experimental reading research by manipulation of text materials, reading tasks, and reading abilities of subjects. Subsequently they discuss these difficulties from the perspective of thinking skills. It has been indicated that reading can be viewed as a higher order skill, because research on reading revealed that complex thinking skills, like inductive reasoning and problem solving, are required throughout the reading process. Increasing the specificity of the links between difficulties and thinking skills will not only enhance insight into the required skills; but it can also facilitate accurate description of lower order skills that help readers to reach automaticity and higher order skills that are needed to organize and monitor appropriate appliance of the necessary skills. The authors also discuss how the difficulties which were selected for elaboration in this chapter are related to the broader reading process.

Hager's (Chapter 13) methodological issues primarily deal with the application of concepts of evaluation research to evaluation of cognitive programmes. After defining 'cognitive programmes', Hager considers the connections between these programmes and basic theories interpreting the programmes as 'systems of technological rules'. This interpretation makes it necessary to empirically evaluate the effectiveness of any programme independent of the status of the basic theory that 'inspired' it. Subsequently, two categories of goals or objectives for such programmes are identified, proximal or near, and distal or far goals. It is argued that different criteria are appropriate with respect to these two categories of goals and that the criteria actually chosen should consider the goals to be assessed. There is further discussion on the sizes of statistical effects associated with the programmes and some basic types of evaluation research are reviewed. Some of the most important consequences of the distinction between comparative and non-comparative evaluations are discussed in greater detail, with special attention given to the most plausible effectiveness hypotheses and the sizes of the statistical effects to be expected. Non-comparative evaluations aim at the programme's effectiveness as such, and no statement can be made on the effectiveness of the control programme. In comparative evaluations, the effectiveness of two or more programmes with the same objectives are compared. The problem of appropriate comparison groups in order to control for different nuisance effects is also addressed. Some general criteria for a programme's effectiveness are proposed and the hypothesis-oriented view underlying the chapter is considered.

Csapó and Hamers (Chapter 14) draw some final conclusions on the work presented in this book. They attempt to outline a strategy for managing the diversity found in the chapters. Furthermore, they address some general issues related to present and future research regarding teaching thinking. They discuss the dilemmas that will influence the next few years of research and how a fruitful balance can be found between the benefits and drawbacks of diversity and consistency in theories, programmes and methodological issues.

References

Adey, P.S., & Shayer, M. (1994). *Really raising standards: Cognitive intervention and academic achievement*. London: Routledge.

Baron, J.B., & Sternberg, R.J. (Eds.) (1987). *Teaching thinking skills theory and practice.* New York: Freeman.

Boekaerts, M., & Simons, P.R.J. (1993). *Leren en instructie* [Learning and instruction]. Assen: Dekker & Van de Vegt.

Brown, D., & Desforges, C. (1979). *Piaget's theory. A psychological critique.* London: Routledge.

Bruner, J. (1990). *Acts of meaning.* Cambridge, MA: Harvard University Press.

Bruner, J., Goodnow, J.J., & Austin, G.A. (1956). *A study of thinking.* New York: John Wiley and Sons.

Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies.* Cambridge, MA: Cambridge University Press.

Case, R. (1985). *Intellectual development: Birth to adulthood.* New York: Academic Press.

Case, R. (1996). Changing views of knowledge and their impact on educational research and practice. In D.R. Olson & N. Torraance (Eds.), *The handbook of education and human development* (pp. 61-82). Cambridge, MA: Blackwell.

Chipman, S.F., Segal, J.W., & Glaser, R. (Eds.) (1985). *Thinking and learning skills.* Hillsdale, NJ: Erlbaum.

Cobb, P. (1996). Constructivism and learning. In E. De Corte & F. E. Wienert (Eds.), *International encyclopedia of developmental and educational psychology* (pp. 338-341). Oxford: Pergamon Press.

Coles, M.J., & Robinson, W.D. (1991). *Teaching thinking. A survey of programmes in education.* London: Bristol Classical Press.

Costa, A.L. (Ed.) (1991a). *Developing minds. A resource for teaching thinking.* Alexandria, VA: Association for Supervision and Curriculum Development.

Costa, A.L. (Ed.) (1991b). *Developing minds. Programmes for teaching thinking.* Alexandria, VA: Association for Supervision and Cuculum Development.

Csapó, B. (1990). Integration of the development of the operational abilities of thinking and the transmission of knowledge. In H. Mandl, E. De Corte, N. Bennett, & H. E. Friedrich (Eds.), *Learning and instruction. European research in an international context* (Vol. 2.2, pp. 85-94). Oxford: Pergamon Press.

De Corte, E., & Verschaffel, L. (1980). Een exploratief onderwijsexperiment met aanvankelijke rekenopogaven [An explorative experiment in education with early arithmetic problems]. *Pedagogische Studiën, 57*, 433-448.

De Groot, A.D. (1965). *Thought and choice in chess.* The Hague: Mouton.

De Koning, E., & Hamers, J.H.M. (1995). *Inductief redeneren 1* [Inductive reasoning I]. Utrecht: Utrecht University ISOR.

De Koning, E., Hamers, J.H.M., & Sijtsma, K. (1996). *Test inductief redeneren. Test en handleiding* [Test inductive reasoning. Test and manual]. Utrecht: Utrecht University ISOR.

Demetriou, A., Shayer, M., & Efklides, A. (1992). *Neo-Piagetian theories of cognitive development.* London: Routledge.

Dewey, J. (1991). *How we think.* Buffalo, NY: Prometheus Books (Original work published in 1910).

Donaldson, M. (1978). *Children's minds.* London: Fontana.

Duijker, H.C.J. (Ed.) (1977). *Encyclopedie van de psychologie* [Encyclopedia of psychology]. Amsterdam: Elsevier.

Dumont, J.J. (1966). *De ontwikkeling van de intelligentie* [The development of intelligence]. Den Bosch: Malmberg.

Duncker, K. (1935). *Zur Psychologie des productiven Denkens* [Towards a psychology of productive thinking]. Berlin: Springer.

Elshout, J.J. (1988). Intelligentie en goed beginnen [Intelligence and a good start]. In G. Kanselaar, J.L. van der Linden, & A. Pennings (Eds.), *Begaafdheid: Onderkenning en beïnvloeding* (pp. 46-54). Utrecht: Spectrum.

Ennis, R.H. (1987). A taxonomy of critical thinking dispositions and abilities. In J.B. Baron & R.J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 131-145). New York: Freeman.

Feuerstein, R. (1980). *Instrumental Enrichment. An intervention programme for cognitive modifiability.* Baltimore: University Park Press.

Fisher, R. (1990). *Teaching children to think.* Oxford: Blackwell.
Frijda, N.H., & Elshout, J.J. (1976). Probleemoplossen en denken [Problem solving and thinking]. In J.A. Michon, E.G.J. Eijkman, & L.F.W. de Klerk (Eds.), *Handboek der psychonomie* (pp. 56-68). Deventer: Van Loghum Slaterus.
Guilford, J.P. (1956). The structure of intellect. *Psychological Bulletin, 53,* 267-293.
Gustaffson, J-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8,* 179-203.
Halpern, D.F. (Ed.) (1992). *Enhancing thinking skills in the sciences and mathematics.* Hillsdale, NJ: Erlbaum.
Hamers, J.H.M., De Koning, E., & Sijtsma, K. (1998). Inductive reasoning in the third grade: Intervention promises and constraints. *Contemporary Educational Psychology, 23,* 132-148.
Hamers, J.H.M., & Overtoom, M. (Eds.) (1997). *Teaching thinking in Europe. Inventory of European Programmemes.* Utrecht: Sardes.
Hamers, J.H.M., & Ruijssenaars, A.J.J.M. (1984). *Leergeschiktheid en leertests* [Learning potential and learning potential tests]. Lisse: Swets & Zeitlinger.
Hamers, J.H.M., Sijtsma, K., & Ruijssenaars, A.J.J.M. (Eds.) (1993). *Learning potential assessment. Theoretical, practical and methodological issues.* Amsterdam: Swets & Zeitlinger.
Idol, L., & Jones, B.F. (Eds.) (1991). *Educational values and cognitive instruction: Implications for reform.* Hillsdale, NJ: Erlbaum.
Jacobs, P.I., & Vandeventer, M. (1972). Evaluating the teaching of intelligence. *Educational and Psychological Measurement, 32,* 235-248.
Jones, B.F., & Idol, L. (1990). Dimensions of thinking and cognitive instruction. Hillsdale, NJ: Erlbaum.
Kail, R., & Pellegrino, J.W. (1985). *Human intelligence, perspectives and prospects.* New York: Freeman.
Kohnstamm, Ph. (1930). Over de 'vormende waarde' der leervakken [About the school subjects]. *Pedagogische Studiën, 11,* 129-150.
Maclure, S., & Davies, P. (Eds.) (1991). *Learning to think, thinking to learn.* Oxford: Pergamon Press.
McGuinness, C., & Nisbet, J. (1991). Teaching thinking in Europe. *British Journal of Educational Psychology, 61,* 174-186.
Newell, A., Shaw, J.C., & Simon, H.A. (1958). Elements of a theory of problem solving. *Psychological Review, 65,* 151-16.
Nickerson, R.S., Perkins, D.N., & Smith, E.E. (1985). *The teaching of thinking.* Hillsdale, NJ: Erlbaum.
Nisbet, J., & Davies, P. (1990). The curriculum redefined: Learning to think thinking to learn. *Research Papers in Education, 5,* 49-72.
Philips, D.C. (1997). Coming to grips with radical social constructivisms. *Science and Education, 6,* 85-104.

Phye, G.D. (Ed.) (1997). *Handbook of academic learning. Construction of knowledge.* San Diego: Academic Press.

Podd'jakov, N.N. (1979). Probeerhandelingen in het sensomotorisch denken [Trials in sensomotoric thinking]. *Pedagogische Studiën, 6*, 460-470.

Presseisen, B.Z. (1987). *Thinking skills throughout the curriculum.* Bloomington, ID: Pi Lambda Theta.

Puskin, V.N. (1975). Onderzoek van het denken als proces [Research on thinking as a process]. In C.F. van Parreren & W.A. van Loon-Vervoorn (Eds.), *Denken. Teksten en analyses Sovjet-psychologie 1* (pp. 130-143). Groningen: H.D. Tjeenk Willink.

Resnick, L.B. (1987). *Education and learning to think.* Washington: National Academic Press.

Resnick, L.B., & Klopfer, L.E. (1989). *Toward the thinking curriculum. Current cognitive research.* Alexandria, VA: Association for Supervision and Curriculum Development.

Salomon, G., & Perkins, D.N. (1989). Rocky roads to transfer: Rethinking mechanism of a neglected phenomenon. *Educational Psychologist, 24*, 113-142.

Schoenfeld, A.H. (1985). *Mathematical problem solving.* New York: Academic Press.

Segal, J.W., Chipman, S.F., & Glaser, R. (Eds.) (1985). *Thinking and learning skills. Vol. 1: Relating instruction to research.* Hillsdale, NJ: Erlbaum.

Selz, O. (1935). Versuche zur Hebung des Intelligenzniveaus [Attempt to raise the level of intelligence]. *Zeitschrift für Psychologie, 134*, 236-301.

Simons, P.R.J., & Verschaffel, L. (1992). Transfer: Onderzoek en onderwijs [Transfer: Research and education]. *Tijdschrift voor Onderwijsresearch, 17*, 3-16.

Snow, R.E., Kyllonen, P.C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 191-210). Hillsdale, NJ: Erlbaum.

Spearman, C. (1973). *The nature of intelligence and the principles of cognition.* New York: Arno Press (Original work published in 1923).

Sternberg, R.J. (1977). *Intelligence, information processing and analogical reasoning.* Hillsdale, NJ: Erlbaum.

Sternberg, R.J. (1984). How can we teach intelligence? *Educational Leadership, 42*, 38-48.

Sternberg, R.J. (1985). *Beyond IQ. A triarchic theory of human intelligence.* New York: Cambridge University Press.

Sternberg, R.J., & Berg, C.A. (1992). *Intellectual Development.* Cambridge: Cambridge University Press.

Van Parreren, C.F. (1990). *Leren op school* [Learning in school]. Groningen: Wolters-Noordhoff.

Van Parreren, C.F., & Van Loon-Vervoorn, W.A. (1975). *Denken. Teksten en analyses Sovjet-psychologie 1* [Thinking. Texts and analyses in Soviet psychology 1]. Groningen: H.D. Tjeenk Willink.

Vosniadou, S., & Ortony, A. (1989). *Similarity and analogical reasoning*. New York: Cambridge University Press.

2

Improving thinking through the content of teaching

B. Csapó

Introduction

Throughout the history of education, there has been a dilemma between the teaching of declarative knowledge ('knowing what': e.g. facts, figures, verbal information) and procedural knowledge ('knowing how': e.g. skills of doing something, thinking). One side of this dilemma received periodically greater emphasis than the other. In the last century, for example, formal systems like mathematics or Latin grammar were considered the best means for cultivating the mind. In this century, the dilemma reoccurred in a refined and more sophisticated form when the methods of teaching thinking (or even improving intelligence) were considered: should thinking be taught directly in separate courses using specific materials (the so-called stand-alone courses) or should it be taught within the framework of the established school disciplines by integrating these efforts into the regular school curricula (the 'infusion' or 'embedding' approaches)?

In recent years, the number of publications discussing theoretical aspects of content-based development of thinking skills has been increasing as well as the variety of experiments, programmes and research projects focusing on fostering thinking in the context of mastering subject matter knowledge. Many arguments in the current literature support the content-related approaches and several authors suggest improving thinking within the context of teaching subject matter knowledge.

The recognition of the importance of knowledge in relationship to thinking gained a new impetus in the 1980s. This view was well formulated in several books and papers. For example, Glaser (1984, p. 97) states: "Much recent work emphasises a new dimension of difference between individuals who display more or less ability in thinking and problem solving. This dimension is the possession and utilisation of an organised body of conceptual and procedural knowledge, and a major component of thinking is seen to be the possession of accessible and usable knowledge." A few years later, Resnick (1987) argued for the benefits of embedding the fostering of thinking skills in academic disciplines.

By the end of the 1980s, the expectations concerning the development of thinking within the framework of regular instruction became apparent and perhaps were stated in the clearest terms by Perkins and Salomon (1989, p. 24) in the conclusion of their paper discussing the degree to which cognitive skills are context-bound: "We forecast that wider scale efforts to join subject-matter instruction and the teaching of thinking will be one of the exciting stories of the next decade of research and educational innovation." In the same year, Resnick and Klopfer (1989) collected related cognitive research under the title 'Toward the thinking curriculum', and a few years later, Nisbet (1993, p. 281) began his paper by stating: "The argument of this paper was that the concept of 'the thinking curriculum' is winning long-overdue recognition in education."

A large number of studies examined and highlighted several aspects of teaching thinking as they relate to subject matter instruction. Perkins (1987) proposed elaborate classroom activities for teaching thinking through the content. Swartz (1987) emphasised the importance of critical thinking and also proposed the infusion of thinking skills into mainstream instruction. Nickerson (1988) reviewed a large number of studies aimed at improving thinking through instruction. Canfield and Ceci (1992) related learning to intellectual development.

As these trends indicate, teaching thinking skills in the framework of subject matter instruction has received a growing attention. However, this focused interest resulted in a still growing diversity of programmes and approaches rather than in a firm and consistent theoretical foundation. Although a large body of theoretical considerations and empirical results have been accumulated and certain convergent tendencies can be observed, practitioners still lack guidelines to develop such programmes.

For ease of expression, the approach described here will be called 'content-based methods'. In short, content-based methods is a way of improving general thinking skills while teaching subject matter knowledge. Another practical simplification will also be applied: instead of the long expression 'teaching thinking skills' the term 'training' will be utilised. In the experimental phase this training may take the form of an intervention in the traditional sense, but as I propose here, content-based methods should be so seamlessly integrated into the regular school instruction that ultimately the regular school instruction itself should be functioning as such 'training'.

This chapter collects the relevant information on this issue. First, the primary practical needs and theoretical considerations that support the integration of teaching thinking with subject matter instruction will be presented. Then some general principles for using the discussed procedures in real life school instruction will be introduced; and examples will be used to describe how such training exercises can be constructed and how they can be integrated into the mainstream curriculum. Finally, the difficulties and perspectives of the content-based approaches will be examined.

A framework for teaching thinking through the content

Why content: Practical considerations and arguments

When we develop a subject matter course or a curriculum that contains 'infused' or 'embedded' opportunities for training thinking skills, we face the difficulties posed by the constraints of the content which we are supposed to use. We may ask the question: 'Why should we put forth so much effort to include the training in the established curricular disciplines, when there would be fewer constraints if we devised a separate course?' The first and most trivial answer to this question is: because the subject matter knowledge is there and the students are required to deal with it and finally to master it anyway. Students spend thousands of hours studying the contents of several subjects. Why not better utilise this time by also improving thinking?

Thinking always needs a content; we think about something. The 'empty thinking' or 'thinking about nothing' does not exist. Separate courses for teaching thinking often use exercises with abstract content without any concrete meaning in the hope that the thinking processes acquired in this way do not stick to some specific concrete situations so they transfer well to any other domain of thinking. However, there is little evidence that these programmes have long-term effects on intellectual development. If the training of thinking is integrated into the curriculum, the information given there can be used to process, by the skills to be practised. In this way, as Resnick (1987, p. 49) notes, "It is ensured that there is something solid to reason about."

In most educational systems, school curricula already contain a huge amount of subject matter knowledge that students are expected to acquire. Pressure is exerted on the schools to accommodate their curricula to the new developments of sciences, newly emerging fields of social studies or activities of creative arts. Describing this knowledge and operationalising the goals of teaching are easy. Subject matter knowledge appears in a concrete form and it traditionally finds its way easily into the curricula. On the other hand, the goals of improving thinking abilities are harder to define and operationalise. They are much less articulated and their position is rather weak when competing for instructional time. Thus, methods must be found for transmitting subject matter knowledge and improving thinking that do not compete but rather cooperate.

Thinking is not only a goal of instruction, a desired outcome that finally appears as a result of specific training, but it is a means of learning that has to be practised throughout the entire learning process. One of the most common experiences of researchers and practitioners alike is that learning is possible without intensive thinking, but if students spare thinking, simple memorisation or rote learning results in inert knowledge that can be used for little. Some main problems frequently mentioned in this context are:

- Since students are not able to mobilise their knowledge in contexts other than in which they learned, their knowledge cannot provide a firm basis for further learning. Thus students' knowledge falls into separated, isolated segments.
- Students are not able to apply their knowledge in real life situations.
- School learning does not affect students' naive theories and misconceptions, even if they learn the content of the subject matter and are able to recite it. Therefore their misconceptions are more likely to influence them then is their science knowledge when they make decisions.

In contrast, meaningful learning results in coherent understanding of content. Understanding requires active processing of the material, following the inherent logic of the subject matter, organising the concepts and facts, drawing conclusions from the information given, and building relationships between already existing knowledge and newly acquired information. In summary, practising thinking in the framework of teaching the subject matter knowledge is necessary not only for improving the quality of thinking but also for improving the quality, accessibility and applicability of knowledge as well.

Several innovations and reforms in education indicate that significant improvements cannot be expected without significant additional efforts. This is so with improving thinking as well; no short-cuts or quick fixes exist. Although not requiring much less effort, modification of already existing courses, practices, and teaching methods in order to foster thinking is more conceivable than introducing new courses and producing totally new materials. Furthermore, school curricula are already full and new programmes can be added only if others are eliminated. In most educational systems, stand-alone programmes would have little chance in competing for the limited educational time against the well-established science, humanities, language and social studies programmes.

Theoretical sources: From Piaget to information processing models of cognition

Among the leading paradigms of psychology that attempted to explain the development and functioning of thinking, at least two must be considered when discussing theoretical backgrounds for using subject matter knowledge to improve thinking. Piaget's theory, despite the controversies and the modifications and alterations that have been proposed since its original formulation, is still one of the most consistent models for explaining the origin and accumulation of knowledge. The organising principle of the other paradigm that consists of a set of models is that it describes cognition as information processing.

Piaget, influenced by the structuralist approaches of his time and his background in biology, described cognitive development as an adaptation process. Adaptation takes place through two different processes. Assimilation is the process by which a child integrates new information into already existing structures. Accommodation is the modification and reorganisation of the existing structures. The latter takes place when the new information cannot be assimilated into the old structures. In this theory, development is discontinuous; it goes through different qualitative stages and finds its equilibrium or its end stage by reaching the stage of formal operations. From the point of view of teaching thinking, one of the most interesting aspects of the theory is Piaget's epistemological consideration of the origin of thinking skills, or in his terms, operations. The internalisation process starts with concrete operations, the physical manipulation of real objects. Then, when the same operational structures are used on different objects of the environments, the structures became detached from the concrete contents of the operations. The operational structures become internal and the child becomes able to carry out the operations not only with concrete physical objects but also with their symbols, including abstract concepts and verbal propositions. Thus, according this theoretical framework, ready-made knowledge is not acquired, but instead, new knowledge is actively constructed.

Besides this constructivist approach, another important feature of Piaget's work is that he and his co-workers always studied children's reasoning in real situations, in 'semantically rich' contexts, and not in solving content-free puzzle-like problems. The Piagetian tradition emphasises the operational side and the universal features of cognition and pays less attention to the differences. Neither the problem of differences between the individuals nor the differences between the specific contents or domains of thinking are elaborated in the theory.

The information processing paradigm emerged after a series of changes in psychology often referred to as the cognitive revolution. The information processing paradigm drew several of its concepts, ideas and models from computer science, especially from artificial intelligence research. The main research areas of cognitive psychology are concerned with detecting, perceiving and coding information; and with the questions of how meaning is attributed to the information, how knowledge is represented in the mind and how it is organised into schemes and mental models.

The results of the research carried out in this framework have changed our view of the role which knowledge plays in human cognition. Consistent findings showed that productive thinking more likely means mobilising previous experiences and existing knowledge rather than pure, computation-like reasoning. Human cognition is much less rational than was generally believed in the past. Content of the problems often plays a more important role than its structure. Learning certain skills in a specific domain provides little chance to use these skills in a new, unfamiliar context. In other words, we have little natural ability to decontextualize our thinking skills which are acquired in a specific context.

Models of the semantic representations of information in memory reinterpreted and reinforced of long held views. Specifically, these models suggest that independent pieces of information (e.g. meaningless words, names, dates - in general stand-alone facts and figures) are harder to learn, less accessible and sooner forgotten than information which is organised into coherent units or schemes (coherent texts, stories, descriptions, theories). Organised knowledge and interrelated sets of information are better maintained and more easily recalled, although, more effort is required to reveal the internal relationships of such schemes; in other words, to understand them.

The Piagetian theory and the results of recent cognitive research seem to contradict each other at first glance. However, several researchers have attempted to further develop Piagetian theory and find a balance or a synthesis of the advantages of the two approaches. Some of these researchers are labelled 'neo-Piagetians'. They offer fruitful frameworks for developing programmes for training students' thinking (e.g., Demetriou, Shayer, & Efklides, 1992). Despite the differences in the two approaches, they provide a consistent message: to ensure that students will access previously acquired knowledge, learning must be an active, constructive process.

Certain terms must now be introduced and an explanation of how they will be used in this chapter will be provided here. 'Domain-specific skills' are the procedural components of competence in a certain domain. They comprise the main body of a specialized domain knowledge. An expert of a particular domain is more likely to possess them than a novice. Doing well in school subjects, as well as in professions requires mastering a number of domain-specific skills. The student who skilfully solves algebraic equations or who is able to carry out chemical experiments; the civil engineer who designs homes for her clients; and the lawyer who is able to select, organise and present the arguments, all possess several domain-specific skills. Domain-specific skills are context-bound and are closely attached to the particular content knowledge of the domain. Such skills are relatively easy to identify and describe.

Domain-free, domain-general, or simply 'general skills' are those that contribute to efficient reasoning in a number of different domains. Higher order thinking skills, inductive, deductive and critical reasoning skills, problem solving skills as well as their sub-skills surely belong to this group. These skills may well be candidates for inclusion in general intelligence models. To define and identify them is more difficult. As used in this chapter, the term 'general skills' is not necessarily very complex. Simpler skills, like those that are often called 'Piagetian reasoning skills' (seriation, class inclusion, logical and combinatorial operations, etc.) may also belong to this domain-free set of skills. Carroll (1993) also considers them as belonging to the factors of intelligence. The structure (operational schemes, patterns of actions, rules) of general skills rather than their content characterise them. General skills may also be context-bound and attached to certain contents, but because of their common structural features, they have the potential to be freed from the particular contents in which they were mastered and then be generalised across the domains.

What the acquired thinking skills are for: The problem of transfer

The most crucial question concerning the programmes that aim to improve thinking is how general the acquired skills are. Are the newly acquired skills strictly context bound or can they be used in a broader area? This is again the question of transferability of thinking skills. To discuss the issue of transfer in general is not the aim of this chapter, but as every training programme needs to explain how the training may have an effect on domains or contexts other than the ones in which trained occurred (see for example Klauer, 1990), the problem of transfer must be dealt with here. Only those aspects that are relevant for the content-based approaches will be addressed here.

The research regarding transfer remains inconclusive. Thus far (as Resnick, 1987, also concluded) for teaching there have been few convincing reports of training programmes which have facilitated truly broad transfer. The way one defines transfer depends largely on one's theoretical position. Those who are closer to the Piagetian paradigm may rely on the larger effects of transfer. Beyond the results of several experiments, practical everyday experiences support this view; after all, to learn everything in every new situation is not necessary. Other views maintain that transfer does not exist or that it occurs very little. Those who are closer to the information processing approach may be more likely to share this view. If transfer does not exist, then the same structures must be relearned in each novel situation.

Depending on how the role of transfer is considered, two types of content-based teaching of thinking must be distinguished here. (a) For those proponents of content-based teaching of thinking who either do not believe in transfer, consider it very limited, or do not think it is important, most of the thinking skills are context-bound and domain-specific. Therefore, they do not care about transfer. They argue for the content-based training of thinking because, in their view, this is the only way to foster thinking. In this framework, thinking skills must be taught in every particular context and domain, because only this can ensure that students become competent thinkers in every possible domain. (b) Other proponents of the content-based methods who recognise the possibility of transfer and aim to teach transferable skills: skills that are learned within one domain, but can be used in others, possibly many different domains. In this model, almost any content area is suitable, because skills can be transferred from anywhere to almost any other content area. If this works, teaching thinking skills in one or in a few domains is sufficient.

In this chapter, another theoretical position is proposed that draws from both above described views and pursue a balance between them. The view argued in this chapter suggests that transfer does exist, although limited, and occurs only in certain circumstances. In some conditions the degree of transfer approaches zero while in other conditions it is significant. Thus, the task is to find those conditions in which transfer works fairly well, and to design training tasks to ensure the best transfer. Thus, the purpose of the research is to find those methods that result in transferable skills.

In the context of content-based methods, transfer should not be considered a yes-or-no, an all-or-nothing phenomenon. Rather, it should be thought of as a measurable, continuous variable, ranging from zero to full transfer. The degree of the transfer may be different for each skill and for every possible pair of domains. Furthermore, the transferability of a skill depends on the conditions under which it was mastered.

Based on the results of a large number of experiments, as well as the theoretical arguments presented in the current cognitive psychology literature, I suggest a somewhat pessimistic view of transfer, that is, it is more productive to hypothese a low level of transfer. Thus, caution and awareness of the limitations of transfer should be taken into account when designing content-based methods for training thinking. Three main plausible limitations should be considered. (a) Even if a skill that is potentially transferable is mastered within one domain, transferability is not a feature that comes automatically with the skill. Thinking skills, especially in the early phase of their development are bound to the content in which they are practised. To make them transferable, further specific training is required. (b) The type and content of the training determines how broad the transfer can be. Skills can be more easily transferred into close, familiar content areas than into distant and unknown fields. (c) The skill itself cannot be transferred into another domain; rather, transfer means an improved ability to learn a skill (with the same or similar structure) in new content areas. The consequence of these constraints for content-based training is that the training exercises must be embedded into every relevant academic subject.

While the very essence of teaching thinking by using the content of learning materials is the transferability of the skills, a more elaborate conception of improving transfer is needed. For this, the sub-domains should be considered as basic units of subject matter that use a consistent set of concepts, facts and domain-specific thinking skills. Within such a unit, transfer is not questionable, because a skill is considered to be acquired if it works for the whole of such a unit. However, the content of the sub-domains is different and transfer between them is not automatic. The topics of the traditional school subjects are such sub-domains. For example, the content of geometry obviously differs from the content of algebra, although both are fields of mathematics. Similarly, mechanics has a content different from optics. In order to make a skill transferable, training in the content of more than one sub-domain is required. This makes it possible to generalise the skill, and to detach the structure of the skill from its actual content. If a skill is trained with materials from only two different sub-domains, then its transfer into any other area cannot be expected. However, the presumption that the skill will work in at least these two content areas is plausible. Furthermore, another plausible assumption is that after the skill is mastered in two different fields, it will be learned in a third field more easily. Extending this reasoning, the more content areas a skill is trained in the easier it can be learned in a new domain.

Several experiments propose an active and conscious decontextualisation to facilitate the transfer of a skill. Metacognitive effects may be used to improve the transferability. Despite these efforts, the skills usually cannot be universally applied; there are always unfamiliar contexts where application (without further training) fails. If the skill was practised in several different content areas, the training should result in a skill which is applicable in several domains. Thus, from a practical point of view, the content-based method is much less risky to suggest since the training will have certain benefits, even if the transfer is not very broad.

Designing teaching thinking materials in the content areas

General principles of teaching thinking through the content

Since Resnick and Klopfer (1989) introduced the term 'thinking curriculum', a new view about teaching thinking has gradually become more and more dominant. The assumption that cultivating the mind should be the primary goal of school instruction is unquestionable. Accordingly, the task of teaching thinking cannot be completed in one or even a few separate courses. Improving thinking has to be a continuous goal for the entire period of compulsory schooling from the very first day to the final years or even further, until the completion of higher education. The question is not whether thinking can be improved at school; but how it can best be accomplished; how every single lesson can contribute to the development of thinking.

Although the type of training discussed in this chapter is different from both the one that teaches domain-specific skills and the one that teaches general skills in separate courses with abstract, domain-free materials; it does manifest elements of both approaches. Such a synthetic approach has already been proposed by Glaser (1984). Glaser first described the advantages of the domain-free methods on the one hand and the training in the context of specific domains on the other hand, and he cited a method that combines these two. But then he goes on to describe a further possibility, a deeper integration of these approaches: "But rather than switching between general and specific, I would also examine a fourth possibility: teaching specific knowledge domains in interactive, interrogative ways so that general self regulatory skills are exercised in the course of acquiring domain-related knowledge" (Glaser, 1984, p.102) .

The content-based method is similar to the training of domain-specific skills in that it uses learning materials to train thinking. It is also similar to many stand-alone programmes because the skills to be trained are analysed carefully and described in detail. Furthermore, it is a more or less direct method because the targeted skills are directly practised. It differs from the direct approaches in two ways. In the content based method, instead of abstract contents, the elements of the teaching materials are used, and the training is not limited to a short period of time.

One of the main characteristics of a domain-independent method is the precise analysis of the structure of the skills to be trained. One example of this precise description of the skills is Klauer's (1993a, 1993b) system which is the basis of a stand-alone programme for developing inductive reasoning. Such descriptions, as will be illustrated later, can be utilised for the content-based methods as well. However, the development of content-based programmes requires further work. The main steps of this process to be illustrated in the next section of this chapter are: (a) defining the goals of the training; (b) identifying and defining the skills to be developed; (c) selecting the teaching materials to be used for the training; (d) analysing the subject matter knowledge and searching for places where specific exercises can be embedded; (e) designing the training exercises; and (f) integrating the exercises into the teaching-learning processes.

The next sections describe a process for a possible implementation of teaching thinking in the content areas. The description of this approach will be given through examples from experimental programmes. Since 1985, in the framework of several research projects, we have been experimenting with modified teaching materials and the examples presented here are from these projects (Csapó, 1990, 1992, 1995).

As it is a principle of the content-based method, the subject matter is the concrete material for the training of thinking skills. Accordingly, the examples presented here are from the subject matters of chemistry, physics and grammar. The purpose of these examples is only to illustrate the possibilities of this method, and the examples quoted here should be comprehensible to those who are not experts in the given disciplines. Thus, both with respect to the given skill and the content in which they are practised, the simplest possible examples are presented. However, I must emphasise, that the described method is not limited to such simple skills or well-known contents. It can be used practically anywhere if (a) the particular skills to be developed can be defined and described and (b) the learning material is complex enough to accommodate well structured problems and exercises. When designing a content-based training programme, the first phases of the work are the same as those for the stand-alone programmes. Therefore, the experiences of the stand-alone programmes can be utilised. The other phases of the work, developing and using the exercises, are different.

Defining the goals of training

The purpose of training: The outcome and the aimed group

To conceptualise the goals of the training, two related aspects must be examined: (a) What do we consider to be the outcome of the training? and (b) Which group should be targeted? As for the first aspect, the changes that can be expected in students' cognition must be clarified as to the ways in which these changes are to be measured. In other words, what are the specific criteria for claiming that training has improved students' thinking? Criteria often mentioned are: (a) Can students use their skills in the same domain in which those skills were acquired (does the training have any effect at all)? (b) Can students use their new skills in other domains (is there any transfer)? (c) Do students perform better on general intelligence tests (is there a broad transfer)? and (d) Have students become better learners (does the training affect their learning abilities)?

In the content-based approach all of these goals and levels of evaluation may be relevant but a new way of evaluating the effects of training can also be proposed. Have the students become more intelligent users of the knowledge they have mastered during their training? As a result of the modified ways of teaching, students can be expected to become more competent users of their knowledge; develop a deeper understanding; become better able to mobilise their knowledge in other contexts when it is appropriate, apply it to new situations, apply their abstract knowledge to everyday situations, and make decisions on the basis of their scientific knowledge instead of their naive theories or misconceptions.

The second aspect regarding the goals of training is determination of the targeted group. In general, three groups can be targeted: those whose skills or abilities are below average, around average and above average. However, the methods are not equally beneficial for each of the three sub-populations. (a) Those who are below average may require remediation. They may have learning difficulties or certain problems of understanding that must be corrected in order to catch up with the average students. If the aim of teaching the subject matter is to achieve a deeper insight and understanding, and the students - because of the lack of understanding - cannot do more than simply memorise the material, the proposed content-based methods are typically for them. The proposed method that stimulates thinking about the material to be mastered may be especially helpful for this group. (b) For those students who are average, the training may enhance development beyond that which would be reached with regular instruction. The average students may be the primary targets of the content-based methods. (c) Those who are above average already excel in acquiring and understanding the learning materials and these methods offer little extra benefit. For them more challenging learning materials should be offered instead of the regular materials with more intensive thinking.

Selecting the skills to be trained

To design a programme for teaching thinking, a taxonomy, or at least an inventory of the thinking skills, is needed. A list of the skills to be developed is also required and several attributes of these skills must be specified.

A universally accepted model or a methodical description of thinking for a system or inventory of thinking skills is still a long way off. However, several inventories or taxonomies can be used as an initial model for designing the training. It is not my intention to collect the available training models nor to list the skills that may be developed through the content of teaching. Instead, in this chapter I will address the attributes of the skills which are imperative to succesful training. At least four attributes of skills should be considered before designing the training: (a) relevance, (b) development, (c) structure, and (d) modifiability of the skills.

As for relevance, only thinking skills of broad relevance should be considered. However, for designing training in certain domains, one must examine whether specific skills have relevant function in the reasoning within the given content domain.

Knowledge of the general 'developmental tendencies' of the given thinking skills is necessary. A theoretical model for the development of the given skills would be helpful. Furthermore, empirical data regarding the development of the skills under natural circumstances is needed. Developmental curves may indicate in what period the changes are fastest under natural circumstances and what is the developmental level that the students reach in average without the intervention. Examining individual differences may also help to make decisions about the goals of training.

'Modifiability' is a crucial feature of the given skill. Experiments, even the results of experiments carried out under other conditions, may be helpful in deciding whether an attempt to improve a skill for certain age groups is worthwhile. Modifiability may also be age dependent, so results of experiments with children of certain ages may not be generalisable to children of other ages without further consideration. It is also important to determine whether there are specific ages during which a skill is especially sensitive to developmental influences and, whether there are 'imprinting like' periods during which fast improvements can occur.

The characteristics listed above are helpful in designing training programmes. A massive body of research data is available for this purpose, but there are only a few skills that are described more or less completely in the terms proposed here. In general, the more we know about a skill, the better our chances are of designing an efficient training programme. Without such detailed information some chance may still exist to design successful training, but such detailed information is indispensable in one aspect, namely one has to know to the structure of the given skills.

From the point of view of designing content-based programmes, the 'structure' is the most important characteristic of the skills. This is the feature by which the skills can be described and identified. Only if the structure of a certain skill is known, can the same skill be embodied in different contents and the same skill (the skill with the same structure) developed in different content areas. The structure of the skills may be described verbally, but formal models are the best representations of the structure of the skills.

Knowing the structure is a necessary condition for designing training programmes and it is the structure which poses the strictest limitations on the content-based methods. The structure of simple thinking skills can be easily described, but the structure of those skills that are often referred to as higher order thinking skills are difficult to describe. Nevertheless, successful attempts to describe the structure of such skills have been made.

Preparing the training materials: Some examples

One of the main differences between the American and the European educational systems is that in most of the European countries the content of teaching is more strictly defined and a larger proportion of teaching materials are centrally or locally prescribed. In some European countries the content of teaching is defined in national core curricula, while in other countries it is prescribed by local governments or educational authorities. Nevertheless, the discussion in this chapter considers the content of teaching as given, defined and organised by disciplinary experts. Accordingly, only minor modifications can be made when teaching content.

In the first step of preparing training exercises, the learning materials (curricula, textbooks, other instructional materials) must be analysed to identify where the skills can be exercised or where such exercises can be placed. This requires that the elements of knowledge (concepts, propositions, etc.) which can be used as materials for constructing training tasks be identified.

The working hypothesis to begin this analysis is as follows: if general thinking skills that can be applied to several activities in several content domains actually do exist; and if such thinking skills are general and relevant; then they must be found in (almost) any larger units of (almost) any content area. That is to say that general thinking skills are already present or they can be incorporated without abandoning the original goals of teaching. Such skills may even foster better acquisition of subject matter knowledge.

Constructing training exercises requires unifying the already prepared structures of skills with the given elements of teaching materials, or in other words, to fill the 'empty structures' with the actual content in order to give the abstract or formal description of the structure a concrete meaning. Three examples will serve to illustrate the design of training materials.

Deductive reasoning

The development, role and place of logical operations in thinking is one of the most controversial issues in cognitive psychology. This is also the field where the Piagetian theory most sharply clashes with the information processing paradigm. On the one hand, formal thinkers - in Piagetian terms - are supposed to solve certain logical tasks whatever the content of the tasks; on the other hand, children solve logical tasks with familiar contents, yet often fail to solve tasks with the same structure but unfamiliar content. How do these problems appear in school learning, and can student's logical reasoning skills be improved? Experiments indicate that teaching formal logic is of little help. Can such skills be developed by using the content of teaching?

Primary school science textbooks are full of propositions of complex logical structures and children are able to learn these correctly. They reproduce the statements when they are asked in the same context and usually they are able to interpret accurately what they have learned in the context of the given content. But even if they know the actual meaning of the complex statements, they are rarely able to generalise the logical structure of the statements and use the same logic in other cases.

For formal descriptions of reasoning skills, binary operations of propositional logic are the best examples. These operations form the central part of the Piagetian logico-mathematical structures, and the system of the sixteen operations is well elaborated and easy to represent formally. In the textbooks, propositions are often connected with such operations (AND, OR, IF ... THEN, etc.), and, especially in science texts, understanding of the exact meaning of these complex statements is crucial.

An example for such an operational structure is: IF (p OR q) THEN r (where p, q, and r are simple propositions).This structure can be embodied in several contents if p, q, and r are substituted with real, meaningful statements. For example, if p = 'the milk is pasteurised', q = 'the milk is boiled', and r = 'the harmful bacteria are destroyed', then the actual content of the operational structure is: If the milk is pasteurised or boiled, then the harmful bacteria are destroyed.

This is a real statement that can be found in a real school textbook. How can a training exercise for logical reasoning be built around this complex statement? First, the logical nature of this statement must be emphasised. Several ways for doing this are available.

One way is the systematic evaluation of the logical truth-table of the complex proposition. We may systematically consider what happens to the bacteria when the milk is: neither pasteurised nor boiled, pasteurised but not boiled, not pasteurised but boiled, both pasteurised and boiled. Then the difference between the actual status of the bacteria and the truth of the whole complex statement can be discussed, as well as, whether the statement itself is true when the p, q, and r propositions are respectively: true-true-true, true-true-false, true-false-true, and so on. Of course, the whole discussion should be about the actual problem, in the terms of pasteurisation, boiling and bacteria. The very essence of the exercise is to prompt the students to think in an organised, structured way about the material to be learned.

Another possibility is to ask students to tell the same thing in different ways. Whether the different wordings mean exactly the same thing can be discussed. The students can be asked how they can prove that this complex statement is true. What experiments should be carried out? What kind of possibilities should be examined to ensure that the statement is true? Why do the students think that the statement could be proved in a certain way? How could the statement be falsified? What would be the facts that could contradict the statement?

These exercises should be organised in a realistic and meaningful way. For example, students may discuss where boiling milk is still practised, where pasteurisation is practised, and what are the advantages of pasteurisation. Then the students could be asked, what happens to the bacteria if pasteurised milk is boiled.

Another way to emphasise the logical nature of a statement is to help students to decontextualise the structure of the skill with some further practise. The students may be asked if they have ever dealt with a statement that is similar to the one under discussion. If they respond with an example, it can be analysed to determine if it is really similar. Students can be asked why they believe it is similar, what similarity in this case means, whether the rules discussed previously apply to this statement, and, whether they can construct similar statements?

A principle of this method of improving thinking is to use neither technical terms nor formalisations. Everything should be expressed in terms of the actual content of learning. Long-lasting results may not be expected from one or even a few exercises. Many such exercises should be carried out during an academic year and the skills should be practised over several years until students reach the optimum level in the given skill.

Of course, deductive reasoning cannot be limited to the logic of propositions. Exercises may be constructed for other types of deductive reasoning as well, in a way similar to the process presented here.

Combinatorial reasoning

According to Piaget's theory of cognitive development, combinatorial reasoning is an integral part of formal thinking (Inhelder & Piaget, 1955). Since the operational structures examined by Piaget and his co-workers (Piaget & Inhelder, 1951) can also be formally represented and organised into a more complete system, the development and structure of combinatorial reasoning can be empirically studied and described (Csapó, 1985). Combinatorial reasoning plays an important role both in school learning and everyday thinking. It is performed when different elements are combined into larger units or constructions and several, usually all possible constructions, are looked for or enumerated and examined. Well-developed combinatorial skills may improve the fluency of thinking when considering different solutions for a problem; finding unusual relationships between certain elements, concepts, propositions; or generating a large variety of patterns from given units.

The example presented here for the training of combinatorial reasoning is from a set of exercises devised from the contents of seventh-grade chemistry. This example illustrates how such exercises may help students to find unusual relationships between given concepts. In this way the ability to make remote associations may be developed as well. In this example, the aimed thinking structure may be formally described as enumerating all possible combinations of two elements of a given set. Let's consider a set of five elements: {A, B, C, D, E}. The combinations may be enumerated in this way: AB, AC, AD, AE, BC, BD, BE, CD, CE, DE.

The textbook, an existing and broadly used chemistry book, that provides the content for devising the training tasks, lists some possible groupings of materials in an introductory section. The groups of materials introduced there are: sources of energy (A), inflammable materials (B), nutritive materials (C), metals (D), and minerals (E). Students could be asked to combine these aspects in every possible way (combinatorial reasoning) and then to discuss the connections between the various concepts in pairs (making the exercise relevant in terms of the given content).

The possible pairs of the five groupings are:

- Source of energy - inflammable material.
- Source of energy - nutritive material.
- Source of energy - metal.
- Source of energy - mineral.
- Inflammable material - nutritive material.
- Inflammable material - metal.
- Inflammable material - mineral.
- Nutritive material - metal.
- Nutritive material - mineral.
- Metal - mineral.

After enumerating the possible pairs, students could be asked what they can say about these relationships. This allows for the collection of many known facts, for example, numerous sources of energy are inflammable; certain nutritive materials are sources of energy for living organisms; salts of certain metals are vital, whereas others are poisonous for living organisms; most of the metals can be found in the form of minerals in Nature, and so on. The unusual combinations of the groupings of materials inspire students to reason in a way that is different from the usual pattern of a given discipline but may be practised across disciplines. These operations offer a new possibility for increasing the consistency of knowledge because they highlight relationships which might otherwise never appear in the teaching-learning processes (Csapó, 1990).

This specific exercise was placed at the beginning of the study of a new topic and used for mobilising students' preliminary knowledge. These types of activities are especially helpful for embedding a new body of knowledge into the students' existing experiences. They allow the students to relate their previous everyday experiences with the framework of knowledge to be acquired. Such an exercise may also be used at the end of a topic. After summarising and organising the newly mastered knowledge according to its internal logic, it may then be viewed from a different perspective. These kinds of exercises may help students cope with the isolation of the knowledge of a specific topic by building a large number of associations with other fields.

In several cases, the combinatorial structures are given in the learning materials and students carry out enumerations with great accuracy by using the given contents. However, the skills of enumerating the possible combinations are strongly attached to the given content and students are unable to decontextualise the operational structure of the skill and are unable to use it outside of the given context. Such examples can be found in grammar or foreign language learning. To form pairs from two sets of elements, for example {S, P} and {1, 2, 3}, such that the first component of the pairs should be chosen from the first set, while the second component from the second set, the enumeration of the possible pairs (the Cartesian product of the sets) is: S1, S2, S3, P1, P2, P3.

If S stands for singular, P for plural, and 1, 2, and 3 for the first, second and third person, this abstract structure becomes the well known pattern for the conjugation of verbs. Students do these enumerations when learning grammar, especially in languages where the cases have different endings. This structure must be reproduced when learning the formal grammar of another language. Students often have to compare the similarities and differences between their first and second (third, etc.) languages. In German, for example, some of the cases are not different but in French the verb 'être' is different in every case. In Hungarian every regular verb has different endings in the different cases. Thus, the above enumerated formal structure can be embodied by the conjugation of verbs of different languages, as the following example illustrates:

Formal structure		German		French		Hungarian	
S1	P1	lerne	lernen	suis	sommes	tanulok	tanulunk
S2	P2	lernst	lernt	es	êtes	tanulsz	tanultok
S3	P3	lernt	lernen	est	sont	tanul	tanulnak

Within the framework of language learning, students usually learn these skills for enumerating such lists and they are even able to recognise the pattern of correspondence, similarities and differences between them. However, students are usually not able to extend the scope of these skills, the operational scheme of the enumeration, beyond the context of grammatical structures. The decontextualisation process may be facilitated by 'translating' the structures into another content that is obviously outside the context of natural languages. For example, if singular is 'translated' into 'square' and plural into 'circle', and then the first, second and third persons are 'translated' into blue, yellow and red colours, students can be asked to construct a pictorial representation of conjugation. They would draw a blue square, a yellow square, ..., and a red circle.

Transferring the skill into a remote domain cannot take place without further efforts. In order to develop a skill with the same structure in physics, students must practice the skill with contents taken from physics. Such an exercise can be created when students are experimenting with the pendulum. In an exercise with the same '2x3' structure, students are provided with a heavy (H) and a light (L) ball that can be put at the end of a short (S), medium (M) or long (L) string. If students are to observe how the time of the pendulum swing depends upon the weight of the ball and the length of the string, the best strategy is to construct every possible variation of the pendulum from these materials. Accordingly, pendulums of HS, HM, HL, LS, LM, and LL should be constructed and then students can perform the necessary measurements and comparisons. Then, the structure can also be extended within the context of physics by designing experiments with more elements (4x5 measurements and so on) and by introducing new dimensions (sets, variables).

Much can be done to extend these combinatorial structures within grammar as well. For example, a new dimension can be introduced by using different tenses. Students usually have difficulties considering more than two dimensions at the same time, and enumerating grammatical structures in unusual order may help to overcome this difficulty.

Latin grammar was taught for centuries in the belief that it cultivated the mind. Maybe it really did, at least to some extent, otherwise it could not have been a practice lasting for centuries. But the benefit of those rigid exercises could hardly have been proportional to the sufferings and boredom of the students. Of course, when examining the possible utilisation of grammatical structures in the training of thinking, I do not intend to revitalise those old practices. On the contrary, I would like to show, that some easy and playful exercises may help to decontextualise the skills that children acquire with almost an imprinting-like ease.

Following the well-structured rules of enumerating combinations and variations does not necessarily mean limitations, rather, it may be the starting point for creation and construction. This was observed in an experiment when combinatorial operations were developed in art education. After some examples of systematic enumeration were presented, children produced a great number of variations of shapes, figures and colours in their drawings (Zombori, 1992).

The above examples show that not only transmitting disciplinary knowledge and fostering cognitive skills can be linked but also the processes and exercises of improving different types of thinking skills. In these examples, students were supposed to compare things, recognise similarities and differences, and eventually find analogies. These are the processes of a more complex skill, inductive reasoning.

Inductive reasoning

Inductive reasoning is different from the other two types of reasoning examined in the previous sections. While the previous two reasoning skills are examples of skills that are rather simple and easy to describe, inductive reasoning is a type of thinking that is often referred to as a higher-order cognitive skill. Its central role in thinking and its relationship to general intelligence is broadly studied in several relationships and contexts; several intelligence tests contain inductive reasoning tasks. Inductive reasoning has been studied as a central component of critical thinking (Ennis, 1987); as one of the mechanisms of hypothesis generating and hypothesis testing (Gilhooly, 1982) and concept development (Egan & Greeno, 1974; Gelman & Markman, 1987; Markman, 1989); and as one of the basic learning abilities (Pellegrino & Glaser, 1982) or learning skills (Ropo, 1987); while current works use inductive tasks to measure learning potential (Resing, 1993; Tissink, Hamers, & Van Luit, 1993).

Holland, Holyoak, Nisbett, and Thagard (1986) presented a comprehensive theoretical analysis of inductive reasoning. Klauer (1993a, 1993b) developed a formal description that is used for defining the structures of the tasks in his training programmes. In terms of Klauer's definition, the essence of inductive reasoning can be identified by comparison processes: attributes of objects and relations between the objects are compared to detect similarities and differences.

Since inductive reasoning is a rather complex cognitive skill, in order to develop it within the framework of teaching subject matter knowledge, inductive reasoning must be dealt with at a higher and more complex level, where larger units and more advanced thinking processes can be identified. In this way, inductive processes can be described at two levels. (a) Formal descriptions, like Klauer's, allow identification of certain forms of inductive reasoning by their structures and embodiment of the same structure in different contents. Here, these contents are the elements of teaching materials. (b) The extensive previous research into induction resulted in a large amount of theoretical and empirical knowledge that can be mobilised to identify larger units of inductive reasoning. These larger units should also be identified in the learning processes, and training exercises should bear the relevant attributes of these larger units of thinking.

At this higher level, for example, such processes can be identified, and trained across several topics within a domain or across several school subjects (Csapó, 1995):

- Generalising rules from measurement results, observations, and everyday experiences; hypothesis formation and hypothesis testing.
- Analogies, where the relationship can be element-set, part-whole, cause-effect, contrast, function, transformation, origin or functional part-whole.
- Series to continue, where the members of a series are connected to each other by the relationship of element-set, part-whole, time, cause-effect or transformation.
- Grouping, organising facts and figures, creating two or more dimensional tables,

system formation.

- Concept formation and concept development, concepts as sets and subsets, comparing everyday and scientific concepts.
- Complex analogies, analogue series, analogue systems, parallel developments, isomorphic phenomena, rules and laws.

Inductive approaches are well known in educational practice; teaching by examples, among others, is an old method. However, a closer look at existing educational practice reveals that students usually are provided with ready-made knowledge, even if the knowledge presented to them is the result of inductive processes. Students are not expected or forced to think actively. In order to advance beyond this practice, teachers have to receive more theoretical support and be provided with applicable methods. The teaching material must be modified and enriched with exercises that require active thinking.

This will be illustrated with an example taken from the chemistry curriculum. (since the exercise must be presented here in a simplified way, the details of the chemical aspects will be omitted). From two related parts of the textbook, two phenomena can be brought together and presented as an analogy: the battery and the corrosion. In this exercise, a pictorial representation of these two chemical phenomena is presented (a battery on the one side and a corroding piece of iron with water and atmospheric oxygen on the other side). Then the students are asked to analyse the two phenomena. At first, they are asked to list the similarities and differences they can observe. At this step they may collect both relevant and irrelevant features. In the next step students have to express their observations in chemical terms (electrolyte, positive electrode, negative electrode, and so on). In this way, they realise that the two phenomena can be described with common terms and they can recognise similarities again, in a more explicit way. They may be asked to find corresponding parts of the two sides of the figure (part-whole analogies). Then they should find the scientific term that names the essence of the common features of the two phenomena (electric cell - concept formation and concept development). When students are asked what makes the two cells work (chemical energy - similarity), a functional analogy can be shown. To return to the differences, the concrete materials may be examined and the chemical reactions in the two cases may be compared, as well as the different voltages, energies, and reaction speeds. Practical applications may also be discussed by collecting the similarities and differences, and comparing beneficial and harmful aspects.

The exercise built on the above scheme can serve several functions in the teaching process, depending upon its place in the curriculum: (a) one of the phenomenon is already known and the other phenomenon can be explained by using the analogy; (b) both phenomena are already known, but the common features can be generalised and a higher concept can be taught; (c) two or more phenomena may be introduced in parallel, to generalise common features.

These suggestions are not strange and are not even new in the context of current science education reforms. Activities like these have frequently been proposed, sometimes under other terms (for example analysis-synthesis). Scientific reasoning skills are considered main components of general intellectual skills (see Voss, Wiley, & Carretero, 1995), and inductive reasoning, especially hypothesis generating and hypothesis testing, is supposed to be best trained in the context of science education. However, in order to teach transferable skills and make the skills usable beyond the context of dealing with sciences, the same type (structurally identical) reasoning skills should be trained in history, grammar, and other subjects. These skills can be used beyond one or a few given contexts.

Here the ideas that thinking skills cannot be separated and that different types of thinking abilities should be practised within the teaching of every particular topic of each subject matter must be emphasised. To show a possible relationship between the types of thinking used in these examples, another exercise can be considered. Students are given six different materials: A, B, C, D, E, F. Among these are materials that can be used as electrodes and electrolytes. The students are asked to construct the best battery (electric cell) from these materials. One strategy is to compose every possible combination of three materials (ABC, ABD, ..., DEF), and then exclude those that cannot form an electric cell. Then the students determine (by theoretical or experimental methods) which cell would produce the best result. Another strategy utilises analogical reasoning, especially finding structural analogies which may be helpful in the decontextualisation processes. Improving the transfer of a given skill from one domain to another may be accomplished through collecting analogous exercises from the different domains.

Integrating the exercises into the instructional process

During the design phase, for analysis, the structure of the exercises designed to practice a given thinking skill must be clearly identified. But for practical use, the exercises must be integrated into the teaching-learning processes. The training exercises must not be artificial or unusual in the given context; otherwise they would become somewhat independent from the given context and they would do no more than the context-free exercises. The exercises must be embedded in the process of acquiring the subject matter knowledge and they must contain real functions in it. The practice of thinking skills has to be consistent with the original goals of teaching and must facilitate better acquisition and understanding.

Thinking exercises can be used to improve knowledge acquisition in several typical ways:

- Thinking exercises can establish relationships between old and new knowledge, in order to integrate new information into the context of the existing body of knowledge.
- Practising thinking skills may be used to build relationships between the different areas of the existing knowledge. It can promote the integration of the knowledge of different sub-domains of a discipline or of several disciplines.

- Training exercises can be used to enrich the connections between the students' previous everyday experiences and their scientific knowledge. Training may facilitate the substitution of naive generalisations, theories, and misconceptions with more appropriate scientific knowledge.
- Training thinking skills may facilitate the use of newly acquired knowledge by connecting it to practical applications through reasoning processes.

The exercises can be integrated into the teaching-learning process in several different ways. The best approach is probably a combination of the different possibilities in order to avoid monotonous, schematic and boring training. The optimal proportion of the different applications largely depends on the age of the students and the type of domain. Some possible ways to integrate the exercises are: (a) interactive classroom work with teacher guidance; (b) group work utilising co-operative learning; (c) student experiments in the laboratory; (d) individual work with worksheets, workbooks; (e) individual projects that require performing several structured exercises; and (f) several forms of homework.

The training exercises can be embedded into the teaching processes at several levels. Since such embedding requires competence in at least two domains (in the given school discipline and in the thinking skills), it can be best accomplished by specially trained experts and working groups. However, the required expertise can be acquired by practising teachers as well. Thinking exercises can be designed for one or more lessons spanning one or more weeks of study. However, more enduring effects can be expected only from longer periods of training, i.e., training which lasts at least a semester if not a whole school year. Such training programmes can usually be developed by groups of teachers and other experts. Depending on the curriculum policy of a given country, such training programmes could be designed at national as well as at local (school district or school) levels.

In an ideal case, training thinking would be consistently designed for a variety of cognitive skills, carried out in several school subjects, and continued for several years or for the whole schooling period. Despite conscious efforts, a curriculum that places as much emphasis on teaching general thinking skills as on teaching subject matter knowledge is still far from reality. Maybe Nisbet was too optimistic when he predicted: "... before the century is out, no curriculum will be regarded as acceptable unless it can be shown to make a contribution to the teaching of thinking." (Nisbet, 1993, p. 290) However, sooner or later this prophecy will be fulfilled.

Concluding remarks: Problems and promises

The idea of training thinking through the content of teaching is not new at all, but as current literature shows, only in the last decade has it attracted attention and inspired comprehensive research projects. Since large-scale experiments and the implementation of these ideas in practice require more time, it is too early to determine whether it can be used in everyday practice and whether the benefits of modified instruction are worth the extra efforts invested in designing training exercises and modifying instructional practices. Even though the time is too short and the number of large-scale experiments are too few to draw strict scientific conclusions about the practical applicability of these approaches; and despite the enthusiasm observable since the 1980s, theoretical analyses and propositions still outnumber the real experiments.

Although the conceptual framework for evaluating large-scale experiments has improved a great deal, and not only well-founded theories exist but also empirical studies, (see, for example, Adey & Shayer, 1994; Shayer, 1992) the evaluation of content-based developmental programmes needs further refinement. New approaches are necessary; for example, to find better assessments of the quality and accessibility of knowledge.

A further problem concerns the replicability of content-based intervention studies. While the content of the regular curriculum is used to design training exercises, these exercises must be produced again and again if the conception of the training is to be transferred from one educational system in another or even from one subject matter to another subject matter. Similarly, successful training programmes cannot be directly exported from one system into another. Consequently, ensuring standard conditions for experiments is difficult, and measurement and comparison of the effects of the training are not possible with the same accuracy as are possible with the stand-alone curriculum-independent training programmes.

From the content-based methods one cannot expect great results in the short run. These methods work better if applied in several school subjects over several years. So, short term effects may hardly be detectable. The efforts invested in devising better curricula may offer a higher return at the societal level; small changes accumulate over the years and, if the improved curricula is taught to large masses of students, the small effects accumulate again.

The examples presented in this chapter are simple. The principles of the training may be generalised and can apply whenever the structure of the targeted skills can be determined. But how far can we go in describing the structure of higher order skills? Whether the growing attention to content-related methods of fostering thinking indicates that the pendulum is now swinging from the general abilities to the direction of content knowledge or whether this recent emphasis on the integrated methods indicates a final or at least a temporary balance between the two sides of the dilemma remains to be seen.

Acknowledgements

The first version of this chapter was drafted while the author was a Fellow at the Center for Advanced Study in the Behavioral Sciences. I am grateful for financial support provided by the Andrew W. Mellon Foundation and the Johann Jacobs Foundation. The research projects that were the empirical background of some ideas and generalizations presented here were supported by the Hungarian National Science Foundation (OTKA-1622, T018577).

References

Adey, P., & Shayer, M. (1994). *Really rising standards. Cognitive intervention and academic achievement.* London: Routledge.

Canfield, R.L., & Ceci, S.J. (1992). Integrating learning into a theory of intellectual development. In R. J. Sternberg & C.A. Berg (Eds.), *Intellectual development* (pp. 278-300). Cambridge, MA: Cambridge University Press.

Carroll, J.B. (1993). *Human cognitive abilities. A survey of factor-analytic studies.* Cambridge, MA: Cambridge University Press.

Csapó, B. (1985). Untersuchung der kombinativen Fähigkeit und ihrer Operationen bei 14-jahrigen Kindern [Examination of combinative ability in 14-year-old children]. In H.J. Henning & P.K.G. Günther (Eds.), *Bremer Beiträge zur Psychologie, Nr.43* (pp. 29-75). Bremen: Bremer Methodenkonferenz.

Csapó, B. (1990). Integration of the development of the operational abilities of thinking and the transmission of knowledge. In H. Mandl, E. De Corte, N. Bennett, & H.E. Friedrich (Eds.), *Learning and instruction. European research in an international context* (Vol. 2.2, pp. 85-94). Oxford: Pergamon.

Csapó, B. (1992). Improving operational abilities in children. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development. Implications and applications for education* (pp. 144-159). London: Routledge.

Csapó, B. (1995, August). *Improving inductive reasoning through the content of teaching materials.* Paper presented at the Sixth European Conference for Research on Learning and Instruction. Nijmegen, The Netherlands.

Demetriou, A., Shayer, M., & Efklides, A. (1992). *Neo-Piagetian theories of cognitive development. Implications and applications for education.* London: Routledge.

Egan, G.E., & Greeno, J.G. (1974). Theory of rule induction: Knowledge acquired in concept learning, serial pattern learning and problem solving. In L.W. Gregg (Ed.), *Knowledge and cognition* (pp. 43-104). Potomac, MD: Erlbaum.

Ennis, R.H. (1987). A taxonomy of critical thinking dispositions and abilities. In J.B. Baron & R.J. Sternberg (Eds.), *Teaching thinking skills* (pp. 9-26). New York: Freeman.

Gelman, S.A., & Markman, E.M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development, 58,* 1532-1541.

Gilhooly, K.J. (1982). *Thinking: Directed, undirected and creative.* London: Academic.

Glaser, R. (1984). Education and thinking. The role of knowledge. *American Psychologist, 39*, 93-104.

Holland, J.H., Holyoak, K.J., Nisbett, R.E., & Thagard, P.R. (1986). *Induction. Processes of inference, learning, and discovery.* Cambridge, MA: MIT.

Inhelder, B., & Piaget, J. (1955). *De la logique de l'enfant á la logique de l'adolescent* [From the logic of the child to the logic of the adolescent]. Paris: Presses Universitaires de France.

Klauer, K.J. (1990). Paradigmatic teaching of inductive thinking. In H. Mandl, E. De Corte, N. Bennett, & E.H. Friedrich (Eds.), *Learning and instruction. European research in an international context* (Vol. 2.2, pp. 23-45). Oxford: Pergamon.

Klauer, K.J. (1993a). *Cognitive training for children. A developmental programme of inductive reasoning and problem solving.* Paper presented at the Fifth European Conference for Research on Learning and Instruction. Aix-en-Provence, France.

Klauer, K.J. (1993b). Training des induktiven Denkens [Training of inductive thinking]. In K.J. Klauer (Ed.), *Kognitives Training* (pp. 141-163). Göttingen: Hogrefe.

Markman, E.M. (1989). *Categorisation and naming in children. Problems of induction.* Cambridge, MA: MIT.

Nickerson, R. S. (1988). On improving thinking through instruction. *Review of Research in Education, 15*, 3-57.

Nisbet, J. (1993). The thinking curriculum. *Educational Psychology, 13,* 281-290.

Pellegrino, J.W., & Glaser, R. (1982). Analysing aptitudes for learning: Inductive reasoning. In R. Glaser, (Ed.), *Advances in instructional psychology* (Vol. 2, pp. 269-345). Hillsdale, NJ: Erlbaum.

Perkins, D.N. (1987). Knowledge as design: Teaching thinking through content. In J.B. Baron & R.J. Sternberg (Eds.), *Teaching thinking skills. Theory and practice* (pp. 62-87). New York: Freeman.

Perkins, D.N., & Salomon, G. (1989). Are cognitive skills context bound? *Educational Researcher, 18,* 16-25.

Piaget, J., & Inhelder, B. (1951). *La genése de l'idée de hasard chéz l'enfant* [The development of the conception of probability in children]. Paris: Presses Universitaires de France.

Resing, W.C.M. (1993). Measuring inductive reasoning skills: The construction of a learning potential test. In J.H.M. Hamers, K. Sijtsma, & A.J.J.M. Ruijssenaars (Eds.), *Learning potential assessment. Theoretical, methodological and practical issues* (pp. 219-242). Amsterdam: Swets & Zeitlinger.

Resnick, L.B. (1987). *Education and learning to think.* Washington, DC: National Academic Press.

Resnick, L.B., & Klopfer, L.E. (Eds.). (1989). *Toward the thinking curriculum: Current cognitive research.* Alexandria, VA: ASCD.

Ropo, E. (1987). Skills for learning: A review of studies on inductive reasoning. *Scandinavian Journal of Educational Research, 31*, 31-39.

Shayer, M. (1992). Problems and issues in intervention studies. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development. Implications and applications for education* (pp. 107-121). London: Routledge.

Swartz, R.J. (1987). Teaching for thinking. A developmental model for the infusion of thinking skills into mainstream instruction. In J.B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills. Theory and practice* (pp. 62-87). New York: Freeman.

Tissink, J., Hamers, J.H.M., & Van Luit, J.E.H. (1993). Learning potential tests with domain-general and domain-specific tasks. In J.H.M. Hamers, K. Sijtsma, & A.J.J.M. Ruijssenaars (Eds.), *Learning potential assessment. Theoretical, methodological and practical issues* (pp. 243-266). Amsterdam: Swets & Zeitlinger.

Vidákovich, T., & Csapó, B. (1988, June). *Changes in student' logical structures after a 10-month period of experimental training*. Poster presented at the Third European Conference on Developmental Psychology. Budapest, Hungary.

Voss, J.F., Wiley, J., & Carretero, M. (1995). Acquiring intellectual skills. *Annual Review Psychology, 46,* 155-181.

Zombori, B. (1992). *Fähig Visualität und Variationen: Zeichenaufgaben zur Förderung kombinationatorischer keiten* [Visuality and variation: Drawing problems for activation of problem solving abilities]. In I. Bak, A. Kárpáti, O. Scholz, & A. Volger (Eds.), *Aesthetische Erziehung in Ungarn* (pp. 88-101). Berlin: Hochschule der Künste.

3

Thinking Science: Science as a gateway to general thinking ability

P.S. Adey

Introduction

'Thinking Science' is a set of activities and methods of teaching that was developed within a series of research projects called Cognitive Acceleration through Science Education (CASE). The activities are set in science contexts and taught by science teachers in secondary schools. They are founded in theories of cognitive development described by Piaget and by Vygotsky, and in particular draw on the notion of cognitive conflict as a spur to equilibration on a higher level, and on metacognitive processes as a necessary element in the realisation of the potential gains available within the Zone of Proximal Development (ZPD). The main aim of CASE was to investigate the possibility of raising the general cognitive (or thinking, or intellectual) level of young adolescents in ordinary school settings. This aim begs two important questions:

- Is there any such thing as a general cognitive level, or is all cognition domain or situation specific?
- Even if a child can be described as having some general cognitive level, is there anything that teachers or parents can do about raising that level, or is it genetically programmed and impervious to educational influence?

So our aim was to shed some light on these two questions also. An intervention programme that has a positive effect on students' general cognitive processing capability would have enormous potential for raising academic achievement in an efficient manner, since improved processing would make all academic learning more effective. The Thinking Science programme has already been shown, in a well-controlled experimental evaluation, to have long term effects on pupils' general thinking ability. This is a fairly remarkable claim, because in spite of the enormous interest world-wide in the potential value of teaching thinking, and in spite of a plethora of 'thinking skill' programmes, rigorous evaluation and reliable reports of success have been notably rare. We believe that the process of raising general thinking levels is necessarily a slow one, and although other thinking programmes may be effective, very few have had the research support necessary for the three- and five-year monitoring needed to detect permanent change.

In this chapter, I will:

- Present the theoretical foundation of Thinking Science.
- Describe (with some examples) the sort of classroom activities which we have developed.
- Give an account of how the programme is evaluated and details of recent results obtained.
- Conclude by reflecting on some of the difficulties involved in effective evaluation which might account for the paucity of good evaluation studies of thinking skills programmes.

The theory base: Piaget and Vygotsky

In order to address the questions raised in the introduction we need, firstly, some way of describing what counts as 'raising cognitive level'. We could use IQ scores, based on some non-verbal test such as Raven's Matrices. We know that such measures have technical reliability and that they correlate quite well with pupils' academic performance in schools. The problem with such measures, however, is that they are technical instruments developed psychometrically, but without a theoretical foundation in the process of cognition or of learning. They work well within their own limitations, but we have little idea how they work.

At the other extreme are well thought-through accounts of the nature of intelligent behaviour, such as that developed by Resnick (1987). She describes higher order thinking as being, inter alia, non algorithmic, complex, yielding multiple solutions, involving nuanced judgement and interpretation, and self-regulated. It is not difficult to agree with such a characterisation of higher level cognitive performance, and yet it is still not sufficient or our purpose, for two reasons. One is that there is no indication within the description of the mechanism by which lower order thinking might develop into higher order thinking (and therefore nothing for the teacher of thinking to get a handle on); and the other is that it would take a great deal of development work to devise an operational way of assessing the attainment of these higher order capabilities.

However, there is available (and has been for many years) an excellent operational description of the process of cognitive development which includes an account of the mechanism by which it occurs naturally (and therefore by the manipulation of which, development might be encouraged) and also a well-established method of assessing the attainment of different levels. I refer to Piaget's account of cognitive development through a series of stages from preoperational to formal operational thinking. For the age group in which we are interested, 11 - 14 year olds, it is Inhelder and Piaget's (1958) description of the schema of formal operations which provides a relevant goal for any programme which aims to help in the development if thinking capability.

For science teachers in particular, the characterisation of the schema of formal operations in terms such as control of variables, proportionality, compensation, equilibrium, probability, correlation, and formal models seems to be immediately applicable. These are all types of reasoning which have many examples within the development of scientific thinking although it would be a mistake to think of Piaget's account of formal operations as being limited to science. It was always intended to be quite generally applicable across all subject domains, although in only a few (such as history and mathematics) has its application been worked out as thoroughly as it has in science.

In considering how teachers might encourage their pupils to develop higher level thinking, we drew on two sources. The first is Piagetian, and emphasises the role of well managed and carefully graded cognitive conflict. The hypothesis here is that the solving of somewhat difficult problems, with the carefully graded help (mostly through questioning) of a teacher or more able peer, leads not only to a solution of that problem but also to the general stimulation of the pupil's cognitive processing mechanism. A steady diet of such experiences, even if of only moderate intensity, will have a permanent and irreversible effect on the subject's cognitive development. The idea here is that conflict, even if only partially resolved during a particular activity, leads the pupil to construct for herself the type of reasoning required. We have described this (Adey & Shayer, 1994) as 'meta-constructivism': putting pupils in the situation where they must construct for themselves not just knowledge, but ways of thinking about variables and relationships.

The second source is Vygotskyan, and emphasises the importance of reflection, and in particular of social exchange, in the development of thinking as well as the development of knowledge. Pupils who are encouraged in class to talk with the teacher or with each other about how they are tackling and solving a problem, or what difficulties they are finding with it, become more conscious of their own thinking processes and this itself promotes cognitive development. Such metacognition is now widely recognised as an essential element in the development of general intellectual abilities (Brown, 1987; Perkins & Saloman, 1989). The well-managed class discussion provides the opportunity for the social dimension of construction; groups of pupils constructing understanding together by talking about their understandings.

To these two major pillars of cognitive conflict (which leads to construction) and metacognition are added a 'prologue' and 'epilogue'. The prologue is what we call concrete preparation. You cannot simply hit children with problems designed to induce cognitive conflict without some careful groundwork. To be effective, such problems must already have in-built conceptual difficulty (as far as the pupil is concerned) but they are also likely to involve the use of new language and a new practical situation. In order to clear the path for concentration on the conceptual difficulty, concrete preparation is the phase of the activity where new words are introduced and practised, and the context of the problem situation made familiar through demonstration and questioning.

The epilogue phase is called bridging. It is the part of the activity where links are built between the type of reasoning being developed during the Thinking Science activity, and the use of that reasoning in other contexts, in other 'regular' science lessons, in other subject areas, and in the world outside school. Since the early days of learning psychology, it has been known that transfer does not occur automatically. For transfer to occur, the possibilities of applying a particular form of reasoning developed in one context to a new context must specifically be explored with the learner. The relationship between these five key elements in Thinking Science activities is shown in Figure 1.

Before turning to some examples to show how these principles are used in the construction of teaching and learning activities, a word is in order about the use of language within Thinking Science. The psychological interrelationship of language and thought has been the subject of debate - sometimes heated - since the 1930s, and at one time was characterised by the opposition of 'language first' and 'thought first' camps. The former proposed language as the essential ability of human kind which allowed higher level thinking processes to develop. The opposite argument was that our use of language became possible because of the development of higher level cognitive functioning. It seems now that such an argument is fruitless, since

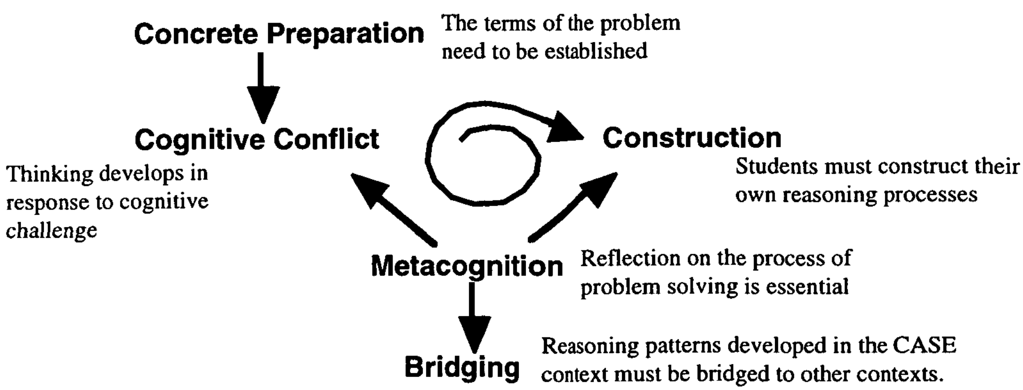


Figure 1: Theoretical structure of Thinking Science Activities

language and thought develop together interdependently, if not in a completely integrated fashion. Words are the tools of thought and it is difficult to imagine the development of higher level thinking without continual recourse to language which employs both a complex structure and an adequate vocabulary. But at the same time, the process of struggling with the use of complex structures and the meanings that may be attributed to new words is a critical part of the development of higher thought processes.

In Thinking Science, words essential to the development of a particular reasoning pattern are systematically introduced and practised in the concrete operation phase. Complex thought cannot be modelled by simplistic language. Teachers who have been brought up on a diet of readability scales and the continual simplification of text to make it 'accessible' sometimes believe that Thinking Science worksheets are going to be too difficult for their pupils. But words such as 'variable', 'relationship', and 'proportionality' are not simply dropped into the text wilfully. They serve an essential purpose in starting students on the way to using them meaningfully and developing the related formal reasoning pattern. What is more they are introduced in concrete contexts, and plenty of practice is given in applying the words in different ways.

To summarise this section, The CASE approach is based on providing concrete preparation for carefully managed cognitive conflict, encouraging metacognitive reflection on children's own problem solving processes, using the language appropriate to the reasoning pattern, and bridging the reasoning being developed in the Thinking Science activity to other science topics and other domains. The reasoning patterns (schema) of formal operations provide the context within which the activities are set and also become the objectives of the course; that is, the type of thinking which is to be constructed by the students.

What do Thinking Science activities look like?

To show how these principles are worked out in practice, three activities will be described in some detail. These are all drawn from the published Thinking Science curriculum materials, available in British (2nd edition), German, and American versions (Adey, Shayer, & Yates, 1995; Adey, Shayer, & Yates, 1993; Adey, Shayer, & Yates, 1992 respectively).

'TS3: Tubes'. This is the third activity in the programme. In the previous activities, the ideas of variable, values of variables, and relationships have been introduced. Students have a box of small tubes. Questioning in a whole class discussion ensures that they identify the variables and values: length of tube (short, medium, long); the width of tube (wide or narrow); and the material of the tube (glass or plastic). Now they are asked to blow across the tubes, and listen to the note produced. The question is this: what affects the note that you get? They have some free exploration time and are asked, if they think they know what affects the note, to explain what they think and why they think it to the teacher or to another student. There is often a need, after some minutes, to call the class together and suggest that they take tubes two at a time.

A child may come up with the claim that the width of tube affects the note. "Show me" says the teacher. The student demonstrates with two tubes of different width that produce different notes. Looking at the tubes, teacher points out that they also have different lengths. "How do you know whether it is the length or the width that affects the note?" Here the teacher is establishing some cognitive conflict, challenging the student to take account of a variable which she had not yet noticed. Typically a child might answer "it is both width and length that affect the note". She does this as it seems a simple way to resolve the conflict, but the teacher perseveres with the questioning, concluding "go and choose another pair of tubes, but this time try to find a pair that will give us a clear answer". Note that the teacher does not direct the student to choose two tubes in which only one variable has altered. The whole point is that the student must construct for herself this control of variables strategy.

In a comprehensive mixed ability class of 12 year olds, it is possible that there will be one or two children who find the whole task so easy that they do not experience much challenge (cognitive conflict). For these the teacher may suggest a higher level task, such as looking for interaction between variables. There may be one or two others who, at the end of the 60 or 70 minute lesson, remain quite confused by the whole exercise and still fail to see the point of controlling variables. The great majority, however, will have experienced (through interaction with the apparatus, worksheet questions, the teacher and with other students) sufficient conflict to have constructed for themselves at least the beginning of a control of variables strategy. The full development of this into an internalised, unconscious, schema which is 'naturally' brought to bear on all experimental situations will still take some time, but essential groundwork has been laid and previous concrete 'change everything and see what happens' schema will have been severely shaken, if not broken up altogether. Even for the least able students who remain confused at the end of the activity there will have been a struggling with the problem and some doubts cast on the ineffective concrete strategy. Even a slight sense of unease at the way in which experimental questions are approached is of value. It is the cognitive struggle which is critical in the promotion of cognitive development, so the ideal objective has been reached if every child experiences some cognitive conflict and goes some way towards finding a resolution satisfactory to her or himself. This is likely to be an end point which differs for each student, according to their ability and their personality.

For a second example, consider 'TS8, The Wheelbarrow', concerned with the reasoning pattern of proportionality. Before this activity, students have explored ideas of scaling by looking at pictures of embryos drawn to different scales, and by using maps of the school environs to estimate real distances. The word 'ratio' has been introduced. The apparatus for TS8 consists of a stick about 8 mm in diameter and just over 60 cm long. It is notched in two places so that a mass hanger and a Newton spring balance can be attached as shown in Figure 2. Illustrations on the worksheet draw clear parallels between this apparatus and the application of lift and load forces in a wheelbarrow.

Students record and tabulate the lift as successive loads are added. With about six pairs of values completed, they use a given grid to draw the straight line graph relating lift to load. From this they are asked to make predictions about what lift force would be needed with extra loads which are not available. The first predictions can be read off the graph by simple extrapolation but then the graph paper runs out and a concrete strategy is no longer available. This is the point of conflict, where a more sophisticated view of the relationship involving the constant ratio of load to effort has to be invented. In the table of lifts and loads, they calculate the ratio for each data pair and discover (sometimes after pooling the whole class's data) that the ratio is constant. This allows them (a) to associate the idea of a constant ratio with a straight line graph, and (b) to use the constant ratio value to calculate what lift would be needed for any new load. They thus have to

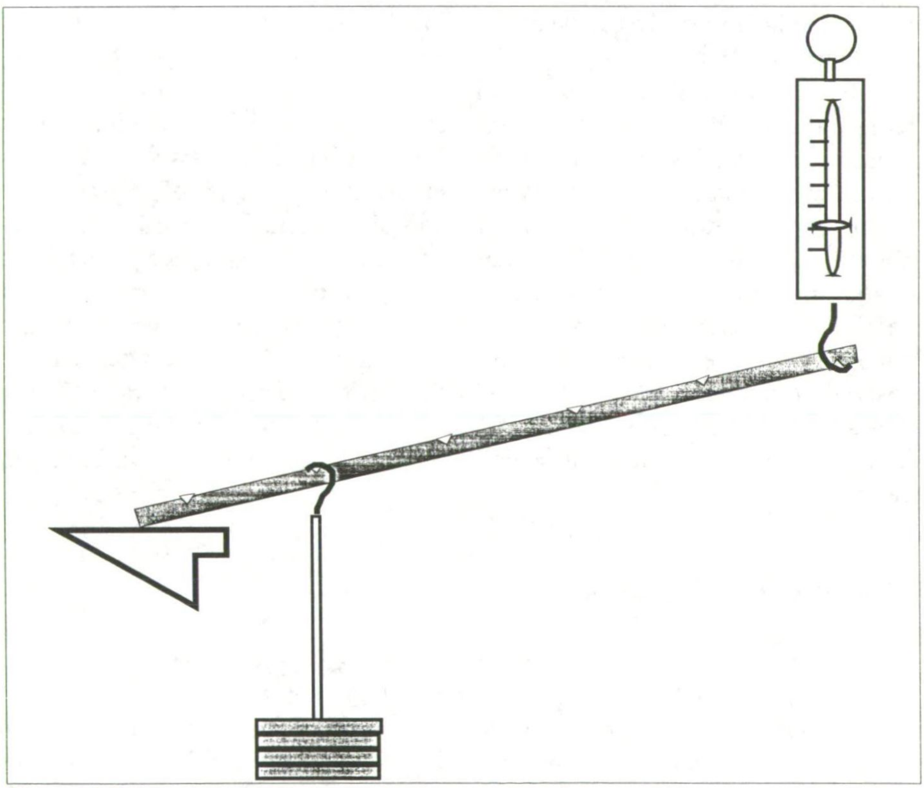


Figure 2 The 'wheelbarrow' apparatus

go beyond the concrete support of the graph and construct a more general formal mathematical model through which they can extrapolate. Cognitive operations on the data must be formalised to achieve a successful solution to the problem.

'TS 18, Treatments and Effects', which is taught in the second year of the programme, provides a final example. This is set in the context of the reasoning pattern of correlation. Each group of students in the class is given a set of 20 cards. Each set of cards shows one organism (rose, wheat, cow, pig, or sheep), and also shows whether or not the organism has received some treatment (e.g. fertiliser, pills to make more milk, etc.,) and whether the animal or plant demonstrates an effect (by growing more, producing more milk, meat, etc.). Students first sort the cards into four piles according to whether they have:

- Not been treated and not shown an effect (A).
- Not been treated but shown the effect anyway (B).
- Been treated but not shown an effect (C).
- Been treated and shown an effect (D).

Students then address the question of whether any effect seen is likely to be the result of the treatment or not. For example, IF the treatment causes the effect, in which of the four piles A, B, C and D would you expect to find large numbers? Discussion in groups leads to the conclusion that you would expect piles A and D to be large, and B and C to be small. The actual numbers in the piles are counted, and the sums of A+D compared with B+C. In the discussion of these results the terms positive correlation, negative correlation, and no correlation are introduced to help students think about what sorts of relationships exist between treatments and effects. This activity models at a simple level the type of experimental evaluation of treatments which is at the heart of much medical, agricultural, and other research. Without an understanding of correlation and associated probabilistic relationships, the majority of popular science reports in newspapers are incomprehensible (witness the recent gross over-reaction by politicians, the media, and the population at large to the relationship between eating beef and Creutzfeld Jakobs Disease). Admittedly this is a rather difficult thinking exercise for average 13 year olds (and even provides a challenge for teachers!) but it does lay the groundwork for important understandings in scientific investigation as well as contributing to general cognitive growth.

Delivery of Thinking Science

Initially our choice of age for the delivery of the Thinking Science programme was determined as much by pragmatic as by psychological reasons. Against a desire to start thinking lessons early were set psychological concerns about the minimum age at which it would be reasonable to expect the promotion of formal operations to be effective, and also the fact that in UK primary schools, teachers are not specialists and might find the Thinking Science requirements daunting. At the time of our original research, these first two years of secondary education were also not overshadowed by any external national assessment, and so schools were somewhat more relaxed about trying an innovation than they would have been in the two years before the General Certificate of Secondary Education examinations. These considerations point to UK years 7 and 8 (US grades 6 and 7, ages about 11-14 years) as the focus age for the intervention. But this also accords well with the age at which the brain seems to be programmed to take advantage of stimulations aimed to promote the development of formal operations (see Adey & Shayer, 1994; Epstein, 1990).

A total of 32 activities make up the complete Thinking Science programme. Taught as a complete course they would provide a very rich diet, and the frequent absence of clear conclusions at the end of each activity might well engender some frustration in both teacher and student. Accordingly, we planned for one TS activity to be done every two weeks, instead of a normal science lesson. With such a programme, it can be seen that the CASE intervention was neither 'infusion' nor 'add-on'. That is, it was neither a set of procedures for the development of thinking which were to be completely integrated with the regular content curriculum, nor was it a completely separate Thinking Skills curriculum. The difficulties with 'add-on' thinking curricula are (a) the problem of persuading school managements to create extra space in already over-crowded timetables, and (b) ensuring that the reasoning developed in the special thinking lessons is adequately bridged to applications within the content areas. On the other hand, whilst 'infusion' methods may represent an ideal, the professional difficulty of maintaining within one lesson both the content and the reasoning objectives is formidable. In fact, it is a specific recommendation to CASE teachers that when they are teaching a Thinking Science lesson they should signal clearly to the students that 'today we are doing something different, we'll get back to the regular science work tomorrow'. As well as contributing to the students' consciousness that their own thinking ability is capable of development, this assists teachers in focusing on the reasoning pattern objectives of the activity, and letting any 'content' objectives wait until the regular part of the science curriculum.

There does remain some concern by science teachers about the time 'lost' from their regular content work to deliver the CASE intervention, which might amount to about 25% of the total time allotted to science in the curriculum. In recent years this has been heightened by the emphasis on examination results and 'league tables' in the UK[1]. Luckily we now have such good evidence for the long term effect of CASE on students' academic achievement, that such fears can often be allayed. Had the national tests which are now given to students at the end of year

[1] The goverment publishes the total examination successes of all schools each year, with the highest scorers at the top of the list

9 (grade 8, age 14) been in existence in the early 1980s when our intervention research was initiated, it is possible that whole CASE project would not have got off the ground. This is a sobering reflection on the short-sighted impact of political initiatives on the educational enterprise.

Evaluation of the effect of CASE on academic achievement

The effect of the CASE intervention on students' cognitive development and academic achievement determined from our original research project has now been widely reported (see for example Adey & Shayer, 1993; Adey & Shayer, 1994). Nevertheless, a summary of that work will be given here before considering more recent evidence.

The 1981-84 experiment

Originally we chose ten schools representing widely different social and geographical environments in England to trial the materials which the authors (Adey, Shayer and Yates) had themselves already taught in two London comprehensive schools. The results which will be described here are for the ten experimental classes (four '11+' starting in year 7 and six '12+' starting in year 8) in seven schools that continued with the programme, more or less as intended, for a period of two years. In each of these schools, one or two classes were designated as experimental, and used the Thinking Science activities as described above once every two weeks for two years. In each school also parallel classes were identified which were matched with the experimental classes for age and ability. The control classes were taught their regular science curriculum without loss of time for the CASE intervention.

All classes, experimental and control, were given a pre-test of cognitive development to act as a base line for measuring any subsequent growth and to make allowance for any initial differences between experimental and control groups. At the end of the two year intervention period, all classes were given post-tests of cognitive development, and also a test of science achievement. This was the end of the intervention programme (and of the funded research), but one year later we revisited the schools to collect information on all of the students' science achievement. One further year later, those classes which had started the CASE intervention in their Year 8 took their General Certificate of Secondary Education (GCSE) examinations, and for all of the students previously designated as experimental and control we collected the grades attained in science, maths, and English. One year on again, those who had started in Year 7 sat their GCSEs and again we collected their grades. We thus had the data which allowed us to compare (a) cognitive growth and (b) academic achievement over a long period of initially matched students some of whom had experienced the CASE intervention and some of whom had simply followed their regular science courses.

In order to allow for individual differences in starting cognitive levels, all data was processed by (a) finding the regression coefficient for each post measurement on pre-cognitive measures for the control groups; (b) using these regression coefficients to predict the value of the post-measures for each experimental child as if s/he was no different from a control child; (c) subtracting the predicted post-measure from the actual post-measure obtained. This difference is the residualised gain score (r.g. score). For any group of students the mean r.g. score is a measure of the extent to which their development or learning has been different from the initially matched control group. This method of analysis takes account of the different measures of academic achievement used in different schools, since each 'experimental' child is compared to norms established by 'control' children from the same school, who have had the same experiences, apart from exposure to Thinking Science. The method is heavy on data-processing requirements but offers a powerful way of making the best use of all data available.

For convenience of comparisons, all results will be reported in terms of r.g. scores. Note that r.g. scores build in comparison to controls and that by definition the mean r.g. score of a control group must be zero. Results for four groups will be considered: boys who started the intervention at the beginning of Year 7 ('11+ boys'), boys who started the intervention at the beginning of Year 8 ('12+ boys'), and the corresponding girls' groups. Table 1 summarises for each group the number of students, mean r.g. score, standard deviation, and (where significant) the significance level and effect size in standard deviation units, for the immediate post test of cognitive development and then delayed science achievement and GCSE grades obtained up to three years after the end of the intervention.

Attention should be drawn to a number of features of these results, some of which are obvious and some of which are not clear from the raw figures:

- The immediate effects seem to be rather limited, but (a) more recent immediate effects obtained on cognitive development have been much larger (see below) and (b) there is a strong correlation on an individual student basis between cognitive gains over the two year intervention programme and subsequent gains in GCSE scores.
- In spite of the moderate immediate effects, there is a long term, and apparently growing, effect of the intervention of students' academic achievement. In principle this is what might be expected from an intervention programme which raises students' general thinking capability. The effect of the raised cognitive levels will be, starting at the end of the intervention, to improve student's ability to benefit from normal classroom instruction. Such improvement is likely to be cumulative as better understood conceptual learning provides a sounder platform for further learning, and so on.
- There seems to be an age/gender interaction effect, in that the intervention is most effective with younger girls and with older boys. Although this notion fits neatly with a model of a cognitive window of opportunity for the promotion of

Table 1 Residualised Gain Scores on Successive Tests after Completion of Two Year CASE Intervention, Based on Pre-Cognitive Tests, September 1984

	Group	Number	Mean gain	Standard deviation	Signifi-cance, p<	Effect size (s.d.)
Immediate post cognitive test July 1987	11+ boys	29	-0.21	0.95	-	-
	11+ girls	27	0.08	1.10	-	-
	12+ boys	65	0.70	1.00	.001	0.75
	12+ girls	52	0.03	0.98	-	-
1 year delayed science achievement July 1988	11+ boys	37	2.72	15.45	-	-
	11+ girls	31	7.02	12.76	.025	0.60
	12+ boys	41	10.46	16.60	.005	0.72
	12+ girls	36	4.18	14.41	-	-
GCSE 1989 Science	12+ boys	48	1.03	1.34	.005	0.96
	12+ girls	45	0.19	1.38	-	-
Maths	12+ boys	56	0.55	1.23	.005	0.50
	12+ girls	54	0.14	1.27	-	-
English	12+ boys	56	0.38	1.27	.05	0.32
	12+ girls	57	0.41	0.96	.01	0.44
CGSE 1990 Science	11+ boys	35	-0.23	1.46	-	-
	11+ girls	29	0.67	1.36	.025	0.67
Maths	11+ boys	33	-0.21	1.59	-	-
	11+ girls	29	0.94	1.26	.005	0.72
English	11+ boys	36	0.26	1.65	-	-
	11+ girls	27	0.74	1.32	.025	0.69

formal operations, which in line with their generally earlier maturity at this age comes earlier for girls than it does for boys, we must be very careful before drawing such a conclusion. For one thing, the 11+ group was actually more able overall than the 12+ group, both age groups starting the intervention at about the same mean level of cognitive development. For another, more recent data does not show anything like the same gender effect.

- The distribution of gains within any group (not shown here, but see Adey & Shayer, 1994) is often bimodal. That is, some of the students make very large gains, around two standard deviations, whilst other make gains little more than the controls. We do not know why this is, but it may possibly be due to the 'fit' of the Thinking Science methods with different learning styles.

More recent results

Results reported in the last section were from the original research experiment, in which we were able to measure effects on particular experimental classes against the results of well-matched control classes in the same schools, with the same teachers. The disadvantages, however, were that the numbers were relatively small because we were only able to collect data from one or two classes in each school, we ourselves were still in the process of inventing the method for training the teachers, and the teachers themselves were working on the project in isolation within their schools.

Following the publication of the long-term effects on GCSE scores in May 1991, there was a great demand from schools for the materials and methods that would enable them to replicate the results. Since then, we have been running a series of two-year in service teacher education courses to introduce the methods. Although we are now collecting much new data, an important difference between this and the original experiment is that now we have a method which we believe works, we cannot ethically deny it to any class just to provide an experimental control. One way of analysing new data is to compare gains made by CASE schools with the national norms established in a large-scale survey of the school population of England and Wales by the Concepts in Secondary Mathematics and Science project (CSMS) in the 1970s (Shayer, Küchemann, & Wylam, 1976; Shayer & Wylam, 1978). The national data provides a control, in the sense of an expectation of 'normal' development against which the development of CASE pupils can be compared. From the first cohort of schools participating in the CASE training programme, we were able to collect pre- and post-test data on levels of cognitive development for 63 classes in 8 schools. Some of these classes made a Year 7 (11+) start on the intervention, some a Year 8 (12+) start, and one school started the intervention in both years. A summary of the effect sizes of the school mean residualised gain scores compared with national norms is shown in Table 2.

We have actually studied the effect size obtained in each of the 63 classes. In one class, there was a significant negative effect, possibly due to some error in the

Table 2 Effect Sizes of Cognitive Development Residualised Gain Scores in 8 Schools which Participated in CASE Training, 1991-93

School	Start age	Effect size (σ units)
1	11+	0.67
1	12+	0.76
2	11+	0.69
3	11+	1.12
4	11+	1.12
5	12+	0.80
6	11+	1.00
7	12+	0.29
8	12+	1.26[1]

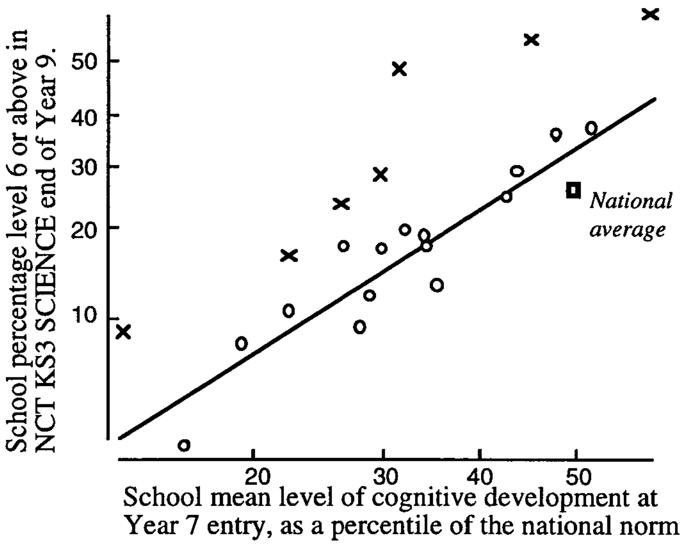
[1] by comparison with previous year 9 group, questionable

administration of the pre-test. In four others there were insignificant negative effects. In three classes there were positive effects of less than 0.3σ. In all of the remaining 58 classes there were significant positive effects of the CASE intervention on children's rate of cognitive development. As we have shown previously, cognitive gains attained over the intervention period are related to subsequent academic gains.

More recently (April 1996) we have been able to collect data on academic achievement of CASE schools, compared with non-CASE schools, for the 'Key Stage 3 National Curriculum Test' (KS3 NCT). In the UK, the government has instituted a series of nationally moderated tests to be given in various subject areas at the end of each 'Key Stage' of education, which means at the ends of years 2, 6, and 9 when children are about 7, 11, and 14 years old respectively. For schools which use Thinking Science in years 7 and 8, the KS3 NCT given at the end of year 9 provides a convenient measure of academic achievement one year after the end of the intervention.

In Figures 3a, 3b, and 3c each point represents one school. The x-axis is the mean score of the school's students at the beginning of Year 7 on measures of levels of cognitive development, expressed as a percentile of the national average. This is a measure of the school's intake ability, which is a reflection of factors such as the socio-economic conditions in the school's environs and whether there are selective schools in the area which cream off the more able students. It so happens

3a: Science



3b: Mathematics

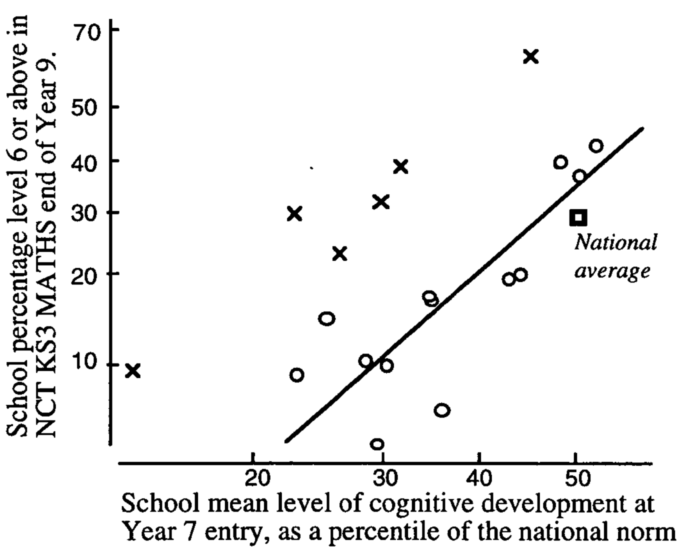


Figure 3 Relationship between school entry cognitive levels and NCT KS3 test performance at end of year 9 for CASE (X) and non-CASE (O) schools

that almost all of the schools for which we have data at present are in the lower half of the intake ability range. The y axis is a measure of success in the KS3 NCT. These tests are scored for National Curriculum levels, which fall on a range from 1 to 10 (or more recently 1 to '8 and over'). The percentage of students attaining level 6 and above at Key Stage 3 is commonly taken as a measure of the success of the school. In order to make the plot linear, all scores have been transformed into logits: ln(%/100-%). This is why the axis scales are not equal-interval.

In each figure, the regression line has been drawn on the basis of the control (non-CASE) schools only. It shows, not surprisingly, that success on the KS3 tests is closely related to the intake ability of the school's pupils. This is not surprising, since success in academic tests must be partly a reflection of general ability, and where teaching methods and expectations are unexceptional, there is bound to be a direct relationship between a school's mean intake level and their mean examination scores. What is striking is that for all of the data we have so far, CASE schools fall above - often far above - the regression line for control schools. Even for English, although points are more widely distributed because of the lower reliability of the assessments, every CASE school falls above the mean for non-CASE schools. The effect is generally equivalent to an addition of about 30 percentile points to school mean academic achievements. It is very clear that the CASE intervention has systematically added greater academic value to students of a given starting cognitive level than is normal for non-CASE schools, and that the effect is on a general function of students which transfers far beyond the science context in which the cognitive intervention programme is delivered. This claim is based especially on the effects shown on the far transfer to English results. These cannot be explained in terms of direct training in thinking set within a particular context, but must arise from effects on general processing mechanisms which can be applied, consciously or unconsciously, by the students in all of their learning.

There are good reasons to suppose that this increased effect should have been expected. In the implementation, unlike in the original experiment, all science teachers in a school are involved in teaching Thinking Science. It becomes part of the culture of the science department and there is much opportunity for mutual support amongst teachers. At the same time, in the implementation we have been able to concentrate more fully on developing our inservice professional development methods.

Conclusions and discussion

I do not believe that Thinking Science is by any means the only way to influence children's general cognitive capability. There are many other programmes which offer equally exciting activities and which seem to have the potential for producing long-term far-transfer effects. Why, then, are so few thinking skill programmes able to offer the kind of evidence for their effect that has been summarised here?

Here are some possible reasons:

- Time. It is in the nature of evaluating general effects that there must be a long term study. Evaluation studies which last, as ours has, for 15 years already are extremely difficult to staff and to fund. If one relies only on graduate students to do your research, or are bound to research grants which are never longer than three years in duration, it is going to be extremely difficult to get good evidence for long-term general effects even if they are there.
- Models. One needs a comprehensive model of the nature of learning and of development which can provide a consistent basis for the design of both activities and assessments. The Piagetian model of cognitive development through maturation and equilibration has recently been through a phase of being seen as unfashionable and the author has had experience of academic referees assuming that work based on Piaget must necessarily be ill-conceived. Nevertheless, the Piagetian model has provided us with just the sort of consistent, comprehensive model which is required. Of course, there are alternative explanations for the effect that CASE is achieving, but unless one has a theoretically derived hypothesis, there is no way that one can test alternatives in an attempt to improve further the effect.
- Generality. Until very recently, educational research has been heavily influenced by the notion that all cognition is 'situated'. The claim is that it is impossible to talk of intelligence in a general way, but only of the intelligent behaviour which a particular individual displays in a particular situation. In this 'Zeitgeist', there is little motivation to look for ways of increasing children's general intellectual power. This situated cognition position is based in large part on the misuse and slanted interpretation of some good research (Adey, 1997). I believe that data of the kind presented in this chapter should help to restore our faith in the ideas that there are general factors which influence intellectual performance, and that these factors are amenable to educational influence.
- Quantitative methods. More specifically, there has been some loss of faith in quantitative methods following a rash of inappropriate behaviourist research in 1950s and 60s which paid too little attention to nature of constructs being investigated or to the quality of interactions between people. But the pendulum has now swung too far in the opposite direction, and it is a sobering experience to search through the programme of, say, the American Educational Research Association annual meeting, looking for the few nuggets of quantitative studies which have been allowed to slip through. We do need to know much more about the quality of thinking skill programmes, but it is also essential to get reliable measures of their effects on learning. The plea with which I conclude this chapter, then, is for more long-term, quantitative, and theoretically-based investigations of programmes which display faith in the possibility of raising general academic achievement through some rather specific modifications in teaching methods and materials.

Acknowledgement

Professor Michael Shayer was the Director of the original CASE projects, and he has also provided all of the most recent data processing results reported here.

References

Adey, P.S. (1997). It all depends on context, doesn't it? Searching for general, educable dragons. *Studies in Science Education, 29,* 45-92.

Adey, P.S., & Shayer, M. (1993). An exploration of long-term far-transfer effects following an extended intervention programme in the high school science curriculum. *Cognition and Instruction, 11,* 1 - 29.

Adey, P.S., & Shayer, M. (1994). *Really raising standards: Cognitive intervention and academic achievement.* London: Routledge.

Adey, P.S., Shayer, M., & Yates, C. (1992). *Thinking Science: U.S. edition.* Philadelphia: Research for Better Schools.

Adey, P.S., Shayer, M., & Yates, C. (1993). *Naturwissenschaftlich denken* [Thinking Science]. Aachen: Aachener Beiträge zur Pädagogik.

Adey, P.S., Shayer, M., & Yates, C. (1995). *Thinking Science: The curriculum materials of the CASE project* (2nd ed.). London: Thomas Nelson and Sons.

Brown, A.L. (1987). Metacognition, executive control, self-regulation and other more mysterious mechanisms. In R. Kluwe & F. Weinert (Eds.), *Metacognition, motivation and understanding.* London: Lawrence Erlbaum.

Epstein, H.T. (1990). Stages in human mental growth. *Journal of Educational Psychology, 82,* 876-880.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking.* London: Routledge.

Perkins, D.N., & Saloman, G. (1989). Are cognitive skills context bound? *Educational Researcher, 18,* 16-25.

Resnick, L.B. (1987). *Education and learning to think.* Washington, DC: National Academy Press.

Shayer, M., Küchemann, D.E., & Wylam, H. (1976). The distribution of Piagetian stages of thinking in British middle and secondary school children. *British Journal of Educational Psychology, 46,* 164-173.

Shayer, M., & Wylam, H. (1978). The distribution of Piagetian stages of thinking in British middle and secondary school children. *British Journal of Educational Psychology, 48,* 62-70.

4

Applying the theory of knowledge to teaching thinking

P.M. Scheinin & J. Mehtäläinen

Introduction

Anyone involved in teaching thinking is faced with an obvious question. We need a reason why thinking should be taught in the first place. In a complex society and culture, a major task of schooling is to transfer accumulated knowledge and values. The transfer of knowledge is, however, not a passive process and knowing is not, in itself, enough. The learner needs to build personal frameworks of interpretation, modes of understanding, and strategies of problem-solving. Nisbet (1993, p. 282) points out that "there are skills and strategies in thinking which we build from experience; the mastery of these skills need not be left to chance" and that "appropriate teaching can help all of us improve our competence."

Accepting this, we still need to find out what this appropriate teaching is. In all likelihood, it will depend on which skills are to be taught, to which area of knowledge these skills are to be applied, and at which age the skills are to be taught. The effectiveness of a programme will depend on the approach and methods chosen. The aim of this chapter is to present one such programme: The Formal Aims of Cognitive Education (FACE) project. We will introduce the theoretical framework of the FACE programme, describe its implementation, and report and discuss our findings.

The FACE project (Voutilainen, 1991) was a school intervention. It was set up to try out one approach to teaching thinking, namely, one based on the philosophical theory of knowledge, i.e., epistemology, and to evaluate the outcome of the intervention. In the FACE project, a systematic attempt was made to integrate the teaching of skilful thinking into most subjects of the curriculum. That is, the teachers tried to teach thinking skills by using the subject matter to demonstrate and exercise thinking skills in practice, while teaching the structure and content of subject areas more or less as normal.

In the second section of this chapter, we briefly describe different approaches to teaching thinking. An overview is given of the effects of gender and development on cognitive abilities. We also discuss some aspects of the theory of knowledge and more specifically the formal aims of cognitive education and their implementation in the classroom. In the third section, the practical aspects of the intervention are presented. The fourth section deals with the research methods, design and procedure of the study. The results of the evaluation process and the effects of the FACE project are presented in fifth section. We describe what the teachers thought about the project and what happened in the classrooms. The focus is on the effects the intervention. The sixth section of this chapter contains conclusions and discussion.

General theoretical orientation

Teaching thinking

Previous research findings suggest that no simple answers can be given to the question of whether cognitive abilities can be improved through specific interventions. By cognitive ability we mean any ability that concerns some class of cognitive tasks. These are tasks in which correct or appropriate processing of mental information is critical to successful performance. The evidence suggests (Carroll, 1993) that general intelligence (i.e., a cognitive ability broad enough to influence most, if not all, cognitive tasks) and other major cognitive abilities may be less malleable than more specific abilities. Knowledge and skill are, on the other hand, trainable, more or less by definition.

A number of studies have independently provided evidence of positive results of various interventions in the domains of intelligence, skilful thinking or metacognition (see Coles, 1993; Nisbet, 1991, 1993). Powerful and replicated transfer and long term effects have also been reported (Adey & Shayer, 1995). Taken together, the evidence suggests that well-planned, intensive, and preferably long-term interventions do have positive effects on the development and application of cognitive abilities. Some of these effects would seem to be general in nature while most can be interpreted as improvement in less general, domain-specific abilities or skills such as those described by Gardner (1983). In any event, it is safe to assume that if quick or easy results were possible, we would know of them by now (Nisbet, 1993).

Programmes intended to improve thinking mainly use one of three types of interventions. One type of intervention has been in the form of separate and relatively short thinking skills courses. In the second type of intervention, a new subject, in which thinking skills are taught, is added to the curriculum. The third type of intervention is based on the assumption that thinking should be taught as part of all subjects in the curriculum, that is, it should be infused throughout the curriculum. An effort may also be made to teach thinking as a part of one or a few school subjects, usually the sciences, reading or composition, or mathematics. However, the teaching of thinking has only seldom been integrated into most subjects in the curriculum, and very rarely throughout all the subjects of the curriculum (Csapó, 1990, 1992; Nisbet, 1993).

Most projects involved in teaching thinking take a psychological approach. A distinction can be made between the information-processing view of thinking, mostly associated with cognitive psychology, and the perhaps more constructivist tradition of seeing thinking as sense-making, mostly associated with the Piagetian or neo-Piagetian frameworks (McGuinness, 1993). Interventions with a generally philosophical approach are far less common (Nisbet, 1991). Probably the best known example of these is Philosophy for Children (Lipman, 1985). Interventions within a philosophical framework generally view thinking as a process of making judgments. Such a viewpoint is part of what has been termed the critical thinking movement (McGuinness, 1993).

Distinctions and categorisations of programmes for teaching thinking skills are far from mutually exclusive. Typically, the various programmes for teaching thinking show characteristics of one or more of the approaches just described. For further examples and discussion, see Chance (1986), Segal, Chipman, and Glaser (1985), Fisher (1991), as well as Hamers and Overtoom (1997).

Evaluating the success of interventions designed to enhance thinking is a complicated matter. The classical approach has been to use tests of various dimensions of cognitive ability and thus to try to measure the effects of the intervention. A promising trend (Nisbet, 1993), is to take into account affective aspects as well. The students' attitudes and their perceptions of their own competencies become important. Coles (1993) has expressly pointed out that one of the factors to be accounted for in evaluating the effects of programmes is the self-concept of the pupil.

One factor arguing for the inclusion of tests of self-concept in the evaluation of programmes for teaching thinking is the consistent finding of a strong correlation between self-concept and performance (Hansford & Hattie, 1982; Marsh, Walker, & Debus, 1992). There is some evidence that points to direct causal effects (Marsh, 1990). The most likely interpretation of this correlation is that there is a dynamic process of interaction between, for instance, school achievement and academic self-concept. Performance would be guided by self-concept and the feedback from the performance would in turn shape the future self-concept. In this way, it is possible that the effects of a programme on future performance are caused partly or even wholly by changes in the self-concept of the participants.

It is self-evident that in many respects we are all different as thinkers. When two people think about the concept of 'democracy', they have different associations and feelings. They may also differ in their cognitive abilities, and in structures of knowledge, attitudes, and values. There may also be differences in their orientation towards the activity called thinking. When trying to improve thinking in pupils, the ideal would be to take all individual differences into consideration. Thus, a combination of different approaches to teaching thinking may well produce the best results. It is, however, probably neither necessary nor possible to test all effects simultaneously. The optimal strategy in research would seem to be to try out various methods and approaches one by one. This implies testing different approaches to teaching thinking separately, in a broad range of programmes, and then meta-analytically summing up the findings of this broad field of research. The findings of the present study should be seen as one brick in a large and rapidly growing structure.

Gender, development, and cognitive abilities

The focus of this study was on the effects of education on cognitive abilities. Some attention needs, however, to be given to the effects of gender and development. This is necessary so as not to confuse the effects of these factors with the effects of the intervention. Even relatively small differences may be confused with the effects of the intervention when the comparison groups have different proportions of male and female participants or when the intervention is long enough for cognitive development to take place.

Gender differences in cognitive abilities have been studied extensively (Guilford, 1971). Though recent reports seem to indicate that, over the past few decades, certain differences related to gender are steadily diminishing, research shows that there are systematic differences between the sexes in various areas of cognitive ability. There are systematic gender differences favouring males in mathematical reasoning, and in spatial and mechanical reasoning, whereas females tend to do better than males on verbally oriented achievement tests (Lim, 1994; Lubinski, Benbow, & Sanders, 1993; Rosen, 1995). Similarly, several studies (Marsh, 1994; Marsh, Barnes, Cairns, & Tidman, 1984; Marsh, Chessor, Craven, & Roche, 1995) have shown gender-specific differences in academic self-concept. Girls have a more positive self-concept than boys in language skills but have a more negative view of their abilities in math.

Human development is generally regarded as a process of interaction between individual characteristics and environmental opportunities. Many different types of developmental theories exist, but they differ in approach or in how they explain development (Anderson, 1992; Guilford, 1971; Mönks & Mason, 1993). For the purposes of this study, it is enough to point out that, at present, the theories seem to agree that cognitive abilities do indeed increase with development.

The theory of knowledge

The FACE project was based on epistemology. Epistemology is the philosophical theory of knowledge. It seeks to define knowledge, distinguish its principal varieties, identify its sources, and establish its limits (Toulmin, 1972). To understand what epistemology has to offer in the classroom we need to identify some basic elements of thinking as a process. These are assumed to be essential to productive thinking:

- Something that requires thinking is needed. For example, this may be a question, a problem given, or a problem to be found. All kinds of questions or problems do not necessarily require much thinking. Questions such as: "What is the main export article of Brazil?", "Why is Brazil a remarkable producer of coffee?", or "What happens to the production of coffee if the earth stops going around the sun?" are quite different in this sense.
- Knowledge, or content, is needed to provide answers. Thinking is always thinking about something. The food of thought is knowledge, which is to be analysed, categorised, evaluated, combined in synthesis, or otherwise transformed.
- In thinking, the pursuit and handling of knowledge has a formal side to it (Voutilainen, 1991; Voutilainen, Mehtäläinen & Niiniluoto, 1990). Here the term 'formal' refers to the form of knowledge. The emphasis is on general rules or overall patterns and principles rather than on the content of knowledge. Piagetian formal operations are not (directly) intended. The formal dimensions of the thinking process are seen to involve general aspects of the use of concepts (concept formation and analysis), reasoning and drawing conclusions (inductive and deductive), explanation (logical, causal, and teleological), as well as establishing the limits of knowledge (criteria of truth, belief vs. truth, logical vs. empirical). The finding and construction of wholes or entities of knowledge may also be seen as a formal aspect of knowledge. Skilful thinking involves knowledge of the formal aspects of thinking and skilful application of this knowledge. These formal aspects fall within the domain of the philosophical theory of knowledge, i.e., epistemology.
- Orientation towards thinking as a purposeful action involves at least three different aspects. First, a will to think and perseverance to come to a conclusion are required, followed by a willingness and ability to imagine, seek alternative premises, and perhaps unfeasible possibilities. This can be seen as creative thinking or mental curiosity. Finally, mental responsibility, often called critical thinking, is needed. This simply means the conscious pursuit of truth (applying the criteria of truth). We need to analyse, for example, our own reasoning and that of others, the concepts we have used, the structures or entities of knowledge we use, as well as our attitudes and values.

None of the previous elements of productive thinking are new in themselves. Not all of them have equal standing in the curriculum, however. While teachers are familiar with the content matter of the subjects they teach, few have a similar familiarity with the formal or epistemological aspects of the same knowledge area. Teachers typically attempt to teach knowledge. Teaching the pupils systematically to understand the formal aspects of knowledge is rare indeed. However, some of the formal aspects of acquiring and handling knowledge are common to most or all branches of knowledge. The formal aspects of knowledge may thus be used as common denominators in integrating different subjects in school.

Concepts, for instance, may serve as useful or as misleading tools of thought, whether the subject is mathematics, language or religion. A concept is not the same thing as the name or symbol that is used to represent it. Neither is it just the sum of the entities belonging to it. In languages we encounter various synonyms. These are words representing the same general content. The meaning or idea behind the content that the synonyms stand for is the concept. In mathematics, the idea of a symbol representing nothing (0) made counting and calculation much more efficient. The spreading of this concept had a vast influence on modern culture. Similarly, we often operate with cloudy concepts or concepts that are based on misconceptions. These hinder our thinking and actions just as the pancake model of the earth hindered exploration.

There are also systematic differences in the methods of fact-finding and the criteria of truth typically used in various subjects, such as, history, physics or biology. The deceptively simple form of questions beginning with "Why...?" hides subject-specific assumptions regarding the type and level of explanation as well as of what constitutes acceptable proof. Usually, the formal similarities and differences between areas of knowledge are not brought to the attention of the pupils.

Throughout the ages, skilled teachers have taught their pupils how to think. The effort has mostly not been planned for and it has seldom been coordinated throughout the curriculum. Yet, it is possible to envision most or all teachers in a school teaching a broad range of school subjects as usual, but bringing a second 'formal' theme of thinking skills into their instruction. The content matter would change from lesson to lesson, but certain aspects of teaching thinking would be common to all subjects and would, thus, become an integrating element throughout the curriculum. This would constitute an effort to systematically and consciously guide and develop the thinking of the pupils.

The formal aims of cognitive education

Most of the formal aspects of knowledge that can be emphasised and demonstrated in school instruction across a range of subjects can be identified as related to one of the formal aspects of knowledge defined above: concept formation and analysis, reasoning, determining appropriate level and type of explanation, and determining the criteria of truth. They include widely accepted basic knowledge of the set theory (as seen in the relationship between a concept, its definition, and the entities belonging to the set that is the extension of the concept), logic, and different forms of explanation.

The different skills, habits and attitudes associated with the formal aspects of productive thinking may be expressed as a set of instructional aims. The formal aims developed by Voutilainen (Voutilainen et al., 1990) have been presented as specific goal statements. These aims can be used as a basis for integrating the teaching of thinking into the curricula. Some examples of the formal aims of cognitive education are presented below.

Concept formation and analysis:
- One should find out the qualities and relations which determine whether or not an entity belongs to the extension of a certain concept.
- One should compare the entities that belong to the extension of the concept with regard to both the qualities that are included in the intension of the concept and those that are not.
- One should find out which broader concepts a concept belongs to and which subconcepts are associated with it.

Reasoning:
- One should learn to distinguish premises and conclusions from each other and to see when the inference is logically binding.
- One should comprehend that if the premises are accepted as true, then a logically binding inference must also be accepted as true.
- One should master the nature of inductive reasoning and understand that inductive generalisations broaden our knowledge of reality beyond observable things and events.
- One should develop a readiness to reject an unreliable generalisation.
- One should accept and learn to use imagination as a tool for thought experiments.

Explanation:
- One should know how to find out whether or not the phenomenon to be explained is an event in reality.
- One should be able to indicate the possible causal relation between two or more events that fulfils both the material and formal prerequisites.
- One should comprehend that explaining a causal relation in a logically binding way presupposes a causal law.
- One should be able to find out whether or not the event to be explained is an act.
- One should be able to indicate the interdependence between the act, the agent and the purpose.
- One should comprehend that an act that has taken place in a different way than was originally planned can be explained if we know the purpose of the agent

and the knowledge and options at his or her disposal.
- One should understand that setting a purpose for an act is always also in itself an act whose explanation can be the subject of enquiry.

Truth:
- One should learn to distinguish propositions that can be either true or false from propositions that can be given no truth-value.
- One should learn to know the criteria for logical and empirical truth.
- One should learn to assess the human factors that tend to make it more difficult to find out the truth.
- One should learn to distinguish an opinion from a statement of fact.
- One should learn to be critical in a constructive way.
- One should learn to use the idea of truthfulness in relation to all knowledge and not just as the opposite of a lie.

Entities of knowledge:
- One should understand that a concept is a whole, that it is formed by separate sub-entities or groups of such, and that the parts of a whole are in a certain relation to each other and to the whole.
- One should learn to look for the systemic or the general level of knowledge that makes a set of beings or occurrences a whole.
- One should see that grasping certain specific and general concepts and their relations in a system is the basis for reasoning, explanation, or the evaluation of veracity.

Applying the formal aims

Let us now take an example from biology to demonstrate the formal principles involved in everyday instruction. We begin with a statement claiming something: 'A hare has protective colouring.' Stripped of its content, this proposition can be formulated as a formal sentence in the following ways: 'X has a trait called A' or 'X belongs to the extension of concept A'. The content of the proposition consists of two concepts representing reality: 'a hare' and 'protective colouring'. The formal aspect involves examining the characteristics and relations which are decisive in determining whether a certain being belongs to the extension of a concept. Propositions can be taught and learned as sentences to be stored in memory. It is, however, also possible to use concept formation to reach a deeper understanding of the phenomenon itself while using the phenomenon as an example to learn something about concepts, or thinking, in general.

To understand the meaning of the concept 'protective colouring', the teacher and pupils examine other creatures that have protective colouring. It should be noted that the aim is not to formulate a list of creatures that have protective colouring to be remembered, but to define the connotation or intension of the concept 'protective colouring' in order to find its extension, that is, its range of use in our thinking. Of course, this does not suffice for a complete understanding of the essence of this concept. To more fully understand it, we have to proceed to an analysis of the concept of 'evolution', the law of natural selection and its basic concepts: heredity, selection, and variation. With what special tools of survival has nature provided mankind? The mechanisms of evolution enable the teacher to make the pupils familiar with some formal aspects of causal explanation. Finally, the question of a hare and its protective colouring may lead the teacher and the pupils to think about the balance between man and nature. To what extent has man disturbed the balance of this system? What were the intentions, and what are the possible consequences? At this phase, the teacher and pupils are analysing the relation between the system, the act, the agent, and his/her purposes, using causal and teleological explanation. Thinking, at least in this case, means both the construction of knowledge (conceptualisation) and the use of knowledge in a new context related to larger wholes or entities. The formal and content aspects of knowledge are, thus, tightly interconnected in instruction.

The programme

In the FACE programme (Formal Aims of Cognitive Education), the formal qualities of concept formation, reasoning, explanation, and truth were demonstrated during the process of learning and instruction. The teachers did not teach these qualities as such, but integrated them into the subject matter. The formal qualities of knowledge directed the thinking of the teacher in the teaching process.

In the early stages of the programme attempts were made to follow the formal aims of cognitive education in most subjects. There were, of course, differences in the applicability and emphasis of the various formal principles in the different school subjects but, as a whole, the system of formal aims was the same for all teachers and throughout the curriculum.

A premise of the programme was that the teachers were experts in their own subject matter and in how to teach it. Therefore, they had to master the formal aims and the rationale for these aims so well that they were able to apply them autonomously. The aims to be applied in each situation remained at the discretion of the teacher. Similarly, the teacher decided which methods to use. The teachers were instructed to plan in advance the goals and methods to be applied, but the final decisions were made in the actual teaching situation.

The original idea was that the teachers would use forms of questions that corresponded to the formal aims. This would mean actual questions with appropriate content. First, teachers would put the questions to the pupils. Then the pupils would gradually be trained to ask such questions of themselves, the teachers, and each other. It was assumed that the pupils would develop a basic set of questions to help them carry out such thinking tasks as analysing concepts, examining the relation between arguments and conclusions, and identifying causal relations, among others. Quite soon it became obvious that this idea would not work. The teachers became too sensitive to some aspects of their teaching asking questions about their work such as: "Is this really an appropriate question?", "Is this question formally correct?", or "What else do I need to ask?".

The teachers were then instructed to emphasise the planning of teaching. Using formal aims and forms of questions, they tried to perceive their subject matter in a new way. They tried to find and construct entities or structures of knowledge that are essential to understanding and learning the subject matter to be taught. These structures of knowledge could be elements such as concepts, events, regularities in systems, or others. Working in this way, the teachers established a basic certainty regarding the formal principles and their application. And, as the teachers had been viewing the contents of their subject matter from a 'formal perspective', they could now feel more relaxed in applying them when they taught. They felt they had more room to improvise.

The intervention was begun in the autumn of 1990. The intervention was implemented in one class of 18 pupils in one junior secondary school. The school in question is the Alppila Junior Secondary School in central Helsinki. The areas from which the pupils came are relatively mixed, socially and economically. It could be argued that the school reflects the country as a whole better than schools in more homogeneous environments. Six of the teachers working with this group were actively involved in the intervention. They taught the following subjects: Finnish language, mathematics, physics/chemistry, geography/biology, history, and religion.

Depending on the content of the subjects to be taught, the pupils received up to 15 hours of instruction based on the formal aims of cognitive education a week. The average amount of FACE-based teaching was approximately four to eight hours a week. The intervention continued for three years (from grade 7 to grade 9, that is, the duration of the junior secondary school) and it ended in the spring of 1993. Many aspects of the FACE project have since been integrated into the curriculum of the school involved.

Method

The purpose of the FACE project was to apply the formal aims of cognitive education as an integrated part of the curriculum, to study the intervention process in practice with a focus on the experiences of the teachers, and to also evaluate the outcome of the project in terms of effects on the pupils' thinking. Thus, an evaluation of the implementation process and a study of its effects were required. The research questions of this study were the following.

Process evaluation:

- Were the formal aims of cognitive education applied during the intervention?

- How did the teachers evaluate the programme?

Effect study:

- Did the pupils' way of thinking (cognitive ability and formal thinking skills) change as a result of the intervention?
- Did the pupils think differently of themselves (cognitive self-concept and self-esteem) as a result of the intervention?

Design

The design of the study can be described as a quasi-experimental combination of a follow-up and a cohort study (Campbell & Stanley, 1973). The intervention effects were evaluated by comparing the pretest and posttest performance of the pupils in the intervention with the performance of pupils in same age-cohort control groups (see Figure 1).

The experimental group (6 girls and 13 boys) was one class. There were two control groups. The pupils in the control groups were from the same school as the experimental group. Both control groups were tested at the time of the pretest, which was administer at the end of the first term of the intervention. Thus, all groups were tested at the same time of the year. This was done to prevent maturation effects from playing a part in the results. As will be remembered age-typical cognitive development was to be expected. The pretest results (in grade 7, 13 to 14 years of age) of the experimental group were compared with those of a control group (35 girls and 44 boys) of the same age-cohort and the posttest results (in grade 9, 15 to 16 years of age) were similarly compared with those of a control group (49 girls and 40 boys) of that age-cohort.

An alternative to the design used would have been to have pupils from several schools in the experimental group and in the control groups. This could not be managed, as the cost of teacher training would have become prohibitive. Neither was it realistic to posttest the control group at the same time as the experimental group was posttested. This was due to the fact that it was not possible to keep the

	Grade		
Group	7		9
Experimental	O_{90}	X	O_{93}
Control 1	O_{90}	-	-
Control 2	-	-	O_{90}

Figure 1 The research design of the study, O_{90} denotes testing in 1990 and O_{93} testing in 1993, X denotes the intervention

ideas of the programme from seeping into the instruction of the control groups. The groups were, after all, from the same school and shared several teachers.

Procedure

Process evaluation procedure

The process evaluation carried out in this study had two goals. The first goal was to examine whether the intended intervention did indeed take place in the classroom. In order to answer this question, a classroom observation study was conducted during the FACE intervention. A sample of lessons was videotaped for each of the subject areas involved in the intervention. The purpose was to examine whether, and to what extent, the formal aims of the FACE programme were being applied in instruction, and also to examine how these aims were being implemented. The analysis of these data is still in process. However, some preliminary results are reported here.

The second goal of the process evaluation was to examine how the teachers evaluated the programme. In order to answer this question, the teachers in Alppila junior high school were interviewed at the end of each school year of the intervention. The aim was to describe the experiences and attitudes of the teachers involved in the project. Special attention was given to the types of formal aims that were incorporated in the instruction of each subject, and attention was also focused on the reactions of the pupils. In a separate preliminary study within the FACE project, Erma (1993) also used a questionnaire to measure the experiences and attitudes of the teachers.

Measurement of programme effects

In order to examine the effects of the intervention, a battery of tests (described below) was used. Some of the tests were used to measure any changes in the pupils' ways of thinking that occurred as a result of the intervention. These tests included a broad range of measures of cognitive abilities (Matrices, Box-Folding, Vocabulary Comprehension, Mechanical Reasoning, and Reading Skill) which were used to indicate possible broad transfer effects. Two tests were designed specifically to measure formal cognitive skills (FACE Reasoning and Definitions). Still other tests were used to examine whether the pupils thought differently of themselves as a result of the intervention (Cognitive Self-Concept and Self-Esteem).

To enhance reliability, standard guessing correction procedures (Choppin, 1988) were applied in all tests except Definitions, Cognitive Self-Concept, and Self-Esteem. In these tests guessing does not effect the results. In the FACE reasoning test, and in the Self-Esteem test, principal component scores were used, and in the test of Cognitive Self-Concept, factor scores were used (as a type of weighted sum). These operations should increase reliability from a plain sum score.

Cronbach's alpha (α) does, however, provide an adequate, if slightly low, estimate of the level of internal consistency. The reliabilities of each test or sub-test were within the limits considered to be acceptable for this type of research (Nunnally, 1978). Short descriptions of the tests used to measure the effects of the study are presented below.

Tests of cognitive ability

Matrices: Matrices of figures which change from left to right and top to bottom according to different principles. The task is to identify the missing figure. This nonverbal test is similar to the Raven Progressive Matrices and is strongly associated with general intelligence (Ministry of Labour, Employment Services Division: R1/P, 48 items, reliability: $\alpha = .90$).

Box-Folding: Finding three-dimensional objects corresponding to a two-dimensional drawing. The items call for perception, visual-spatial processing, and reasoning. The test is strongly associated with general intelligence (40 items, reliability: $\alpha = .90$).

Vocabulary Comprehension: Lists of five words each, in which four of the words share a concept and the fifth does not. The task is to identify the one word which does not fit. This verbal test is similar to the verbal component of many intelligence tests, and it is strongly associated with general intelligence (Ministry of Labour, Employment Services Division: V1/P, 70 items, reliability: $\alpha = .81$).

Mechanical Reasoning (DAT): Pictures which depict principles of physics applied to everyday situations. The task is to answer short questions about each picture. Possible answers are presented as multiple-choices. Technical knowledge, visual-spatial processing, and reasoning are called for. The test is also associated with general intelligence (68 items, reliability: $\alpha = .84$).

Reading Skill: Finding answers to simple questions based on texts of 100 to 140 words. The items call for rapid decoding and text comprehension (Ministry of Labour, Employment Services Division: V3/P, shortened version, 24 items, reliability: $\alpha = .81$).

Tests of formal cognitive skills

Definitions: Defining everyday words such as 'chair'. The success with which the pupils include the essential components of the concept is evaluated qualitatively. The test is strongly verbal (Six items, reliability: $\alpha = .69$, inter-rater agreement on the items was between 86.2% and 100% for close agreement and between 48.3% and 86.2% for exact agreement).

FACE Reasoning: Verbal problems of reasoning involving concepts, recognition of contradiction, and critical thinking. The idea is to determine which sentence in a set of multiple sentences is in accordance with a text (71 items, reliability: $\alpha = .83$).

Factors of cognitive self-concept and self-esteem

Cognitive Self-Concept and Self-Esteem: A questionnaire was designed to measure cognitive self-concept and self-esteem. The former, which focused on general intelligence (cognitive ability and creativity, reliability: $\alpha = .91$), orderly thinking (concentration or confusion of thought, reliability: $\alpha = .88$), mathematical ability (ability, skill and speed; reliability: $\alpha = .86$) and linguistic ability (ability and skill, reliability: $\alpha = .63$), was measured with 170 items. Self-esteem (self acceptance, sense of self-worth, sense of meaning and manageability; reliability: $\alpha = .86$), was measured with 38 items.

Results

Results of the process evaluation

In an observation study based on the video-taped sample of lessons, we attempted to find out if the formal aims of cognitive education were being applied in the instruction received by the pupils during the intervention. The preliminary results indicate that, in general, the instruction during the intervention appeared to fulfil one or more of the formal aims of cognitive education. There were differences observed in the extent to which these aims were fulfilled across lessons for individual teachers, as well as differences among teachers.

A more precise analysis of the sampled lessons shows that, all in all, 10% of the class time during the observed lessons was non-instructional. The questions asked and the answers given during the lessons were categorised. 46% of the questions asked were identified as serving a formal aim (concept analysis, 16%; reasoning, 21%; final or causal explanation, 9%; evaluation of truthfulness, 0%). The applicability of the formal aims varied across different subjects and subject material. Language instruction was found to be well suited for demonstrating concept formation (20-35%). Depending on the pupils' skills in the language, questions demanding reasoning could also be included (2-23%). In the sciences, history, and religion, questions concerning concept formation were relatively few (1-11%), while reasoning was well represented (10-35%), and final or causal explanation represented some 0% to 22% of the questions. For obvious reasons, none of the questions in mathematics represented final explanation. These differences in frequency of the types of questions are significant.

Of the total number of questions asked, 43% either were not answered, needed specification, received an answer that was not audible or could otherwise not be coded. The answers were categorised according to whether they contained elaboration or not. Straightforward answers (right or wrong) represented 38% to 72%, while elaboration was relatively rare (0-14%), again depending on the subject. These differences in frequency of the two types of answers are highly significant.

Using a questionnaire and teacher interviews, attempts were made to find out how the teachers evaluated the programme. Erma (1993) reported results of the questionnaire showing that junior high school teachers believed that the way they thought about instruction and how they taught had changed due to their involvement in the FACE project. The results indicated that the teachers felt that their conception of teaching had changed during the experiment. They also claimed that they managed to include formal aims in their teaching. They had incorporated epistemological themes into lessons in their subject areas; they claimed that this was a new approach for them, and that it had a positive effect on the thinking of the pupils.

Some general findings concerning the teacher interviews are available, and they clearly support the findings of Erma (1993). In every interview, regardless of teacher, subject, or the year in which the interview was conducted, certain general principles of applying the formal aims in instruction identified. These general principles are described below:

- Based on formal aims and using questions, the teachers guided their pupils in searching for and constructing entities or systems of knowledge. The pupils were asked to question concepts, analyse arguments and conclusions, give reasons for their own arguments and conclusions, find and analyse causal relations and more general invariance in events, and analyse attitudes and values that may lead to the actions of people.
- The teachers used the teaching methods they were used to. For some teachers, the main method was continuous classroom conversation, while for others, it was group work, work in pairs or other methods of instruction. More important from the teachers' point of view was the quality of the questions they asked, and the quality of the tasks they gave to the pupils. There were, however, some common features in the instruction provided by all teachers. Each teacher had increased the amount of discussion in the classroom, and the type of questions asked during instruction had changed. The focus was less on content, and more on understanding and thinking.
- Teachers tried to give reasons for studying the entities of subject matter to be taught, and they tried to connect what was taught with everyday life.
- Integration across the subject areas and contexts was attempted. The rationale for such integration was twofold: Its purpose was to extend the field of application of knowledge pupils already had learned, and to make this knowledge 'come alive' by connecting it to larger contexts in real life.
- For the above-mentioned changes to take place, it was necessary to reorganise the curriculum. Some material was left out or restructured, while other material was given additional emphasis. The subject matter was used to demonstrate and exercise thinking, and the instruction was set up accordingly.
- As the teachers evaluated the performance of their pupils during lessons, in tests, or through home work, they tried to emphasise that being correct was not the most important thing. More important was that the pupils showed that they had at least tried to think, not just remembered or copied.

- The teachers were worried. They frequently expressed doubts such as: "Am I doing the right things in the right way?", "Am I giving the pupils enough to think about?", "Is it alright to drop traditional parts of the curriculum and concentrate on certain issues?", or "Have I remembered all essential details?". Partly, this worry was due to the pressures caused by the experiment itself. The teachers knew that they were doing something that had not been done previously. At the end of the intervention, the teachers were quite convinced that they really had been implementing the formal aims. Still, they also felt that they could and should have done much more.

Effects of the intervention

We wanted to know if the pupils' ways of thinking had changed and if the pupils now thought differently of themselves as a result of the intervention. Two-factor analyses of variance (group and gender) were used to test for differences between the experimental group and the control groups both before and after the intervention. The second factor in these analysis was gender, since differences in test performance could be expected for boys and girls (Guilford, 1971; Lim, 1994; Lubinski et al., 1993; Rosen, 1995) and the groups were unequal in distribution of boys and girls.

One-factor analyses of variance were used to test for differences between the two age-cohort groups used as control groups. These were carried out to examine and control for age-typical development. Similarly, repeated measures analyses of variance were used to test for changes in the performance of the experimental group over the course of the intervention. The results are presented in the form of effect sizes (Glass et al., 1981; Hedges & Olkin, 1985; McGaw, 1988) in Table 1.

With regard to the results presented in Table 1, it is important to note that the experimental group did not differ significantly from the control group on any of the tested dimensions in the pretest. The programme had already been implemented for about three months when the pretests were carried out. It is therefore possible that some of the nonsignificant differences between the experimental groups were in fact due to the intervention. As the table shows, the intervention effects on performance in the tests of cognitive ability were nonsignificant. It may be noted, however, that all of tests of cognitive ability, except for the Box-Folding test, showed a nonsignificant but positive difference between the posttest results of the experimental group and the control group.

In the Vocabulary Comprehension test, the pretest difference between the experimental groups was large enough to explain posttest differences. The differences in developmental effect sizes between the experimental groups were large enough to indicate possible broad transfer effects in the Matrices, Mechanical Reasoning and Reading Skills tests.

Positive intervention effects were evident in the tests measuring formal cognitive skills. Both tests were verbal. The FACE Reasoning test measured text

Table 1 Results of the FACE Programme

Tests	Pretest[a]	Development in[b] Exp. group	Controls	Posttest[c]	Gender[d]
Tests of cognitive abilities:					
Matrices	0.21	0.45*	0.36*	0.30	-0.11
Box-Folding	-0.01	0.48*	0.51*	-0.04	0.13
Vocabulary Comprehension	0.30	0.51*	0.61*	0.13	-0.42
Mechanical Reasoning	-0.03	0.60*	0.29*	0.28	0.89
Reading Skill	0.02	0.84*	0.44*	0.43	-0.31
Tests of formal cognitive skills:					
Definitions	0.16	0.29	0.00	0.42*	-0.50*
FACE Reasoning	0.40	0.62*	0.46*	0.57*	-0.19
Factors of cognitive self-concept and self-esteem:					
SC of Intelligence	0.32	0.15	-0.29	0.78*	0.25
SC of Orderly Thinking	0.05	0.10	-0.44*	0.60*	-0.20
SC of Mathematical Ability	0.36	-0.22	0.14	0.02	0.26
SC of Linguistic Ability	0.42	-0.15	0.17	0.11	-0.09
Self-Esteem	-0.35	-0.11	-0.40	-0.08	-0.23

* $p < .05$

Note The results are presented in terms of effect sizes. The greater the effect size is, the more the experimental group differs from the controls (pretest[a] and posttest[c]), the posttest results differ from the pretest results (development[b]) or the boys differ from the girls (gender[d]). The reported gender related effect sizes are from the posttest situation comprehension, reasoning, recognition of contradiction, and critical thinking, while the Definitions test measured understanding and analysis of concepts. In both tests, the experimental group improved its performance beyond the age-typical development seen in the controls. In the FACE Reasoning test, the experimental group actually scored as well as pupils of the senior high school who were one year older. This is all the more notable as admission to senior high school is based on earlier school achievement, while junior high is compulsory.

The Self-Concept test included items of cognitive self-concept and self-esteem. The cognitive self-concept was evaluated on four components or dimensions of interest, namely, self-concept of intelligence, orderly thinking, mathematical ability, and linguistic ability. In the posttest, the self-concept of the experimental group was clearly more positive than that of the control group on the dimension of intelligence and orderly thinking. No intervention effect was observed for the other dimensions of self-concept.

Significant gender-related effects were observed in performance on the Mechanical Reasoning test and on the Definitions test. Similar, but nonsignificant, effects were found in the self-concept dimensions of orderliness of thinking, mathematical ability, and linguistic ability. All the tests of cognitive ability and the FACE Reasoning test showed significant improvement over time, i.e., the effects of age-typical development or maturation, while change in the self-concept area was mainly insignificant.

Conclusions and discussion

The philosophical theory of knowledge has not previously been systematically used in school interventions (Chance, 1986; Chipman et al., 1985; Fisher, 1991; Nisbet, 1991). A general philosophical approach (Lipman, 1985) is less rare but still not as common as psychological approaches to teaching thinking. Many, if not most programmes do, however, include some aspects of it. The FACE project was a school intervention project based on the theory of knowledge, i.e., epistemology. In the project, a set of formal aims of cognitive education were formulated and applied. An attempt was made to integrate the teaching of skilful thinking to most of the subjects in the curriculum. Other distinguishing features of the project were its three-year duration and the full scale evaluation study accompanying the intervention.

The fact that gender-specific differences and age-related development were found in the results of the tests administered, and that these were consistent with those in prior research (Anderson, 1992; Guilford, 1971; Lim, 1994; Lubinski, et al., 1993; Mönks & Mason, 1993; Rosen, 1995), provides support for the validity of the tests used in this study.

There are, on the other hand, some built-in factors threatening the validity of the type of quasi-experimental design used in this study (Campbell & Stanley, 1973).

These include a testing effect, as only the experimental group was retested. The time which elapsed between the first and the second testing session, and perhaps also the relatively large number of tests used, may be seen to lessen the risk of remembering previous responses. However, the experimental group did have more experience with testing.

Many specific events could also have occurred between the measurements, thus masking an experimental effect or posing as one. This threat is made smaller by the fact that the follow-up of the intervention process did not turn up such occurrences. An additional concern in the design is that it was not possible to randomly assign pupils for the intervention and control groups. This leads to some doubts about the generality of the results, and thus, as was explained in section three, efforts were made in the choice of the experimental school to counter this threat. In addition, to minimise the possible effects of nonrandom assignment, the groups were tested at a relatively early stage of the intervention. The groups did not differ significantly at this time, which we see as a sign that the selection process had produced groups which were reasonably similar on the dimensions to be studied.

Some relatively large effect sizes were, however, found between the groups in the pretest. These may be interpreted either as a sign that the experimental group was indeed cognitively somewhat more able than the control group or as a sign that the intervention had an effect even before the first testing. This was a situation we were not prepared for. In the first case, some of the positive posttest results would at least partly not be due to the intervention. In the second case, the results would indicate that improvements in the thinking of the pupils were more rapid than we expected.

Differential loss from the comparison groups was not observed. Other threats to validity, such as maturation or reactions to experimental arrangements rather than to the treatment, are relatively unproblematic in this design. We conclude that the results of this study have a level of reliability and validity typical of quasi-experimental research. The results can be trusted to the extent that the previous problems and cautions are kept in mind. It is in any case obvious that the results should be tested in an independent control study.

It is in itself important that the teachers reported that a significant change took place in their way of teaching and thinking of teaching. The actual process of instruction also seems to have changed. The teachers agreed on this and the observation study showed that the formal aims of cognitive education were being applied. These may be seen as the necessary prerequisites for a successful intervention. On the other hand, we observed that the answers given by the pupils in class were predominantly not elaborated. This situation is probably typical of school instruction, in that it is a natural consequence of the need to transfer information and to check that it has been received. A more Socratic dialogue might have led to larger changes in the thinking of the pupils.

The change in the thinking of the pupils was measured using a broad range of tests. The results indicate that schools may indeed improve the way their pupils think as a result of applying the theory of knowledge to how the subject matter is taught. We conclude that the formal cognitive skills of the pupils improved due to the intervention. The experimental group performed better in the tests that demanded analysis of concepts, comprehension and critical evaluation of complex and partly misleading information, as well as reasoning based on this information. In this sense, the pupils became more skilful thinkers. Also, several nonsignificant effects indicated possible transfer to a broader range of cognitive abilities. The nonsignificant effects are promising from the point of future interventions, although they are, of course, not to be seen as conclusive evidence.

We also conclude that the programme had a positive effect on how the pupils think of themselves. The effects of the programme on the self-concept of the pupils seem to have been strongest for the more general aspects of thinking. After the intervention, the pupils thought more positively of their cognitive ability and productivity of thinking, but not in the more specific areas of mathematical or linguistic ability. The reason for this may be that the achievements of the pupils are evaluated frequently from a relatively early age. Knowledge of these accumulated results probably plays a prominent role when the pupils form their academic self-concept. The self-concept of the pupils may, therefore, be more easily influenced in the areas of cognitive competence that have not traditionally been evaluated in school.

The intervention effect on self-esteem was not significant. This is hardly surprising. Self-esteem is probably too central a dimension of personality to be readily influenced by anything less than relatively dramatic changes in the environment at home, at school, or in peer group relations. A change in the self-concept or self-esteem of the pupils is potentially just as important as a change in ability. Choices and actions are influenced by what one thinks of oneself. The application of knowledge and skill can, thus, be inhibited or enhanced by a person's self-concept and self-esteem.

The instruction provided during the intervention was not similar to the tests that were used to measure the intervention effect. The teachers did not use the tests as training material during the lessons or for evaluation afterwards. The content to which the skills of thinking were applied during the lessons was very different from the content of the tests. Finally, the testing situation was different from normal school work. All of this indicates that the results were not due to the pupils becoming test-wise, and that a true intervention effect was evident.

There are no theoretical or empirical reasons to believe that the FACE approach could not or should not be extended to younger or older pupils than those involved in the experiment, or to adult education. The FACE programme was deliberately limited in its approach to simplify the experimental design. There will probably be benefits of synergy in developing a programme that combines an epistemological approach with other effective approaches to teaching thinking (Nisbet, 1993). Further intervention studies of this kind, preferably with larger experimental groups, are clearly in order.

References

Adey, P., & Shayer, M. (1995, August). *Teaching intelligence: Transfer is the evidence, but what is the mechanism?* Paper presented at the Sixth European Conference for Research on Learning and Instruction. Nijmegen, The Netherlands.

Anderson, M. (1992). *Intelligence and development.* Oxford: Blackwell.

Campbell, D.T., & Stanley, J.C. (1973). *Experimental and quasi-experimental designs for research.* Chicago: McNally.

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge, MA: Cambridge University Press.

Chance, P. (1986). *Thinking in the classroom: A survey of programs.* New York: Teachers College Press.

Choppin, B.H. (1988). Correction for guessing. In J.P. Keeves (Ed.), *Educational research, methodology, and measurement* (pp. 384-386). Oxford: Pergamon Press.

Coles, M.J. (1993). Teaching thinking: Principles, problems and programmes. *Educational Psychology, 13,* 333-344.

Csapó, B. (1990). Integrating the development of the operational abilities of thinking and the transmission of knowledge. In H. Mandl, E. De Corte, N. Bennett, & H.F. Friedrich (Eds.), *Learning and instruction: European research in an international context* (Vol. 2.2, pp. 85-94). Oxford: Pergamon Press.

Csapó, B. (1992). Improving operational abilities in children: Results of a large scale experiment. In: A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development. Implications and applications for education* (pp. 144-159). London: Routledge.

Erma, T. (1993). *Opettajien kokema ammatillinen kasvu - tutkimus tiedollisen kasvatuksen formaalit tavoitteet (FOTA) - koulukokeilusta* [The professional growth of teachers - a study of the formal aims of cognitive education (FACE) project]. Helsinki: University of Helsinki.

Fisher, R. (1991). *Teaching children to think.* Hemel Hempstead: Simon and Schuster Education.

Gardner, H. (1983). *Frames of mind.* New York: Basic Books.

Glass, G.V., McGaw, B., & Smith, M.L. (1987). *Meta-analysis in social research.* Beverly Hills: Sage Publications.

Guilford, J.P. (1971). *The nature of human intelligence.* London: McGraw-Hill.

Hamers, J.H.M., & Overtoom, M.Th. (1997). *Teaching thinking in Europe: Inventory of European programmes.* Utrecht: Sardes.

Hansford, B.C., & Hattie, J.A. (1982). The relationship between self and achievement/ performance measures. *Review of Educational Research, 52,* 123-142.

Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Lim, T.K. (1994). Gender related differences in intelligence: Application of confirmatory factor analysis. *Intelligence, 19,* 179-192.

Lipman, M. (1985). Thinking skills fostered by philosophy for children. In J.W. Segal, S.F. Chipman, & R. Glaser (Eds.), *Thinking and learning skills. Vol. 1: Relating instruction to research.* Hillsdale, NJ: Lawrence Erlbaum.

Lubinski, D., Benbow, C. P., & Sanders, C. E. (1993). Reconceptualizing gender differences in achievement among the gifted. In K. A. Heller, F. J. Mönks, & A. H. Passow (Eds.), *International handbook of research and development of giftedness and talent* (pp. 693-707). Oxford: Pergamon Press.

Marsh, H.W. (1990). Causal ordering of academic self-concept and academic achievement: A multivariate, longitudinal panel analysis. *Journal of Educational Psychology, 82,* 646-656.

Marsh, H.W. (1994). Using the national longitudinal study of 1988 to evaluate theoretical models of self-concept: The self-description questionnaire. *Journal of Educational Psychology, 86,* 439-456.

Marsh, H.W., Barnes, J., Cairns, L., & Tidman, M. (1984). Self-description questionnaire: Age and sex effects in the structure and level of self-concept for pre-adolescent children. *Journal of Educational Psychology, 76,* 940-956.

Marsh, H. W., Chessor, D., Craven, R., & Roche, L. (1995). The effects of gifted and talented programs on academic self-concept: The big fish strikes again. *American Educational Research Journal, 32,* 285-319.

Marsh, H.W., Walker, R., & Debus, R. (1992). Content specificity of relations between academic achievement and academic self-concept. *Journal of Educational Psychology, 84,* 35-42.

McGaw, B. (1988). Meta-analysis. In J.P. Keeves (Ed.), *Educational research, methodology, and measurement* (pp. 678-685). Oxford: Pergamon Press.

McGuinness, C. (1993). Teaching thinking: New signs for theories of cognition. *Educational Psychology, 13,* 305-316.

Mönks F.J., & Mason, E. J. (1993). Developmental theories and giftedness. In K. A. Heller, F. J. Mönks, & A.H. Passow (Eds.), *International handbook of research and development of giftedness and talent* (pp. 89-101). Oxford: Pergamon Press.

Nisbet, J. (1991). Projects, theories and methods: The international scene. In M. Coles & W. Robinson (Eds.), *Teaching thinking* (2nd ed., pp. 25-34). Bristol: Bristol Classical Press.

Nisbet, J. (1993). The thinking curriculum. *Educational Psychology, 13,* 281-290.

Nunnally, J.C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Rosen, M. (1995). Gender differences in structures, means, and variances of hierarchically ordered ability dimensions. *Learning and Instruction, 5,* 37-62.

Segal, J.W., Chipman, S.F., & Glaser, R. (1985). *Thinking and learning skills.* Hillsdale, NJ: Lawrence Erlbaum.

Toulmin, S. (1972). *Human understanding: General introduction and part I.* Oxford: Clarendon Press.

Voutilainen, T. (1991). The project FACE. In S. Maclure & P. Davies (Eds.), *Learning to think: Thinking to learn* (pp. 137-144). Oxford: Pergamon Press.

Voutilainen, T., Mehtäläinen, J., & Niiniluoto, I. (1990). *The conception of knowledge*. Helsinki: The Government Printing Centre.

5

Training domain-specific abilities: The case of experiential structuralism

A. Efklides

Introduction

This book focuses on 'thinking skills' which, according to the editors of this volume, forms the basis on which the teaching of thinking rests. Looking at the title of this paper, it is clear that it comes from a different research tradition, namely that of intelligence. Therefore the question one has to answer first is if skills are equivalent to abilities. A second issue pertains to the domain-specificity of abilities (and/or skills) and its implications for the teaching of thinking. If skills (and/or abilities) function at various levels of generality, as the term 'domain-specific' implies, then the question is at which level of generality will the teaching of thinking take place.

Of course thinking is not the same as intelligence; thinking is one of the paramount manifestations of intelligence. From this point of view, the work on the teaching of thinking is relevant to the teaching of intelligence and cognitive acceleration issue. However, there are a number of differences between thinking and intelligence which need to be pointed out. Thinking has a conceptual and a procedural aspect whereas intelligence is basically procedural. Intelligence also involves metacognitive functions such as metacognitive knowledge and monitoring (Brown, Bransford, Ferrara, & Campione, 1983; Sternberg, 1985). Furthermore, intelligence has also to do with the person's resources to adapt to novel and complex conditions, and this is the feature which leads us to ascribe the term 'ability' to intelligent behaviour. Therefore, the conceptualization of thinking as intelligence implies that we stress certain aspects of it, one of which is 'ability'.

We are coming now to the issue of skill and ability. Skills, such as reading, writing or other sensory-motor complex responses, are usually differentiated from abilities, such as verbal, spatial, and reasoning abilities. Of course, both terms are used in order to denote procedural aspects of thinking in contrast to conceptual aspects of it. However, their basic difference lies in that skills are considered amenable to training and practice, whereas abilities are not necessarily so. Abilities are latent factors underlying performance and they are often considered to be constrained by developmental factors. Thus, whereas skills are acquired, practiced and integrated in skill hierarchies and systems (Anderson, 1981, 1983; Fischer, 1980), abilities develop. Research on the structure of intelligence has also shown that abilities are hierarchically organized (Carroll, 1993; Cattell, 1971; Gustafsson, 1988; Spearman, 1927). Of course, there is still a lot of discussion about the number of abilities one can identify, their exact structure, and the sources of developmental change. Nevertheless, if we concentrate on the similarities of the two concepts, what we have is procedural knowledge structures which form hierarchical systems, which may change along with development. The source of developmental change is to be found either in changes of the person's capacity resources or in experience and instruction.

If we accept this rationale for skills and abilities, then the issue of the levels of generality at which skills and abilities function becomes relevant. Research on skill acquisition and architecture of cognition (Anderson, 1981, 1983) has shown that skills form production systems of hierarchical nature. Depending on the range of applicability of the production systems underlying the various skills, we may assume that there are general and specific skills. However, no theorist in this tradition has claimed that there may exist such a general skill, which will be able to control all lower-order skill structures, as intelligence theorists do. Furthermore, skill theory cannot adequately accommodate individual differences and developmental change phenomena. Only Fischer's skill theory (Fischer, 1980; Fischer & Farrar, 1988) tried to do this. It claims that what changes with development is the skills representation and structure. Thus, at the lowest level there are simple skills (or task-specific skills), which later on get mapped and coordinated into systems of skills. At the highest level there are systems of systems of skills. This theory explains how thinking comes to master complex skill structures in a domain and leaves no room for an overarching general ability which is responsible for individual differences in the development of all kinds of skills.

Intelligence theories, by contrast, both psychometric and developmental, do claim there is a general ability factor, the nature of which they try to define. Psychometric research has convincingly demonstrated the existence of abilities of various levels of generality, i.e., broad and narrow abilities, organized hierarchically. At the top of the hierarchy there is one general factor (g), presumably tapping general intelligence (Carroll, 1993; Gustafsson, 1988). Developmental research, on the other hand, has also led to identification of domain-specific abilities as well as domain-free or general systems (Demetriou & Efklides, 1987b, 1994). Such a hierarchical system identifies structures at the task level (task-specific abilities), broader structures underlying performance on a whole domain (domain-specific abilities) and a general ability level at the top of the hierarchy. In this sense, the model of one general factor and broad and narrow abilities of lower order is common to the two traditions of thinking. If we follow the same strategy as before and focus on the similarities of skill-based and ability-based theories of intelligence, then what we find is that both theoretical viewpoints identify task-specific and domain-specific structures and their major difference is the acceptance or not of a general ability factor. What are the implications of the above state of affairs for the teaching of thinking?

If one leaves no room for a general factor, then the teaching of thinking can only take place at the task-specific or the domain-specific level. If one accepts the general factor, then teaching can take place also at this general level. This rationale could lead to some clear-cut results, if only psychologists could agree on the definition of the ability domains and the nature of the domain-free or general abilities. In fact, there is very little consensus over these issues, and this may explain why there are so many and so diverse approaches to the teaching of thinking.

In this book one can clearly distinguish two main approaches to the teaching of thinking: the first identifies general intelligence with inductive ability and, by cultivating it, hopes to achieve generalization and transfer to all lower-level abilities (see Klauer, this volume). The second approach represents the intervention at the skill-specific level, such as mathematical, reading comprehension, and text production skills. The assumption in this case is that these are specialized skills, and therefore training the one will not transfer necessarily to the other. It is important, however, to note that the above skills, no matter how important they are, by no means do they exhaust the possible number and type of skills people use in their activities, and particularly in school.

Evidently, there are conceptual domains such as mathematics, physics, history and so forth, which are educationally and culturally defined rather than procedurally as the skills and abilities approach demands. Therefore, if we want to apply the skill-schema, we need to look across conceptual domains and find out which skills are used by more than one domains. In this respect, reading comprehension and text production are language-related skills, which are used independently of conceptual domains. Mathematics and physics also share with each other mathematical skills. Consequently, mathematical and language-related skills are broader than task-specific skills used within conceptual domains.

But are these general skills the only ones a person uses when dealing with a particular knowledge domain? We would argue that they are not. This is the point where the abilities-related research becomes relevant. For example, what processes do we use in order to understand history and biology? Only language-related skills? In fact, comprehension of historical events or biological facts also requires the ability to identify complex causal structures and to analyze the role of each factor in the causal chain. This kind of reasoning is called scientific thinking, and it is often thought to be pertinent only to experimental sciences. However, it seems to be also a prerequisite of historical thought (Demetriou, Gustafsson, Efklides, & Platsidou, 1992). This kind of reasoning evidently is also relevant to physics but not to mathematics where causality, in the form found in physical phenomena, is not present (Efklides, 1992; Metallidou, 1996). Where does this approach lead us to?

If we overcome the educational model of conceptual domains and try to identify modes of thinking which cut across conceptual domains, then we come to a conceptualization of domains in terms of basic relations or categories of mind, such as quality (concepts, categories), quantity, causality, and space. Another basic domain is the semantic world, that is, the representation of states and events. These categories of mind form reality domains, which tend to be represented by different symbolic systems, require specialized processes and they are subjectively felt as different. Furthermore, there is evidence that the domain-specific processes (i.e., abilities) develop in relative independence from each other (Demetriou & Efklides, 1987a, 1987b). These are the principles, which according to our theory, Experiential Structuralism, govern the organization of mind (Demetriou & Efklides, 1987b, 1994). Thus, each of the reality domains comprises a number of abilities which form a Specialized Structural System (SSS) and which develop at particular age periods. When the individual comes across a conceptual domain, he/she employs abilities, which may come from one or more of the SSSs, depending on the nature of the task at hand. A brief outline of Experiential Structuralism is the following.

Theory

In order to understand the structure of the human mind, the theory proposes three levels of description: first the hardware or capacity characteristics of the cognitive system; second the SSSs, and third the metacognitive system. The capacity characteristics of the cognitive system are domain independent and they are defined in terms of speed of processing and working memory (both control processes and storage) (Demetriou & Efklides, 1994). The metacognitive system (called hypercognition) is also domain-free, and it has to do with both the representation of the person (and the others) dealing with particular problems in particular domains, the representation of mental processes, such as memory, and the monitoring and regulation of behavior. The metacognitive system is the interface between the person and reality as well as between SSSs themselves. Metacognition involves both on line awareness as well as reflection on cognitive processes and related experiences. Thus it builds models of the world, persons (including one's self) and cognition.

Five SSSs have been identified by Experiential Structuralism. The qualitative-analytic, the quantitative-relational, the causal-experimental, the spatial-imaginal, and the verbal-propositional. The first part of the name of the SSSs denotes the reality domain each of them applies to and the second the procedural character of the SSS:

- The qualititative-analytic (QA) SSS addresses categorical structures and it involves inductive, analogical and analytical processes, such as identification of properties or relations and comparison processes.
- The quantitative-relational (QR) SSS addresses the quantifiable aspects of reality structures. It is relational in nature, because it relates items within a dimension as well as dimensions between them. It involves abilities such as quantitative specification and representation, that is, counting, measurement, and arithmetic operations; dimensional-directional construction, i.e., dimensional representation, specification of the form of the dimension (linear, curvilinear, etc.), construction of the concept of variable; dimensional-directional coordination, namely, specification of inter-dimensional relations, such as proportional and other complex relations.
- The causal-experimental (CE) SSS processes causal relations. It involves combinatorial abilities; which allow the person to define the possible coexistence structure within which the cause-effect relation is to be found; hypothesis formation abilities, which lead to specific predictions about the causal relation possibly existing; hypothesis testing abilities, such as experimentation, and model construction abilities, which allow the interpretation of the results of experimentation.
- The spatial-imaginal (SI) SSS is directed to the processing of images and spatial relations. It depicts elements such as form, location and orientation. It involves abilities such as image construction and reconstruction, image integration, image transformation, mental rotation and coordination of perspectives.
- The verbal-propositional (VP) SSS concerns the verbally encoded semantic representations and the relations between them rather than the relations between the objects depicted in them. It is propositional in nature and it involves lexical abilities as well as inference drawing abilities, which are based on formal aspects of the propositions, such as truth and falsity, and the relations between propositions, such as logical relations.

The studies reported here focused on two SSSs only: the quantitative-relational and the causal-experimental. This was deemed necessary because we wanted to establish, first, if the abilities of the SSSs are trainable (when and how) and if they transfer from one SSS to the other. Once this is clear, then we can go on to determine the possible 'transferability distance' between combinations of SSSs and/or combinations of abilities, which share symbolic or other characteristics between them.

Study 1

This study (Efklides, Demetriou, & Gustafsson, 1992) aimed to test the trainability of cognitive abilities and the role of general and domain-specific abilities in cognitive change. The specific hypotheses tested were: first, training effects are a function of both general (fluid intelligence, Gf) and specialized (SSS-specific) abilities; second, the various SSSs, due to their differential constitution, are differentially amenable to training effects; third, training one SSS would not transfer to the other and, fourth, training would interact with age and the person's cognitive (developmental) level.

Method

Design

In order to test the above hypotheses, an intervention study was designed. It involved two experimental groups and one control group. The first experimental group received training on the QR SSS (Quantitative-Relational Treatment Group, QRTG). The targeted ability was proportional reasoning. The second experimental group received training on the CE SSS (Causal-Experimental Treatment Group, CETG). The targeted ability was experimentation. The control group (Control Treatment Group, CTG) received no training at all. The idea guiding the selection of the targeted abilities was that they develop during the age period covered in the study, that is, adolescence.

All subjects were tested before and after training with the same battery of tasks. The battery consisted of four QR and four CE tasks. At the pretest all subjects were also required to solve a set of four tests addressed to fluid intelligence (Gf). Three of these tests were inductive in nature and, therefore, they may be considered as indicative of the Qualitative-Analytic (QA) SSS. The fourth task was a field independence measure, which is a good indicator of Gf (Gustafsson, 1988). Therefore this set of tasks was indicative of a more general factor than the QA SSS. For this reason, these tasks were used as predictors of training effectiveness and not as measures of transfer effects from the QR and CE to the QA SSS.

Subjects

The study involved 1028 subjects. Four age groups were represented in the sample: 10, 12, 14, and 16 years old. In all, 509 girls and 519 boys took part in the experiment. Regarding their socioeconomic origin, 313 came from high SES families of urban residence; 375 came from low SES families of urban residence; 340 came from low SES families and lived in rural areas.

Tasks

The two sets of tasks (QR and CE tasks) were constructed so that they had the same structure, and differed only in terms of the ability required for their processing. The structure of the tasks resembled Fischer's (Fischer & Farrar, 1988) hierarchy of skill levels. Thus, although all four tasks in a set tapped the same ability, they differed in structural complexity and respective difficulty. Each of the tasks in a set corresponded to one of the four developmental levels of the tier of abstract thought, namely, the level of single abstract sets (from now on called level 1), the levels of abstract mappings (level 2), abstract systems (level 3), and systems of abstract systems (level 4). In this way the cognitive level of the person in each SSS could be determined, depending on the most difficult of the four tasks he/she had successfully solved. Furthermore, change of cognitive level could be determined both within an SSS and across SSSs. Change of cognitive level rather than simple quantitative increase of performance scores was the criterion for the success of the intervention.

There were three sets of tasks: the QR, the CE and the Gf tasks.

- Quantitative-relational SSS tasks.

  Four problems involving proportional relationships were addressed to the quantitative-relational SSS (QR1 - QR4):

  - QR1: Subjects were presented with a two times two table showing a relationship between watering frequency (twice and four times/month) and yield (2 and 6 kgs/hectare for plant A and 3 and 6 kgs/hectare for plant B). The task was to select from a number of alternatives which plant is more affected by watering and to explain why (i.e., produce the calculations necessary to justify one's choice). In this task, two variations had to be co-ordinated into a single set that forms an abstraction.
  - QR2: Two tables like the one in QR1 (i.e., a double table) were presented, showing the effects of watering on plants A and B in two areas I and II. Thus, in this task two single sets/abstractions had to be combined.
  - QR3: Two double tables were presented showing the effects of watering on plants A and B in areas I and II, when fungi are and when fungi are not present. Thus four single (or two double) sets of data, representing a system of abstractions, had to be combined to solve the task.
  - QR4: Four double tables were presented showing the effects of watering on plants A and B in areas I and II, when fungi are present, with or without use of fungicide, and when fungi are not present, with or without use of fertilizer. Thus, four double or eight single sets of data had to be combined to solve the task. This task represents a system of abstract systems.

- Causal-experimental SSS tasks.

  Four problems were constructed involving the design of experiments in order to test hypotheses (CE1-CE4). As stated above, these tasks were structurally equivalent to the QR tasks in the sense that they also tapped the four skill levels of the tier of abstract thought:

  - CE1: A simple hypothesis was given ('the increase in watering frequency increases the productivity of plants') and the subject was asked to use plants A and/or B and two watering frequencies (twice a month or four times a month) to design an experiment to test the hypothesis (single abstraction). A table was presented in which the subject had to fill in the appropriate plant and watering, following the principle of 'all the others being equal ...'.

- CE2: A hypothesis was given about the interaction between two factors ('Watering increases the productivity of plant A, but does not affect the productivity of plant B'). An experiment, integrating two single ones, had to be designed to test the above hypothesis (abstraction mapping).
- CE3: In this task, the experiment to be designed had to test two interaction hypotheses, regarding the effects of watering on A in areas I and II and on B in areas I and II (abstract system). Thus a three-way design (plant x area x watering) had to be proposed.
- CE4: In this task, yet another factor, fertilization, had to be taken into account. The solution of the task required design of a four-way experiment (plant x area x fertilizer x watering). Such a design captures the interaction of two abstract systems, therefore it is a system of systems.

- Fluid intelligence tests.

A set of four tests was used to measure fluid intelligence (Gf), three of which were selected from the Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, & Harman, 1976):

- The Letter Sets (LS) test contained fifteen items in which five sets of four letters were presented. The task was to find the rule which related four of the sets to each other and to mark the one which did not fit the rule.
- The Figure Classification (FC) test included fourteen items presenting two or three groups of figures. Each group of figures contained three geometrical figures which were alike in accordance with some rule. The task was to discover these rules and to classify each of eight given figures to one of the groups.
- The Hidden Figure test (HF) presented sixteen items in which the task was to decide which of five geometrical figures was embedded in a complex pattern. The HF test is constructed to measure field independence, or flexibility of closure, but is known to be a good indicator of g as well (Gustafsson, 1988). The other two tests are classified as measures of the inductive factor, which in turn more or less coincides with Gf (cf. Gustafsson, 1988). Inductive ability was also tested with a number series test.
- The Number Series test (NS) contained twenty items in which a series of five or six numbers was given, and the task was to add two more numbers to the series (Gustafsson, Lindstrom, & Bjorck-Akesson, 1981).

Training tasks

The training consisted of three parts: an introduction, explaining that the subject had made a mistake at one of the tasks of the previous testing and now he/she would be given instructions how to solve it. The importance of being able to handle this kind of problems was also stressed. At the second part, a problem similar to the ones of the pretest, but applying to a different situation, was presented; the solution was then given in a step by step fashion. For example, the following form of instruction was given for the QR training: In order to find out which plant, A or B, had the more productivity change, you have to compute the rate of productivity change for each plant and then compare the two outcomes. In order to do this, you need to divide the productivity of the plant when it is watered 4 times with the productivity of the plant when it is watered 2 times. That is , Plant A → 16:4=4, i.e., four times increase; Plant B→20:5=4, i.e., four times increase. Therefore A=B in terms of productivity change. A similar procedure involving the principle underlying the solution of the problem and the steps for implementing it was used in the CE tasks. Finally, at the third part, the subject was presented with a new problem (similar to the previous one) and was asked to solve it. Once the problem had been solved, feedback was provided. The feedback consisted in the detailed solution of the problem. The subjects were asked to study it and correct their mistakes. When the subjects finished this procedure, they were given the posttest.

Procedure

All subjects were tested before and after the training period with the QR and CE tasks. All testing was carried out in groups in the pupils' regular classrooms. The pretest session lasted for approximately two school hours, and comprised the eight SSS tasks, and the Gf tasks. The training session was held about two weeks later, followed by administration of the eight SSS tasks as posttests. The training session lasted approximately half an hour. The training leaflets were personally addressed to each subject according to their assignment to the experimental groups and the level achieved at the pretest. Thus, each subject was given training at the cognitive level next to his/hers.

The control treatment group received no training at all; subjects were instructed that they would be given no training and that they should do their best to try to attend to the details of the tasks now that they were familiar with their requirements. It was particularly stressed that they must try to improve their performance. Control group and experimental group subjects were tested in the same classroom. No time limit was imposed at any of the three phases of the experiment.

Results

As mentioned above the first aim of the analyses was to determine the trainability of domain-specific abilities and the role of Gf in cognitive change. For this sake, the data were analyzed with confirmatory factor analysis, i.e., with structural modeling techniques. Structural modeling is a method of testing the goodness of fit between a theoretical model and actual data. In this way, one confirms rather than explore the structure of the data. The model tested each time may involve first order latent factors, closely related to the particular tasks used, and second-order or higher-order latent factors, which capture the variance shared by more than one narrow factors.

What is important with this method is the possibility to identify relations between the various factors themselves. In the case of training studies this method may reveal the effects of pretest factors on posttest performance. That is, it may show whether posttest performance is accounted for only by the respective task-specific factor and/or by the respective pretest factor and the Gf factor. The lower the correlation between the pretest and respective posttest factor, the more substantial the effect of the training procedure is because training introduced a source of variability. Therefore, this method reveals not only training effects but also their possible sources.

The LISREL VI program was used (Jöreskog & Sörborn, 1986). In these models the four Gf tasks were assumed to be related to one latent factor; the four QR and four CE tasks of the pretest were assumed to be related to a QR factor (PreQR) and a CE factor (PreCE), respectively. The QR and CE tasks of the posttest were assumed to be related to the respective PostQR and PostCE factors. This model was fit separately for each age group, with the estimates constrained to be equal for all the treatment groups within an age group. The results confirmed the existence of the Gf, PreQR and PreCE factors, and thus supported the existence of both general and specialized factors in cognitive organization. The existence of the PostQR and PostCE factors was also confirmed.

In order to identify the effects of training and its possible interaction with Gf, we investigated a) the degree of correlation between the pretest and posttest factors, and b) the correlations between Gf and both the pretest and posttest factors. These correlations are shown in Table 1. The correlations between pretest and posttest factors were generally high, indicating a high degree of stability over the two testing sessions.

The stability was highest in the case of QR factors, and particularly so (around .80) in the non-trained groups (CETG and CTG). The correlations tended to be a little lower in the QRTG. This finding implies that QR abilities change with training, although this is a small effect. What is even more important is that QR abilities do not change without training. Looking at the correlations between pre CE factors and post CE factors in the three treatment groups, it is evident that the correlations tend to be lower (around .60 and .70). Therefore, there is less stability from one testing to the next, and change occurs in all treatment groups and not only the CETG. These results are very important because they show that the various cognitive abilities are differentially amenable to change and, most importantly, training interacts with the ability trained. It seems that QR abilities are more difficult to change and require more specialized training than CE abilities, which may change even without specialized training. In order to identify the role of Gf in the change of QR and CE abilities, a path model was tested. In this model, Gf was the only latent independent variable which was related to PreQR and PreCE factors, and these in turn were related to PostQR and PostCE factors. The possible direct relationship between Gf and PostQR, PostCE factors was also tested. This model was tested in the three treatment groups and it was found that at no age level was there a significant direct relationship between Gf and PostQR. On the

Table 1 Correlations between Pretest and Posttest Factors

	CTG				QRTG				CETG			
Factors												
Treatment / Age	10	12	14	16	10	12	14	16	10	12	14	16
PreQR / PostQR	.76	.93	.85	.91	.72	.75	1.00	.72	1.00	.99	.71	.84
PreCE / PostCR	.84	.61	.77	.71	1.00	.51	.94	.89	.78	.71	.66	.89

contrary, PostCE was significantly related to Gf within all age groups except 10-years-olds. These results support the previous conclusions, namely that postQR abilities are a function of preQR abilities only, and do not easily change. Of course, one should take into account the fact that the training procedure was very short. Nevertheless, CE abilities, with the same short training period, were more variable. In fact, PostCE abilities were a function of both Gf and PreCE abilities, and, therefore, they were determined by both specialized and general factors.

Finally, the path analysis showed no transfer effects from one SSS to the other. This indicates that SSSs preserve their autonomy after training, as expected. The fact, however, that CE abilities are a function of Gf may imply that at least in the case of the CE SSS, there might be relations between inductive abilities and causal reasoning.

Discussion and conclusions

The structural analysis presented above showed that training had limited effects. This was to be expected due to the short duration of the training provided. However, our goal was to identify the possible paths (that is, through general intelligence and the SSSs or only through the SSSs) through which cognitive change may occur rather than test effects of a particular long-term intervention method. Nevertheless, the correlations between the pretest and posttest factors showed that there were some changes in the rank ordering of the subjects' level of performance in certain ages in the various treatment groups. The correlations shown in Table 1 indicated that age was a factor that differentiated the relationship between pretest and posttest factors. This finding was further investigated by differentiating age from cognitive level effects. This was deemed necessary because age per se does not explain why training may or may not be effective. Cognitive level, however, indicates the person's capability to handle domain-specific tasks and his/her potential for further development. According to the construction of the tasks, each of them corresponded to a different cognitive level. Thus, level 0 denoted the subject's inability to solve correctly even the simplest of the tasks presented; level 1 denoted ability to handle correctly tasks involving abstract relations within a single set of data; level 2 denoted ability to relate two sets of data; level 3 denoted ability to deal with data forming a single system, and level 4 ability to deal with systems of systems. In other words, levels 1 and 2 are indicative of relational thinking whereas levels 3 and 4 indicate systemic thought. The subjects' cognitive level was determined according to the most difficult task within each SSS- battery they had solved correctly. The characterisation of subjects' cognitive level according to the most difficult of the SSS-specific tasks they had solved correctly, was based on rating scale analyses (Demetriou, Efklides, & Platsidou, 1993; Demetriou, Platsidou, Efklides, Metallidou, & Shayer, 1991) which had shown very high person- and item-separation reliability indexes. This implies that individuals occupying a given position on the scale, succeeded on items below this position and failed on those above.

As regards age effects, it was found that training was least effective at the age of 10 years. Only about 35% of them skipped level, that is, they moved to a higher level than their pretest one (31.73% for the CETG and 38% for the QRTG). The rest of the age groups in the CETG were successful at about 55% (53.37%, 56.17%, and 58.87% for 12, 14, and 16 years olds, respectively). The percentage for the QRTG were: 72.39, 69.15, and 80.66 for the respective age groups. The effect for the 10 years old was expected because children of this age are at the verge of acquiring the abilities addressed in this study. It is interesting that although at the pretest some of them did solve some of the problems, the abilities were not yet established, and consequently their posttest performance deteriorated. This finding replicates Kuhn, Amsel, and Loughlin's (1988) data showing that the task-relevant strategies may exist well before they are consistently used by the child. It is plausible that the child is not yet aware of the strategy used and, therefore, cannot match it with its formal presentation provided in the training context. Training was most effective at the age of 16, when adolescents have both the potential and the metacognitive background which allows them to monitor their own thinking vis-a-vis the solution provided in the training.

Our data also showed that, regardless of age, the person's cognitive level did have an effect on the progress made after training. Subjects of level 0 exhibited the highest rate of success of training (64.74% in the CETG, and 81.63% in the QRTG). The respective percentage for level 1 and level 3 subjects was 42.85 and 59.08 in the CETG, and 73.38 and 65.60 in the QRTG. Training was least effective with level 2 subjects. Only about 30% of them (33.47% in the CETG and 37.83% in the QRTG) progressed to the next level of thinking. The lack of progress of level 2 subjects implies that moving from the relational type of thinking to systemic thinking was difficult, and probably requires much more training than the one provided. On the contrary, change within the levels of either the relational (level 0 and 1) or the systemic thought (level 3) was easier to achieve.

Finally, in order to investigate the change of cognitive level within each age and treatment group we compared subjects of the same cognitive level across the four age groups. In this way we matched subjects in terms of their pretest cognitive level in the two SSSs and examined their gainings at the posttest. It was found that in the main change of cognitive level of the ability trained, occurred at the age of 12 and 16 years. Specifically, as regards QR abilities in the QRTG level 0 subjects (Mean pretest performance, M = 0.524) progressed to level 1 (Mean postest performance, M = 1.00) at the age of 12, and level 1 subjects (M = 1.368) progressed to level 2 (M = 2.421) at the age of 16. Spontaneous change of QR cognitive level 0 in the two treatment groups which did not receive training on QR abilities, namely the CETG and CTG, occurred at the age of 14. (Mean pretest performance, M = 0.917 and 0.688, and mean posttest performance, M = 1.208 and M = 1.438 for the CETG and CTG, respectively.) Similar trends were detected in the case of CE abilities in the CETG. These findings imply that training accelerated cognitive development for about two years. It seems that at the age of 12 level 0 subjects had the capacity required to meet the demands of level 1 tasks but they did not use their capacity spontaneously. Training helped them reach their potential. Non-trained subjects showed spontaneous acquisition of the same level two years later.

It is also interesting that training did not have the same accelerating effect in all age groups. At the age of 14, training had no particular effect on the ability trained. This implies that there are upper limits of the person's potential, which cannot be overcome by training alone. At the age of 12 and 16 there seem to be more general changes of potential, and training helps the person to take advantage of the new potential and formalize it within the particular SSS trained. Unfortunately, spontaneous change of cognitive level after the age of 16 could not be detected in this study, because older subjects were not included in it.

It can be concluded that this study clearly showed that QR and CE abilities are different in nature, as proposed by the theory. Cognitive change, however, may be a function of both general (Gf) and specific abilities or specific ability only. The QR SSS depended on its own specialized abilities, whereas CE abilities depended on both Gf and domain-specific abilities. The reasons why this happens are not clear.

Furthermore, our study showed that change of cognitive level through training is possible. Yet, training had limited effects, and these effects were achieved only in particular age groups, namely 12 and 16 years old. Change of cognitive level occurred even in non-trained groups, only at a later age. Training seems to be accelerating cognitive change; still, it cannot enhance abilities at such a point as to overcome developmental constraints pertaining to the type of ability trained and its structural complexity.

Finally, there were no clear transfer effects from one SSS to the other. However, there were some indications that training had a generalization effect, in the sense that in certain cases trained subjects performed better as regards non-trained abilities than control subjects, who had no training at all.

Study 2[1]

Study 1 showed that when we deal with thought abilities which are developmentally constrained, only part of the cognitive change observed may be attributed to training or practice. The limited success of training, however, could be a result of the particular training procedure used. The training procedure involved both step by step algorithmic presentation of the task solution as well as explanation of the general strategy required for the solution of related tasks. For example, in the case of QR training, we were explaining the use of ratio to estimate the rate of change, and in the case of CE training we were explaining the control of variables schema, i.e., the strategy of 'all other things being equal...'. This approach led to lengthy instructions, particularly in the case of level 3 and level 4 tasks. Therefore, the training approach adopted might be more suitable for level 1 and level 2 tasks, where the number of items to be handled was small, and less suitable for level 3 and level 4 tasks, where the processing load was high. In the latter case, subjects had to accommodate both lengthy instructions and complex sets of data.

Furthermore, since the training provided was both algorithmic and metacognitive, it was not clear which of the two aspects of training was more effective and if the form of instruction interacted with the ability trained. It may be the case that QR abilities are more algorithmic in nature whereas CE abilities are more metacognitively controlled. It may also be the case that age or gender interact with the instructional form.

Finally, the Gf tasks used in Study 1 were largely inductive in nature, and although they differed in terms of symbolic modality, they all required analytic abilities to be solved. Therefore, the role played by Gf in the change of CE abilities might be due to the fact that analytic abilities are probably required for the identification of the factors involved in the causal structure. Consequently, if general ability is defined not as Gf but as the broad factor underlying various forms of cognitive performance, then the direct relationship between Gf and CE might not be found. This assumption had to be further investigated, by using different types of the tasks than the ones used in Study 1.

In order to investigate the questions posed by Study 1 and understand better the mechanism and the factors influencing cognitive change, Study 2 was carried out.

Method

Subjects

The sample comprised 1127 subjects of 12, 14, 16, and 20 years of age. Specifically, there were 356, 416, and 314 students of 7th, 9th and 11th grade, respectively,

[1] This study was carried out by A. Efklides, M. Papadaki, G. Papantoniou, M. Koutsioumba, A. Demetriou, and G. Kiosseoglou

and 41 university students. The small number of university students was due to the fact that, despite the large number of them tested at the pretest (120), only a limited number of them had performance low enough to be selected for training. Both genders were about equally represented. All subjects came from low and upper middle class families.

Tasks

The QR and CE tasks used in Study 1 were also used in this study.

The General Intelligence tests (G) represented the verbal, the imaginal and the numeric symbolic systems. All tests, except the Number Series test were selected from the Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, & Harman, 1976). They were all time constrained tests.

Two of the tests addressed language-related abilities. They tapped semantic fluency. The first was the Synonyms test (SYN) and involved 10 items. The task was to produce as many synonyms as possible to each of the words presented. As synonyms were accepted words which had a relevant meaning and would be used in place of the word given in various contexts. The second semantic fluency task was the Opposites test (OP), which involved ten items and required production of words with opposite meaning to that of the words given.

There were two tests addressed to the numeric abilities. They tapped fluency in the use of arithmetic operations. These were: the Number Series test (NS) used in Study 1, and the Number Facility test (NF), which involved a large number of additions, subtractions, multiplications, and divisions.

Finally, there were three tasks addressing the imaginal abilities. They measured figural fluency, figural flexibility and visualization. The Symbols test (SYM) is a figural fluency task. It involved five items, which gave a word or phrase and required the subject to draw up to five different symbols to stand for it. The Toothpicks test (TP) is also a figural flexibility task. It involved five items which tapped spatial arrangements of a set of toothpicks. The subject was asked to present up to five different arrangements according to sets of specified rules. Finally, the Paper Folding test (PF) is a visualization task. It involved ten items which required mental folding and unfolding of pieces of paper.

Procedure

The same procedure as the one used in Study 1 was followed. As the pretest the General Intelligence (fluency) tasks were administered as well as the QR and CE tasks. The posttest was administered right after the training procedure. The training session and the posttest took place approximately one month after the pretest. The control group received no training.

Training

There were three forms of training: the algorithmic, the metacognitive, and the computer-assisted:

- The algorithmic training (ALG) had exactly the same form as the step by step presentation of the problems used in Study 1. For example, the QR training had the following form. In order to find out which plant, A or B, had the more productivity change, you need to divide the productivity of the plant when it is watered 4 times with the productivity of the plant when it is watered 2 times. That is, Plant A → 16:4 = 4, i.e., four times increase; Plant B → 20:5 = 4, i.e., four times increase. Therefore, A = B in terms of productivity change. The respective algorithmic CE training had the following form. In order to test the hypothesis about the effect of light on the productivity of plants A and B you need to make the following experiment:

	Plant	Light		Plant	Light
a 1.	A	Dark	b 1.	B	Dark
2.	A	Lighted	2.	B	Lighted

- The metacognitive training (MET) was verbal in nature and focused on the general process (or strategy) rather than on the details of the problem-solving procedures. For example, the QR metacognitive training had the following form. In order to find out which plant had the more productivity change, you must compute the rate of productivity change for each plant and then compare the two outcomes. The respective CE metacognitive training had the following form. In order to test the hypothesis about the effect of light on the productivity of plants A and B, you need to make an experiment in which you keep all the other factors the same and vary only the light.
- The computer-assisted (C-A) training made use of the algorithmic presentation, although accommodated for computer use. The computer application we used involved presentation of the problem and a number of questions, each of them tapping part of the solution of the problem. Only one question at a time was presented. Three or four alternative answers were provided to each question, and the subject had to select the one he/she thought was the correct one. There was immediate feedback on every selection made in the form of right - wrong. The selection procedure was terminated only when the subject made the correct choice. At this point, the feedback was more extensive and it included the principle on which the correct answer was based. In this way we wanted to make sure that even in case the correct selection was random or for wrong reasons, the subject would be informed about the reasons underlying the correct answer.

The training procedure of Study 2 lasted approximately the same time as the training of Study 1, that is about half an hour. However, this training differed from the one effected in the previous study in another respect: subjects were trained not necessarily on the level next to theirs. Instead, there were two levels of training: Level 1 training was administered to subjects who had not solved correctly the level 2 task , i.e. the QR2 or CE2 task for the respective treatment group. Level 2 training was administered to subjects who had solved correctly the level 2 task. The training tasks were similar in structure but differed in content from the respective level 2 and level 4 tasks of the pretest. Thus, subjects were trained either one or two levels above their own.

Results and conclusions

Training effects

Structural modelling analysis using the EQS (Bentler, 1993) statistical programme was applied in order to identify the possible training effects in the three treatment groups and the role of G (in terms of fluency) in cognitive change. The PreQR, PreCE, PostQR and PostCE factors as well as the G factor were first confirmed. However, when we introduced the path model, it was found that no single model could fit the data of all three treatment groups. This means that the training procedure used in this study produced results different from those of Study 1. A second difference from Study 1 is that in this study the G factor was identified with semantic fluency (see Figure 1) whereas in Study 1, the G fluid factor was inductive in nature and more related to the PreQR factor.

For expository reasons, first, the common core of the three treatment group models will be presented, and then, the peculiarities of each particular model and their meaning will be discussed. The model is presented in Figure 1.

As expected, the G factor loaded the PreQR and PreCE factors, and the PreQR, PreCE factors explained part of the variance of the PostQR and PostCE factors, respectively. However, what is worth noting in this model is the path connecting the PreQR with the PreCE factor, and the respective path between the PostQR and PostCE factors. These paths suggest that there are relations between the SSSs, other than those captured by the G fluency factor. These paths were stronger in the CETG, which means that in this group subjects used their QR abilities in order to solve the CE tasks. The correlation between PostQR and PostCE (.41) was weaker than the correlation between the PreQR-PreCE factors (.69), which indicates that the training of the CE abilities led to a relative independence from the QR abilities. This finding suggests that training leads to differentiation and specialization of thought structures.

As regards the peculiarities of the models fitting the three treatment groups, it is interesting that they were all related to the role of the G factor in the formation of the posttest factors. It was found that the G factor was related only to the factor corresponding to the SSS trained; that is, in the QRTG it was related to the PostQR factor, in the CETG it was related to the PostCE factor, and in the CTG it was related to both the PostQR and PostCE factors. This finding implies that subjects mobilized both their general ability and SSS-specific ability in order to respond to the training provided. The control group, which had no training at all, relied on general ability for the solution of all the posttest tasks.

These results, along with the previous finding showing a moderate relationship

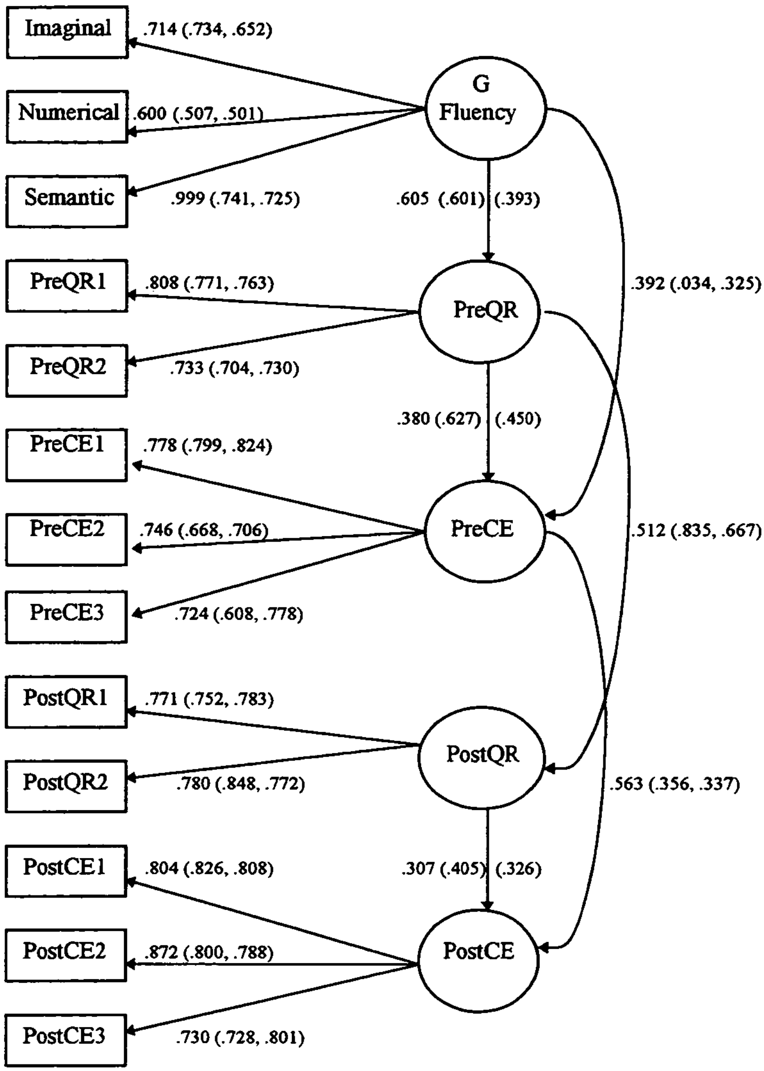


Figure 1 The common core model best fitting the data of Study 2

Note The symbols QR and CE stand for Quantitative-Relational and Causal-Experimental and the symbols Pre- and Post- denote pretest and posttest. The numbers out of the parenthesis represent the loadings of the QRTG. The loadings in the parenthesis represent the CETG and CTG, respectively

between postQR and PostCE factors, suggest that the ways through which cognitive change may occur are multiple. One way is to use the existing domain-specific structures and modify them in order to capture more complex relations (see Study 1). A second way is to use one's general ability and/or fluency in handling language, images and numbers along with domain-specific abilities (see Study 1). A third way is to use abilities of another SSS, along with the general and domain-specific abilities (Study 2). This may be considered as evidence of transfer of training. It is evidence of the synergistic nature of human abilities and activation of multiple task-related abilities, when the person does not have in his/her repertoire the specific structure best-suited for the processing of the task at hand. Therefore, our study showed that there is no single mechanism of cognitive change and, besides domain-specific abilities, there are general abilities of inductive or semantic nature which get involved in the acquisition of higher or more complex modes of thinking.

Effects of the form of training

The previous analyses led to the conclusion that the ability trained and the conditions of training (training a specific ability or not) may have an effect on the mechanism of cognitive change used. The next question to be answered was if the form of training applied (algorithmic, metacognitive, or computer-assisted) influenced the mechanism of change used. In order to answer this question, we tested the already identified model for each treatment group in the three sub-groups of training form within each treatment group.

The consistent finding in both the QRTG and CETG groups was that in the case of algorithmic training the importance of the G factor in the formation of the trained ability, i.e., the PostQR and PostCE factors in the respective treatment groups, weakened. The significance of the G factor was retained in the metacognitive and the computer-assisted training. This is a very interesting finding, which shows that the 'teaching' method may pose different demands on cognitive resources. The step-by-step algorithmic training poses the least demands on general ability, probably because the person applies a ready-made procedure; the person does not need to interpret the solution described, and simply applies it as it is. In the case of metacognitive training, the person has to interpret the verbal description of the strategy offered to the procedural form needed in order to process the particular tasks at hand. The computer-assisted training used in this study probably required 'integrational' processes, as well as proceduralisation, because the multiple-choice format may have prevented the direct grasp of the strategy presented. Therefore, the various forms of training seem to require different processes for their processing, and, consequently, their effectiveness may vary as a function of their processing demands.

Effectiviness of training

The results presented up to now regarded the effects of training on cognitive organization. In order to compare the three treatment groups with regard to the effectiveness of training the effect sizes were computed.

As shown in Table 2 trained subjects, either in the QRTG or in the CETG, tended to perform better than the CTG on the non-trained SSS as well as the trained one. This was particularly evident in the QRTG, in which performance on CE tasks improved significantly more than in the CTG. This may be indicative of transfer of training (as stated before) from the QR SSS to the CE SSS but not vice versa. Control group subjects improved at the posttest but not as much as the trained subjects.

The effects shown above were further analyzed in terms of effect sizes. Effect sizes were computed as *d*, following Rosenthal and Rubin (1986). As shown in Table 3, the effect size for the QRTG (vs. CTG) was .474 and .451 for the Level 1 training and .420 for Level 2 training (but only for level 4 tasks). These findings imply that Level 1 training, which was addressed to subjects functioning at the verge of the relational level of thinking, was effective for this type of thinking.

Level 2 training, on the other hand, was significant only for level 4 tasks and tended also to improve performance on the tasks of the relational level of thinking. The respective effect size for the QR tasks in the CETG (vs. CTG) was non significant. This finding suggests that QR abilities require specialized training in order to improve.

The effect size for CE tasks in the CETG (vs. CTG) and for Level 1 training was significant only in the case of level 2 tasks (d=.253). There was no transfer to the experimentation schema, Level 2 training helped them apply it to more complex sets of data and improve their overall performance. This was not the case of QRTG

Table 2 Mean Performance as a Function of Treatment Group, SSS, Testing and Task

		SSS				QR			
			Pretest tasks				Posttest tasks		
TG	N	1	2	3	4	1	2	3	4
QRTG	437	1.291	1.034	.478	.215	1.492	1.327	.636	.467
		(.857)	(.876)	(.669)	(.411)	(.780)	(.802)	(.780)	(.692)
CETG	426	1.200	.937	.521	.268	1.315	1.131	.589	.420
		(.868)	(.842)	(.673)	(.548)	(.859)	(.849)	(.750)	(.625)
CTG	235	1.289	1.009	.570	.170	1.255	1.064	.723	.328
		(.853)	(.887)	(.672)	(.377)	(.884)	(.877)	(.803)	(.612)

		SSS				CE			
			Pretest tasks				Posttest tasks		
TG	N	1	2	3	4	1	2	3	4
QRTG	437	1.087	.918	.490	.204	1.325	1.213	.822	.419
		(.814)	(.679)	(.719)	(.548)	(.804)	(.744)	(.838)	(.708)
CETG	426	1.296	1.089	.481	.115	1.315	1.282	.812	.441
		(.810)	(.733)	(.669)	(.319)	(.797)	(.752)	(.810)	(.731)
CTG	235	1.14	1.009	.519	.145	1.243	1.162	.706	.315
		(.812)	(.734)	(.730)	(.353)	(.850)	(.751)	(.792)	(.587)

where Level 2 training was effective only for level 4 tasks.

As regards transfer of training from one SSS to the other, it was found that training QR abilities helped improve CE abilities d=.495 (for Level 1 training) and d=.041 and d=.394 for level 1 and level 4 tasks, respectively (in the case of Level 2 training). There was no similar transfer of CE training to QR abilities, except for one case (Level 2 training, level 2 task d=.580). Therefore, the training of the QR SSS had an effect on both QR and CE abilities, whereas the training of the CE SSS was effective only for CE abilities.

Finally, as regards the effect of training form the ANOVAs performed showed that training form was not significant in the CETG, nor did it interact with age and gender. There was a marginal interaction with level of training, $F(2,400)=3.09$ $p=.047$. That is, algorithmic training was generally more effective than the other two forms of training, particularly so for Level 2 training ($M_{ALG}=1.336$, $M_{MET}=1.099$, $M_{C\text{-}A}=1.125$). In the QRTG there was a marginally significant age by training form by task interaction, $F(12,1203)=1.80$ $p=.044$. Algorithmic training ($M_{ALG}=1.070$) tended to be equally effective with metacognitive training ($M_{MET}=1.030$) and both of them were better than computer-assisted training ($M_{C\text{-}A}=0.925$). Furthermore, computer-assisted training was not effective at all in younger subjects (12 and 14 years old).

Table 3 Effect Sizes as a Function of the Ability Trained, Treatment Group, Level of Training and Task Level

			Task level			
Ability		Level of training	1	2	3	4
QR	QRTG vs. CTG	I	d=0.474 p<.001	d=0.451 p<.001	d=-0.007 n.s.	d=0.086 n.s.
		II	d=0.368 n.s.	d=0.390 p<.05	d=-0.172 n.s.	d=.042 n.s.
QR	CETG vs. CTG	I	d=0.143 n.s.	d=0.074 n.s.	d=-0.100 n.s.	d=0.050 n.s.
		II	d=0.090 n.s.	d=0.580 p<.01	d=-0.447 p<.05	d=0.220 n.s.
CE	QRTG vs. CTG	I	d=0.214 n.s.	d=0.063 n.s.	d=0.495 p<.001	d=-0.004 n.s.
		II	d=0.410 p<.05	d=0.366 n.s.	d=0.187 n.s.	d=0.394 p<.05
CE	CETG vs. CTG	I	d=0.192 n.s.	d=0.253 p<.05	d=0.100 n.s.	d=-0.131 n.s.
		II	d=0.498 p<.05	d=0.258 n.s.	d=0.465 p<.05	d=0.387 p<.05

It can be concluded, then, that changing the measurement of G (from inductive to fluency tasks) led to the identification of clear relations between the SSSs, and particularly effects of QR to CE SSS as well as effects of G fluency on the ability trained. However, the effect of G was moderated by the form of training applied, with algorithmic training making less use of G fluency than the other two forms.

General discussion

When we approach the teaching of thinking from the point of view of intelligence, the first thing we have to establish is the possibility of cognitive intervention and cognitive acceleration and, second, the possible constraints on this endeavor. It is understandable that the theory of intelligence one endorses, creates its own limitations, because no existing theory can sufficiently account for all the aspects of intelligent behavior and the development of intellectual abilities. What is even more important, theories do not always define the mechanism(s) of cognitive change, and even if they do, this does not mean that cognitive change through teaching follows the same road and uses the same mechanisms as the ones advocated by the theories (see Brainerd, 1983; Goossens, 1992). Therefore, cognitive intervention research should expand its scope and try to study the possible ways through which cognitive change may be induced, their cognitive demands and their interaction with developmental and individual differences factors.

Experiential Structuralism was the point of reference of the present research programme which aimed to answer some of the above questions. Structural modeling techniques were used for the identification of the underlying mechanism of change and it was found that there is 'synergism' of cognitive abilities, general and domain-specific. Furthermore, domain-specific abilities of one SSS may be used for the processing of tasks addressed to another SSS. Yet, this is not a reciprocal relationship, because in our study, whereas QR abilities contributed to CE abilities, the opposite was not true. Therefore, the concept of domain-specificity is a useful concept and it may show which specialized abilities change only with the use of domain-specific training and which benefit from instruction to other domains as well.

This finding shows that, despite the fact that in the 1980s a lot of studies reported lack of transfer (see De Corte, 1987), there may be transfer of training in certain cases, and we need to look for it. The role of general ability needs also to be reconsidered, because our research has shown that fluid intelligence (inductive ability) is also involved in cognitive change. This finding is in accordance with Klauer's work (see this volume), but it needs to be further investigated, because the effect of fluid intelligence is not the same across abilities.

The teaching method we use is also important because the various methods pose different demands on cognitive resources and may hinder our efforts for cognitive change. In our research we found that algorithmic training was the less demanding in processing resources and, consequently, more effective with younger subjects and with abilities not well developed. Metacognitive training was equally effective in the case of QR abilities but not in the case of CE abilities. In the latter case, subjects could not easily 'translate' the 'rule' into a procedural schema. Finally, the computer-assisted training we used was most ineffective in younger ages.

Finally, our work showed that there were also individual differences effects. These are important findings because they indicate that we should use different training forms depending on the ability to be trained and the persons' individual characteristics. One particular characteristic of children that proved significant is their cognitive developmental level, which represents a major constraint on the progress made. Therefore, although cognitive intervention work cannot display great successes, beginning to realize the factors that limit our efforts is a major step forward.

In conclusion, Experiential Structuralism, being an intelligence theory which integrates general and domain-specific abilities as well as a developmental perspective, allows the testing of critical hypotheses regarding cognitive change and the teaching of thinking. However, the effectiveness of the training of the domain-specific abilities advocated and its generalization to other conceptual and skill domains, has yet to be proven.

References

Anderson, J.R. (Ed.) (1981). *Cognitive skills and their acquisition.* Hillsdale, NJ: Erlbaum.

Anderson, J.R. (1982). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Bentler, P.M. (1993). *EQS: Structural equations program manual* (2nd ed.). Los Angeles, CA: BMDP Statistical Software.

Brainerd, C.J. (1983). Varieties of strategy training in Piagetian concept learning. In M . Pressley & J.R. Levin (Eds.), *Cognitive strategy research: Educational applications* (pp. 3-27). New York: Springer-Verlag.

Brown, A.L., Bransford, J.A., Ferrara, R.A., & Campione, J.C. (1983). Learning, remembering, and understanding. In P.H. Mussen, J.H. Flavell, & E.M. Markman (Eds.), *Handbook of child psychology* (Vol. 3, pp. 77-106). New York: Wiley.

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge: University of Cambridge Press.

Cattell, R.B. (1971). *Abilities: Their structure, growth and action.* Boston, MA: Houghton-Miffin.

De Corte, E. (1987). Introduction. *International Journal of Educational Research, 11*, 603-606.

Demetriou, A., & Efklides, A. (1987a). Towards a determination of the dimensions and domains of individual differences in cognitive development. In E. De Corte, H. Lodewijks, R. Parmentier, & P. Span (Eds.), *Learning and instruction: European research in an international context* (pp. 41-51). Oxford: Pergamon.

Demetriou, A., & Efklides, A. (1987b). Experimential structuralism and neo-Piagetian theories: Toward an integraded model. *International Journal of Psychology, 22*, 679-728.

Demetriou, A., & Efklides, A. (1994). Structure, development, and dynamics of mind: A meta-Piagetian theory. In A. Demetriou & A. Efklides (Eds.), *Intelligence, mind and reasoning* (pp. 75-109). Amsterdam: Elsevier.

Demetriou, A., & Efklides, A. (1997). Dynamic patterns of intra- and inter-individual change in the acquisition of complex thinking abilities from 10 to 17 years of age: A longitudinal study. *Scientific Annals of the School of Psychology, Aristotle University of Thessaloniki, 2,* 69-108.

Demetriou, A., Efklides, A., & Platsidou, M. (1993). The architecture and dynamics of developing mind. *Monographs of the Society for Research in Child Development, 58,* (5-6, Serial No. 234).

Demetriou, A., Gustafsson, J. E., Efklides, A., & Platsidou, M. (1992). Structural systems in developing cognition, science, and education. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development: Implications and applications for education* (pp.79-103). London: Routledge.

Efklides, A., & Demetriou A., & Gustafsson, J.-E. (1992). Training, cognitive change and individual differences. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development: Implications and applications for education* (pp. 122-143). London: Routledge.

Efklides, A., Demetriou, A., Moisiadis, C., & Marketos, A. (1992). Cognitive abilities and mathematical performance of 7th and 8th graders. *Psychology: The Journal of the Hellenic Psychological Society, 1,* 11-29.

Ekstrom, R.B., French, J.W., & Harman, H.H. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review, 8*, 477-531.

Fischer, K.W., & Farrar, M.J. (1988). Generalizations about generalization: How a theory of skill development explains both generality and specificity. In A. Demetriou (Ed.), *The Neo-Piagetian theories of cognitive development: Toward an integration* (pp. 137-171). Amsterdam: North-Holland.

Goossens, L. (1992). Training scientific reasoning in children and adolescents: A critical commentary and quantitative integration. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development: Implications and applications for education* (pp. 160-182). London: Routledge.

Gustafsson, J.-E. (1988). Hierarchical models of individual differences in cognitive abilities. In R.S. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 35-71). Hillsdale, NJ: Erlbaum.

Gustafsson, J.-E., Lindstrom, B., & Bjorck-Akesson, E. (1981). *A general model for the organization of cognitive abilities*. Göteborg: University of Göteborg.

Jöreskog, K. G., & Sörborn, D. (1986). *LISREL VI: Analysis of linear structural relationships by maximum likelihood and least square methods*. Mooresville, IN: Scientific Software.

Kuhn, D., Amsel, E., & O' Loughlin, M. (1988). *The development of scientific thinking skills*. New York: Academic.

Metallidou, P.I. (1996). *Cognitive abilities and their relation with cognitive-affective aspects of the self in adolescence: Structure and development*. Thessaloniki: Aristotle University.

Rosenthal, R., & Rubin D.B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin, 99*, 400-406.

Snow, R. (1994). A person-situation interaction theory of intelligence in outline. In A. Demetriou & A. Efklides (Eds.), *Intelligence, mind and reasoning* (pp. 11-28). Amsterdam: Elsevier.

Spearman, C. (1927). *The abilities of man*. New York: MacMillan.

Sternberg, R.J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge, MA: Cambridge University Press.

6

Fostering higher order reasoning skills: The case of inductive reasoning

K.J. Klauer

Introduction

A remarkably great number of cognitive strategies has been identified by researchers. All of these strategies can be ordered according to the range of their applicability: Whereas some strategies like the method of loci in memory techniques or the carry-over method in subtraction are appropriate in narrowly limited situations only, there are strategies which are applicable in nearly all situations where a subject is required to solve a problem or even to perform a task. Attentional strategies, for instance, are beneficial whenever an assignment has to be carried out and means-end analysis, backward planning or forward planning are possible whenever a problem has to be solved. Higher-order reasoning strategies are higher located on the generality dimension, they are more generally applicable.

In the beginning of cognitivistic research, many scientists believed that highly general strategies corresponded to high intelligence so that the more intelligent person would master more of these widely applicable strategies which would enable them to solve problems more efficiently. Research, however, in areas like problem solving or expert-novice comparisons showed that this expectation fell short of. Actually, it was found that experts or skilful problem solvers own a body of highly specialised knowledge enabling them to solve a problem directly and without any detours. Facts as these led researchers like Newell (1980) or Anderson (1987) to distinguish between strong and weak strategies. Means-end analysis, for instance, is a weak strategy though it is widely applicable. In a concrete problem it does not ensure at all that the problem will be solved. Today, researchers even assume a power-generality trade-off (Friedrich & Mandl, 1992; Perkins & Salomon, 1989). They assume that a strategy is the weaker the more generally applicable it is and, conversely, the stronger the less broad applicable it is. Actually, in a given problem situation like, for instance, the tower of Hanoi, it might be called for to make use of a certain bit of information in order to being able to solve the problem.

Such facts seem to open up rather pessimistic perspectives as far as teaching of higher-order reasoning strategies is concerned. The conclusion drawn in this chapter is to rely on strategies of middle-high generality because they certainly are appropriate only in certain domains but fairly strong within their fields of application. Such rather strong strategies are domain-specific, but the domains might be more or less important. The decision to confine ourselves to the study of domain-specific strategies implies two consequences. One is to leave too specialised strategies to teaching of subject-matter, i.e. to teaching of the various disciplines since these strategies prevail within subjects like reading, mathematics, geography, and so on. The second consequence is to abstain from intensively teaching very general means like reflective-thinking, general problem-solving, or attentional strategies at least to normal subjects because teaching such strategies may not lead to sufficiently large effects within the broad field where they should be beneficial.

One such domain-specific strategy seems to be the strategy of inductive reasoning. It is of medium broad applicability because it is limited to the induction of regularities, i.e. to the induction of concepts, principles, rules, and laws. It is evident that most of the subject matter in academic fields makes intensive use of regularities of these kinds so that mastering the strategy should be helpful in acquiring knowledge in those fields. Moreover, it will be shown that inductive reasoning has a close relationship to fluid intelligence which also enables us to acquire new knowledge and to solve problems. Finally, it will be shown that the strategy of inductive reasoning includes another strategy which is weak but far-reaching and which supports acquiring of declarative knowledge as such.

Theoretical background

To avoid misunderstandings, it might be useful to distinguish between inductive reasoning and inductive inferring. Inductive reasoning is aimed at detecting generalizations or regularities. If, for instance, a number of objects are given and if it is found that all of them are toys made of wood, then a generalization or regularity has been discovered. Should we extend this generalization to all toys by stating that all toys are made of wood, then we would have made use of an inductive inference, although falsely in this case. An inductive inference is going beyond the scope of experience by asserting something about the totality of relevant objects, for instance about all elements of an infinite set of objects. Philosophy and logic study the conditions that render an inductive inference valid or invalidate it. An inductive inference moves from specifics to a general statement whereas a deductive inference moves -under certain conditions- from generality to a specific assertion. In this contribution, the philosophical question of validity of such an inference is not dealt with. Instead, it deals with inductive reasoning, that is with discovering regularities or generalizations or commonalities in a given set of objects. In this way, inductive reasoning never exceeds our experience. One can assume that no inductive inference is possible without preceding inductive reasoning but it is sure that inductive reasoning is possible without any inductive inference.

Definition of inductive reasoning

The process of inductive reasoning leads to a product, namely a regularity or generalization. Psychology, however, is interested in the process itself, in the question how the process proceeds and which steps are essential such that a generalization is arrived at. It is evident that it is not possible to describe all of the individual varieties how people come to inductions in the infinite number of possible inductive problems. However, it seems possible to postulate theoretically which steps are the necessary basic or essential ones that must be taken if a generalization is to be discovered. Figure 1 provides such a theoretical definition which describes the essential steps considered to be necessary for inductive reasoning to take place. It postulates the relevant steps that are deemed to be necessary and sufficient in order to attain a generalization.

Such a theory includes a prescriptive approach to inductive reasoning: If the definition given actually specifies the necessary and sufficient steps of an induction process, then teaching of these steps should improve inductive reasoning of the trained subjects given that the subjects are not yet perfect in inductive reasoning. This latter hypothesis is most promising from a point of view of educational psychology so that it stimulated a number of research designed to both, testing the hypothesis and fostering the inductive reasoning of children.

The definition of Figure 1 is given in form of an incomplete mapping sentence as they are used in the tradition of Guttman. It combines three facets (or factors in a variance analytical sense), the facets A, B, and C. The facets deal with 3, 2, and 5 features, respectively. To generate an item requiring inductive reasoning, three features have to be combined, one out of each facet. That is the reason why, according to the definition given, the set of possible inductive-reasoning items consists of 3 x 2 x 5 = 30 subsets.

As to the inductive procedure, essential steps to find out regularities and irregularities are specified in Figure 1 by detecting similarity or difference or both,

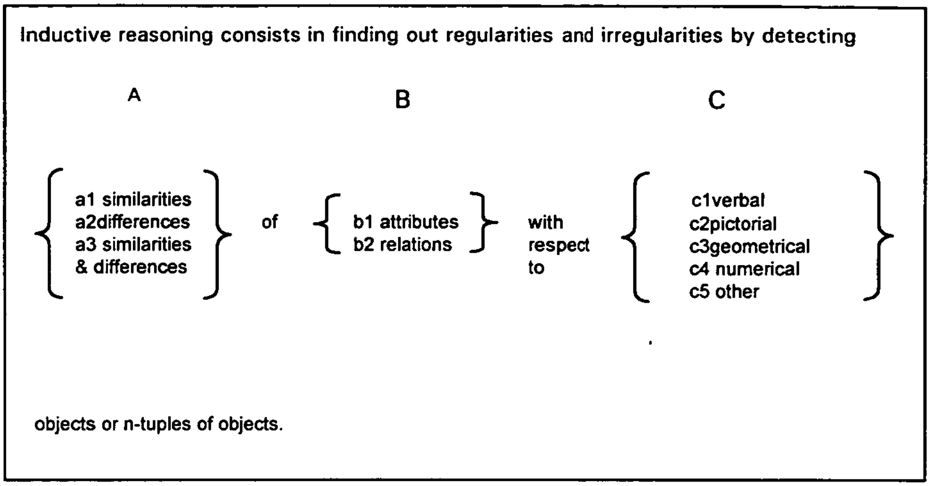


Figure 1 Definition of inductive reasoning

similarity and difference, i.e. by facet A. Detecting similarity means finding out commonalities, detecting difference means finding out variance, disturbances of order. Now, similarity can be defined as partial identity. Objects are partially identical when they have one or more attributes or features in common and pairs, triples etc. of objects are partially identical when they have one or more relationships among them in common (facet B). Hence, it is supposed that inductive reasoning requires the subjects to look for similarity and/or difference with respect to attributes of or with respect to relationships among objects. According to this definition, inductive reasoning always implies abstract thinking: Comparing objects with respect to certain attributes requires abstraction as well as comparing pairs of objects with respect to relations holding between them.

In terms of formal logic, attributes are predicates with one argument and relationships are predicates with two or more arguments. Hence, attributes and relationships exhaust all possibilities of talking about objects. This indicates a rather far-reaching aspect of inductive reasoning which will be dealt with later on.

Finally, facet C of Figure 1 refers to classes of material the objects in question belong to. Here are classes depicted as it could be if a test of inductive reasoning should be constructed. Using this definition, a content valid test of inductive reasoning could be developed representing the various subsets of items in prespecified proportions. However, the material class could be divided into other subclasses, for instance according to subject matter taught in schools as English, math, history, social studies etc. In this case, items requiring inductive reasoning with subject matter would result.

The basic subsets of inductive problems

Facets A and B are the central ones, because they specify the essential processes

Table 1 Types of Inductive Reasoning Problems

Subset	Facet Identification	Problem Formats	Cognitive Operation Required
Generalization (GE)	a_1b_1	class formation class expansion finding common attributes	similarity of attributes
Discrimination (DI)	a_2b_1	class exclusion	differences in attributes
Cross Classification (CC)	a_3b_1	4-fold scheme 6-fold scheme 9-fold scheme	similarity & difference in attributes
Recognizing Relationships (RR)	a_1b_2	series completion ordering series analogy	similarity of relationships
Differentiating Relationships (DR)	a_2b_2	disturbed series	differences in relationships
System Construction (SC)	a_3b_2	matrices	similarity & difference in relationships

indispensable in inductive reasoning and because facet C only refers to classes of material. Facets A and B constitute 3 x 2 = 6 subsets of inductive problems. As Table 1 shows, each of the six subsets requires certain specific processes.

The first column contains the names given to the six subsets of inductive problems as well as the abbreviations to designate the subsets. Each subset can be solved by specified procedures depicted in the last column. The names of the subsets are chosen such that they allude to the processes required. The second as well as the last column refer immediately to the definition of Figure 1. Only the column 'Problem Formats' needs some annotations. Here are typical problem formats given as they can be met in intelligence and other cognitive tests involving inductive reasoning. Whereas three typical item formats can be found for Generalization (GE) and for Recognizing Relationships (RR) problems, only one typical item format with Discrimination (DI) or with Differentiating Relationships (DR) could be found. For instance, a typical DI-item is the class exclusion demanding to find out one object that does not fit in with the others because it does

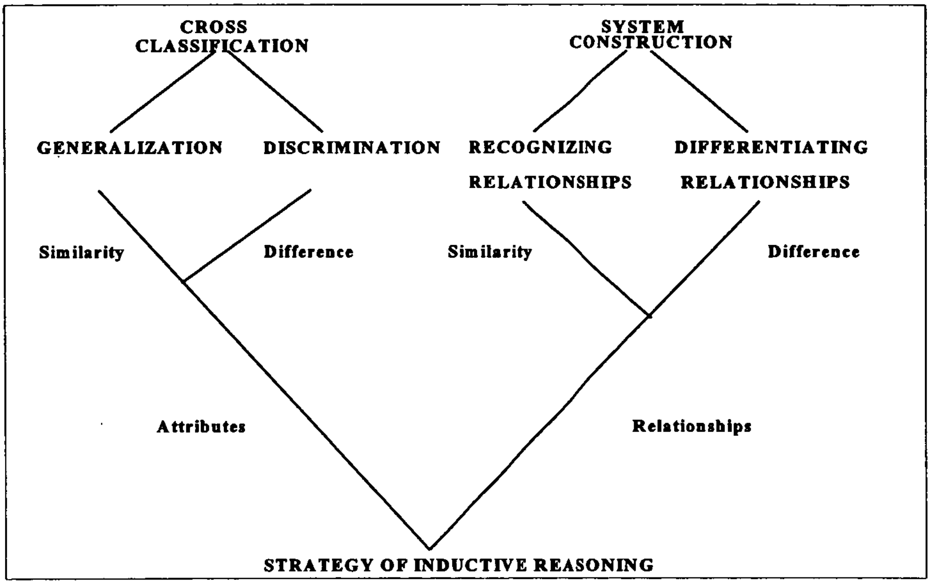


Figure 2 The genealogy of tasks in inductive reasoning

not share the critical feature. The column 'Problem Formats' should not be misunderstood in the way that it depicts exhaustively all possible inductive problem formats. Actually, there are much more formats possible and particularly used in different domains of subject matter.

Traditionally it is assumed that inductive reasoning is required by the item formats classification, analogy, series, and matrix (cf. Büchel & Scharnhorst, 1993; Goldman & Pellegrino, 1984; Greeno, 1978; Holzman, Pellegrino, & Glaser, 1983; Sternberg & Gardner, 1983; Van de Vijver, 1991; Whitely, 1980). Generalization, discrimination and cross-classification are varieties of classification tasks. They deal with scanning attributes. Analogies, series and matrices deal with scanning relationships. Thus, it is evident that the definition presented here does fit in the framework traditionally considered to be the inductive frame of reference but it additionally provides an explanation why these and similar varieties of items require inductive reasoning.

There are certain relationships between the classes of inductive reasoning problems which are represented by Figure 2. The figure shows two branches, the attributional branch and the relational branch, and both branches are broken up again into two sub-branches each of which are bound together in the tops. Formally, the figure is characterized by three symmetry axes due to corresponding procedures. A GE item, for instance, corresponds to a RR item insofar as in both cases detecting similarity is demanded. The only difference consists in whether the same attribute or the same relationship has to be discovered. In this way, each of the three left-hand classes has a corresponding right-hand class. These relationships can give rise to transfer hypotheses such that, for instance, training to solve GE problems should lead to transfer in RR problems and vice versa. Moreover, Figure 2 allows one to predict additional transfer, for instance should a training in GE problems furthermore transfer to DI as well as to CC problems. Transfer hypotheses of this kind have been tested in the first 'wave' of our research in inductive reasoning which yielded so encouraging results that we decided to develop structured programmes for a training in inductive reasoning. These programmes enabled different researchers to test various additional hypotheses.

Transfer Hypotheses

Intra-Inductive Transfer

Different transfer hypotheses can be derived from the theory. Based upon the genealogy in Figure 2, several intra-inductive hypotheses are possible. Take, for instance, a discrimination training with, say, verbal material. Such a training should transfer to numerical or figural discrimination tasks or vice versa. Moreover, it should transfer to cross-classifications because discriminating is a prerequisite of cross-classifications. Also transfer on generalization problems would be expected since generalization requires the antonymous procedure which subjects should use in order to test their solution of a discrimination problem. More complex transfer could be expected as transfer on the other 'branch' of Figure 2. In this case, a shift from scanning attributes to scanning relationships (or in the reverse order) is required which may imply an additional difficulty.

In view of the definition of Figure 1, inductive items can vary according to the three facets of processes, predicates, and material. Putting things together, an overall hypothesis of intra-inductive transfer can be derived such that the amount of transfer monotonously increases with the amount of common facets between training and testing items or that it decreases in a monotonous way with the amount of different facets between training and test problems (hypothesis 1). It should be highest when both kinds of items share all facets, for instance with transfer from figural generalization items in the training situation and such items in the test situation - provided the figures in both situations are different, otherwise we would not deal with transfer at all. On the other hand, it can be hypothesized that intra-inductive transfer would be smallest if training and test items only share one facet. If they share none of the facets A and B, we would deal with transfer on non-inductive problems, that is with extra-inductive transfer. Whether or not such transfer on non-inductive tasks is conceivable is an open question. Certain possibilities will be discussed below.

Furthermore, hypotheses concerning intelligence tests as dependent variables are possible. Inspecting the item formats depicted in Table 2, it is easy to realize that many intelligence tests contain a smaller or larger proportion of items requiring inductive reasoning. Actually, ever since Spearman (1923) psychologists have been interested in inductive reasoning. Spearman himself was convinced that inductive reasoning plays a major role with respect to general intelligence. Later on, different researchers found overwhelming empirical evidence for this assumption (cf. Gustafsson, 1984; Snow, Kyllonen, & Marshalek, 1984; Thurstone, 1938; Undheim & Gustafsson, 1987). However, factor g of general intelligence must not be identified with a factor enabling to reason inductively. It also provides intellectual processes which are not inductive in nature, for instance deductive reasoning. Possibly, the same holds true for fluid intelligence as it is conceived by Cattell (1963). But fluid intelligence is normally measured using test items requiring inductive reasoning. This concerns particularly Cattell's Culture Fair Tests (CFT) and Raven's Progressive Matrices though the latter only make use of one item format, the matrix, and although both tests only apply meaningless geometric-figural material.

Accordingly, another intra-inductive hypothesis can be formulated, namely that a training in inductive reasoning should transfer to intelligence tests if the tests consist of inductive-reasoning items (hypothesis 2). Learning of subject matter in academic settings correlates with typical measures of inductive reasoning and fluid intelligence (see Csapó, 1989). That is the reason why even transfer from a training in inductive reasoning on acquiring knowledge in schools could be expected. One can assume that this indirect transfer will be smaller in amount than the direct transfer on measures of fluid intelligence, particularly since the correlation between fluid intelligence and learning is only medium high.

A substantial proportion of subject matter in various fields is also governed by regularities, i.e. by generic concepts, by rules, laws and general structures (Csapó, 1989; Curtis & Reigeluth, 1984; Greeno, 1978; Holzman et al., 1983; Norman, Gentner, & Stevens, 1976). In the typical case, declarative knowledge contains generic concepts to be learned, moreover rather often rules, laws or other generalizations. To understand declarative knowledge, schemata have to be retrieved or even modified. In all these instances inductive reasoning should be beneficial so that training in inductive reasoning should lead to substantial transfer effects in acquisition of declarative knowledge. Hence it can be assumed that there is a direct transfer effect of a training to reason inductively on acquiring subject matter which demands inductive reasoning. For these kinds of subject matter a definitely higher transfer effect should be expected because of the joined direct and indirect transfer.

Trans-inductive transfer

Even if there were no regularities at all to be learned, a certain transfer effect could also be expected with learning of declarative material: How declarative knowledge ever might be stored, for instance as propositions or as a mental map, it is an assembly of elements, characterized by attributes and tied together by relationships. In any case, acquiring declarative knowledge demands scrutinising the relevant attributes of the objects given and examining the important relationships among the objects. Now, inductive reasoning as it is conceived in Figure 1 implies careful inspecting of both attributes and relationships, even if they are singular events. That is the reason why a training to reason inductively also should lead to a smaller transfer effect in those parts of declarative knowledge that are non-inductive in nature.

Summarising, we can assume distinguishable influences of a training of inductive reasoning on acquisition of declarative knowledge: The indirect path via fluid intelligence to learning, and the direct path from inductive reasoning to acquiring regularities, and its 'side-effect' leading to acquiring connections between elements, their attributes and relationships. Thus, an improvement in inductive reasoning also should enhance learning of singular facts.

Taking these different influences together, one can assume that there should be a remarkable transfer of inductive reasoning on the acquisition of declarative knowledge, possibly even a larger one than on measures of fluid intelligence (hypothesis 3).

Training programmes

In order to test these hypotheses, three training programmes have been developed:

- Programme I is designed to teach inductive reasoning to children between 5 and 8 years (Klauer, 1989a).
- Programme II addresses 10 to 13 years old children (Klauer, 1991a).
- Programme III is designed for 14 to 16 years old youths (Klauer, 1993a).

Programme I is nonverbal insofar as all problems are presented with concrete objects, pictures of concrete objects or with symbols so that children are not required to read or write at all. Of course, they have to understand the instructions and questions. The manual of ProgammeI is available both in an English language version (Klauer & Phye, 1994) and in a Dutch version (Klauer, Resing, & Slenders, 1996). Detailed lessons in Dutch language have been developed and experimentally tested by De Koning and Hamers (this volume). Moreover, these authors describe ProgammeI in some detail. Programmes II and III make use of texts in German language so that they cannot be used with children not speaking German. Nevertheless, the Dutch manual of Progamme I published by Klauer et al. (1996) also contains information and advice concerning Progamme II.

Programme III is especially designed for learning-disabled subjects. Programme I can be used with normal children, with younger bright children as well as with older learning-disabled children. As will be shown in the sequel, many empirical evidence is available demonstrating that all these groups take advantage from the programmes. Programme II is particularly successful with normal and bright children, i.e. with children participating in regular or in higher streams or courses. It is too difficult for learning-disabled children.

As far as the construction principles are concerned, all three programmes are similarly built. All of them consist of 120 problems, each problem presented on a separate plate. In all cases, the 120 problem cards are distributed to 10 lessons of 12 items each. It is, of course, not necessary to teach exactly these 12 items per

Table 2 Stages of a Paradigmatic Training

Phases	Material	Training Goals
Declarative Knowledge	Concrete or familiar problems with clear attributes and relationships. Paradigmatic problems (i.e., problems suited for detecting the relevant aspects)	Knowing attributes of objects and relationships between objects. To know that each object has certain attributes shared by other objects. Knowledge that an object is related to other objects and that pairs of objects can share relations. Knowing that these pieces of knowledge are helpful in solving problems
Procedural Knowledge	A great number of problems of each type to ensure enough repetitions	Internalising and automating the declarative and procedural knowledge
Strategic Knowledge	Problems from diverse areas, progressively dissimilar to the paradigmatic problems	Transfer in various contexts and practice: (a) recognition of the basic problem structure, even in unfamiliar contexts, (b) spontaneous application of solution and control strategies adapted to the particular problem

lesson though this distribution proved to be useful commonly. There may, however, be reasons to teach less or more problem plates dependent on the abilities of the children. Each programme offers 20 problems of the six basic problem patterns depicted in Table 1 or Figure 2.

The first lesson starts with the easiest kinds of inductive problems (Generalization GE and Recognizing Relationships RR) but with attending to attributes as well as to relationships. This technique of contrasting follows Piagetian principles and prevents the children from building a set such that, for instance, only attributes have to be looked for. The distribution of items of one kind starts with massed practice first and distributed practice in longer intervals later. In this way, each lesson has a special emphasis on particular problems but repetition of earlier problems always takes place. The last two lessons are devoted to a general repetition of all problem types.

Programmes II and III do not make use of concrete material. Instead, they refer to simple everyday-life problems in the introductory phase. With this exception, programmes II and III are also constructed as ProgammeI. Important to realize is, however, that subject matter as it is taught in schools also belongs to the everyday lives of these older pupils. A great proportion of the problems of these programmes are taken from various subject matter the children meet in school. In this way, both programmes make use of 40 verbal, numerical and figural problems, respectively.

It is intended, that the pupils learn to transfer the paradigms to problems they encounter in school as well as outside of school.

Testing the theory: Review of research

Intra-inductive transfer

According to the statements in the previous section, it is expected that the amount of transfer increases in a monotonous way with the amount of common facets between training and transfer tasks. This is equivalent to the following formulation of hypothesis 1: The amount of transfer monotonously decreases with transfer distance, if transfer distance is an increasing function of the relative frequency of different facets.

Referring to Tversky (1977), the author developed such a measure of transfer distance which is applicable whenever sets of training and transfer problems are defined using the facet approach (Klauer, 1989b). Based upon n = 20 training studies with a pretest-posttest design and training and control groups, 73 effect sizes could be calculated as well as the corresponding 73 transfer distances. Effect size d was estimated using formula $d = (M_{exp} - M_{control})/s_{control}$ and transfer distance was calculated using a measure derived in Klauer (1989b). In these twenty studies, the problems of the training sessions typically belonged to one or more of the six classes of the inductive reasoning family whereas the test problems belonged to the same or another class of inductive reasoning items. In no case, however, the two sets of problems were identical so that transfer was always required. Thus, the prerequisites were given to test hypothesis 1. The product-moment correlation between transfer effect and transfer distance was $r=-0.50$ ($p<0.001$), and the regression coefficient was $b=-0.32$ ($p=0.001$). The result was additionally confirmed by a cross validation. The 37 pairs of values with odd numbers were used to calculate a multiple correlation and the prediction equation. Based upon this equation, the expected transfer effects d' were calculated for the 36 pairs of values with even numbers. Now, a correlation $r_{dd'}= 0.52$ ($p=0.001$) between the actual and the predicted effect sizes resulted indicating that the theoretical predictions describe reality to a considerable amount. Thus, hypothesis 1 can be accepted.

Note that these findings imply a domain-specific effect of the training. In most cases, the training material was highly specific because only one or two classes of items were included in the training. It is important to see that such a training does not transfer equally well to any kind of inductive reasoning problems. Instead, transfer decreases linearly when transfer distance increases, i. e. whenever the transfer problems get more and more dissimilar to the training problems. Several conclusions can be drawn from this result:

- The facets as they have been described theoretically actually play a major role in determining the difficulty of an inductive problem.

- Transfer between two problems increases linearly whenever the facet descriptions of the problems get more similar and decreases whenever the facet characteristics get more dissimilar.
- Training to solve certain kinds of inductive-reasoning problems entails domain-specific transfer even within the set of inductive problems.
- The training effect cannot completely be explained by more general factors as motivation, warming up, novelty, special attention, Hawthorne effects etc. Such effects should have led to more general instead of domain-specific effects.
- If one wants to improve inductive reasoning generally, it seems to be inevitable to make use of a representative sample of all kinds of inductive problems.

The results reported so far belonged to the first wave of our research in inductive reasoning which was distinguished by the use of only one or two classes of inductive problems. The results were so encouraging that - following the last conclusion just mentioned - the three training programmes described above had been developed. Each of the programmes offers a representative or content valid sample of inductive problems, suitably constructed for the children or youths the individual programme is designed for.

In a second wave of experiments hypothesis 2 was tested. It states that training with one of the programmes leads to positive transfer on intelligence tests if the tests only consist of inductive items.

The latter condition holds true for all variants of Raven's Progressive Matrices (CPM, SPM, APM), furthermore for Cattell's Culture Fair Tests (CFT), for the German version of the Columbia Mental Maturity Scale (CMMS) as well as for subtests of some other test instruments. Table 3 provides an overview of the pertinent research where Progamme I has been made use of and where it has been contrasted to a no-training control group. The table contains the effect sizes yielded with 17 experiments. The last column of Table 3 contains the effect sizes d_{corr}, corrected for possible pre-experimental differences ($d_{corr} = d_{post} - d_{pre}$). There is only one negative effect size d_{corr} (-0.05) indicating that the control group gained slightly more than the training group. However, this difference is by no means statistically significant. Of the 23 effect sizes, 22 have a positive sign. Even the simple sign test would establish a significant training effect. Note that not in each training experiment a significant effect could be found, however 13 of the 17 experiments led to at least one significant effect. Hammill and Larsen (1974) proposed that a training programme - if its application in practice were to recommend - should lead to significant effects in at least 50 % of the cases, when contrasted to a not trained group. This criterion is certainly met by Progamme I.

A few of the experiments allowed us to calculate two effect sizes, be it due to two dependent variables or to two different training methods. Two effect sizes, however, bring about a problem when synthesising the results by a meta-analysis. Introducing the 23 effect sizes of Table 3 into a meta-analysis would artificially increase the number of subjects involved and would ignore that some of the effect sizes are not from independent studies. In order to avoid these difficulties, two

Table 3 Contrasts between Training Groups and Untrained Control Groups (Programme I)

Experiment	Subjects	Test	N	d_{corr}
Bornemann a (Klauer & Phye, 1994)	kindergarten children	CAT-K	27	1.12*
Bornemann b (Klauer & Phye,1994)	kindergarten children	CPM CAT-K ST 2+3	20	0.54 0.48
Bornemann c (Klauer & Phye, 1994)	kindergarten children	CPM CAT-K ST 2	22	0.80* 1.08*
Bornemann d (Klauer & Phye, 1994)	1st graders	CPM CFT	279	0.80* 0.65*
Johnen (Klauer & Phye, 1994)	kindergarten children	CAT-K ST 2+3 CPM	19	-0.05 1.66*
Ziesemer (Klauer, 1992b)[a]	1st graders	CFT 2	20	1.00*
Alizadeh et al. (Klauer, 1991a)[a]	kindergarten children	CPM	50	0.41*
Kolmsee (Klauer, 1991a)	1st graders	CAT 1-3	20	0.45
Beck et al. (1993)	school kindergarten	CAT-K	140	0.11
Masendorf (1994)	learning disabled children	CFT 2	20	1.29*
Angerhoefer et al. (1992)[a]	learning disabled children	CFT 2	20	0.24
Phye & Sanders (1992)	kindergarten children (USA)	WISC-R similarities	15	0.82*
Hamers et al. (1995)	ethnic minority children (NL)	SPM	28 27	0.86* 0.47*[b]
Tomic (1995)	1st (Dutch: 3rd) graders (NL)	CFT	34	0.74*[c]
Beck et al. (1995)	ethnic minority (Turkish) children	CFT ST 2-5	60	0.67*
Windg.-Fischer (Klauer, 1996b)	2nd graders	CFT 2	45	0.76* 0.59*[b]
Tomic & Klauer (1996)	1st (Dutch: 3rd) graders (NL)	CFT 1 ST 2-5	34	0.18

*Groups significantly different $p \leq 0.05$ [a] see Klauer & Phye (1994)
[b] another training method [c] d (instead of d_{corr}) estimated using the F-value
ST: Subtest; CAT: German version of the Cognitive Abilities Test

meta-analyses have been calculated. Both are based upon one single effect size per experiment.

Whenever there are two effect sizes available in an experiment, meta-analysis 1 is based upon the first and meta-analysis 2 is based upon the second effect size. Both meta-analyses (as well as the following ones) make use of the random effects model where the resulting overall effect-size estimation is referred to as delta (Hedges & Olkin, 1985). Here are the results:

- Meta-analysis 1: n = 17 studies, N = 853, delta = 0.58 ± 0.07, $p < 0.001$.
- Meta-analysis 2: n = 17 studies, N = 852, delta = 0.55 ± 0.07, $p < 0.001$.

Synthesising all of the accessible studies comprising more than 800 children leads to a clear conclusion. Programme I yields effect sizes of about half a standard deviation. In the mean, a trained child outperformed a not trained child by half a standard deviation. One can expect effect sizes of this order whenever intelligence tests are given which consist of inductive reasoning items like classifications, analogies, series, and matrices.

In both meta-analyses, the seemingly considerable variance among the effect sizes is completely explained by sampling errors. Hence, there is no need to look for moderating variables. Most of the training experiments used only small samples of children so that considerable variation was to be expected. Moreover, there are different dependent variables which fact of course also contributes to variation of effect size measures.

So far, 17 studies have been performed with Programmes II and III. Table 4 gives an overview of these training experiments and their effects on measures of intelligence. Again, about 800 children participated as trained or not trained subjects and again, one effect size turned out to be negative, however not significantly and small in amount. Out of the 17 studies, twelve yielded a significant training effect so that the criterion of Hammill and Larsen (1974) has been met again. The two meta-analyses produced comparable results.

- Meta-analysis 1: n = 17 studies, N = 797, delta = 0.47 ± 0.07, $p < 0.001$.
- Meta-analysis 2: n = 17 studies, N = 797, delta = 0.46 ± 0.07, $p < 0.001$.

The estimated mean effect sizes delta are slightly slower than those concerning ProgammeI but the difference seems to be negligible. It can be concluded that both training programmes entail a mean improvement of about half a standard deviation. Again, according to the meta-analyses, the variance between the effect sizes of Table 4 can completely be explained by sampling variation so that there is no need to search for moderating variables as, for instance, kind of dependent variable or kind of subjects involved.

'Use it or lose it', is a sound principle in strategy learning. It can be supposed that a learned strategy will be lost if it is not used regularly (Duffy & Roehler, 1989). Hence, one cannot expect a strategy to be maintained over a long period of time if it is not used meanwhile. A short strategy training of about ten hours will probably not last very long unless students make use of it for themselves or teacher encourage students to apply it with their subject matter. In any case, it would be

Table 4 Contrasts between Training Groups and Untrained Control Groups (Programmes II and III)

Experiment	N	Programme	Subjects	Test	d_{corr}
Igelmund (Klauer, 1991a)	20	II	fourth graders	CAT[a]	0.43*
Benicke (Klauer, 1992a)	24	II	5th & 6th graders	APM	0.15
Conrad (Klauer, 1996a)	50	II	5th & 6th graders	SPM	0.80*
Hintermaier (Klauer, 1993b)	61	II	6th graders higher track	SPM/ APM	0.48*
Esenwein (Klauer, 1993d)	61	II	6th graders higher track	CFT	0.59*
Munz (Klauer, 1993c)	66	II	6th graders higher track	CFT	0.31
Kuschel (Klauer, 1992a)	30	II	5th & 6th graders grammar school	SPM/ APM	0.82*
Jackmuth (Klauer, 1994)	71	II	6th graders grammar school	APM	0.78*[b] 0.38*[b]
Hellenbrandt (Klauer, 1995b)	32	II	5th graders grammar school	CFT	0.23
Beck, Lübking & Meier (1995)	68	II	6th graders comprehensive school	CFT BBT	-0.04 0.15
Meiss (Klauer, 1996c)	82	II	6th graders grammar school	SPM	0.31*
Golzem (Klauer, 1996c)	60	II	8th graders grammar school	CAT[a]	1.13*
Igelmund (Klauer, 1993e)	32	III	learning disabled	SPM	0.67*
Esser (Klauer, 1993e)	32	III	learning disabled	CFT	0.19
Souvignier (1996)	29	III	learning disabled	SPM	0.45*
Werk (Klauer, 1995a)	34	III	learning disabled	SPM	0.32*[c]
Rademacher (Klauer, 1995a)	45	III	learning disabled	SPM	0.59*[c]

*$p \leq 0.05$ [a] only the inductive subtests
[b] different training procedures
[c] contrasted to a not inductively trained group
BBT: German test of school relevant cognitive abilities

Table 5 Durability of the Training Effects (Interval in Months)

Experiment	Subjects/ Treatments	Test	N_1	d^1_{corr}	Inter-val	d^2_{corr}	N_2
Johnen (Klauer & Phye, 1994)	kindergarten children/ Progamme I vs kindergarten	CPM	19	1.15*	4	0.99*	19
Tomic & Klauer (1996)	first graders/ Progamme I vs classes	CFT 1 ST 2-5	34	0.18*	4	0.31*	34
Igelmund (Klauer, 1993e)	learning disabled/ Progamme III (2 training methods) vs classes	SPM	46	0.76* 0.56*	5	1.24* 0.12*	38
Hager & Hasselhorn (1993)	school kindergarten/ Progamme I vs classes	CFT 1 ST 3-5	30	0.06	5	0.62*	28
Bornemann (Klauer & Phye, 1994)	first graders/ Progamme I vs classes	CPM	279	0.89*	6	0.90*	219
Conrad Klauer (1996a)	5th & 6th graders/ reading training plus Progamme II or Progamme II alone vs classes	SPM	45	0.77* 0.80*	6 6	0.81* 1.09*	41
Bornemann (Klauer & Phye, 1994)	kindergarten/ ProgammeI vs kindergarten	CPM	22	0.89*	7	0.34	19
Igelmund (Tomic & Klauer, 1996)	first graders/ ProgammeI vs perceptual training	CPM	23	1.13*	9	1.51*	22
Bornemann (Klauer & Phye, 1994)	first graders/ Progamme I vs classes	CPM	31	1.10*	22	0.97*	31

*Training group significantly better than contrast group ($p \leq 0.05$);
ST: subtest; d^1_{corr}: corrected effect size immediately after training

valuable when even a short training of ten hours would lead to effects lasting at least some months. Table 5 gives an overview of the durability of the effects on intelligence tests after four to 22 months. N_1 and d_1 refer to the results immediately after training, N_2 and d_2 to the results several months later.

Because there is more than one effect size in some cells, two meta-analyses will be performed for each test time. Immediately after training, the following values resulted.

- Meta-analysis 1: n = 9 studies, N = 535, delta = 0.78 ± 0.11, p < 0.001.

- Meta-analysis 2: n = 9 studies, N = 535, delta = 0.75 ± 0.10, $p < 0.001$.

For the delayed posttest the given months later, these results were yielded.

- Meta-analysis 1: n = 9 studies, N = 451, delta = 0.85 ± 0.10, $p < 0.001$.
- Meta-analysis 2: n = 9 studies, N = 451, delta = 0.74 ± 0.13, $p < 0.001$.

Between 83 % and 100 % of the variance of the effect sizes can be explained by sampling errors. Hence, there is no or only little systematic variance among the effect sizes of each column of Table 5.

Taking all in all, one can conclude that the effects of the training programmes are not too transient phenomena but that they last at least for several months.

Trans-inductive transfer

In a third wave, a number of studies have been performed where a training has been ensued by a lesson on a certain subject matter and by a subsequent criterion-referenced test. This kind of experiments allows one to test hypothesis 3 which expects that a training in inductive reasoning also transfers to learning of subject matter and even to a greater extent to learning of subject matter than to tests of measuring fluid intelligence.

Hypothesis 3 can be tested in connection with Table 6. This table gives an overview over 13 training experiments. Comparing the inductively trained with the corresponding not trained group allows one to assess the training effect to learning of the subject matter. Table 6 provides information about the subject matter in question, the effect sizes yielded with a test of fluid intelligence and the effect sizes concerning learning of the material. Accordingly, two meta-analyses were calculated.

- Meta-analysis concerning intelligence: n = 13 studies, N = 599, delta = 0.47 ± 0.08, $p < 0.001$.

This result fits well into the results of Table 4. Since nine of the experiments of Table 6 also made use of Programmes II or III, similar results could be expected.

- Meta-analysis concerning learning of subject matter: n = 13 studies, N = 599, delta = 0.74 ± 0.09, $p < 0.001$.

Even a short inspection of Table 6 shows that considerable transfer to learning cannot be doubted. The meta-analysis concerning the effect sizes on learning displays a considerably high mean effect size. The effect on learning is substantially higher than that on intelligence as can be evidenced by Wilcoxon's signed rank test for paired values ($T = 17$, $N = 13$, $p < 0.03$, two-sided). One can conclude that hypothesis 3 stands the empirical test.

Finally, with both meta-analyses the hypothesis that the effect sizes are homogeneous is not to be refused. The variability of the effect sizes can completely be traced back to sampling errors so that moderator or suppressor variables probably do not play a substantial role. That means, among others, that the training is comparatively effective irrespective of the ubjects involved, the various kinds of learning material or subject matter.

Table 6 Transfer of a Training to Reason Inductively on Fluid Intelligence and Learning of Subject Matter

Experiment	Subjects (grade)	N	Subject matter	Test	d_{corr} Intelligence	d Learning
Igelmund (Klauer, 1993e)	learning disabled (9)	46	mathematics, operators	SPM	0.53*[a]	0.66*[a]
Esser (Klauer, 1993e)	learning disabled (8-9)	36	mathematics, operators	CFT	0.19	1.10*[b]
Benicke (Klauer, 1992a)	high school (5)	36	mathematics, symmetry	SPM	0.15	0.55[b]
Conrad (Klauer, 1996a)	high school (5-6)	60	reading comprehension	SPM	0.79*[a]	0.81*[a]
Hintermaier (Klauer, 1993b)	advanced track (6)	61	foreign language learning ability	APM	0.48*	0.65*
Esenwein (Klauer, 1993d)	advanced track (6)	61	spelling rule	CFT	0.59*	0.90*
Munz (Klauer, 1993c)	advanced track (8)	66	physiology of hearing	CFT	0.22	0.88*
Jackmuth (Klauer, 1994)	grammar school (6)	70	biology: digestion, predatory animals	APM	0.58*[a]	0.84*[a]
Hellenbrandt (Klauer, 1995b)	grammar school (5)	32	grammar: syntax	CFT	0.23	0.60*
Igelmund (Tomic & Klauer, 1996)	first graders	23	math: relations between numbers	CPM	1.13*[a]	0.80*[a]
Tomic (Tomic & Klauer, 1996)	first graders	34	math: relations between numbers	CFT[c]	0.54*	0.31
Werk (Klauer, 1995a)	learning disabled (9-10)	34	physics: inertia of mass	SPM	0.32*	1.37*
Rademacher (Klauer, 1995a)	learning disabled (9-10)	40	physics: inertia of mass	SPM	0.59*	0.34

[a] Mean of training conditions, [b] d refers to the standard training,
[c] without subtest 'Maze', * $p \leq 0.05$

Discussion

Theory of inductive reasoning

The theory of inductive reasoning depicted earlier turned out to be useful in enabling one to produce empirically sound predictions. Thus, it can be concluded that its main statements are empirically based. In particular, it can be assumed that regularities can be detected by comparing objects with respect to common features or by comparing pairs, triples etc. of objects with respect to relationships commonly holding among them. Generalizations or rules can be found by looking whether or not attributes or relationships are shared among objects or pairs, triples etc. of objects. And such a regularity cannot be taken for granted if no such commonality can be discovered.

Note, however, that the theory does not imply assumptions about the actual processes subjects are performing when they reason inductively. There may be a great variety of different mental processes performed by different subjects or by the same subject under different conditions. On the other hand, the theory assumes that anybody will finally make use of comparing processes, i.e. of scanning objects with respect to common or different attributes and/or relationships, if he or she is successful in detecting a regularity or in disproving an only seeming regularity. Looking for similarity and difference among objects or n-tuples of objects is taken as a necessary part of the process of inductive reasoning.

The procedure to test the theory was to train subjects to better perform these comparing processes. Actually, that procedure generally led to an improvement of the subjects' performances when solving inductive problems was called for.

Transfer of training

Three aspects are theoretically of importance as far as transfer is concerned. First: It has been shown that transfer can be predicted based upon an inspection of the facet characteristics of the problems involved in training and test situation. Obviously the facet characteristics do not describe only surface attributes of the items but such features which determine the kind of how the items are to be solved. If, for instance, attributes have to be scanned, then fairly different processes are called for in comparison to the situation where relationships are to be looked for. The same holds true if either similarity or difference has to be recognized. Hence, the facet characteristics mirror essential ways of how to solve a given problem. The more two problems differ with respect to their facet descriptions, the more different solving procedures are optimal and the less transfer will take place.

Second: To a minor extent and less systematically, transfer on non-inductive problems can also be expected due to concomitantly ongoing processes. Careful inspecting of visually presented material, step by step analysing of complex problems, allocation of attention and other mental resources are systematically improved during a training in inductive reasoning. Such broader skills might also be transferred though one can assume that the probability of such more general transfer will depend on the surface similarity between training and test situations. The less this similarity is, the less transfer of that kind will probably occur. However, this way farther transfer cannot be excluded. Taking this kind of transfer into consideration, one can expect that - at least in some cases - the transfer effect of a training in inductive reasoning might be composed of two components, a larger inductive component and a smaller more general component.

Third: Taking all in all, a domain-specific transfer effect of such a training is to be expected. In the typical case, the effect on inductive skills will be higher than the effect on more general skills, particularly if the more general skills cannot easily be transferred to any pertinent problem.

As far as inductive reasoning is concerned, though, one must not underestimate its possible range of applicability. In everyday life, in social situations as well as with academic learning generalizations of various kinds have to be detected. If only the training was sufficiently broad, the subjects should have learned to apply the skills of inductive reasoning to a broad range of inductive problems.

Inductive reasoning and fluid intelligence

It has amply been shown that all three programmes foster such intelligence tests which are known as marker variables of fluid intelligence. This is another confirmation that fluid intelligence enables one to reason inductively. However, whether or not the reverse is also true, namely that fostering inductive reasoning also improves fluid intelligence cannot be stated for sure. All of the intelligence tests used so far as dependent variables consisted more or less exclusively of items requiring inductive reasoning. Hence it is possible that there are certain non-inductive but fluid intelligence tests and it has not yet been proven whether or not such performances are also fostered due to a training. This question seems still to be open. Learning, however, is commonly agreed to be under the influence of fluid intelligence, too, and there cannot be any doubt that learning has been improved by a training.

Impact of the training on learning

Our original theory had to be refused: This theory assumed that a training to reason inductively immediately fosters fluid intelligence which in turn enhances learning of subject matter. The idea that the influence on learning was only mediated by the influence on fluid intelligence cannot be hold any more, because learning is definitely more improved by the training than fluid intelligence. There must be an additional immediate influence of the training on learning.

As has been supposed above, certain possible effects have to be taken into consideration. Many a subject matter consisting of declarative knowledge require inductive reasoning. They contain regularities or generalizations not only in form of rules or laws but also in form of generic concepts. Moreover, in a schema-theoretic view declarative knowledge is understood by integrating it into available schemata or by differentiating and modifying such schemata. Now, schemata are generalizations. Hence, subsuming an individual case to a schema requires inductive reasoning. Inductive reasoning is required as well when modifying a schema or creating a new one. Taking all this into consideration, one can conclude that there are good reasons to expect a positive influence of inductive reasoning on acquiring declarative knowledge.

Even if there are non-inductive components of declarative knowledge, a smaller positive impact of a training in inductive reasoning on these components could also be expected. If the training facilitates encoding and processing of those components which are amenable to inductive reasoning, then subjects will have more mental resources available for other material which might be learned as well and more or less incidentally.

Duration of the effects

It could be shown that the training effects lasted at least between four and 22 months after the end of the training. Nevertheless, it seems to be wise not to overemphasise these results, since any cognitive skill that is not used continuously will fade out in time. However, one can conclude that many subjects actually made use of the inductive skills even after the end of the training: Should that not have been the case, the effects would have disappeared already, probably as a matter of weeks.

We did not expect that the effects of such short a training would last as long as they obviously do. Two measures seem appropriate in order to stabilise the effects in time. First, there is the possibility to repeat the training some years later and at a higher level of development because there are three programmes available which are constructed according to the same principles but for children of different ages. The second possibility is to ask teachers that they make use of the principles realized in the training programmes when teaching their regular lessons (cf. Klauer, 1991b). This latter possibility implies an adequate training of teachers in teaching inductive reasoning.

Recommendations for educational practice

The three programmes presented here belong to those few programmes which meet the criterion established by Hammill and Larsen (1974). They lead to significant results in more than half of the training experiments though, in the typical case, small-scale experiments were used. Under these circumstances one can take the responsibility for using the programmes even within school settings, particularly since it has been shown that the programmes lead to a remarkable transfer to regular school learning. As has just been mentioned, it would be wise to repeat the training at a later time using one of the programmes for older pupils.

Should it be possible that teachers actually apply the principles of inductive reasoning to their classroom teaching, then it might be possible to dispense with a special training. In this case the programmes could be reserved for children with special educational needs as, for instance, high-ability children (cf. Klauer, 1992c), learning-disabled children (cf. Angerhoefer, Kullik, & Masendorf, 1992), ethnic-minority children (cf. Hamers, De Koning-De Jong, & Pennings, 1995) and so on.

References

Anderson, J.R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94*, 192-210.

Angerhoefer, U., Kullik, U., & Masendorf, F. (1992). Denk- und Rechenförderung lernbeeinträchtigter Kinder: Multivariate Änderungsbeurteilung mittels Prädiktions-KFA [Fostering thinking and arithmetic with learning disabled children]. *Psychologie in Erziehung und Unterricht, 39*, 190-195.

Beck, M., Lübking, M., & Meier, U. (1995). Die Bielefelder Studien zum Denk training von Klauer [The Bielefeld studies with Klauer's training to reason]. In W. Hager (Ed.), *Programme zur Förderung des Denkens bei Kindern* (pp. 294-308). Göttingen: Hogrefe.

Beck, M., Lüttmann, B., & Rogalla, U. (1993). Wenn Du denkst, Du denkst... Eine Untersuchung der Effektivität des Klauer'schen Denktrainings [If you think you think... A study of Klauer's reasoning training]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 25*, 297-306.

Büchel, F.P., & Scharnhorst, U. (1993). Training des induktiven Denkens bei Lern- und Geistigbehinderten [Training of inductive reasoning with mildly and severely mentally reatarded children]. In K. J. Klauer (Ed.), *Kognitives Training* (pp. 95-123). Göttingen: Hogrefe.

Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1-22.

Csapó, B. (1989). Integrating the development of the operational abilities of thinking and the transmission of knowledge. In H. Mandl, E. de Corte, N. Bennett, & H. F. Friedrich (Eds.), *Learning and instruction* (Vol. 2.2, pp. 85-94). Oxford: Pergamon Press.

Curtis, R.V., & Reigeluth, C.M. (1984). The use of analogies in written text. *Instructional Science, 13*, 99-117.

De Koning, E., & Hamers, J. H. M. (1995a). *Programma inductief redeneren I. Achtergronden en toelichting* [Programmme inductive reasoning 1: Backgrounds]. Utrecht: ISOR.

De Koning, E., & Hamers, J. H. M. (1995b). *Programma inductief redeneren I. Lessen* [Programme inductive reasoning 1: Lessons]. Utrecht: ISOR.

Duffy, G.G., & Roehler, L.R. (1989). Why strategy instruction is so difficult and what we need to do about it. In C.B. McCormick, G. Miller, & M. Pressley (Eds.), *Cognitive strategy research* (pp. 133-154). New York: Springer.

Friedrich, H.F., & Mandl, H. (1992). Lern- und Denkstrategien [Learning and thinking strategies]. In H.F. Friedrich & H. Mandl (Eds.), *Lern- und Denkstrategien, Analyse und Intervention* (pp. 3-54). Göttingen: Verlag für Psychologie.

Goldman, S.R., & Pellegrino, J.W. (1984). Deductions about induction: Analyses of developmental and individual differences. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 149-197). Hillsdale, NJ: Erlbaum.

Greeno, J.G. (1978). Natures of problem-solving abilities. In W.K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 5, pp. 239-270). Hillsdale, NJ: Erlbaum.

Gustafsson, J.-E. (1984). A unifying model for the structure of intellctual abilities. *Intelligence, 8,* 179-203.

Hager, W., & Hasselhorn, M. (1993). Induktives Denken oder elementares Wahrnehmen? [Inductives reasoning or elementary perception?]. *Empirische Pädagogik, 7,* 421-458.

Hamers, J. H. M., De Koning-De Jong, E. & Pennings, A. (1995). *Stimulering van de denkvaardigheden van allochtone leerlingen* [Stimulating thinking skills of minority children]. Utrecht: ISOR.

Hammill, D. C. & Larsen, S. C. (1974). The effectiveness of psycholinguistic training. *Exceptional Children, 41,* 5-14.

Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic.

Holzman, T.G., Pellegrino, J.W., & Glaser, R. (1983). Cognitive variables in series completion. *Journal of Educational Psychology, 75,* 603-618.

Klauer, K.J. (1989a). *Denktraining für Kinder I* [Training to reason for children I]. Göttingen: Hogrefe.

Klauer, K.J. (1989b). Die Messung von Transferdistanzen [Measuring of transfer distances]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 21,* 146-166.

Klauer, K.J. (1991a). *Denktraining für Kinder II* [Training to reason for children II]. Göttingen: Hogrefe.

Klauer, K.J. (1991b). Erziehung zum induktiven Denken. Neue Ansätze der Denkerziehung [Educating to reason inductively. New ways in cognitive education]. *Unterrichtswissenschaft, 19,* 135-151.

Klauer, K.J. (1992a). 'Bottom up' oder 'top down'? Über die Transferwirkungen zweier Strategien zum Training des induktiven Denkens [Bottom up or top down? On two strategies of training to reason inductively]. *Sprache & Kognition, 11,* 91-103.

Klauer, K.J. (1992b). Problemlösestrategien im experimentellen Vergleich: Effekte einer allgemeinen und einer bereichsspezifischen Strategie [Problem solving strategies in an experiment. Effects of a general strategy and a specific strategy]. In H.F. Friedrich & H. Mandl (Eds.), *Lern- und Denkstrategien* (pp. 57-78). Göttingen: Hogrefe.

Klauer, K.J. (1992c). Teaching inductive thinking to highly able children. *European Journal for High Ability, 3,* 164-180.

Klauer, K.J. (1993a). *Denktraining für Jugendliche* [Training to reason for youths]. Göttingen: Hogrefe.

Klauer, K.J. (1993b). Über die Auswirkungen eines Trainings zum induktiven Denken auf zentrale Komponenten der Fremdsprachenlernfähigkeit [On the effects of a training to reason inductively on central components of the ability to learn foreign languages]. *Zeitschrift für Pädagogische Psychologie , 7,* 1-9.

Klauer, K.J. (1993c). Über den Einfluß eines Trainings zum induktiven Denken auf den Erwerb und die Nutzung der Lernstrategie des 'Networking' [On the influence of a training to reason inductively on the acquisition and use of the learning strategy 'networking']. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 25,* 333-352.

Klauer, K.J. (1993d). Induktives Denken beeinflußt das Rechtschreiblernen Inductive reasoning influences learning to spell]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 25,* 352-365.

Klauer, K.J. (1993e). Denken und Lernen bei Lernbehinderten: Fördert das Training des induktiven Denkens schulisches Lernen? [Thinking and learning with learning disabled subjects: Does a training to reason inductively foster learning in school?]. *Heilpädagogische Forschung, 29,* 50-66.

Klauer, K.J. (1994). Transferiert der Erwerb von Strategien des induktiven Denkens auf das Erlernen eines schulischen Lehrstoffs? [Does the acquisition of strategies to reason inductively transfer to learning of subject matter?]. *Zeitschrift für Pädagogische Psychologie, 8,* 15-25.

Klauer, K.J. (1995a). Weitere Erprobung des 'Denktrainings für Jugendliche' in der Oberstufe der Schule für Lernbehinderte [Further test of the 'Training to Reason for Youths' with learning disabled students]. *Heilpädagogische Forschung, 21,* 157-170.

Klauer, K.J. (1995b). Induktives Denken erleichtert die Konstruktion analoger Satzstrukturen [Inductive reasoning facilitates the construction of analogical sentence patterns]. *Sprache & Kognition, 14,* 221-227.

Klauer, K.J. (1996a). Denktraining oder Lesetraining? Über die Auswirkungen eines Trainings zum induktiven Denken sowie eines Lesetrainings auf Leseverständnis und induktives Denken [Training to reason or training to read? On the effects of a training to reason inductively and a training to read on reading comprehension and inductive reasoning]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 28,* 67-89.

Klauer, K.J. (1996b). Teaching inductive reasoning: Some theory and three experimental studies. *Learning and Instruction, 6,* 37-57.

Klauer, K.J. (1996c). Begünstigt induktives Denken das Lösen komplexer Probleme? [Does inductive reasoning support solving of complex problems?]. *Zeitschrift für Experimentelle Psychologie, 43,* 85-113.

Klauer, K. J., & Phye, G. (1994). *Cognitive training for children. A developmental program of inductive reasoning and problem solving.* Seattle, Wash: Hogrefe & Huber.

Klauer, K.J., Resing, W.C.M., & Slenders, A.P.A.C. (1996). *Cognitieve training voor kinderen. Ontwikkeling van het inductief redeneren bij kinderen* [Cognitive training for children. Development of cognitive reasoning with children]. Göttingen: Hogrefe.

Masendorf, F. (1994). Förderungstypen des induktiven Denkens und des räumlichen Vorstellens bei lernbeeinträchtigten Kindern: Eine metatypologische Mehrstichproben-KFA [Types of fostering learning disabled children by inductive reasoning and spatial perception]. *Psychologie in Erziehung und Unterricht, 41,* 14-21.

Newell, A. (1980). One final word. In D.T. Tuma & F. Reif (Eds.), *Problem solving and education* (pp. 57-84). Hillsdale, NJ: Erlbaum.

Norman, D. A., Gentner, D. R., & Stevens, A. L. (1976). Comments on learning schemata and memory. In D. Klahr (Ed.), *Cognition and instruction* (pp. 177-196). Hillsdale, NJ: Erlbaum.

Perkins, D.N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher, 18,* 16-25.

Phye, G.D., & Sanders, C.E. (1992). Accessing strategic knowledge: Individual differences in procedural and strategic transfer. *Contemporary Educational Psychology, 17,* 211-223.

Schneider, W., & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84,* 1-66.

Snow, R.E., Kyllonen, P.C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 47-103). Hillsdale, NJ: Erlbaum.

Souvignier, E. (1996). *Die Förderung räumlichen Denkens bei lernbeeinträchtigten Schülern* [Fostering spatial reasoning with learning disabled children]. Köln: Universität Köln.

Spearman, C. (1923). *The nature of intelligence and the principles of cognition.* London: MacMillan.

Sternberg, R.J., & Gardner, M. (1983). Unities in inductive reasoning. *Journal of Experimental Psychology: General, 112,* 80-116.

Thurstone, L.L. (1938). *Primary mental abilities.* Chicago, IL: University of Chicago Press.

Tomic, W. (1995). Training in inductive reasoning and problem solving. *Contemporary Educational Psychology, 20,* 483-490.

Tomic, W., & Klauer, K. J. (1996). On the effects of training inductive reasoning: How far does it transfer and how long do the effects persist? *European Journal of Psychology of Education, 11,* 283-299.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327-352.

Undheim, J.O., & Gustafsson, J.-E. (1987). The hierachical organization of cognitive abilities: Restoring general intelligence through the use of linear structural relations (LISREL). *Multivariate Behavioral Research, 22,* 149-171.

Van de Vijver, F. (1991). *Inductive thinking across cultures: An empirical investigation.* Tilburg: University of Tilburg.
Whitely, S. E. (1980). Modeling aptitude test validity from cognitive components. *Journal of Educational Psychology, 72,* 750-769.

7

Teaching inductive reasoning: Theoretical background and educational implications

E. De Koning & J.H.M. Hamers

Introduction

Official education in the Western civilization arose from the increasing amount and complexity of knowledge and skills required. The implicit, incidental learning of customs, traditions, ways of thinking and skills changed into an explicit intentional form of learning. Plato, the Greek philosopher, laid the foundation for the content of schooling, based on an analysis of the structure of knowledge (Nisbett, 1993). The logical structure of knowledge can be described by abstract, formal rules. Plato saw schooling as a way of teaching students these rules in domains where they are clearly evident, namely in logic and mathématics. The teaching of formal rules in these domains would, in Plato's view, induce a direct transfer of their application into specific domains.

Aristotle, who was Plato's student, expanded this view by making the formal rules of deduction that were to be taught more explicit (Fong, Krantz, & Nisbett, 1993). Applying these rules to a general statement clearly leads to a true or untrue conclusion. The statement can take the form of two premises: 'if p then q' - 'if you wear glasses then you can see well.' The rules are the modus ponens: 'p therefore q' - 'you wear glasses, therefore you see well!' - and the modus tollens: 'not q therefore not p' - 'you don't see well, therefore you don't wear glasses.' This example shows that logical and correct conclusions do not always have to be valid. Apart from this conditional form of deductive reasoning, there are also two related syllogistic forms. The first, linear form has the structure: 'if $p > q$ and if $q > s$, then $p > s$' - 'if you are taller than Jean and Jean is taller than Peter, then you are taller than Peter.' Applying this rule is also called transitive reasoning and it forms the basis of the skill for making series. The second, categorical form of syllogistic deductive reasoning has the structure: 'if $p = q$ and if $s = p$, then $s = q$' - 'if all children are human and if all babies are children, then all babies are human.' This form of reasoning is the basis for making categories.

Training in the use of the formal, deductive rules has occupied a central place in the standard curricula up to the 20th century (Nisbett, Fong, Lehman, & Cheng, 1993). Although the Romans and Humanists added the teaching of grammar and of Latin and Greek, respectively, little has changed in the fundamental thinking that knowledge has its own internal structure and that it is important to teach students the formal rules underpinning this foundation. One can ask whether people reason by applying formal rules. Or, do people form mental representations of formal rules unrelated to context, which guarantee a broad transfer? Essentially, this concerns the relation between man (psyche) and the structure of knowledge (logic).

In this chapter, we will first discuss this relation. Philosophical schools such as empiricism and rationalism form the foundation for two distinctive paradigms in which psychologists give shape to their research into the relation between man and knowledge. This originally led to contrasting views about how people use knowledge, how they reason, and how education can take advantage of these factors. Research into each of these paradigms has led to a blurring of the previously rather sharp boundaries. In this respect, the so-called pragmatic deductive- and inductive reasoning schemes (Nisbett, 1993) were used, which can give substance to this synthesis. These context-free schemes appear, in general terms, to represent the abstract level of human reasoning. They are responsible for the orderly processing and application of knowledge and the organization and reorganization of (stored) knowledge.

Secondly, we will discuss the balance between induction and deduction, and the importance of induction for young children who must absorb a relatively large amount of knowledge to be able to order their environment. The probabilistic character of induction demands the use of 'intuitive statistical schemes' (Nisbett, 1993), in which the balance between the acquired knowledge and the adaptation of the knowledge structure based on the new knowledge is closely monitored. Adults employ the inductive pragmatic schemes of the 'law of large numbers' and 'regression' (Fong, Krantz, & Nisbett, 1986; Nisbett, Krantz, Jepson, & Kunda, 1983). It is assumed that young children first have to form an image of 'group structures' and 'row structures', which can be expressed as nominal and ordinal/rational scales (see Halford, 1993), if we use statistical terminology.

Thirdly, we will discuss Klauer's (this volume) inductive reasoning programmes which are linked to the concept of group and row structure schemes. Training of school classes with visual, numerical and verbal inductive reasoning problems has revealed the general character of the reasoning schemes (Hamers, De Koning, & Sijtsma, 1998).

Fourthly, we describe theories and models in the domain of 'understanding texts'. Research in this domain is marked by a change in paradigms too. A reading comprehension programme (De Koning & Hamers, 1996) gets its place in the paradigm shift by applying the reasoning schemes to the understanding of texts.

In general the description of theoretical issues of the teaching of inductive reasoning is accentuated in this chapter. The disposition of 'teaching reasoning' will be described by making a historical 'walk' along some philosophical schools and psychological theories. The philosophical parts may not be elaborated enough to satisfy rigorous philosophers. However, a more thorough description does not fit in the main lines of this chapter. In this context it is worth noting that Csapo (this volume) tackles the same subject although he follows a slightly different approach to the problem. The empirical parts of the chapter serve the pupose of examples of possible studies. The results of the various (pilot) studies described in this chapter give an indication of the role that group and row structure schemes can play in the relation between man and knowledge. Klauer (this volume) presents a more elaborate overview of the empirical foundation of his inductive reasoning programmes.

Man and knowledge

Rationalism and Empiricism → Piaget's theory of cognitive development and Learning theory

The way in which the relation between man and knowledge is defined in philosophical schools has had far-reaching consequences for the views on the cognitive development of children and for the way in which education is set up (Case, 1996). Empiricists such as Locke and Hume, and rationalists such as Kant, see man and knowledge as two separate entities. Knowledge is considered as information with its own logical structure which students will master in the course of their development.

Rationalists see the acquisition of knowledge as a process of (active) knowledge construction: 'Knowledge is acquired by a process in which order is imposed on sensory data by the innate human faculty of reason, not merely detected in these data' (Case, 1996, p.78). Piaget's theory of cognitive development (Piaget, 1970), rooted in rationalism, lies closest to the classic view that people (adults) employ formal rules. Piaget assumed a logical structure of knowledge and he assumed that adults apply deductive rules to abstract, even invalid, premises. According to his view, children develop by continually interacting with and manipulating their environment. In doing so, they form mental representations (schemes) concerning the relations within and between people and/or objects and/or events. Or children induce order into their environment by classification and seriation, which they then use as a basis for further actions. Piaget therefore assumed a certain balance between induction and deduction. The nature of their actions and mental representations shows a fixed sequence, related to age. This discontinuous development is related to biological maturity and, thus, cannot be influenced by training. He assumed that most human environments meet children's minimum needs for environmental input of this sort, and thus have equal developmental potential. An important difference with the classic (Greek-Roman) education is, therefore, Piaget's notion that teaching formal rules is not worthwhile. A theme such as teaching reasoning would not fit into his theory.

Empiricists see the acquisition of knowledge as a process of more or less passive knowledge reconstruction: 'The process of knowledge acquisition is one in which the sensory organs first detect stimuli in the external world, and the mind then detects customary patterns or conjunctions in these stimuli.' (Case, 1996, p.75). In these behavioral theories of learning the mind is considered as a 'tabula rasa' which can be filled with knowledge by inducing stimulus-response associations (conditioning of behavior) and stimulus-stimulus associations (accumulation of facts). This view is supported by the disappointing results from the research into transfer (Gick & Holyoak, 1987; Hayes & Simon, 1977; Thorndike, 1919). Several researchers (Griggs & Cox, 1982; Manktelow & Evans, 1979; Reich & Ruth, 1982) concluded that deduction can only be guided by domain-specific or context-related rules. In this tradition, knowledge is seen as a complex of separate domain contents. Education should be directed at each of these domains. Cumulative learning and development are therefore equivalent concepts according to these early empiricists (Case, 1996). And, as such, research in the empiricist tradition took a step further away from the classic view of the content of education. A theme such as teaching reasoning would not fit into domain-tied thinking.

It can be concluded that empiricists and rationalists differ in the way in which they define knowledge, in terms of domain-specific or domain-transcended structures, respectively. In addition, there is a difference in the way in which empiricists and rationalists see people's acquisition of knowledge (also known as the performance-competence dilemma). Both schools are, however, united in their view that the process of induction (either passive or active) plays an important role in the way in which children learn about their environment.

Piaget's theory of cognitive development and Learning theory → Neo-Piagetians' theories of cognitive development and Information processing theories

As a sequel to the research schools in the previous section, in this section we consider the elaborations of the Piagetian tradition and the learning theory in neo-Piagetian and the information processing theory, respectively. The question whether a theme as 'teaching reasoning' fits in these theories is a central one. We try to answer this question by looking at the research designs adopted by both theories. Although they are very different, they both use comparable designs: within children across tasks (research into domain-specificity), between children within tasks (research into general development sequences), within children within tasks (research into training possibilities and environmental factors).

Along the lines of rationalism, neo-Piagetians assume that even if tasks are founded on identical logical knowledge structures, students of the same age (and development phase) may vary in the extent to which they correctly perform the tasks (Sternberg & Berg, 1992). This means that, although knowledge can be described in terms of an underlying logical structure, its mental representation does not apparently take the form of formal rules. Research further shows that children of different ages (and different mental representations) do not always perform the same task in accordance with expectations. These results have led to the conclusion that there are in fact fixed age-related sequences of mental representations which are different in nature: a fixed long-term developmental pathway across domains or contexts. However, students' performance in these tasks does not always have to conform with their competence. It is assumed that there is a short-term development which varies as a function of specific child parameters (learning style, practice) and environmental parameters (support, socioeconomic status, culture) (Dodwel, 1960; Kofsky, 1966; Lunzer, 1960). Minimal levels of these parameters determine the lower limit performance. In this way it is possible to define a developmental range. The performance level will approach the competence level depending on the quality of the help in the form of, for example, a clear task structure or instruction. Bruner (1982) distinguished three performance levels: (a) the functional level that is reached without others' help, (b) the optimal level that is reached if well directed help is offered, and (c) the scaffolded level that is reached if the solution is found by true cooperation.

Research into the performance range in the various developmental phases demands firstly an accurate analysis of tasks in terms of the sort and number of dimensions that students have to work on simultaneously. In this way it appeared possible, for example, to link the order of development phases to the capacity of the working memory (see Pascual-Leone, 1970). Secondly, it was attempted to make explicit, accurate descriptions of the way in which children learn different tasks in various domains. Although various researchers lay emphasis on different processes, there is a repetitive mechanism with a phase demanding attention for difficult (upper limit) tasks because the solution is not evident - problem solving and exploration (Case, 1985) - and an integration of lower level skills must occur (Fisher, 1980). Following this phase is a period in which such tasks can be solved with minimal attention because the method for solving the task becomes consolidated, associated and automatic (Case, 1985). Fisher (1980) emphasized here the increasing differentiation and chaining. Both these phases occur independently of the help children receive.

It can be concluded that neo-Piagetians, in comparison with the traditional

Piagetian view, have moved in the direction of the empiricists concerning the domain-specificity of knowledge representation. The knowledge to be learned is nonetheless considered to have a logical structure. This means that the research tasks call on the capacity to classify and serialize in order to induce the necessary knowledge. Although they retain the notion of biologically determined and discontinuous, development phases that can be qualitatively distinguished from each other, there is more room for variation in the development depending on the support that children receive. This shift can be typified as a gradual transfer from a competence model to a performance model. Research in the neo-Piagetian tradition here approaches the classic view on the content of education. A theme such as 'teaching reasoning' can fit in with this thinking that is, to a certain level, partly determined by the 'environment'.

Along the lines of empiricist tradition in the sixties, there arose an increasing interest for the way in which knowledge in the mind is represented and for the cognitive processes which generate and reorganize the representations. The cognitive revolution (Gardner, 1985) introduced the S-O-R model (stimulus - organism - response) in which the working of the computer was regularly used as a metaphor for the information processing in the mind. In the research into knowledge representation, much attention has been paid to the way in which information was coded in the mind. The most concrete representations are images (Kosslyn, 1990; Paivo, 1971) which form, in a spatial and temporal sense, an analogue representation of the attributes of and relationships between specific objects and of events such as: 'Mimie the cat is black.' The most abstract representations are propositions (Anderson & Bower, 1973; Pylyshyn, 1973). They have an arbitrary relation with real objects and events because they show the underlying meaning of knowledge in the form of relations between concepts such as: 'Color/animal.' Intermediate codes are more abstracted, spatially and temporally analogical representations in the form of the prototype sort (Komatsu, 1992; Rosch & Mervis, 1975) such as: 'Any cat/any color.' Or in the form of the semantic representations of concepts related to each other and expressed in language such as: 'Cat-pet-stroke/black-color-fur-stroke.'

What is common to these representation codes is that it is assumed that there is always a sort of network that links units of codes to each other. Depending on the code, the network which arises is more or less generally applicable, or more or less domain-specific. Examples of concrete, analogue represented networks are scripts (Schank & Abelson, 1977), representations of stereotyped series of events in a setting, and frames (Minsky, 1975) representations of stereotyped categories of things in a setting. These sorts of networks form the basis for episodic (domain-specific) knowledge. Concepts are linked together in more abstract, arbitrarily represented networks. Examples of these are schemes (Komatsu, 1992; Rumelhart & Ortony, 1977), semantic networks (Collins & Loftus, 1975) and mental models (Johnson-Laird, 1983, 1989). At this level, detailed specific attributes and relations are no longer used and similarities and differences are looked for. The representations of these consist of a hierarchical organization of knowledge in the form of part-part and part-whole relations. The representation of the pragmatic 'if p - then q' rules also belong to this level, in the form of contractual and causal schemes (Cheng, Holyoak, Nisbett, & Oliver, 1993; Morris & Nisbett, 1993). These schemes are not related to specific domains but depend on the type of relation between the premises. Examples of contractual relations are 'obligation': If 'precondition' (when I am 18 years old) then 'action' (then I can drive a car); or alternatively 'permission': If 'action' (if someone is driving a car) then 'precondition' (then he must be 18). These networks form the basis of semantic knowledge.The most abstract networks are formed by representations of propositions. This code is completely separated from the context. People with propositional networks are thought to reason according to formal (logical) rules. This means that on the basis of even arbitrary (and possible invalid) premises, logically correct conclusions can be drawn.

Within the information processing theories much research has been done particularly with the 'between persons, within task design' (novice-expert research). The experimental subjects are previously selected on the basis of their expertise in a certain domain, and thus often on the basis of age. The results show that children represent their knowledge in a more abstract way as they grow older (Kosslyn, 1976). However, even as adults not everyone achieves the most abstract form of representation. Moreover, the abstraction levels may differ per content domain. Certain domains, for example, strongly visual or strongly verbal ones, probably demand certain representations in the form of image code or semantic code, respectively (Sternberg, 1996).

It can be concluded that there is a shift from a purely empiricist paradigm towards a more rational approach to the relation between man and knowledge. The picture of a passively absorbing person who continuously stores away very small units of information by making links between stimuli, and between stimuli and response (learning theory), changes into a more actively constructive person who uses the knowledge representation as a reference frame in which to organize the knowledge to be absorbed (information processing theory). The shift can be characterized as a gradual transfer from a performance model to a competence model. Research in the empiricist tradition thus approaches the classic view of the content of education. A theme such as 'teaching reasoning' can fit in with this thinking that is to a certain level domain-free.

Induction processes and induction schemes

The process of induction is fundamental to the orderly information processing and to the (re)ordering of the representation of knowledge independent of the code of the representation. Formally this reasoning process is given as 'All observed instances of p are q, therefore all p are q.' This forms the basis for categorizing by the construction of part-part and part-whole relations. A semantic translation of this formal rule 'All observed cats are black, therefore all cats are black' shows that the induction process has a probabilistic nature. There are two other, related forms of induction. The first, the foundation for serialization, can be given as 'All observed p cause q, therefore q is caused by p.' Elements are hereby related to each other. The second has the form of an analogy 'P is (causes/belongs) to q, just as r is (causes/belongs) to s.' Solving analogies requires application of both classification and seriation.

Here too there is the question of whether people form representations of these formal rules. The form that human induction is thought to take varies according to the degree that the researcher fits into the traditional empiricist paradigm or has shifted towards a more rationalistic approach. In the first case, the induction process is comparable to the forming of increasingly stronger associations by frequent exposure to the combined stimuli (Halford, 1993). This process is called priming (Anderson, 1982, 1983) and is applied in education in the form of repeated practice. More rationally based research aims to reveal the inductive processes which form the basis of more complex knowledge absorption. That is to say, emphasis is put on analysis of the ways in which knowledge is absorbed, processed and represented, and on training possibilities.

Mostly two main types of information processes are distinguished (Bisanz, Bisanz, & Korpan, 1994). 'Performance' processes belong to the first type. They ensure the actual processing of information. Sternberg (1977) used the computer metaphor to deduce universal induction performance process steps from the way in which people solve analogies such as u:w - x:?. (a) Encoding of attributes of elements (u and w); (b) inference of relations between the attributes of elements (u and w); (c) mapping of the elements (u and x); (d) application of the relation u-w to x-? and (e) checking the solution in the justification phase (x-z). These performance processes guarantee an orderly processing and representation of knowledge. Knowledge already represented can also be organized more economically into larger units (chunking) by means of these processes. In contrast, represented knowledge will also influence the way in which performance processes work. A larger amount of knowledge, which is gradually represented in a more abstract way, may, for example, make a richer and/or quicker encoding possible (Carey, 1985; Gelman, 1989). The second type of 'metacognitive' processes is based on this knowledge and therefore has an active constructive function in the sense that it directs and monitors the performance processes.

The inductive reasoning process is thus characterized by both 'bottom-up' (performance) and 'top-down' (metacognition) processing (Keith, Holyoak, & Nisbett, 1988). The top-down processing is especially important in induction given the probabilistic character: in contrast to computers, people make use of top-down short-cuts to reach reliable and valid knowledge (Holyoak & Nisbett, 1988).

Nisbett et al. (1983) and Fong et al. (1986) studied two pragmatic, inductive top-down short-cut reasoning schemes based on statistical rules. The first scheme, based on the 'rule of large numbers', reveals that the sample values of parameters approach population values as a function of the size of the sample. This scheme restrains people, for example, from seeing relations between elements if these are not in fact always present (if you are a low-SES child, then you need extra help in school). In fact this scheme forms the basis of the well known 'representative heuristic' (Holland, Holyoak, Nisbett, & Thagard, 1986): seeing relations between observed characteristics of elements is limited by this scheme. The second scheme is based on the 'regression' or 'base-rate mechanism'. That extreme values are no longer found during repeated observations of the same elements or during repeated observation of other, similar elements is represented by this scheme (if you are a white, upper class child, then you have a good score on inductive reasoning tests). The 'availability heuristic' (Holland et al., 1986) which relates elements to each other on the basis of the presence of conspicuous and/or easily identified attributes is regulated by this scheme.

What is common to these schemes is that the generalizing character in induction (by method of agreement) must be accompanied by a discriminating attitude (by method of difference (Mill, 1887)), which monitors the variability in knowledge. Or, classification on the basis of eliminating a part of the conditional premise (Holland et al., 1986) must be in balance with discrimination on the basis of adding restrictions to a condition or action premise. This process of 'tuning' in fact safeguards the equilibrium between what the represented knowledge can explain and where the representation should be amended. Apart from the content-related metacognitive processes, these domain-free scheme-based metacognitive processes seem to play a role in information processing too. Research in adults (Nisbett et al., 1993) has revealed the importance of both these schemes and their application by means of training.

Research (Keil & Batterman, 1984) shows that young children do not yet employ these schemes. It is assumed that prior to the development of the schemes, in which it is principally the certainty with which perceived connections are indeed real, children must become aware of the basis of such connections: categories, and relations between categories. Still using statistical terminology, it means that children must form a representation of a 'group structure' expressed on a nominal scale and of a 'row structure' (see Halford, 1993) expressed on an ordinal or rational scale. Group and row structures are general since they are not tied to content domains or to representation codes. They can therefore be considered as a sort of pragmatic scheme and can be used as a starting point for teaching inductive reasoning to young children. Klauer' s inductive theory (this volume), which is the basis of his programme for inductive reasoning (Klauer, 1989), follows on from the idea of the group and row structure schemes. The next paragraph is dedicated to the description of the programme.

	Group - tasks (Attributes)		Row - tasks (Relations)	
	item - class	item - types	item - class	item - types
Similarity	Generalization of attributes	- class formation - class expansion - finding common attributes	Generalization of relations	- ordering series - series completion - analogies
Dissimilarity	Discrimination of attributes	- class exclusions	Discrimination of relations	- disrupted series
Dissimilarity and Similarity	Cross-classification	- 4,6,9 fold schemes	System-construction	- matrix-figures with complex analogies

Figure 1 Taxonomy of types of tasks

Training programme for inductive reasoning: Theory, instruction and research results

Klauer (1989) defines inductive reasoning as systematic and analytic comparison aimed at discovering regularity in apparent chaos and irregularity in apparent order. Regularities and irregularities of nominal level are recognized through the attributes of elements (for example, color) and the relations among elements (for example, size) represent the ordinal/rational level. Elements are people, animals, objects and situations. In order to validate his theory Klauer designed six types of inductive reasoning tasks with which to train students (Figure 1).

The tasks in Figure 1 are characterized by finding similarities, dissimilarities, or a combination of both. This sequence indicates that attention is paid to the tuning process (or in Piagetian terms, to the mechanism of equilibration) and to the importance of the increase in the number of dimensions to be discovered (see Pacual-Leone, 1970). Further, the masking of the attributes and relations increases in the course of the training. This reflects the skill that children develop in the course of their normal development to increasingly resist being distracted by perceptually conspicuous attributes (Carey, 1985). Apart from the masking arising from a reduced perceptual conspicuousness of attributes and relations, Klauer has attempted to embed the dimensions in tasks with illustrations of daily situations. This yielded block-, object-, situation- and abstract problems, which fit in the inductive reasoning group and row structure schemes.

Figures 2 and 3 show examples of each of the types of tasks included in the programme on inductive reasoning (Klauer, 1989; Klauer & Phye, 1995). Figure 2 shows three types of tasks for grouping objects on the basis of attributes (generalization, discrimination, and a combination in the form of cross-classification) and Figure 3 shows three types of tasks for seriation of objects based

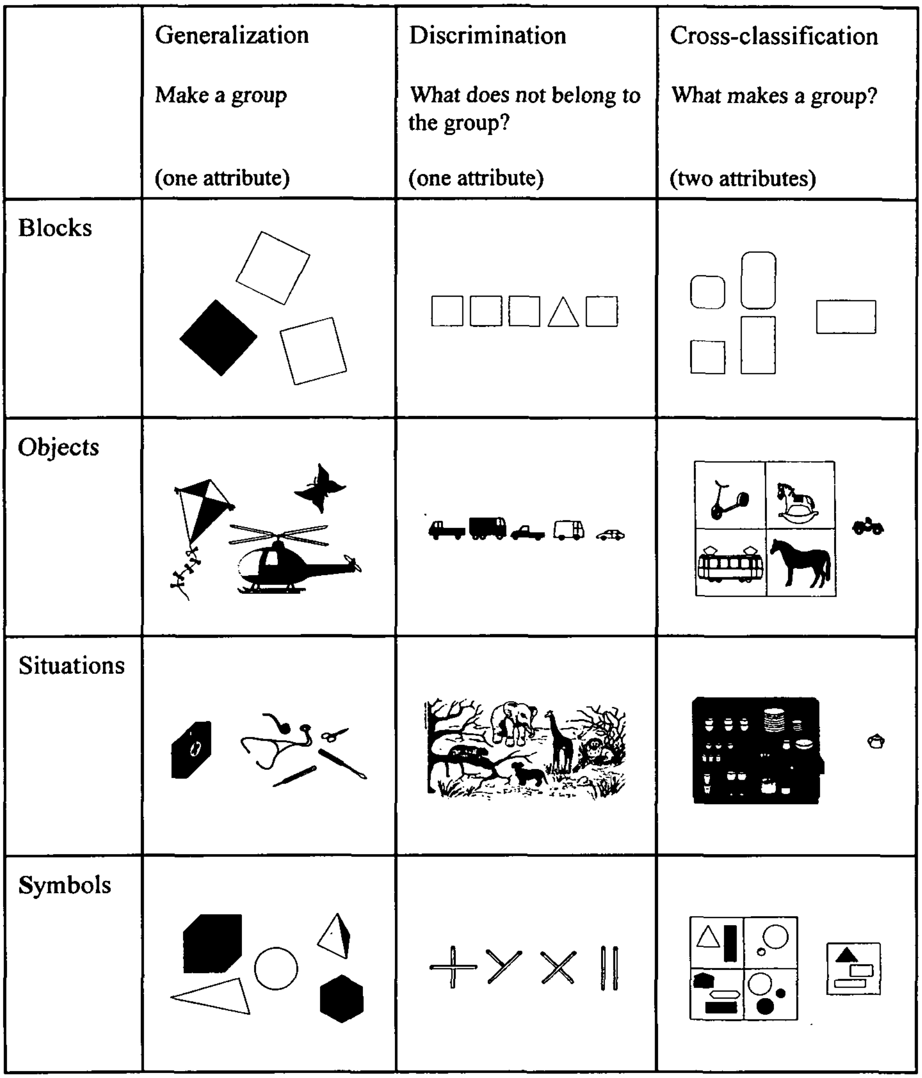


Figure 2 Three types of tasks for grouping objects on the basis of their atrributes

on their mutual relationships (generalization of relationships, discrimination of relationships, and a combination in the form of system construction). Mastering cross-classification and system-construction reflects the possession of inductive reasoning group and row structure schemes. The programme consists of 120 tasks. Each type of tasks consists of 20 reasoning problems.

Klauer (this volume) validated his theory of inductive reasoning in an empirical way in a number of phases. In the first phases he studied the coherence between the task types by means of training studies using the design type 'within child, between types of tasks.' The studies proved the suspected coherence of the tasks. The

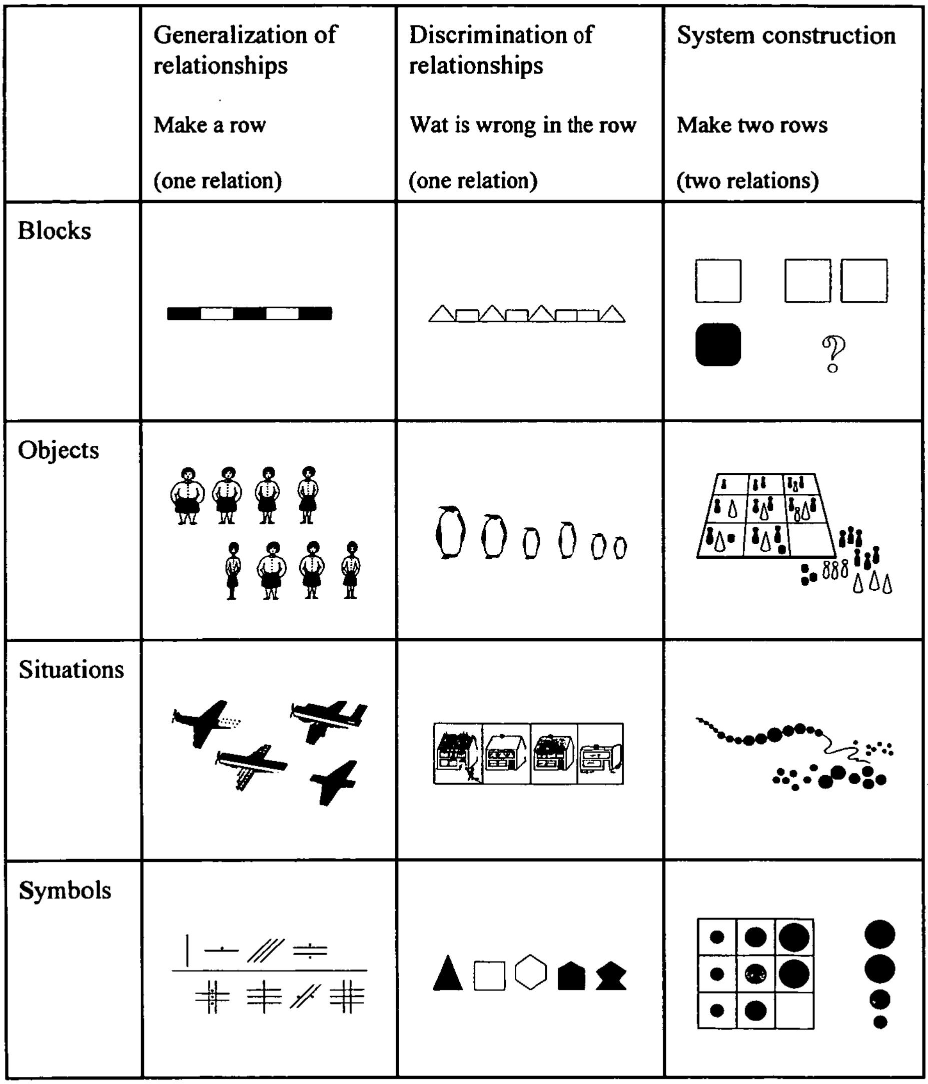


Figure 3 Three types of tasks for seriation of objects based on their mutual relationships

coherence within the group tasks and within the row tasks was greater than between the two sorts of tasks. This supports the view that there appear to be two sorts of schemes. In a test (De Koning, Hamers, & Sijtsma, 1996) consisting of problems of the same type as in this training programme, research has also shown the unidimensionality of the group and row problems on psychometric grounds in a sample of approximately 1000 subjects (De Koning, Hamers, & Sijtsma, in preparation).

In a second phase of research with the design types 'within or between child, between task', Klauer investigated the coherence between training tasks and tasks in intelligence tests. These tests primarily contain tasks which demand inductive reasoning. In addition, he studied the coherence between the training tasks and the knowledge absorption of school subjects such as geography, biology. The results show significant effects from his training programme.

In his programme Klauer (1989) distinguishes three training methods which vary in the degree in which the training leader or the student bear the responsibility for solving the problems. The choice of method depends on the cognitive level of the student. Relatively more intelligent students need less help in solving the training tasks than less intelligent students. They are better able to determine for themselves what type(s) of tasks there are and which solution processes need to be adopted.

In each of the three training methods, three phases can be distinguished (introduction, application, automatization). They must increase students' awareness of the task types and accompanying solution processes and offer the possibility of speeding up their mental operations. There are also three central elements in each training method. Firstly, model problems are used. The simple unmasked appearance of these problems clearly illustrates a certain type of problem and therefore a certain course of solution processes. Secondly, the emphasis lies on the comparative step in the solution process: finding similarities and differences in the attributes of objects and relations between objects. The comparison can be based on accurate (possibly exhaustive) analyses of each of the (pairs of) objects, or on a more global inspection of all the (pairs of) objects. A third element is that students must learn to distinguish the six types of problems and the accompanying solution processes.

De Koning and Hamers (1995) have worked out and adapted these general guidelines for teachers to implement the programme in school classes with respect to the instruction. There are three central points in this adapted programme:

(a) Teaching group and row schemes. The mental representations of these schemes are the foundation of the metacognitive (top-down) guiding of the performance (bottom-up) processes. This means that distinguishing the six types of problems is limited to distinguishing the two main classes, namely group and row problems. In the programme's introductory phase these schemes are visualized by problems using concrete blocks which need to be classified or serialized. Although increasingly masked, the group and row structures remain evident throughout the whole programme in the form of the problems. To provide an overview the pictures and blocks are collected after the problem has been tackled: group problems on the right of the blackboard and row problems on the left. In addition, before and during the solving of new problems, 'model group' and 'model row' problems which have already been tackled are regularly shown for comparison.

(b) Explicit teaching of the sequence of bottom-up steps in the inductive reasoning process. The steps 'encoding - inference - application - justification' described by Sternberg (1977) for solving analogies were taken as a starting point. For practical

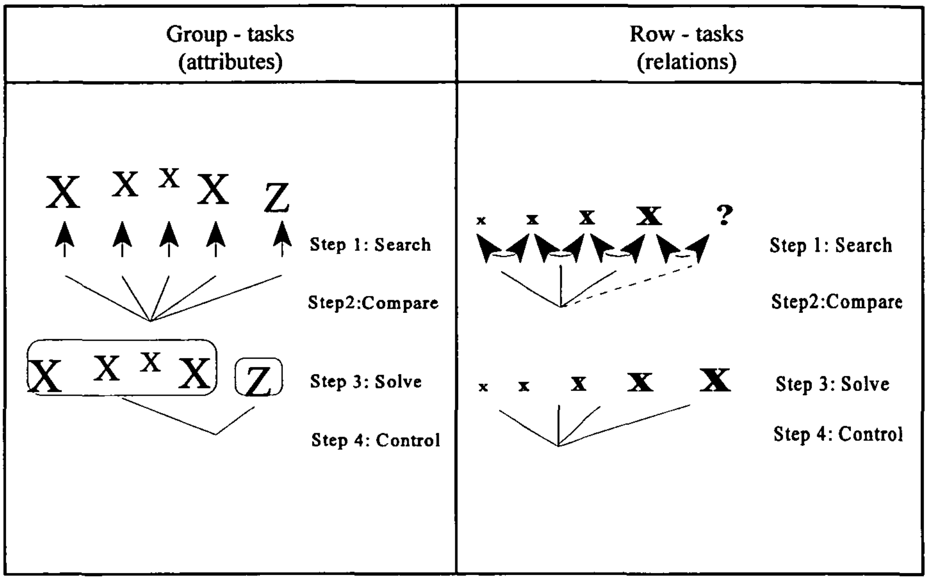


Figure 4 Applying inductive performance processes on group tasks and row tasks

application in both group and row problems the steps are translated for teachers as: search for the relevant attributes or relations; compare similarities or dissimilarities in attributes or relations; solve the problem on the basis of comparison; check the solution. Although the steps for group and row problems are identical, there is a difference in the number of objects that has to be analyzed simultaneously in the search and comparison processes (Figure 4). This figure shows that for group problems the search can be limited to analyzing the attributes of separate objects. For the comparison, a minimum of two objects must be studied. Searching for relations in row problems requires analysis of two objects and the comparison requires three. Studies (see Hamers, De Koning, & Sijtsma, 1998) have shown that the students' results are better if the emphasis during the instruction is placed on the accurate and extensive analysis of the objects during the search and comparison steps.

(c) Intersubjectivity and active participation in the classroom dialogue. It is important for the actual classroom instruction to answer the question 'In what way do the students master the above schemes and solution steps?' Research (Case, 1985; Fisher, 1980) to reveal the way in which people learn has shown that a phase of cognitive effort in order to be able to solve an unknown problem is followed by a phase in which the steps taken to solve the problem are practiced and automatized. Klauer' s three instruction phases (1989) reflect this sequence. In addition, research (see page 159) has shown that the environment has an influence on the development of knowledge. However, this has not yet answered the question as to how the environment, casu quo the teachers or students can enhance the development of knowledge.

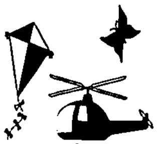
Vygotsky threw light on the relation between student - teacher - task with his view of the relation between man and knowledge acquisition. In contrast to rationalists and empiricists Vygotsky saw knowledge as a socially constructed product. He considered knowledge as something that develops in the interaction between people. People and knowledge were not considered as two separate entities. This is like the development of children: in interaction with more knowledgeable other people. This development moves from the social, intermental plane on which collaboration with knowledgeable other people is central, to the psychological, intramental plane (Wertsch & Kanner, 1992). Vygotsky saw this transition as a basic mechanism for the intellectual development. Language plays an important role in the interaction. The performance range which is bounded by development levels, with or without collaboration, is called in his theory the Zone of Proximal Development (ZPD) (compare this to the 'performance levels' defined by Bruner, 1982).

To be able to develop the potential possibilities present in students Vygotsky thought there were two requirements which the student - teacher relation needs to satisfy (Wertsch & Kanner 1992). The first requirement concerns the inter-subjectivity, that is the tasks for all the students must make a demand on prior knowledge and/or personal experiences. This requirement is met by the programme (De Koning & Hamers, 1996; Klauer, 1989) by using differently colored building blocks of basic shapes and pictures of everyday objects and situations. Each student then has the same prior knowledge and experience when the most abstract problems are tackled.

The second requirement concerns the active participation in dialogue, guided or organized in accordance with certain principles (intermental functioning), gradually moving to handing over the strategic responsibilities to the students (intramental functioning). The lessons in the programme are therefore conducted in a scheme-based and process-based dialogue with the whole group. The students sit in a semi-circle so that they all have a good view of each other, of the teacher, and of the problem which is displayed in front of the class. The teacher' s ability to ask the right scheme-based and process-based questions is crucial to the dialogue. This means, for example, that correct solutions without scheme-based and process-based justification are not considered good answers. Awareness of the group and row schemes and the mental steps puts students on a 'meta-level': they are provided with instruments to monitor their own problem-solving behavior and to adapt it if necessary. During the programme, the responsibility for learning shifts from teacher to student in three phases. This means that the students take an increasingly prominent role in the conducted dialogues. At the end of the programme for example, the students take on the role of teacher in turn so that it is clear where their knowledge is still not sufficient.

Language, as an instrument in the dialogue, fulfils the role of the cognitive catalyst for reaching the meta-level. The explicit labeling of both types of schemes

PROBLEM TABLE 4: GENERALIZATION TASK



Aim:	discovery of an attribute (looking for regularity)
Question:	what have these objects in common
Essential attribute:	flying in the air
Problem table:	* table with a helicopter, butterfly, kite * table 1 and table 3

Instruction:

Recognizing problem type	*Again we are going to form groups, but we are not using blocks this time. Look at this table. First we name the objects on the table. This is a ...* (point to all the objects and give them a name).
Question	*These three objects fit together. They look more or less like each other. We should find out ...* why they fit together.
Step 1: search	*What do we have to do? All right, we are going to describe each object, just as we did with the blocks. What can you tell me about this ...* (point to a object) *and this ...*
Step 2: compare	*What do we have to do now? Yes we have to compare the objects. We look now for similarities* (point to all three the objects).
Step 3: resolve	*Well, the question was: why is it that these object belong to one group? Yes, that's right, because they all can fly. And what must be the other reason? All right, they also belong together because they have a tail.*
Step 4: check	*Now we must check our solution. The helicopter .....* flies in the air, *the butterfly ....* flies in the air *and the kite ...* flies in the air. *Yes, it's okay, they belong together in one group* (same for the tail).
Reflection	Describing in general terms: the concept attribute *Now I want to tell you something. We saw that all these objects (point) are a bit the same: they all fly in the air. We say: these objects have the same attribute. The attribute of the kite is that it flies in the air. The attribute of the butterfly is that it flies in the air. The attribute of the helicopter is that it flies in the air. An attribute says something about an object. For instance, whether it can fly, or what colour is has.* Compare with blocks of tables 1 and 3 *In the case of the blocks, we looked at the color. So we grouped them by color* (point). *Who knows another attribute of the blocks? Very good, the form of the blocks.* Description of the problem type *Objects with the same attribute fit together in one group. Objects that do not share the same attribite do not fit together in one group. Does a car fit in this group? No, a car doesn't have the same attribute: it cannot fly. So a car does not belong to this group*

PUT THE PROBLEM PLATE ON THE RIGHT HAND SIDE OF THE BLACKBORD AT THE GROUP-TASKS

Figure 5 Part of the teacher's manual with respect to the inductive reasoning problem 'Generalization' (group and row problems) and of the mental steps (search, compare, solve, check) and of central concepts such as attributes and relations, is emphasized by adding these terms to the students' vocabulary. As an illustration, Figure 5 shows an example of instructions for a picture which is dealt with in the first programme phase.

Research by Hamers et al. (in press) shows that teachers can use the instruction principles described above in implementing a programme for first grade classes of low-SES students (6 and 7 years of age). In addition, it appears that the inductive reasoning skills of these students can be improved significantly.

The inductive (top-down) group and row schemes and the bottom-up search, compare, solve and check processes are independent of the content domains and of the representation code. Klauer' s programme described above (1989) covers the visual domain. In a second programme (Klauer, 1990) aimed at students in grade 5 (10 and 11 years of age), the problems are not only visual but also verbal or numerical. The internal coherence of the problems and the external construction validity have also been demonstrated by means of empirical research. Jeurissen, Hamers, and De Koning (1997) have also compiled a classroom version with analogy to the above instruction principles for this programme. A study of 86 (low-SES) students (51 students in two experimental classes and 35 students in one control class) has demonstrated the implementability of the instruction as well as significant learning benefits (Hamers, Jeurissen, & De Koning, 1997).

Another study has looked at the general applicability of the above schemes and problem-solving processes using a programme in the verbal comprehension domain (De Koning & Hamers, 1996). This domain was chosen because of the problems low-SES children have in understanding written information. In the following sections we will discuss this study in more detail. The description of the study is preceded by a short overview of theories and models in the domain of understanding texts.

Man and text

Inducing order in the written verbal domain (i.e., reading) takes place less directly than in the visual domain, for example, because the verbal, arbitrary symbols must first be translated into understandable units. Letters have to be decoded to sounds and synthesized to words, words are put together into sentences which are compiled into a coherent text. This section deals with inducing order in sequential sentences or parts thereof.

There has been a shift in the theory and model development concerning the understanding of written information. In traditional models emphasis is put on understanding the information in the text. According to the empiricist view on the relation of man and knowledge, the reader is envisaged as someone who reconstructs knowledge (bottom-up). More recent models center on the integration of information in the text with knowledge already present (Kintsch, 1974, 1988; Van Dijk & Kintsch, 1983). In general, as more knowledge is available, the possibility of reconstruction (bottom-up) and active construction (top-down) increases. This opinion fits in with the rationalist view since the top-down - bottom-up balance shows strong similarities to the concepts of 'tuning' in information processing theories and Piaget's 'equilibrium'.

The models vary not only in their views on the relation between man and text but also in the assumed codes of representation. The most abstract representations are in the form of propositions. In order to understand the texts, the propositions need to be linked into a logical (macropropositional) network. Since reading a text is a sequential process, it is assumed that such a network is developed in steps (bottom-up or bottom-up and top-down) (Gernsbacher, 1996). Relations between propositions are based on, for example, common noun arguments (Gernsbacher, 1996; Goldman, Varma, & Coté, 1996; Kintsch, 1974) and links between predicates (Turner, Britton, Andraessen, & McCutchen, 1996). Making these links, however, does seem to assume the presence of less abstract representation codes. Examples are semantic codes or analogue prototypical imaginal codes (Garnhem, 1996; Yuill & Oakhill, 1991). Some researchers (Johnson-Laird, 1982) assume that there is a transition from one representation code to another during reading, while others (Garnhem, 1996) think that the developed network does not have a propositional nature. Garnhem (1996) assumes that the connection is determined by the way in which the attributes of persons, events and objects are grouped and related to each other. For the study to be discussed here, we assume that students in grade 2 (7 and 8 years of age) of primary school do not yet make representations in a propositional code. It is expected that semantic and imaginal codes will play an important role (see Yuill & Oakhill, 1991). Independent of this code a coherent (network) representation must be developed from the main message in the text.

To understand texts a certain amount of knowledge is necessary about the structure and content. At sentence level, there must be a basic linguistic awareness about the syntactic structure of sentences. Independent of whether children represent this information in abstract syntactic categories (object-verb-subject) or more semantic categories (doer-instrument-recipient, Bock, 1990; Fromkin & Rodman, 1988), research into young children's speech and fluency (Sternberg, 1996) shows that they have an implicit idea about the structure even at a young age. It may therefore be assumed that students make the important distinction between noun phrases and verb phrases in understanding the meaning of the sentences. However, research by Yuill and Oakhill (1991) showed that poor comprehenders do not read according to syntactic categories. Results of samples matched for decodeability show that poor comprehenders as compared to good comprehenders read more by word for word than in the grammatical, semantically relevant sequences.

As far as the content level goes, the meaning of the words first has to be represented in networks. Unknown words or words with a double meaning in the text can be discovered by searching for clues in the text which activate certain parts in these networks, for example. The ease with which these links arise is a function of their separation in the text, among other factors (Sternberg & Powell, 1983). Secondly, it may be that knowledge about the text's theme is already represented to a greater or lesser degree. These representations can enhance the speed (Gernsbacher, 1996) or extent of comprehension. The bottom-up processes are then facilitated by the top-down effects of the representations. Alternatively, the bottom-up processes can lead to adjustments in the existing representations.

The top-down - bottom-up balance is the foundation for development of implicit, explicit and/or elaborative inferences. By this we mean the processes by which relations between text parts or relations between text parts and representations already present are made. Implicit relations are relations which are not made in the text itself but which are important for making a coherent representation. Explicit relations in the text are indicated by certain keywords, such as pronouns (John eats an apple, he likes 'it') and connections such as 'then' (temporal relation) and 'because' (causal relation). Elaborative relations are relations between the information in the text and the knowledge representation already present. If these connections are not made, inconsistencies and gaps in the text cannot be solved and, moreover, comprehension failures will not be noticed. Research (Oakhill, 1996; Yuill & Oakhill, 1991) has shown that poor comprehenders have difficulty making these connections. They appear to monitor their comprehension progress insufficiently. Yuill and Oakhill (1991) pointed out that a limited capacity of working memory could possibly be the cause of this.

The theoretical and empirical issues mentioned in this and the former paragraphs served as guidelines for the development of a reading comprehension programme.

Training programme reading comprehension: Theory, instruction and research results

The aim of the reading comprehension programme (De Koning & Hamers, 1996) is to train students in constructing new knowledge representations from text. The training aims at providing insight into the structure of texts and using one's own knowledge. The programme is based on three core principles. According to the first core principle, students are made aware of sentence structure and combinations of sentences into text parts. The sentence structure is analyzed in a noun phrase (who's it about, what's it about?) and a verb phrase (what is said about it?). The semantic codes are used in the reading instruction (WHO and WHAT) instead of the syntactic categories. Pronouns and connectives are used sparingly but are well targeted in the sentences. These words demand explicit inferences and can be dealt with on the 'who-what' sentence level because the distance between the text parts to be connected is kept small.

<table>
<tr><th></th><th>Similarity</th><th>Dissimilarity</th><th>(Dis)similarity</th><th>(Dis)similarity</th></tr>
<tr><td>Attributes</td><td>Generalisation of attributes</td><td>Discrimination of attributes</td><td>Generalisation of attributes + Discrimination of attributes</td><td rowspan="2">Generalisation of attributes + Discrimination of attributes + Generalization of relationships + Discrimination of relationships</td></tr>
<tr><td>Relations</td><td>Generalization of relationships</td><td>Discrimination of relationships</td><td>Generalization of relationships + Discrimination of relationships</td></tr>
</table>

Figure 6 Taxonomy of types of tasks

The text structure is separated into two main categories, namely group texts and row texts, in analogy to Klauer' s (1989, 1990) programmes. In group texts, the sentences in the text deal with a certain theme (for example, a birthday). In row texts, the relation between the sentences is central, and emphasis is put on analyzing the sequence of the sentences in a text. In both cases the aim is to induce the way in which the sentences in the text are ordered or why there is a mistake in their order. Ordered texts consist of sentences which all deal with the same theme (generalization problem) or which flow into each other in a logical sequence (relation recognition problem). In unordered texts, texts with gaps or inconsistencies, there are sentences which do not deal with the main theme (discrimination problem) or in which the sequence is disturbed by an abnormal relation (relation distinction problem). Figure 6 provides an overview of the different types of problems.

To identify the main theme and the correct sequence, the same inductive reasoning steps are followed as described earlier: search for the relevant attributes or relations; compare similarities or dissimilarities in attributes or relations; solve the problem on the basis of comparison; check the solution. Although the steps for both types of problems are identical, there is a difference in the number of sentences that have to be analyzed simultaneously in the search and comparison processes (see Figure 4). For group problems the search can be restricted to analyzing the attributes of separate sentences, while for the comparison a minimum of two sentences have to be studied. Searching for relations in row problems requires an analysis of two sentences, while for the comparison three sentences are required. As the programme advances, the different problem types are combined in longer texts. In brief: the induction process takes place after the structure of each sentence has been analyzed. The induction process is successful if students are able to compose a good, summarizing title for the text that has been analyzed.

The second core principle encompasses relieving the working memory by introducing a scheme in which the above mentioned sentence and text structures are visualized. In addition, converting arbitrary, verbal symbols into semantic, meaningful images is facilitated by using pictures. Moreover, research shows (Yuill & Oakhill, 1991) that the mental processing of images uses less memory capacity than the processing of sentences or parts of sentences. Figures 7 and 8 give examples of two texts with accompanying schemes from the first, simplest phase.

According to the third core principle the texts change from a concrete everyday content to one that is more abstract and unknown. The texts in the first half of the programme involve situation sketches which link up to the students' present representations. This underpins the forming of implicit and elaborative inferences because there is a facilitating and monitoring influence from the content-relevant top-down processes on the bottom-up processes. The second half of the programme deals with expository texts. It is assumed that many students still have insufficient knowledge of all the words in the text and of the theme. By analogy with the build-up of Klauer's programmes (this volume), this leads to an increased masking of the relevant attributes and relations during the programme. The balance between bottom-up and top-down processing thus acquires the equilibrium necessary for understanding school texts (for example, geography and history). In learning to analyze the sentence and text structures, students are expected to be able to induce content order, know when their comprehension falters and how to solve this. In fact, students are brought on to the meta-level.

The programme comprises sixteen lessons. The instruction for the programme is based on the same basic points as those discussed in Klauer's programmes. All the actions take place within the framework of a scheme-based and process-based dialogue which satisfies the requirements of intersubjectivity and active participation (see page 173). The responsibility for learning shifts gradually from the teacher to the students. After a short phase of classical treatment of problems, there is a phase in which the students solve the problems in small groups. In the last phase a student fulfils the role of teacher in his or her group. A classical discussion always follows. In this programme, too, explicit attention is paid to giving verbal labels to the sentence structure (WHO and WHAT), to the schemes (group and row problems), to the mental steps (search, compare, solve, check) and to the central concepts (attributes and relations).

The first research results reveal the impression that the programme produces the expected positive learning effects. A formative evaluation study of the implementation of the reading comprehension programme (pre-test - post-test - control group design) will be described as an example. Summative studies are still being carried. The following issues were investigated: a) Can the programme be implemented in the classroom? b) Does the reading comprehension of students who follow the programme improve to a greater extent than that of students who do not follow the programme? and c) Does the inductive reasoning skill of students

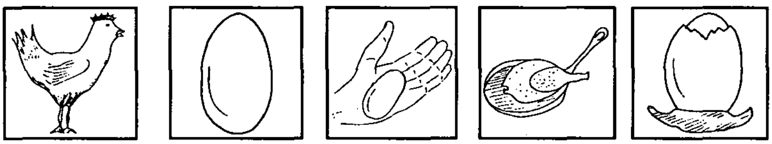
Example text 'Discrimination of relationships'

1. john goes to his chicken.
2. he looks for an egg in the chicken-coop.
3. john gives the egg to his uncle.
4. uncle fries the chicken in the pan.
5. john likes his egg.

what seems wrong here?
think of a title.

READING

SENTENCE	1	2	3	4	5
WHO	john	john	john	uncle	john
WHAT	goes to chicken	looks for egg	gives egg to uncle	fries chicken	likes egg



THINKING

SEARCH later later ? ?

COMPARE later

SOLVE (what seems wrong here) later : uncle fries egg

CONTROL

THINK OF A TITLE

Figure 7 Example of a text and shema of the first, simplist phase of the programme for reading comprehension

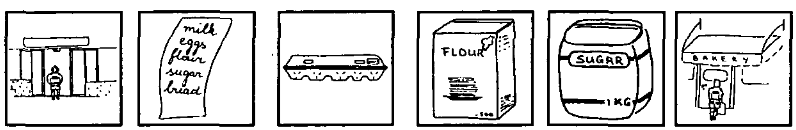
Example text 'Generalization of attributes'

1. john goes to a shop.
2. uncle gave a note to john.
3. the errands are writtten on it.
4. john must buy eggs.
5. and flour for pancakes.
6. he must buy sugar as well.
7. john goes to the baker.

where would you go?
think of a title.

READING

SENTENCE	1	2+3	4	5	6	7
WHO	john	uncle	john	john	john	john
WHAT	goes to shop	gave note with errands	must buy eggs	flour for pancakes	must buy sugar	goes to baker



THINKING

SEARCH buy something
green grocer? which shop? bread
supermarket? supermarket? supermarket? supermarket? supermarket?

COMPARE supermarket?

SOLVE (where would you go) supermarket

CONTROL

THINK OF A TITLE

Figure 8 Example of a text and schema of the first, simplest phase of the programme for reading comprehension

who follow the programme improve to a greater extent than that of students who do not follow the programme? The sample consisted of second grade students (7 and 8 years of age) from primary education, living in a backward social-economic home situation. In the experimental group (n=10), the reading comprehension programme was used, while the control group (n=14) participated in regular second grade activities. There was no significant age difference between the groups. The groups took the same pre-tests and post-tests. The interval between pre-test and post-test was ten weeks, and the sixteen lessons were given during this period.

A standardized reading comprehension test was used for the pre-tests and post-tests which is often used in schools to measure progress in this domain. A test for inductive reasoning was also used (Raven's Standard Progressive Matrices (SPM); Raven, 1958). The SPM tasks are meaningless and abstract geometrical problems, in contrast to the training tasks, which are meaningful verbal problems taken from children's everyday life and from school-type subjects. The SPM was used as a far transfer test. The teacher of the experimental class was trained in two sessions, on the programme's content, structure and instruction. All the lessons were given by this teacher, with an independent observer present.

The answer to the first question was affirmative. The results of the observations showed that the teacher was able to deal with the main didactical requirements of the programme. However, it was not easy for her to continually look past the familiar reading comprehension domain to the underlying mental processes. It demanded extra effort from her to leave out routine treatments or to adapt them. The effort did eventually provide deeper insight into the process of comprehensive reading.

With respect to the second and third research questions: in order to control for absence of randomization, the pre-test scores of the SPM and the reading comprehension test were included as covariate in the analyses of variance. Table 1 shows the means and standard deviations of the pre-tests and post-tests. Double-sided F-tests showed that the test group differed significantly (5% level) from the control group on the SPM post-test score and on the reading post-test score. The SPM pre-test (but not the reading pre-test) had a significant influence on the post-test score as a covariate.

Depending on sample size, significant differences in combination with small

Tabel 1 Means and Standarddeviation of the SPM Raven and Reading Comprehension pretest and posttest scores

	SPM Raven				Reading Comprehension*			
	Pre-test		Post-test		Pre-test		Post-test	
	M	SD	M	SD	M	SD	M	SD
Exp. Group	27.6	11.0	34.6	4.5	98.0	21.6	112.2	12.0
Ctr. Group	27.3	7.9	29.2	7.2	93.6	9.8	99.9	7.4

* standardized scores

effect sizes may cast doubt on confirmation of the hypotheses (Hager, this volume). The corrected effect sizes, calculated on the basis of pooled standard deviations, are reasonable: 0.84 and 1.01 for the SPM and the reading comprehension test, respectively.

Discussion

In this chapter the general applicability and the trainability of two inductive reasoning tools - the pragmatic inductive schemes and inductive reasoning processes - are incorporated into psychological theories and have been studied empirically in our first series of 'classroom research projects'. Use has been made of Klauer's induction theory and the domain-free programmes based on it. Applying the same tools in the verbal comprehension domain resulted in the development of a domain-tied programme for reading comprehension (De Koning & Hamers, 1996). Figure 9 gives an overview summarizing the basic principles of both types of programme.

Figure 9 in fact shows the didactic triangle: the student, the teacher and the knowledge (tasks in the programmes). In the center of Figure 9 are both the types of inductive reasoning tools that are used automatically to an increasing degree and are integrated during the training. This process runs parallel with an increasing awareness of these tools, via 'mastery' and 'awareness' to 'self-regulation', thus

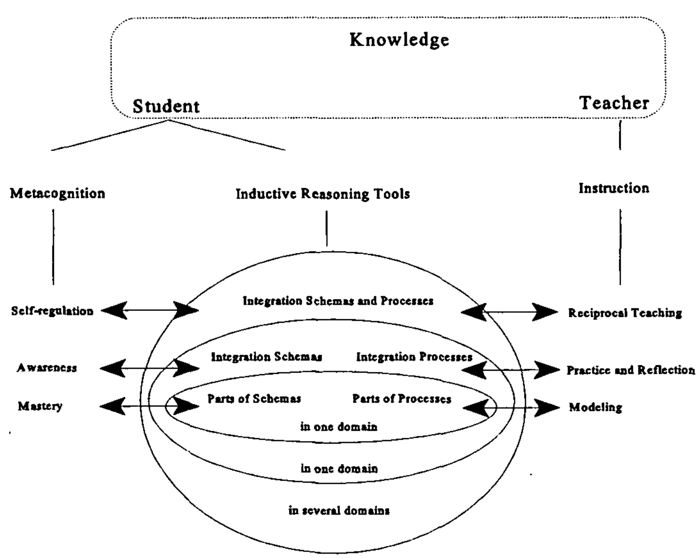


Figure 9 Overview of the basic principles of both types of programmes

bringing the student to a meta-level. This awareness comes with a shift from mainly bottom-up processing to a balance between bottom-up and top-down processing.

The teacher plays a key role in bringing this about by generating a 'scheme-based' and 'process-based' dialogue in the class. In this dialogue students gain an increasingly greater responsibility in directing, monitoring and adapting the learning. In the last phase it is assumed that the students can apply the inductive reasoning tools to different domains. The theoretically predicted and empirically demonstrated transfer leads us to think that the students have not only learned a skill but may have also built up a certain inductive reasoning competence, which we could describe as a general ability to (re)organize knowledge.

There are, however, lots of questions still left to be answered. Firstly, from this kind of result, it is not possible to say what particular aspect of the training was helpful, nor do we have a detailed picture of how the effects arise. It is therefore unclear whether slight changes in the instruction could make it even more helpful. For instance, it may be more effective to induce students into gradually making their own images, instead of providing images for them. The second question refers to the long-term effects. Our own research covers a maximum follow-up period of four months. Klauer (this volume) has been able to show longer effects from his training experiments. In an upcoming research project both kinds of programmes will be applied during a two-year investigation. Thirdly, it is not yet clear what type of programmes (curriculum-free or curriculum-tied) are most effective in training students. There are two streams of teaching thinking skills (Hamers & Overtoom, 1997). Researchers of the first stream believe that thinking skills can be taught explicitly, independent of the regular curriculum ('across-the-curriculum' approach). It is assumed that there are certain more or less universal thinking skills, which can be generalized and applied to school subjects. Researchers from the second stream believe that the learning of thinking skills should be embedded in the school subjects ('infusion' or 'within-the-curriculum' approach). A programme that fits in the 'across-the-curriculum' approach is decribed, for example by Klauer (this volume). Programmes that fit in the 'within-the-curriculum' approach are decribed, for example by Verschaffel (this volume) and Chanquoi (this volume); (see also Halpern, 1992; Resnick & Klopfer, 1989). In the studies presented in this chapter, students were first trained in inductive reasoning using Klauer's (1989, 1990) adapted programmes. The programmes can be characterized as 'across-the-curriculum' because they stimulate the general thinking processes of inductive reasoning. The content of the tasks, however, is meaningful and recognizable from daily situations (Klauer, this volume; Klauer & Phye, 1995).

However, do we need such 'across-the-curriculum' programmes in schools? The answer to this question determines how thinking will be taught to a great extent. If thinking is taught in an 'across-the-curriculum' course, objectives for thinking skills and strategies will be the basis of the programme. If thinking is taught in the context of a school subject, e.g., a programme on reading comprehension, content objectives will be the basis. There is considerable debate as to which context is more effective for teaching at-risk students (Hamers & Overtoom, 1997). Proponents of the first approach argue that low-achieving students may experience overload if they have to learn both content and skills simultaneously. A second reason for the use of such a curriculum-free programme is that it is also easier for the teacher to focus the instruction on the reasoning process because the content of the tasks is not so important. Proponents of the second approach argue that programmes should be content-related because a substantial part of skills and strategies is content-specific, and these skills and strategies do not transfer easily to other areas (see also Csapó, this volume).

In our studies both kinds of programmes were used with children identified as 'at risk' (low-SES children) in the remedial and compensating sense. In both cases the teachers appeared to be able to get the class to start reasoning. In both cases the trained students showed significantly better learning and transfer results than the untrained students. To answer the 'within- or across-the-curriculum issue' both programmes are now being studied together.

References

Anderson, J.R. (1982). Acquisition of cognitive skill. *Psychological Review, 89,* 369-406.

Anderson, J.R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J.R., & Bower, G.H. (1973). *Human associative memory.* New York: Wiley.

Bisanz, J., Bisanz, G.L., & Korpan, C.A. (1994). Inductive reasoning. In R.J. Sternberg (Ed.), *Thinking and problem solving* (pp. 181-213). San Diego: Academic Press.

Bock, K. (1990). Structure in language. Creating form in talk. *American Psychologist, 45,* 1221-1236.

Bruner, J.S. (1982). The organization of action and the nature of adult-infant transaction. In M. Cranach & R. Harre (Eds.), *The analysis of action* (pp. 313-327). Cambridge: Cambridge University Press.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

Case, R. (1985). *Intellectual development: Birth to adulthood.* Orlando, Fl: Academic Press.

Case, R. (1996). Changing views of knowledge and their impact on educational research and practice. In D.R. Olson & N. Torrance (Eds.), *The handbook of education and human development* (pp. 75-99). Cambridge, MA: Blackwell.

Cheng, P., Holyoak, K.J., Nisbett, R.E., & Oliver, L. (1993). Pragmatic versus syntactic approaches to training deductive reasoning. In R.E. Nisbett (Ed.), *Rules for reasoning* (pp. 165-203). Hillsdale, NJ: Erlbaum.

Collins, A.M., & Loftus, E.F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407-429.

De Koning, E., & Hamers, J.H.M. (1995). *Programma Inductief Redeneren 1* [Programme Inductive Reasoning I]. Utrecht: Utrecht University ISOR.

De Koning, E., & Hamers, J.H.M. (1996). *Programma Begrijpend Lezen. Experimentele versie* [Programme Reading Comprehension. Experimental version]. Utrecht: Utrecht University ISOR.

De Koning, E., Hamers, J.H.M., & Sijtsma, K. (1996). *Test inductief redeneren. Test en handleiding* [Test inductive reasoning. Test and manual]. Den Haag: HCO.

De Koning, E., Hamers, J.H.M., & Sijtsma, K. (in preparation). Designing a test for inductive reasoning: Integrative use of several data analyses methods.

Dodwel, P.C. (1960). Children's understanding of number and related concepts. *Canadian Journal of Psychology, 14*, 191-205.

Fisher, K.W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review, 87*, 477-531.

Fong, G.T., Krantz, D.H., & Nisbett, R.E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology, 18,* 253-292.

Fong, G.T., Krantz, D.H., & Nisbett, R.E. (1993). The effects of statistical training on thinking about everyday problems. In R.E. Nisbett (Ed.), *Rules for reasoning* (pp. 91-135). Hillsdale, NJ: Erlbaum.

Fromkin, V., & Rodman, R. (1988). *An introduction to language*. Fort Worth, TX: Holt, Rinehart and Winston.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution.* New York: Basic Books.

Garnhem, A. (1996). The other side of mental models: Theories of language comprehension. In J. Oakhill & A. Garnhem (Eds.), *Mental models in cognitive science. Essays in honour of Phil Johnson-Laird* (pp. 108-125). London: Erlbaum.

Gelman, S.A. (1989). Children's use of categories to guide biological inferences. *Human Development, 32*, 65-71.

Gernsbacher, M.A. (1996). The structure-building framework: What it is, what it might also be, and why. In B.K. Britton & A.C. Graeser (Eds.), *Models of understanding text* (pp. 289-311). Mahwah, NJ: Erlbaum.

Gick, M.L., & Holyoak, K.L. (1987). The cognitive basis of knowledge transfer. In S.M. Cormier & J.D. Haman (Eds.), *Transfer of learning: Contemporary research and applications* (pp. 132-143). San Diego: Academic Press.

Goldman, S.R., Varma, S., & Coté N. (1996). Extending capacity-constrained construction integration: Toward 'smarter' and flexible models of text comprehension. In B.K. Britton & A.C. Graeser (Eds.), *Models of understanding text* (pp. 73-113). Mahwah, NJ: Erlbaum.

Griggs, R.A., & Cox, J.R. (1982). The elusive thematic materials effect in Wason's selection task. *British Journal of Psychology, 73,* 407-420.

Halford, G.S. (1993). *Children's understanding. The development of mental models.* Hillsdale, NJ: Erlbaum.

Halpern, D.F. (1992). *Enhancing thinking skills in the science and mathematics.* Hillsdale, NJ: Erlbaum.

Hamers, J.H.M., De Koning, E., & Sijtsma, K. (1998). Inductive reasoning in third grade: Intervention promises and constraints. *Contemporary Educational Psychology, 23,* 132-148.

Hamers, J.H.M., Jeurissen, M., & De Koning, E. (1997). *Stimulering van denkvaardigheden van leerlingen in achterstandssituaties* [Stimulating thinking skills of low-SES children]. Den Haag: HCO.

Hamers, J.H.M., & Overtoom, M.Th. (1997). *Teaching thinking in Europe. Inventory of European programmes.* Utrecht: Sardes.

Hayes, J.R., & Simon, H.A. (1977). Psychological differences among problem isomorphs. In N.J. Castellan, Jr., D.B. Pisoni, & G.R. Potts (Eds.), *Cognitive theory,* (Vol 2, pp. 21-41). Hillsdale, NJ: Erlbaum.

Holland, J.H., Holyoak, K.J., Nisbett, R.E., & Thagard, P.R. (1986). *Induction. Processes of inference, learning and discovery.* Cambridge, MA: MIT Press.

Holyoak, H.J., & Nisbett, R.E. (1988). Induction. In R.J. Sternberg & E.E. Smith (Eds.), *The psychology of human thought* (pp. 50-91). New York: Cambridge University Press.

Jeurissen, M., Hamers, J.H.M., & De Koning, E. (1997). *Programma inductief redeneren 2* [Programme inductive reasoning 2]. Utrecht: Utrecht University ISOR.

Johnson-Laird, P.N. (1982). Propositional representations procedural semantics and mental models. In J. Mehler, E.C.T. Walker, & M.F. Garret (Eds.), *Perspectives on mental representation: Experimental and theoretical studies of cognitive processes and capacities* (pp. 111-131). Hillsdale, HJ: Erlbaum.

Johnson-Laird, P.N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

Johnson-Laird, P.N. (1989). Mental models. In M.I. Posner (Ed.), *Foundations of cognitive science* (pp. 469-499). Cambridge, MA: MIT Press.

Keil, F.C., & Batterman, N. (1984). A characteristic - to - defining shift in the development of word meaning. *Journal of Verbal Learning and Verbal Behaviour, 23,* 221-236.

Kintsch, W. (1974). *The perception of meaning in memory.* Hillsdale, NJ.: Lawrence Erlbaum.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A constructive-integration model. *Psychological Review, 95,* 163-182.

Klauer, K.J. (1989). *Denktraining für Kinder I* [Cognitive training for children I]. Göttingen: Hogrefe.

Klauer, K.J. (1990). *Denktraining für Kinder II* [Cognitive training for children II]. Göttingen: Hogrefe.

Klauer, K.J. (1991). Denktraining für Jugendliche [Cognitive training for the youth]. Göttingen: Hogrefe.

Klauer, K.J., & Phye, G.D. (1995). *Cognitive training for children*. Seattle: Hogrefe & Huber.

Kofsky, E. (1966). A scalogram study of classificatory development. *Child Development, 37*, 191-204.

Komatsu, L.K. (1992). Recent views on conceptual structure. *Psychological Bulletin, 112*, 500-526.

Kosslyn, S. (1976). Using imagery to retrieve semantic information: A developmental study. *Child Development, 47*, 434-444.

Kosslyn, S. (1990). Mental imagery. In D.N. Osherson, S.M. Kosslyn, & J.M. Hollerbach (Eds.), *Visual cognition and action. An invitation to cognitive science* (Vol 2, pp. 73-97). Cambridge, MA: MIT press.

Lunzer, E.A. (1960). Some points of Piagetian theory in the light of experimental critisism. *Journal of Child Psychology and Psychiatry, 1*, 191-202.

Manktelow, K.I., & Evans, J.St.B.T. (1979). Facilitation of reasoning by realism: Effect or non-effect? *British Journal of Psychology, 70*, 477-488.

Mill, J.S. (1887). A *system of logic*. New York: Harper & Brothers.

Minsky, M. (1975). A framework for representing knowledge. In P.H. Winston (Ed.), *The psychology of computer vision* (pp. 211-277). New York: McGraw-Hill.

Morris, M.W., & Nisbett, R.E. (1993). Tools of the trade: Deductive schemas taught in psychology and philosophy. In R.E. Nisbett (Ed.), *Rules for reasoning* (pp. 228-256). Hillsdale, NJ: Erlbaum.

Nisbett, R.E. (1993). *Rules for reasoning*. Hillsdale, NJ: Erlbaum.

Nisbett, R.E. , Fong, G.T. , Lehman, D.R., & Cheng, P. (1993). Teaching reasoning. In R.E. Nisbett (Ed.), *Rules for reasoning* (pp. 15-54). Hillsdale, NJ: Erlbaum.

Nisbett, R.E., Krantz, D.H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90*, 339-363.

Oakhill, J. (1996). Mental models in children's text comprehension. In J. Oakhill, & A. Garnhem (Eds.), *Mental models in cognitive science. Essays in honour of Phil Johnson-Laird* (pp. 190-210). London: Erlbaum.

Pascual-Leone, L. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica, 32*, 301-345.

Paivo, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart and Winston.

Piaget, J. (1970). Piaget's theory. In P.H. Mussen (Ed.), *Carmichael's handbook of child development* (pp. 703-732). New York: Wiley.

Pylyshyn, Z.W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin, 80*, 1-24.

Raven, J.C. (1958). *Standard progressive matrices*. London: Lewis.

Resnick, L.B., & Klopfer, L.E. (1989). *Towards the thinking curriculum. Current cognitive research*. Alexandria, VA: Association for Supervision and Curriculum Development.

Reich, S.S., & Ruth, P. (1982). Wason's selection task: Verification, falsification and matching. *British Journal of Psychology, 73,* 395-405.

Rosch, E.H., & Mervis, C.B. (1975). Family resemblences: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573-605.

Rumelhart, D.E., & Ortony, A. (1977). The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro, & W.E. Mantague (Eds.), *Schooling and the acquisition of knowledge* (pp. 99-135). Hillsdale, NJ: Erlbaum.

Schank, R.C., & Abelson, R.P. (1977). *Scripts, plans, goals, and understanding.* Hillsdale, NJ: Erlbaum.

Sternberg, R.J. (1977). *Intelligence, information processing and analogical reasoning. The componential analysis of human abilities.* Hillsdale, NJ: Erlbaum.

Sternberg, R.J. (1996). Cognitive psychology. Orlando: Halcourt Brace.

Sternberg, R.J., & Berg, C.A. (1992). Intellectual development. Cambridge: Cambridge University Press.

Sternberg, R.J., & Powell, J.S. (1983). Comprehending verbal comprehension. *American Psychologist, 38,* 878-893.

Thorndike, E.L. (1919). *Principles of teaching.* New York: A.G. Seiler.

Turner, A., Britton, B.K., Andraessen, P., & McCutchen, D. (1996). A predication semantics model of text comprehension and recall. In B.K. Britton & A.C. Graeser (Eds.), *Models of understanding text* (pp. 33-71). Mahwah, NJ: Erlbaum.

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

Wertsch, J.V., & Kanner, B.G. (1992). A sociocultural approach to intellectual development. In R.J. Sternberg & C.A. Berg (Eds.), *Intellectual development* (pp. 328-349). Cambridge: Cambridge University Press.

Wood, D.J. (1980). Teaching the young child: Some relationships between social interaction, language, and thought. In D.R. Olson (Ed.), *The social foundations of language and thought* (pp. 66-80). New York: Norton.

Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension. An experimental investigation.* Cambridge: Cambridge University Press.

8

Thinking skills in realistic mathematics

J.M.C. Nelissen

Introduction

One of the most enduring ideas concerning mathematics instruction is the following: mathematics consists of a set of indisputable rules and knowledge; this knowledge has a fixed structure and can be acquired by frequent repetition and memorization. In the past twenty-five years, far-reaching changes have taken place in mathematics instruction. More than in any other field, such changes were influenced by mathematicians who had come to view their discipline in a different light. Their observations went a long way towards stimulating a process of renewal in mathematics instruction. New consideration was given to such fundamental questions as: how might mathematics best be taught, how might children be encouraged to show more interest for mathematics, how do children actually learn mathematics, and what is the value of mathematics?

According to Goffree, Freudenthal, and Schoemaker (1981), the subject of mathematics is itself an essential element in 'thinking' through didactical considerations in mathematics instruction. Moreover, the notion is emphasized that knowledge is the result of a learner's activity and efforts, rather than of the more or less passive reception of information. Mathematics is learned, so to say, on one's own authority. From a teacher's point of view there is a sharp distinction made between teaching and training. To know mathematics is to know why one operates in specific ways and not in others. This view on mathematics education is the basic philosophy in this chapter (Von Glazersfeld, 1991). In order to understand current trends in mathematics education, we must consider briefly the changing views on this subject.

The philosophy of science distinguishes three theories of knowledge. Confrey (1981) calls these absolutism, progressive absolutism and conceptual change. In absolutism, the growth of knowledge is seen as an accumulation, a cumulation of objective and empirically determined factual material. According to progressive absolutism a new theory may correct, absorb, and even surpass an older one. Proponents of the idea of conceptual change have defended the point of view that the growth of knowledge is characterized by fundamental (paradigmatical) changes and not by the attempt to discover absolute truths. One theory may have greater force and present a more powerful argument than another, but there are no objective, ultimate criteria for deciding that one theory is incontrovertibly more valid than another (Lakatos, 1976). Mathematics has long been considered an absolutist science. According to Confrey (1981), it is seen as the epitome of certainty, immutable truths and irrefutable methods. Once gained, mathematical knowledge lasts unto eternity; it is discovered by bright scholars who never seem to disagree, and once discovered, becomes part of the existing knowledge base.

Leading mathematicians however have now abandoned the static and absolutist theory of mathematics (Whitney, 1985). Russell (in Bishop, 1988) once explained that mathematics is the subject in which we never know what we are talking about, nor whether what we are saying is true. Today mathematics is more likely to be seen as a fluctuating product of human activity and not as a type of finished structure (Freudenthal, 1983). Mathematics instruction should reveal how historical discoveries were made. It was not (and indeed is still not) the case that the practice of mathematics consists of detecting an existing system, but rather of creating and discovering new ones. This evolving theory of mathematics also led to new ideas concerning mathematics instruction. If the essence of mathematics were irrefutable knowledge and ready-made procedures, then the primary goal of education would naturally be that children mastered this knowledge and these procedures as thoroughly as possible. In this view, the practice of mathematics consists merely of carefully and correctly applying the acquired knowledge. If, however, mathematicians are seen as investigators and detectives, who analyse their own and others' work critically, who formulate hypotheses, and who are human and therefore fallible, then mathematics instruction is placed in an entirely different light. Mathematics instruction means more than acquainting children with mathematical content, but also teaching them how mathematicians work, which methods they use and how they think. For this reason, children are allowed to think for themselves and perform their own detective work, are allowed to make errors because they can learn by their mistakes, are allowed to develop their own approach, and learn how to defend it but also to improve it whenever necessary. This all means that students learn to think about their own mathematical thinking, their strategies, their mental operations and their solutions.

Mathematics is often seen as a school subject concerned exclusively with abstract and formal knowledge. According to this view, mathematical abstractions must be taught by making them more concrete. This view has been opposed by Freudenthal (1983) among others. In his opinion, we discover mathematics by observing the concrete phenomena all around us. That is why we should base teaching on the concrete phenomena in a world familiar to children. These phenomena require the use of certain classification techniques, such as diagrams and models (for example, the number line or the abacus). We should therefore avoid confronting children with formal mathematical formulas which will only serve to discourage them, but rather base instruction on rich mathematical structures, as Freudenthal calls them, which the child will be able to recognize from its own environment. In this way mathematics becomes meaningful for children and also makes claer that children learn mathematics not by training formulas but by reflecting on their own experiences.

In the 1970s, the new view of mathematics, often referred to as mathematics as human activity, led to the rise of a new theory of mathematics instruction, usually given the designation: realistic. As it now appears, this theory is promising, but it is not the only theoretical approach in mathematics instruction; three others can be distinguished: the mechanistic, the structuralist and the empirical (Treffers, 1991). In the following we just give a brief characteristic of each approach, because it is beyond the scope of this chapter to discuss the three schools in extension:

- The mechanistic approach reflects many of the principles of the behaviouristic theory of learning; the use of repetition, exercises, mnemonics, and association comes to mind. The teacher plays a strong, central role and interaction is not seen as an essential element of the learning process. On the contrary, mathematics class focuses on conclusive standard procedure.
- According to the second approach - the structuralist - thinking is not based on the children's experiences or on contexts, but rather on given mathematical structures. The structuralist tends to emphasize strongly the teacher's role in the process of learning.
- The outstanding feature of the third trend - the empirical - is the idea that instruction should relate to a child's experiences and interests. Instruction must be child-oriented. Empiricists believe that environmental factors form the most important impetus for cognitive development (Papert, 1980). Empiricists emphasize spontaneous actions.

Realistic mathematics instruction as progressive mathematization

In this section we present five features which characterize realistic mathematics.

At first we are dealing with learning in a context and second with the use of models. The third point (the mathematical subjects are not atomized but interwoven) is not of so much relevance for this book, while the three characteristics of the process of mathematization (construction, reflection and interaction) are analysed in the following sections.

The new realistic approach to learning and thought processes in children has far-reaching consequences. Mathematization is viewed as a constructive, interactive and reflective activity. To begin, the point of departure for education is not learning rules and formulas, but rather working with contexts. A context is a situation which appeals to children and which they can recognize in theory. This situation might be either fictional or real, and forces children to call upon the knowledge they have gained by experience - for example in the form of their own informal working methods - thereby making learning a meaningful activity for them. A well-chosen context can induce an active thought process in children, as the following example shows.

Let us start to give children of, say, 11 years the following formal and bare problem, not presented in a context: 6 ÷ 3/4. Many of them will have a great deal of trouble finding a solution (Streefland, 1991). Some will answer, for example: 2/4, 3/24 or 4 1/2. They manipulate at random with the given numbers, for instance 6 ÷ 3 = 2 , so 6 ÷ 3/4 must be 2/4. This child views fractions as whole numbers and so do other students (Lesh et al., 1987). But some students will calculate that 6 x 4 = 24 and that 24 divided by 3 equals 8. It is true that the latter answer is correct, but when these children are questioned more closely, it turns out that they understand almost nothing about the operation which they themselves have just performed. They just remembered a rule they learned by heart, they know that the given solution is correct however they don't know why.

Now, the same children are next given the following context problem which is accompanied by a picture: a patio is 6 metres long; you want to put down new bricks and the bricks you are going to use measure 75 centimetres in length (3/4 of a metre). How many bricks will you need for the length? This problem is the same as the previous one, but it has now been presented within a context, a picture of a patio and the bricks to put down. This presentation elicits a child's own, informal approach: measuring out. This approach provides insight into the problem, something which the symbolic form (6 ÷ 3/4) did not do. Some students even manipulated and took the measure in reality, this means they measured out step by step 75 centimetres and after 8 steps they counted 6 metres. So the answer must be 'eight', they concluded. This example demonstrates that working with contexts - which, if carefully constructed, can be considered paradigmatic examples - can form the basis for subsequent abstractions and for conceptualization. That is because thinking must achieve a higher, abstract level and at that level these particular contexts no longer serve a purpose. That is not to say that a process of decontextualisation occurs, but rather recontextualisation. The children continue to work with contexts, but these contexts become increasingly formal in nature;

they become mathematical contexts. Their connection with the original context, however, remains clear. The process by which mathematical thinking becomes increasingly formal is called the process of progressive mathematization. Contexts, thus, have various functions. They may refer to all kind of situations and to fantasy situations (Van den Heuvel-Panhuizen, 1996). It is important that the context offer support for motivation as well as reflection. A context should indicate certain relevant actions (to take measures in the example above), provide information which can be used to find a solution-strategy and/or a thinking-model.

Of course, leaving the construction to the students does not guarantee the development of successful strategies. However it guarantees that students get the opportunity to practice mathematician's thinking and problem solving processes. Strategies are tried, tested and elaborated in various situations.

In the previous discussion we have not argued that a student presented with 'bare' numerical tasks (like 6 ÷ 3/4) will necessarily fail to solve the problem. Hence we were not suggesting either that students who are given context problems will necessarily produce the right solution. In recent research there is found a strong tendency of children to react to context problems ('word problems') with disregard for the reality of the situations of these problems. Let us give two examples of items used in research (Greer, 1997; Verschaffel et al., 1997):

- 'An athlete's best time to run a mile is 4 minutes and 7 seconds. About how long would it take him to run 3 miles?'
- 'Steve has bought 4 planks of 2.5 metre each. How many planks of 1 metre can he get out of these planks?'

In four studies, discussed by Greer (1997), the percentage of the number of students demonstrating any indication of taking account of realistic constraints is: 6%, 2%, 0% and 3%. The student's predominating tendency to apply rules clearly formed an impediment to thoroughly understanding the situation.

Verschaffel et al. (1997) confronted a group of 332 students (teachers in training) with word problems and found they procuced 'realistic' responses in only 48% of cases. Moreover the pre-service teachers considered these 'complex and tricky word problems' as inappropriate for (fifth grade) children. The goal of teaching word problem solving in elementary school, after their opinion, was "...learning to find the correct numerical answer to such a problem by performing the formal-arithmetic operation(s) 'hidden' in the problem" (Verschaffel et al., 1997, p. 357).

When solving word problems students should go beyond rote learning and mechanical exercises to apply their knowledge (Wyndhamn & Säljö, 1997). Their research showed that students (10-12 years of age) gave in most cases logically inconsistent answers. The authors interprete these findings by claiming that the students focus on the syntax of the problem rather than on the meaning. That means that the well-known rule-based relationship between symbols results in less of attention being paid to the meaning. The students follow another 'rationality', that is, they consider word problems as mathematical exercises "... in which a algorithm is hidden and is supposed to be identified." (Wyndhamm & Säljö, p. 366). Hence they do not know or realize that they are expected to solve a real life problem.

Reusser and Stebler (1997) discuss another interesting research finding namely the fact that pupils 'solved' unsolvable problems without 'realistic reactions'. For example:

- 'There are 125 sheep and 5 dogs in a flock. How old is the shephard?' (Greer, 1997).

A pupil questioned by the investigators gave as his opinion: 'It would never have crossed my mind to ask whether this task can be solved at all'. And another pupil said: 'Mathematical tasks can always be solved'. One of the author's conclusions is that a change is needed from stereotyped and semantically poor, disguised equations to the design of intellectually more challenging 'thinking stories'. What we need are better problems and better contexts. Finally, Reusser and Stebler (1997) - following Gravemeijer (1997) - give as their interpretation of the research findings that the children are acting in accordance with a typical school mathematics classroom culture.

Second, the process of mathematisation is characterized by the use of models. Some examples are schemata, tables, diagrams, and visualizations. Searching for models - initially simple ones - and working with them produces the first abstractions. Children furthermore learn to apply reduction and schematization, leading to a higher level of formalization. We will demonstrate, once again this using the previous example. To begin, children are able to solve the brick problem by manipulating concrete materials. For instance, they might attempt to see how often a strip of paper measuring 3/4 of a metre fits in a 6-metre-long space. At the schematic level, they visualize the 6-metre-long patio and draw lines which mark out each 3/4 of a metre or 75 centimetres. The child adds 75 + 75 + 75...until the 6 metres have been filled. The visualization looks as follows:

------/------/------/------/------/------/------/------/
3/4 3/4 3/4 3/4 3/4 3/4 3/4 3/4

An example of reasoning on a formal-symbolic level is as follows: 75 centimetres fits into 3 metres 4 times. We have 6 metres, so we need 2 x 4 = 8 bricks. The formula initially tested can also be applied, but this time with insight: 1/4 metre fits 4 times into 1 metre, so it fits 24 times into 6 metres. But I only have 3/4 of a metre, so I have to divide 24 by 3, and that makes 8. At this formal level, moreover, the teacher can also explore the advantages and disadvantages of the two methods with the children.

Third, an important element of realistic mathematics instruction is that subjects and curricula (such as fractions, measurement and proportion) are interwoven and connected, whereas in the past, the subject matter was divided - and so atomized.

Fourth, two other important characteristics of the process of mathematization are that it is brought about both by a child's own constructive action and by the child's reflections upon this action.

Finally, learning mathematics is not an individual, solitary activity, but rather an interactive one.

Construction

Learning mathematics is a constructive activity, an aspect which has been emphasized by many authors (including Bruner, 1986, 1996; Cobb, 1994; Cobb et al., 1997; Resnick & Klopfer, 1989; Steffe, Cobb, & Von Glazersfield, 1988). Children construct internal, mental representations. These might be concrete images, schemata, procedures, working methods at the abstract-symbolic level, intuitions, contexts, schemata of solutions, or thought experiments. To make clear how individuals construct different kinds of representations we present now a example (this example is above the cognitive level of school children).

Suppose we were able to tighten a rope around the equator so that it is taught and lies flat on the surface. We cut the rope and insert a metre-long piece of rope between the two cut ends. Once again we tighten the rope so that it is taught all around. The question is: How far above the ground is the rope now? The mental representations of many adults will contain various elements. To start, they will ask themselves what the circumference of the earth is: 44,000 kilometres. The rope - they can picture it before their very eyes - must therefore measure the same in length. Another metre - they reason to themselves - scarcely matters in proportion to that enormous distance. Probably the rope barely lifts off the ground. These mental representations actually consist of concrete images which form a basis for solving the problem, conceived of as the relationship between that one metre and the entire circumference of the earth. A mathematically trained problem-solver will construct entirely different representations. He or she will immediately dismiss any concrete facts and reduce the earth to a circle, focusing in on the relationship between the radius and the circumference. This relationship is then converted into a formula, 2pi R. The representation is created by converting a concrete problem into an adequate mathematical formula. The rope lies about 16 centimetres above the surface of the earth.

Learning mathematics as a constructive activity means that a child's own discoveries are taken seriously. This does not mean that their discoveries are always on the mark, but they do give the teacher a recognizable handle from which he or she can begin to teach. The teacher learns the general outlines of the representations of children and can adjust his approach accordingly. But what is the function of representations in mathematics instruction? The representational point of departure and the 'representational view of mind' seems to require some constructivist comment. What is criticized and rejected is the metaphor of the mind as a mirror that reflects a mathematically prestructured environment unaffected by individual and collective human activity (Von Glazersfeld, 1991). Correct, internal representations are constructed by confronting them with external representations. These are socially and culturally determined (Cobb, 1994). Children do indeed actively construct their own mathematical knowledge, but their purpose is to participate increasingly in taken-as-shared mathematical practices. These practices are played out both in the classroom as in society and science.

If rules and procedures are prescribed prematurely and one-sidedly, blocking a child's own representations, problems will ensue. The following recorded fragment of conversation serves as a concrete illustration. Henry, a good pupil of 9 years of age, is busy working out subtraction problems in his mathematics workbook. The book gives the following formula to complete the problems:

94 - 52 = ... - ... - ... = ... - ... = ...

Researcher:	How do you do that?
Henry:	First you subtract 50, that's 44, and then you subtract 2, making 42.
Researcher:	Do you always have to do it that way?
Henry:	Yes.
Researcher:	Can't you subtract 2 first?
Henry:	That's not allowed.
Researcher:	But why not?
Henry:	Because the book says. (He points out the following example: 54 - 31 = ., 54 - 30 - 1= ., 24 - 1 = 23)
Researcher:	What if I subtract 2 first anyway?
Henry:	But that's against the rules.
Researcher:	Will the answer be different if you subtract 2 first?
Henry:	Maybe.

Let us now look at an example in which children are given a chance to develop their own constructions. A group of 10 to 11-year old students was asked how they would go about solving the following problem. There are a number of bottles on a table, and each bottle has a different shape. None of them has labels, so no one can tell just how much each bottle can hold. How would you figure out which bottle can hold the most water? The children were asked to present their ideas and talk about one another's ideas. One child suggests weighing the bottles. No, another says, hold them under water and see how much the water rises. A third suggests dumping the contents on the floor and seeing which puddle is the biggest. This is good example of a practical situation in which children are constructing knowledge, taken-as-shared (Cobb, 1994). Note that here is qualified the word 'shared'. The children's solutions do not match precisely, but they are considered 'compatible' and are therefore worth discussing. And this is what happened. The children were criticizing and commenting each other solutions until one child proposed to use a glass as a measure. All the children insisted with this idea but now rose the question of how big the glass should be. At the end of the discussion they decided to choose a small glass but not too small. The idea that children's own constructions form the point of departure for the teaching-learning process in mathematics instruction is one of the fundaments of the realistic school. Confrey (1985) argues that a person's knowledge is necessarily the product of his/her own constructions or mental acts. Thus s/he can have no direct or unmediated knowledge of any objective reality. Knowledge is created by means of images or representations and these are products of our mental actions (Gardner, 1987). But if their own constructions are so very important, children should be allowed to nurture their own constructions (whatever their quality may be). It is not necessarily that this would lead to anarchy and blocked communication during mathematics class, because constructions arise through interaction with other children and with the teacher. Bruner (1986) too asserts that constructivism is not a sort of cultural relativism or an homage to the proposition that 'anything goes'. Neither should constructions be understood in Piagetian terms. Piaget (1976) was concerned with individual constructions which arise from the subject's own position and which are the result of intrinsic and autonomous processes; 'mathematical practices' have relatively little influence on them.

Lo, Grayson, Wheatly, and Smith (1990) discuss the close relationship between construction and interaction in the following fashion: "From a constructivist's perspective learning occurs when a child tries to adapt her functioning schemes to neutralize perturbations that arise through interactions with our worlds" (p. 116). Two important aspects, constructions and interactions, are important in the above statement. Although construction of knowledge is a personal act, it is by no means an isolated activity as many people's interpretations of constructivism imply. Constructivists recognize the importance of social actions as 'the most frequent source of perturbations' (Salomon, 1989). Interactions, thus, lead to construction, because the process of interacting often causes perturbations in the normal pattern of behaviour - particularly when unexpected problems arise - which the person involved will try to resolve by seeking his or her own solutions (constructions). A study conducted by Saxe (1988) investigated how young Brazilian candy sellers, who generally had little or no schooling, had learned mathematics. These children had learned to fix cost prices, to calculate skillfully in cash amounts, and to think in ratios (3 pieces of candy is 500 cruzeiros, 1 piece is 200), but when confronted with classroom problems - for example, reading and comparing double-digit numbers - they were at a complete loss. They had created their own constructions in the process of solving problems encountered in daily social interaction. Constructions, in turn, may once again lead to interactions, in the sense that constructions are 'tested' in interactions: do my ideas make sense, are they valid?

The construction of internal mental representations is one of the features of the process of learning mathematics. We conceive the development of internal representations as a process of signification (Kirshner & Whitson, 1997; Walkerdine, 1997). So we do not make a distinction between an externally represented world and an internally representing world. Representation is looked upon as a process in which new signs in a cyclic process of signification constantly emerge. An internal representation (signifier) transforms and is the basis (signified) for the construction of a new internal representation (signifier). Hence a person constructs internal representations on the basis of internal representations.

The process of learning mathematics distinguishes itself from the process of learning other school subjects to the extent that in mathematics, constructions - in the sense of internal representations which children formulate based on knowledge gained through experience - consistently show a closer correspondence with external representations than in other school subjects (Cobb et al., 1987; Freudenthal, 1983). Children gain experience in the use of measurements, numbers, ratios, and fractions and construct (intuitive) representations. In theory these representations form a basis upon which the teacher can build, although this is not always the case. Particularly this is not the case when children learn to operate with mathematical symbols. Many errors are based on the default nature of natural language encoding processes, as Kaput (1987) has stated. Kaput discusses the well-known Student-Professors problem (Clement, 1982). At a certain university, for every 6 students there is one professor. Write an algebraic equation that expresses the relation between the number of students and professors. Consistently the natural language overrides the algebraic rules as is shown by the high error rates (40-80%) across age and the predominance of the '6s = p' error, typified by Kaput as the 'reversal error'.

The representations which the children construct concerning physical phenomena -also known as preconceptions, misconceptions or intuitive ideas - generally deviate so far from actual physical reality that they are useless as a basis for conceptualization. For example, children associate energy with eating a Mars bar; in their minds, evaporation is the same as disappearing, heat means feeling nice and warm, and light is a ray which goes from the eye to an object, rather than the other way around (Van der Valk, 1989). These representations are useless in instruction, but the teacher must be familiar with them in order to understand the problems that arise in conceptualization. This applies as well to representation in other school subjects. For example, when studying history, children have a great deal of trouble forming representations based on their own experience. One child, for instance, regularly confused 'Enlightenment' with 'more light'.

Interaction

Realistic instruction in mathematics is not only constructive but also interactive. Several authors have pointed out the importance of this interactive, or social, dimension of learning. Bishop (1988) has argued to replace 'impersonal learning' and 'text teaching' with 'mathematical enculturation', thereby emphasizing the relationship between education and culture. Pimm (1990) uses the term 'mathematical discourse', while Salomon (1989) speaks of 'cognitive partnership'.

Granott and Gardner (1994) constructed a theoretical framework of interaction, based on the view of multiple intelligence approach. After their opinion the effect of interaction depends on two dimensions. The first dimension is the relative expertise: from symmetric ('parallel activity' for instance) till asymmetric ('apprenticeship'). The second dimension is the degree of collaboration. 'Scaffolding' is for instance an example of collaboration of a high degree, while 'imitation' is in fact an independent activity (no collaboration).

Interactive teaching has also been called 'cooperative learning' (Slavin, 1986), 'classroom discourse' (Cazden, 1988), 'mutual instruction' (Glaser, 1991) 'guided construction of knowledge' (Mercer, 1995) and 'interactive instruction' (Treffers & Goffree, 1985). Bruner (1986) a proponent of 'discovery learning' - learning on one's own - in the 1960s, revised his ideas several years ago: "My model of the child in those days was very much in the tradition of the solo child mastering the world by presenting it to himself in his own terms. In the intervening years I have come increasingly to recognize that most learning in most settings is a communal activity, a sharing of the culture. It is this that leads me to emphasize not only discovery and invention but the importance of negotiation and sharing." (Bruner, 1986, p. 127). There is no contradiction however between invention and sharing, the contrary is true: both activities influence each other. If a person makes his own invention, it is worthwhile and even in many cases necessary to discuss this invention. And this discussion is the basis for new inventions.

Nowadays, this view of (cognitive) development and learning is classified as social-constructivism, a classification which meshes with the realistic approach to mathematics instruction. In some studies (Driver, Asoko, Leach, Mortimer, & Scott, 1994; Roazzi & Bryant, 1994) is defended the point of view that learning and thinking always take place in a social situation. Learning, they say, is situated learning (Kirshuer & Whitson, 1997), cognition is social cognition. Bruner's (1986) designation for the acquisition of knowledge is 'negotiation of meaning'. Not only words, concepts, gestures, and rituals, but also numbers, symbols, images, visual and graphic representations, etc. have a whole range of meanings. In the case of children, these meanings are frequently highly subjective. In response to the question "How old are you?", one child was heard to answer "I'm four, but when I ride in the bus I'm three" (in the Netherlands children under the age of four can ride public transportation free of charge). Another child believes that when teachers roll the dice, they get double sixes more often than children do. Teaching, says Bruner, means negotiating meaning. You say that zero is "nothing", but what then does zero degrees mean on the thermometer? A child does not believe that her face has a surface. "Why not?" asks the teacher. "Because it isn't length times width," the child responds. By applying the Socratic method, the teacher was gradually able to convince the child that the concept 'surface' was not exclusively linked to the algorithm l x w. Bruner's tribute to Vygotsky (1977) is not at all surprising. According to the latter, a child's higher psychic functions (such as language and thought) first take shape as a social (interactive) activity and only later as an individual activity. Language first functions as a means of communication; afterwards it becomes internalized and serves an individual, self-regulatory function. One of the key concepts in Vygotsky's theory is that education should anticipate actual development. He refers in this connection to the 'zone of proximal development', and it is this idea which inspired Bruner's 'negotiation of learning'. Both are concerned with interactive instruction, which Freudenthal (1984) typifies as 'anticipatory learning' and Van Parreren (1988) as 'developing education'.

Realistic mathematics instruction is interactive, even though children must naturally be given the chance to work independently. As demonstrated in studies carried out by Doise and Mugny (1984), however, the point is that allowing children to experience various perspectives - in other words, showing them that there are other children with other ideas about how to solve a mathematics problem - will stimulate their thinking. Mechanistic (and individualistic) mathematics instruction can exclude such experiences because children are required to comply with the procedures given in their textbook. Discussion is restricted because the essence of instruction lies in teaching irrefutable procedures. Realistic instruction, on the other hand, is based on the exchange of ideas, not only, as in the past, between teacher and pupils, but also between the pupils themselves. Interaction stimulates reasoning, using and analysing arguments, thinking about own solutions and the solutions of others, so interaction reinforces the thinking ability. Currently, social interaction in the classroom is receiving much attention where an important line of research focuses on effects of small group work (Hiebert, 1992). It goes without saying that there should be a 'genuine' occasion for discussion. That is why the point of departure for realistic instruction is frequently a problem in context; again, this emphasizes how tightly interwoven context and interaction are.

A simple example is the following: a teacher asks his class (6 and 7-year-olds) to think up as many ways as they can of doing the sum 5 + 6. The children are allowed to discuss this among themselves and together they came up with several methods: counting from 6 on up; adding 5 + 5 to reach 10 and then adding 1; counting the fingers on both hands and then adding 1; adding 6 + 6 and subtracting 1, etc. Teacher and pupils then discuss which method is the handiest and why. This process leads the children to reflect spontaneously on their own actions: they are forced to compare their methods with those of the other children and consider which is the best (this example makes clear that in realistic mathematics instruction the students not only are confronted with context problems, but with bare problem as well).

Reflection

According to Hiebert (1992), reflection or metacognition can be defined as the conscious consideration of one's experiences, often in the interests of establishing relationships between ideas or actions. It involves thinking back on one's experiences and taking the experiences as objects of thought. With respect to terminology, reflection is seen most frequently in Russian research reports (Davydov, Lompscher, & Markova, 1982; Nelissen & Tomic, 1996; Stepanov & Semenov, 1985; Zak, 1984). The terms self-monitoring or self-regulation are also applied (Glaser, 1991). It would be most tempting to spend a great deal of time discussing the many questions, controversies and dilemmas which have arisen in the literature concerning the concept of reflection - consider, for example, the discussion concerning the extent to which reflective skills are general or contingent on context (Perkins & Salomon, 1989).

We generally do not reflect while performing a routine task, for the simple reason that there is no cause to do so then. There is, however, reason for reflection whenever we are confronted with a problem for which there is no immediate solution at hand. Reflection begins when we ask ourselves how best to approach the problem: 'Should I do it this way or that way?' (planning). Once we have set to work, other questions arise: "Is this working?" (self-monitoring), perhaps even "Can I do it?" (self-evaluation). Other obvious questions are "Will this succeed?" (anticipation) and, finally, "Am I happy with this?" (evaluation). If the solution turns out to be a dead end, then we are forced to ask ourselves "Shouldn't I try something else?" (consider switching methods). These are, in brief, the most important elements of reflection during the process of problem-solving. Reflection plays a significant role in learning to solve mathematical problems, and indeed in human action in general. Through reflection students learn to analyse their own actions critically and also become less dependent on their teacher. Their thinking becomes more systematic, however this is not the case with all students. Some students must be stimulated frequently. Reflection also allows them to investigate problem-solving methods and procedures for general applicability, and increases the flexibility of their thinking. The most important aspect, however, is that reflection builds self-confidence by allowing pupils to discover what they really think and why they think it. Without this knowledge, every result might seem - and in fact might very well be - serendipitous, an awareness that does little to build up confidence: the pupil might not be so lucky the next time around.

That reflection is closely tied to the mathematical learning process and to mathematical thinking can be deduced from the proposition, discussed above, that mathematization is a constructive activity. This activity, in turn, is permanently linked to interaction, as we have seen. We can imagine the connection between construction, interaction and reflection in the following manner: constructive thinking implies that interaction takes place concerning our own constructions (representations). We must naturally be able to test our own constructions and find out how valuable they are. By exploring - and anticipating - the ideas and criticisms expressed by others, we gain greater insight into our own ideas. Knowing what these ideas are and how we ultimately came up with them is called reflection. We internalize the dialogue which we originally conducted with others, turning it into a dialogue 'with ourselves'. Reflection, thus, is nothing less than 'internalized dialogue': from primarily inter-individual to intra-individual activity. Through reflection, we continue to create new constructions, each time at a higher level. In short, reflection is development.

There is a relationship between reflection and the process by which pupils solve mathematical problems. In one study, the reflective thinking and mathematical problem-solving skills of two groups of students were compared (Nelissen, 1987). One group had been taught mathematics according to the realistic method (84 students) and the other according to the mechanistic method (60 students). One striking result was that the students in the first group were more flexible in their thinking than those in the second group, specifically because they were better able to switch strategies whenever necessary. They were less likely to concentrate purely on algorithmic solutions, and were able to develop strategies on the basis of their own experience. They tended to check their own approach without prompting and were aware of their own thought processes. In general, the children who were better able to solve problems were also better at reflection, in the 'realistic' group this was 43% of the students, while in the control-group this was 10% of the students.

A number of factors might serve to explain this close, positive relationship: (a) the school curriculum followed by children in the experimental group was based on problem-solving within a rich context. Instead of being given fixed, standard procedures to learn, pupils in these schools were allowed to think up their own constructions. Through interaction they were encouraged to reflect on their own approach. In this way, problem-solving and reflection were stimulated in relation to each other. Note that the children were not given direct, separate training in reflection. Research has shown that the training of functions in a separate programme has only a limited effect (Derry & Murphy, 1986); (b) by commenting regularly on each other's actions, the 'realistic' children were able to generate a reflective attitude which may have had a positive impact on their problem-solving skills; (c) the children in the experimental group were taught the concepts, models and procedures they needed to solve problems and engage in reflection. To be able to reflect on a specific subject, they needed to acquire domain-specific knowledge; (d) reflection will only prove beneficial after children have come to view the actions they are reflecting on as meaningful. The children in the 'realistic' group found mathematics and problem-solving a meaningful activity. They were therefore more inclined to reflect on problem-solving than the children in the 'mechanistic' group. Children in the latter group saw little reason to apply their own reality to learning mathematics, because this reality was continually supplanted by prescribed standard algorithms. A study conducted by Stepanov and Semenov (1985) revealed that in order to be able to reflect on the process of problem solving, children must first see their own actions during this same process as meaningful. Meaning must therefore be given due attention during instruction, if children are to find reflection a meaningful activity. Reflection, in turn, is vital if the mathematization process is to run smoothly.

Solving mathematical problems

One important objective in mathematics instruction is that children be able to apply the concepts and skills they have acquired with reasonable success. Problem-solving is considered by many to be one of the most important areas of application. That is why so many researchers are interested in whether, and if so, how children solve mathematics problems. Where upon are processes of problem solving based, on declarative knowledge or on knowledge of procedures?

The respective opponents and proponents of 'declarative representations' and 'procedural representations' have been engaged in a vehement debate since the early 1970s (Gardner, 1987). Adherents of the first believe that the knowledge base is the most important factor in problem-solving, while adherents of the second relate success largely to the use of procedures and strategies. In the 1980s the dispute concerned whether such knowledge or such (reflective) procedures were general or domain-specific. This controversy led to yet another split in both camps. Although the debate rages on, its resolution seems to be in sight, specifically because human thought is increasingly being characterized as modular. Learning is therefore by no means a 'content blind' process; neither, according to Gardner (1987) are there such things as 'general cognitive architectures', as Piaget (1976) suggested. Several leading authors appear to share this opinion: Bonner (1990), Resnick and Klopfer (1989), and Schoenfeld (1989). All of these above mentioned researchers tend, albeit from different backgrounds, to maintain that problem-solving (in mathematics) will be most successful when based on a well-organized selection of domain-specific knowledge, but that the use of procedures which are (once again) domain-specific is also indispensable. Experts tend to use domainspecific procedures and principles while novices are more likely to choose general strategies. For this reason novices often fail to solve the problems (Caillot, 1991).

Problem-solving in mathematics, then, is also characterized by a specific mathematical approach, involving the use of domain-specific concepts, tools (procedures) and ways of thinking. A child who has not mastered this approach will have difficulty when solving mathematical problems. A few examples follow. A classroom of six- and seven-year-olds were given the following problem: 2 friends live next door to each other, one at number 3 and one at number 5. How many houses do they live in? The children answered: 8 houses. Numbers mean little to these children, except that they can be added up. Assigning a meaning to numbers is an important component of a mathematical approach. Many children (eleven or twelve years old) have difficulty solving the following type of problem: a walkman costs $150 after the price has gone up by 20%. What was the original price? The answer most frequently given is: 20% of $150 is $30; the original price was therefore $120. This answer is, of course, incorrect. Insight into this problem can, however, be provided by means of a diagram showing that $150 does not equal 100% but 120%:

1/5	1/5	1/5	1/5	1/5	to be added:	1/5
20%	20%	20%	20%	20%		20%

5/5 is 100%; 6/5 then is 120%. So 5/6 x $150 = $125.

Schematization - or visualization - is an important mathematical strategy, a tool for solving problems. Other such tools are: estimating, simplifying problems, testing, changing perspectives and conducting a thought experiment. Gravemeijer (1988) has classified mathematization tools according to characteristics derived from mathematics itself. For example, structuring and generalizing are related to the category 'generality'. Proofs and predictions are connected to 'certainty' because this is for a mathematician very important, symbolization and formalization belong to 'precision' and reduction and constructing algorithms to 'conciseness'.

Resnick and Klopfer (1989) identify knowledge that plays an important role in problem-solving as 'organizing schemata', concepts which are 'powerful' and must be actively acquired. In other words, it is rich, flexible, 'generative knowledge'. This knowledge, or specialist knowledge, forms the basis upon which we can construct the first representation of a problem which we must solve. This representation is the starting point for a successful problem-solving process. Bransford, Sherwood, Vye, and Rieser (1989) warn against the danger of rote knowledge or, as they put it, 'inert knowledge'. Because children do not consider such formal knowledge 'real' or meaningful, they will not be able to apply it or only do so blindly, particularly in mathematics. Children acquire inert knowledge in mathematics when they are forced to learn formulas such as 'To divide by a fraction, multiply by its opposite' or to do plain problems involving meaningless numbers (leading to the type of problem discussed in the example above). In this connection, realistic mathematics instruction makes use of the models, schemata and concepts - called 'conceptual models' by Lesh (1985) - which form the core of the mathematical approach.

This approach is similar to Davydov's (1977) idea of teaching children to work with theoretical concepts - which can be seen as concepts essential to a specific discipline - instead of concepts based on observation or empiricism. In this connection, Davydov has argued for the formation of theoretical thinking, similar in certain respects to Freudenthal's (1984) development of mathematical thinking or attitude. In fact, the two are so similar that both have pointed out the same flaws in the empirical, inductive teaching approach. In this approach, knowledge comes into being through observation or empiricism. Here we can recognize the influence of the empiricist school, which, as we have seen, places very little emphasis on vertical mathematization - which is precisely what Freudenthal and Davydov do wish to emphasize. Mathematical or theoretical insights are required in order to be able to understand reality correctly (including learning tools such as the number line or abacus). Here, however, the similarity between Freudenthal and Davydov ends. According to Freudenthal, Davydov introduces theoretical concepts too early on in education; moreover, Freudenthal rejects the distinction made by Davydov between theoretical and empirical concepts. In Freudenthal's view, theory is always inherent in empiricism; all action is implicitly theory bound.

Research into solving mathematics problems often focuses on initial mathematics problems and word problems. See for example De Corte, Verschaffel, & Greer (1996), Span, De Corte, and Van Hout-Wolters (1989) and Verschaffel and De Corte (1990) for a report on studies carried out on learning and problem-solving in mathematics. Following in the footsteps of Riley, Greeno and Heller (1983), Verschaffel and De Corte (1990) explored which internal representations children formed as a result of problems which they were given to solve, and which role the semantic features of these problems played in this. They also wished to discover how these representations formed the basis for actions, in particular problem-solving procedures. Their research revealed that semantic factors played a role in forming problem representations. Many children did indeed tend to take their lead from the meaning of words (some of which were printed by chance) in the text ('together' or 'with each other') and to base their solution procedures on these words. It was also shown that many verbal, nonpropositional, grammatical characteristics influenced the choice of procedure and not only cognitive schemata, as authors such as Resnick (1983) and Riley et al. (1983) have suggested. Finding the correct solution, these researchers believe, depends on the formation of adequate representations constructed from part-whole schemata.

There are, however, several objections which might be raised (Van Luit, 1994) to this emphasis upon the part-whole schema. To begin, there is no relationship with a child's previous experiences, such as counting (Van Mulken, 1992). Second, this schema is based exclusively on the cardinal interpretation of numbers, whereas in truth numbers may appear in other forms: 9 is 6 + 3, but 9 is also 3 x 3 or the root of 81. Third, a child gets into trouble if he or she comes across a problem which does not contain a part-whole relationship; for example, John is 5 years old and his friend is 6.

A more general comment on research into solving word problems is that the influence exerted by semantic structural features has been given too much emphasis, at the expense of studying the influence which the nature and the size of numbers have on the choice of solution procedure (Van Mulken, 1992; Verschaffel & De Corte, 1990). This is known as 'number sensitivity'. An example is: 62 - 18 = ? If we take account of the nature of the numbers in this problem, the obvious strategy is to subtract 20 first and then add 2. The problem 62 - 33 = ?, however, requires a different approach, for example: 63 - 33 = 30, 30 - 1 = 29. For a lot of children this is not a easy problem, because it supposes much flexibility in thinking. In teaching children to solve word problems, one should emphasize neither mastery of procedures (such as the use of the part-whole schema) nor semantic features, but rather the structure of the problem, specifically the structure of the numbers. Proponents of realistic mathematics instruction argue for flexibility in choosing problem-solving methods or learning to choose them.

Some of the above mentioned researchers maintain that training in meta-cognitive skills - such as planning the course of a solution, setting up schemata, guessing the solution in advance - can simplify the approach taken to problems. However, in none of the studies was training conclusively shown to be successful, although progress was noted. Some researchers point out, no doubt correctly, that the weaker children lacked a certain knowledge background. It was remarkable that the children were able to master a heuristic (for example, making an estimate), but that this did not automatically lead to their choosing the required, formal operation.

In problem solving the context can be important. The function of contexts is to elicit knowledge which children have gained through experience and which they can use once again in forming internal problem representations. This is not, however, a hard and fast rule. Sometimes meanings which have been acquired through experience are directly contrary to the mathematical meaning of concepts (and, as we remarked earlier, in physics this is the rule rather than the exception (Keil, 1989). Walkerdine (1988) drew attention to this in her series of carefully conducted observations. An association with the child's experience, she claims, means an association with contexts, and these are highly domain-specific. We cannot simply assume, therefore, that 'transfer' is affected simply by the insertion of mathematical relations into a 'meaningful context'. The author illustrates this by giving the following example. Children first learn the concept 'more' in a pedagogical situation: 'Just two mouthfuls more, mmm!' In mathematics, however, 'more' is the opposite of 'less', and that is an entirely different concept. There are, then, two entirely different 'discursive practices'; in other words, it is not always the case that children solve problems better by using knowledge based on their own experience.

Discussion and conclusions

There have been radical changes in the approach to mathematics instruction in the past few years. These changes came about because mathematicians began to view their own discipline differently, leading to new research on teaching methodology. This research was supported by new developments in educational psychology. Glaser (1991) analyses these developments. He points out the tendency to relate learning and thinking to specific domains. This approach has been defended by various authors. For example, in a series of studies (Keil, 1989) it is shown that indeed it is plausible that concepts develop largely in specific domains. Research has also demonstrated the importance of reflection.

Experts - adults or children - develop skills in order to plan and monitor their actions and predict what the results of their efforts will be. In an experiment, Glancey (1986) introduced the knowledge representation (including heuristic rules) of experts to a group of pupils. On the basis of these representations, the pupils learned to formulate hypotheses, recognize errors and, in particular, to better organize their knowledge base. Glancey's most original idea was to have the pupils observe experts and question them concerning their methods. The observation strategies and the questions were analyzed in advance, and an analysis was also done of how knowledge might be restructured during the learning process. Here interaction and reflection go hand in hand. The idea of learning as a process of continual restructuring, of increasing architectonics and deeper insight, has also been expounded by White and Frederiksen (1986). They furthermore emphasize establishing a link with the pupils' naive, intuitive models, as in mathematics instruction.

Constructivism is another tendency which Glaser (1991) has frequently come across in analyzing research reports. Although the emphasis in the past was on monitoring the pupils' learning processes, nowadays the pupils has more control over his own learning environment, and many studies have attempted to gain insight into how pupils construct their learning environment in order to be able to learn something. A new image of the pupil is coming into being; they are no longer 'good boys and girls' who learn everything by rote, but children motivated to explore and seek explanations.

It was not only among Anglo-Saxon cognitive psychology that is undergoing significant changes; the Russian Cultural Historical School has long been concentrating on research themes related to the new developments in mathematics instruction. The main focus is on reflective thinking (Stepanov & Semenov, 1985; Zak, 1984), interaction (Davydov et al., 1982), and the development of domain-specific concepts (Davydov, 1977), while education and instruction are in essence seen as the active interrelation of symbolic systems and meanings which a culture has brought forth (Leont'ev, 1980; Van Oers, 1987). For a more thorough comparison between the concepts presented by the Russian Cultural Historical School and realistic teaching methodology, readers are referred to Nelissen and Tomic (1995, 1996).

So, in the past few years a number of interesting new themes relevant to mathematics instruction have received a great deal of attention. We have argued that learning is a process which rests upon children's own constructive activity. Learning takes place in a social context (Bruner, 1996; Slavin, 1986). Mathematical ideas are not merely abstract; they are contained within language and concepts. Learning is a process in which the child masters its cultural heritage, by learning particular sets of symbols. If children are able to put their ideas into words, they will have a better grasp of their own way of thinking. We have also stated that when children are able to put their ideas into words, their teacher will gain greater insight into their thought processes. If their constructions prove to be unusable, the teacher can confront the children with alternative approaches, for example through problem-oriented questioning. When children discuss their ideas with one another (Mercer, 1995), they not only have to state their opinions more concisely, they also have to listen, think along with others and try to understand what the other children actually mean to say. Learning - and cognitive development - is increasingly viewed as a process in which metacognition (or reflection) fulfills a regulatory function.

Solving mathematics problems requires learning domain-specific rather than general knowledge. This knowledge is well-structured and flexible, and encompasses a knowledge of both content and procedures and reflective knowledge. In mathematics, tools and modes of thinking that typify mathematics should be maintained, and used to solve problems. The overriding concern is to maintain a flexible choice of tools, and that choice requires domain-specific reflection. To be able to reflect, however, knowledge of content is once again necessary; one can only reflect on the use of tools, strategies and concepts if one knows them. By now many of these new ideas have filtered through to the practice of teaching mathematics, in part because new programmes are being developed, implemented and supervised which are based on the realistic approach, and in part because teacher training now focuses on the new realistic didactic. It is, ultimately, the teachers themselves who must put these new ideas into practice.

Although there is a high degree of consensus among researchers in mathematics instruction, particularly on initial mathematics problems and word problems, up to the present theory has preceded carefully collected empirical data. If we also agree, on the basis of research findings, that construction, interaction and reflection are essential for learning mathematics, then the practice of mathematics instruction should be altered radically. Teachers are being asked to master a new approach to instruction and to their pupils. Among other things, this means a new approach to testing, to explaining, to cooperating and discussing, to working independently, to thinking intuitively, to understanding and developing concepts.

References

Bishop, A.J. (1988). *Mathematical enculturation. A cultural perspective on mathematics education.* Dordrecht: Kluwer.

Bonner, J. (1990). De implicaties van de cognitieve theorie voor het ontwerpen van instructie: Een heroverweging [Implications of cognitive theory for instructional design: A reconsideration]. In J.S. ten Brinke (Ed.), *Paspoort onderwijspraktijk* (Vol. 3, pp. 453-463). Houten: Bohn Stafleu Van Loghum.

Bransford, J., Sherwood, R., Vye, N., & Rieser, J. (1989). Leren denken en probleem oplossen [Learning to think and problem solving]. In G. Ackers (Ed.), *Paspoort Onderwijspraktijk* (pp. 155-190). Houten: Bohn Stafleu Van Loghum.

Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

Bruner, J. (1996). *The culture of education*. Cambridge: Harvard University Press.

Caillot, M. (1991). Learning thinking through the new technologies. In S. Maclure & P. Davies (Eds.), *Learning to think: Thinking to learn*. Oxford: Pergamon.

Cazden, C. (1988). *Classroom discourse: The language of teaching and learning*. New York: Heineman.

Clement, J. (1982). Algebra word problems solutions: Thought processes underlying a common misconception. *Journal for Research in Mathematics Education, 16,* 16-30.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher, 23,* 7, 13-20.

Cobb, P., Yackel, E., & Wood, T. (1992). A constructivist alternative to the representational view of mind in mathematics education. *Journal for Research in Mathematics Education, 14,* 83-95.

Cobb, P., Gravemeijer K., Yackel E., McClain K., & Whitenack J. (1997). Mathematizing and symbolizing: The emergence of chains of signification in one first-grade classroom. In D. Kirshner & J.A. Whitson (Eds.), *Situated cognition* (pp. 151-233). Hillsdale, NJ: Lawrence Erlbaum Associates.

Confrey, J. (1981). *Conceptual change analysis: Implications for mathematics and curriculum*. Toronto: Wiley.

Confrey, J. (1985). Towards a framework for constructivists instruction. In L. Streefland (Ed.), *Proceedings of the Ninth Conference for the Psychology of Mathematics Education* (Vol. 1, pp. 477-483). Noordwijkerhout: PME.

Davydov, V.V. (1977). De introductie van het begrip grootheid in de eerste klas van de basisschool [Introduction of the quantity concept in first grade elementary school]. In C.F. van Parreren & J.M.C. Nelissen (Eds.), *Rekenen. Teksten en analyses Sovjetpsychologie* (Vol. 2, pp. 1-61). Groningen: Wolters-Noordhoff.

Davydov, V.V., Lompscher, J., & Markova, A.K. (1982). *Ausbildung der Lerntätigkeit bei Schülern* [Development of learning in pupils]. Berlin: Volk und Wissen.

De Corte, E., Greer B., & Verschaffel L. (1996). Mathematics teaching and Learning. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 491-549). New York: MacMillan.

Derry, S.J., & Murphy, D.A. (1986). Designing systems that train learning ability: From theory to practice. *Review of Educational Research, 56,* 1, 1-40.

Doise, W., & Mugny, G. (1984). *The social development of the intellect*. Oxford: Pergamon.

Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23,* 7, 5-13.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht: Reidel.

Freudenthal, H. (1984). *Appels en peren / wiskunde en psychologie: Gebundelde opstellen* [Apples and pears / mathematics and psychology: Collected essays]. Apeldoorn: Van Walraven.

Gardner, H. (1987). *The mind's new science. A history of the cognitive revolution.* New York: Basic.

Glancey, W.J. (1986). From Guidon to Neomycin and Hercules in twenty short lessons: ONR final report 1979-1985. *AI Magazine, 7*, 40-60.

Glaser, R. (1991). Wedergeboorte van de leertheorie in onderwijskundig onderzoek [Rebirth of learning theory in educational research]. In J.S. ten Brinke (Ed.), *Paspoort onderwijspraktijk* (pp. 1.1-1.32). Houten: Bohn Stafleu Van Loghum.

Goffree, F., Freudenthal, H., & Schoemaker, G. (1981). Wiskundeonderwijs. Een micro-didactische beschouwing over het thema 'begrijpen' [Mathematics education. A micro pedagogical consideration on 'understanding']. In H.P.M. Cremers (Ed.), *Losbladig onderwijskundig lexicon* (pp. 1200.3-1200.77). Alphen a/d Rijn: Samsom H.D. Tjeenk Willink.

Granott, N., & Gardner, H. (1994). When minds meet: Interaction, coincidence and development in domains of activity. In R.J. Sternberg & R.K. Wagner (Eds.), *Mind in contexts. Interactionist perspectives on human intelligence* (pp. 133-152). Cambridge, MA: Cambrigde University Press.

Gravemeijer, K.P.E. (1988). *De grondslagen van het programma Rekenen en Wiskunde* [Foundations of the programme Arithmetic and Mathematics]. Rotterdam: OSM.

Gravemeijer, K. (1997). Commentary solving word problems: A case of modelling? *Learning and Instruction, 7*, 389-307.

Greer, B. (1997) Modelling reality in mathematics classrooms: The case of word problems. *Learning and Instruction*, *7*, 293-307.

Hiebert, J. (1992). Reflection and communication: Cognitive considerations in school mathematics reform. *International Journal of Educational Research, 17*, 439-456.

Kaput, J.J. (1987). Towards a theory of symbol use in mathematics. In: C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 159-197). Hillsdale, NJ: Lawrence Erlbaum.

Keil, F.C. (1989). *Concepts, kinds and cognitive development.* Cambridge, MA: MIT.

Kirshner, D., & Whitson J.A. (Eds.) (1997). *Situated cognition.* Hillsdale, NJ: Lawrence Erlbaum.

Lakatos, I. (1976). *Proofs and refutations: The logic and mathematical discovery.* London: Cambridge University Press.

Leont'ev, A.N. (1980). Activiteit als psychologisch probleem [Activity as a psychological problem]. *Pedagogische Studiën, 57*, 324-343.

Lesh, R. (1985). Conceptual analysis of mathematical ideas and problem solving processes. In L. Streefland (Ed.), *Proceedings of the Ninth Conference for the Psychology of Mathematics Education* (Vol. 2, pp. 73-97). Noordwijkerhout: PME

Lesh, R., Behr M., & Post T. (1987). Rational number relations and proportions. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics.* Hillsdale, NJ: Lawrence Erlbaum.

Lo, J.J., Grayson, H., Wheatly, J., & Smith, A. (1990). Potential mathematics learning opportunities in grade three class discussion. In G. Booker, P. Cobb, & T. De Mendicuti (Eds.), *Proceedings of the Fourteenth Conference for the Psychology of Mathematics Education* (pp. 113-121). Mexico: PME.

Mercer, N. (1995). *The guided construction of knowledge.* Clevedon, Phil: Multilingual Matters.

Nelissen, J.M.C. (1987). *Kinderen leren wiskunde. Een studie over constructie en reflectie in het basisonderwijs* [Children learn mathematics. A study concerning construction and reflection in elementary school education]. Gorinchem: De Ruiter.

Nelissen, J.M.C., & Tomic, W. (1995). Reflection in education with an example from mathematics. *Educational Practice and Theory, 17,* 53-70.

Nelissen, J.M.C., & Tomic, W. (1996). Reflections in Russian educational psychology. *Educational Foundations, 10,* 35-57.

Papert, S. (1980). *Mind storms. Children, computers and powerful ideas.* New York: Basic.

Piaget, J. (1976). *Genetische epistemologie* [Genetic epistemology]. Meppel: Boom.

Perkins, D.N., & Salomon, G. (1989). *Are cognitive skills context bound? Educational Researcher, 18,* 1, 16-25.

Pimm, D. (1990). Certain metonymic aspects of mathematical discourse. In G. Booker, P. Cobb, & T. de Mendicuti (Eds.), *Proceedings of the Fourteenth Conference for the Psychology of Mathematics Education* (pp. 129-136). Mexico: PME.

Presseisen, B.Z., Sternberg, R.J., Fischer, K.W., Knight, C.C., & Feuerstein. R. (1990). *Learning and thinking styles: Classroom interaction.* Washington: NEA.

Resnick, L.B. (1983). A developmental theory of number understanding. In H.P. Ginsburg (Ed.), *The development of mathematical thinking* (pp.109-151). New York: Academic Press.

Resnick, L.B., & Klopfer, L.E. (1989). *Toward the thinking curriculum: Current cognitive research.* Alexandria, VA: ASCD.

Reusser, K., & Stebler R. (1997). Every word problem has a solution. The social rationality of mathematical modeling in schools. *Learning and Instruction, 7,* 309-327.

Riley, M.S., Greeno, J.G., & Heller, J.I. (1983). Development of children's problem-solving ability in arithmatic. In H.P. Ginsburg (Ed.), *The development of mathematical thinking* (pp.153-196). New York: Academic Press.

Roazzi, A., & Bryant, P. (1994). Social class, context and cognitive development. In P. Light & G. Butterworth (Eds.), *Context and cognition*. New York: Harvester Wheatsheaf.

Salomon, G. (1989). Technology for the mind: Partner and nourisher. In P. Span, E. De Corte, & B. van Hout-Wolters (Eds.), *Onderwijsleerprocessen: Strategieën voor de verwerking van informatie* (pp 67-84). Amsterdam: Swets & Zeitlinger.

Saxe, G.B. (1988). Candy selling and math learning. *Educational Researcher, 17*, 1, 14-21.

Schoenfeld, A.H. (1989). Teaching mathematical thinking and problem solving. In L.B. Resnick & L.E. Klopfer (Eds.), *Toward the thinking curriculum: Current cognitive research* (pp. 83-104). Alexandria, VA: ASCD.

Slavin, R.E. (1986). *Developmental and motivational perspectives on cooperative learning*. San Francisco, CA: AERA.

Slavin, R.E. (1996) *Education for all*. Amsterdam: Swets & Zeitlinger.

Span, P., De Corte, E., & Van Hout-Wolters, B. (1989). *Onderwijsleerprocessen. Strategieën voor de verwerking van informatie* [Teaching learning processes. Strategies for information processing]. Amsterdam: Swets & Zeitlinger.

Steffe, L., Cobb, P., & Von Glazersfield, E. (1988). *Construction of arithmetical meanings and strategies*. New York: Springer Verlag.

Stepanov, S.J., & Semenov, I.N. (1985). Problemy psichologičeskogo izučenija refleksii I trorčestva [Psychology of reflection: problems and investigations]. *Voprosy Psichologii, 3*, 31-40.

Streefland, L. (1991). *Fractions in realistic mathematics education*. Dordrecht: Kluwer.

Treffers, A. (1987). *Three dimensions. A model of goal and theory description in primary mathematics education Wiskobas*. Dordrecht: Reidel.

Treffers, A. (1991). Didactical background of a mathematics programme for primary education. In L. Streefland (Ed.), *Realistic mathematics education in the primary school* (pp. 21-57). Utrecht: Freudenthal Institute.

Treffers, A., & Goffree, F. (1985). Rational analysis of realistic mathematics education. In L. Streefland (Ed.), *Proceedings of the Ninth Conference for the Psychology of Mathematics Education* (Vol. 2, pp. 97-123). Noordwijkerhout: PME.

Van den Heuvel-Panhuizen, M. (1996). *Assessment and realistic mathematics education*. Utrecht: Freudenthal Institute.

Van der Valk, T. (1989). Het energie-begrip van leefwereld naar natuurkunde [The concept of energy from subjective view to physics]. *Didaktief, 19*, 4, 22-25.

Van Luit, J.E.H. (Ed.) (1994). *Research on learning and instruction of mathematics in kindergarten and primary school*. Doetinchem: Graviant.

Van Mulken, F. (1992). *Hoofdrekenen en strategisch handelen* [Mental arithmetic and strategic action]. Leiden: University Press.

Van Oers, B. (1987). *Activiteit en begrip. Proeve van een handelings-psychologische didactiek* [Activity and understanding. Exemplar of an action psychological pedagogy]. Amsterdam: Vrije Universiteit.

Van Parreren, C.F. (1988). *Ontwikkelend Onderwijs* [Developing Education]. Leuven: Acco.

Verschaffel, L., & De Corte, E. (1990). De techniek van de computersimulatie in het onderzoek naar het leren oplossen van vraagstukken over optellen en aftrekken [The method of computer simulation in research into learning to solve addition and subtraction problems]. *Pedagogisch Tijdschrift, 15*, 259-269.

Verschaffel, L., De Corte E., & Borghart I. (1997). Pre-service teacher's conceptions and beliefs about the role of real world knowledge in mathematical modelling of school word problems. *Teaching and Instruction, 7*, 339-359.

Von Glazersfeld, E. (1991). *Radical constructivism in mathematics education.* Dordrecht: Kluwer.

Vygotsky, L.S. (1977). *Denken und Sprechen* [Thought and language]. Frankfurt am Main: Fischer.

Walkerdine, V. (1988). *The mastery of reason.* London: Routledge.

Walkerdine, V. (1997). Redefining the subject in situated cognition theory. In Kirshner, D. & J.A. Whitson (Eds.), *Situated cognition.* Hillsdale, NJ: Lawrence Erlbaum.

White, B.Y., & Frederiksen J.R. (1986). *Progressions of quantitative models as a foundation for intelligent learning environments.* Cambridge, MA: Bolt, Beranek & Newman.

Whitney, H. (1985). Taking responsibility in school mathematics education. In L. Streefland (Ed.), *Proceedings of the Ninth Conference for the Psychology of Mathematics Education* (pp. 123-142). Noordwijkerhout: PME.

Zak, A.Z. (1984). *Razvitije teoretičeskogo myšlenija mladšich školnikov* [Development in theoretical thinking in young children]. Moskou: Pedagogika.

9

Realistic mathematical modelling and problem solving in the upper elementary school: Analysis and improvement

L. Verschaffel

Introduction

There is nowadays a clear consensus that the acquisition of mathematical problem-solving and reasoning skills and the ability to apply these skills in real-life situations, constitutes a major objective of mathematics education at the elementary school level. In the United States, for instance, the National Council of Teachers of Mathematics (NCTM) stated in its 'Agenda for Action' that "problem solving must be the focus of school mathematics" (NCTM, 1989, p. 1). According to this and other reform documents pupils should be able to apply these problem-solving skills not only on non-routine problems within the domain of mathematics itself (= pure mathematical problems), but also - and at the elementary school level even primarily - in various kinds of context-related problem situations encountered in everyday life or in other curricular domains at school (= mathematical applications) (Burkhardt, 1994; De Corte, Greer, & Verschaffel, 1996).

Unfortunately, despite all of this emphasis on applied problem solving within the international community of mathematics educators, the ability to solve application problems remains one of the most difficult aspects of mathematical performance for pupils to develop. Indeed, there is a large number of studies showing that most pupils, even those who are considered as mathematically strong, have great difficulty with various aspects of the solution of non-routine application problems (De Corte et al., 1996; Lester, Garofalo, & Kroll, 1989; Schoenfeld, 1992; Van Haneghan, Barron, Young, Williams, Vye, & Bransford, 1992).

This chapter addresses the issue of solving mathematical application problems - traditionally presented in the format of word problems - in upper elementary school children. We begin with a brief overview of the different phases and aspects constituting expertise in solving mathematical applications. Afterwards we document some typical shortcomings in elementary school children's skills to solve such problems, and we relate them to characteristics of the current practice and culture of elementary school mathematics. Then three evaluation studies about the effects of instructional programmes for teaching and learning (realistic) mathematical modelling and problem solving in upper elementary school children, are exemplary described and discussed. Finally, some suggestions for future research and development are given.

A model of solving mathematical application problems

By the end of the elementary school, pupils must have acquired the inclination and the ability to apply their acquired formal mathematical knowledge and skills effectively in a wide range of problem situations (see e.g., NCTM, 1989). Applying mathematics to solve a situational problem can be usefully thought of as a complex process involving several phases, with different emphases on the kind of knowledge and aptitudes needed during a particular phase. Although the fine details vary, most models of competent mathematical problem solving involve the following phases: (a) formulating and understanding the problem situation; (b) constructing a mathematical model which describes the essence of those elements and relations involved in the situation that are of interest; (c) rearranging the mathematical model and/or operating on it to identify the unknown element(s) in the model; (d) interpreting and evaluating the outcome of the computational work in terms of the practical situation that lied at the basis of the mathematical model, and communicating the result(s). This multi-phased process of mathematical modelling and problem solving has to be considered as cyclic, rather than as a linear progression from givens to goals (Burkhardt, 1994; De Corte et al., 1996; Schoenfeld, 1992).

There is also a fairly broad consensus that the skilled and successful execution of such a complex and multi-phased problem-solving process requires the integrated and effective application of several categories of aptitudes (De Corte et al., 1996; Lester et al., 1989; Schoenfeld, 1992):

- A well-organized and flexibly accessible knowledge base. This knowledge base involves first of all a variety of mathematical symbols, concepts, principles, rules, procedures, algorithms, etc. With respect to the solution of situational application problems, the required knowledge base involves also real-world

knowledge about the context or situation involved in the problem, which is indispensable for understanding the problem and for selecting those aspects of the problem situation that need to be included in the mathematical model, as well as for interpreting and evaluating the outcome(s) of the operations within this model in terms of the original problem situation.
- Heuristic methods, i.e. systematic search strategies for problem analysis and transformation, such as making a sketch or a schema of the problem situation, decomposing the problem into parts, working backwards, etc. The application of such heuristics by the problem solver does not guarantee, but can significantly enhance the finding of a proper solution, because they induce an active and systematic approach to a problem for which the subject has no ready-made solution.
- Metacognition, which involves two closely related aspects: awareness about one's own cognitive functioning (such as being aware of the limits of one's short-term memory) and activities related to the regulation of one's own cognitive processes (such as planning a solution process, monitoring an ongoing solution process, checking and, if necessary, correcting an answer, and reflecting on one's learning and problem-solving activities).
- Beliefs and affects, such as beliefs, attitudes, and self-confidence with respect to mathematics and mathematics learning and teaching in general and to specific parts of it (like mathematical word problems) in particular.

While these different components (can) intervene and interact in complex ways in all phases of the problem-solving process, certain kinds of aptitudes are more important than others in a particular phase. For instance, heuristics and metacognition are especially important during the initial stages of the problem-solving process, when subjects try to build an appropriate representation of the problem and to set up plans for transforming it into a suitable mathematical model, as well as in the final stages of the solution process, in which they have to interpret and check the outcome of the calculations. Mathematical knowledge and skills, such as knowledge of basic number facts, procedures for mental and written arithmetic, formulas and techniques for solving equations, etc., on the other hand, are crucial in the middle phase of the problem-solving process, wherein subjects have to operate upon the mathematical model.

Pupils' difficulties with context-related mathematical applications problems

The consensus in current reform documents that problem solving is a major goal of mathematics education, contrasts sharply with the research findings showing that many learners in school do not, or at least insufficiently, master the different aptitudes required to approach mathematical application problems in an efficient and successful way. We will report some evidence to support this statement, mainly obtained in ascertaining studies about the problem-solving behaviour and skills of upper elementary school children with respect to arithmetic word problems. (For more detailed descriptions, see De Corte et al., 1996; Lester et al. 1989; Schoenfeld, 1992; Van Haneghan et al., 1992).

Insufficiencies in the domain-specific knowledge base

It is clear from research that many problem-solving deficiencies in (upper) elementary school children can be attributed to lack of domain-specific knowledge and skills. These deficiencies in pupils' domain-specific knowledge base relate to a wide variety of content-related resources that they can or must bring to bear on the problem to be solved (such as mathematical concepts, arithmetic facts, algorithms, problem schemes, knowledge about the problem context, etc.). Furthermore, errors due to domain-specific insufficiencies can not merely be attributed to the absence of necessary pieces of knowledge or skill, but also to unstable or even erroneous knowledge systems. As an illustration of how misconceptions can negatively interfere with pupils' problem-solving ability, we refer to the so-called 'multiplication makes bigger/division makes smaller' misconception, which has been observed in upper elementary school children (but also in older pupils and adults) in a diversity of countries (De Corte, Verschaffel, & Van Coillie, 1988). In several types of multiplication and division word problems with decimal numbers smaller than 1, such as 'Coffee costs 280 Bfr. per kg. How much does 0.75 kg. of coffee cost?', this misconception typically leads to inversed-operation errors (i.e., dividing the two given numbers instead of multiplying them). Like several other kinds of mathematical misconceptions documented in the literature, this misconception can be conceived as a typical result of the overgeneralisation of rules that are valid for learning tasks and problem situations involving whole numbers, but not for tasks and problems going beyond whole numbers (Greer, 1992; Verschaffel & De Corte, 1997b).

Lack of valuable heuristic and metacognitive skills

Several studies have shown that many pupils and pupils attack application problems in a school context in a very superficial and mindless way. Typically, their problem-solving activity is restricted to the execution of one or more computations with the numbers given in the problem statement, with little or no attention to the other aspects of the competent problem-solving model described before.

Striking examples of such a superficial 'number-crunching' approach for application problems can first of all be found in the research literature about the solution of one-step word problems involving the four basic arithmetic operations (Greer, 1992; Sowder, 1988; Verschaffel & De Corte, 1997a). It is a well-documented finding that many pupils attack these problems by means of superficial coping strategies like:

- Simply guessing what operation to perform with the given numbers.
- Selecting the numbers contained in the problem and performing either the operation that was most recently taught or drilled in the classroom, or the

operation at which (s)he feels most competent (e.g., always adding).
- Looking for key words in the problem statement which tell which operation to perform (e.g., if the problem contains the word 'altogether' then add).
- Looking at the numbers; they will 'tell' what operation to use (e.g., 'If it's like 78 and 54, then I'd probably either add or multiply. But if they are 78 and 3, it looks like a division because of the size of the numbers').

The absence of valuable heuristic and metacognitive skills in pupils' problem-solving behaviour is also evidenced in a number of other investigations in which pupils have been confronted with less familiar and more complex problem situations which explicitly appeal to higher-order thinking.

Some of these studies have focused on upper elementary school pupils' spontaneous use of particular valuable heuristics. A major outcome of these investigations is that pupils often do not spontaneously apply heuristics in problem solving. For instance, De Corte and Somers (1982) found that traditionally schooled sixth-grade pupils only rarely made use of estimating as a heuristic strategy for solving word problems involving complex calculations (e.g., calculations with large numbers, with decimals, etc.). Similarly, Van Essen (1991) administered a series of unfamiliar arithmetic word problems to fifth-graders, and found that answers were checked in only 5% of the cases. And if verification occurred, it was mostly restricted to the correctness of the arithmetic calculation(s) performed. More recently, De Bock, Verschaffel, and Janssens (1998) administered a set of geometry problems about length and area of similar plane figures to a large group of 12-13-year olds. Half of these problems required linear proportional reasoning, but in the other half proportional reasoning led to an erroneous answer, as in the following example: 'A farmer needs 8 hours to manure his square piece of land with a side of 200 m. How many hours will he roughly need to manure a similar square piece of land with a side of 600 m?'. While the proportional items were almost always solved correctly, the non-proportional items elicited less than 5% correct responses. Almost all errors on the latter kind of problems were based on the improper use of linear proportional reasoning (e.g., '3 x 200 = 600; therefore 8 x 3 = 24' for the above-mentioned example problem). Although making a sketch of the problem situation would have been extremely helpful to detect the inappropriateness of stereotyped solutions based on linear proportional reasoning and to detect the nature of the non-linear relationship in the problem, pupils made little or no use of this valuable heuristic.

With respect to self-regulation, Schoenfeld (1992) collected a large set of videotapes of high school and college pupils solving complex and unfamiliar mathematical problems in pairs. The problem-solving protocols were analysed by parsing them into so-called 'episodes', defined as periods of time during which the pair of pupils is involved in essentially the same activity. Different kinds of activities - more or less corresponding to the different activities mentioned in the multi-phase model of problem solving presented in the first section of this chapter - were distinguished. Schoenfeld observed that in about 60% of all solution attempts, self-regulatory activities such as analysing the problem and monitoring the solution process - which are so characteristic for the expert approach to problem solving - were completely absent. The typical approach used by the pairs can be summarized as follows: reading the problem, quickly deciding about the approach to follow, and then proceeding it without considering any alternative, even if no progress is made at all.

Lack of sense-making during the solution of mathematical word problems
Another stream of research has focused on the absence of real-world knowledge and sense-making during pupils' modelling and interpreting of school arithmetic word problems. In the late seventies, some French and German researchers tested elementary school children's disposition toward (non-)realistic modelling using absurd problems like 'There are 26 sheep and 10 goats on a ship. How old is the captain?', and found that large numbers (up to 60 %) of them worked out these unsolvable problems by combining the numbers given in the problems to produce answers without being aware of the meaninglessness of the problems and of their solutions (Baruk, 1989; Radatz, 1983). In Radatz' (1983) study, for example, the percentage of children who worked out the absurd problems in such a way grew from the early to the middle years of the elementary school, before going down (but only slightly!) in the upper grades.

In the US only 24 % of a national sample of 13-year-olds was able to solve correctly the following problem which appeared in the Third National Assessment of Educational progress (Carpenter, Lindquist, Matthews, & Silver, 1983): 'An army bus holds 36 soldiers. If 1,128 soldiers are being bussed to their training site, how many buses are needed?' Most pupils from the sample calculated '1,128 divided by 36' correctly as '31 1/3' or '31 remainder 12', but only 1/3 of these pupils gave the answer appropriate to the situation being modelled, namely 32. The others gave '31' as the answer or non-whole numbers like '31 1/3' or '31 remainder 12'.

More recently, Greer (1993) and Verschaffel, De Corte, and Lasure (1994) confronted large groups of pupils (10-13-years olds) with a set of word problems, half of which were standard items (S-items) that could be unambiguously solved by applying an obvious arithmetic operation(s) with the given numbers, while the other half were problematic items (P-items) for which the appropriate mathematical model was less obvious and less indisputable, at least if one seriously takes into account the realities of the context evoked by the problem statement. Examples of P-items used in these studies are 'John's best time to run the 100 metres is 17 seconds. How long will it take him to run 1 km.' and 'Steve has bought 4 planks of 2.5 metres each. How many planks of 1 metre can he saw out of these planks?'. The analysis of the pupils' reactions to these P-items yielded an alarmingly small number of realistic responses or comments based on realistic considerations (e.g., responding the above-mentioned runner-item with 'This problem is unsolvable, because John will not be able to run constantly at his record speed' instead of the stereotyped reaction '17 x 10 = 170 seconds', or responding the planks-item with 'eight planks' instead of the stereotyped response 'ten planks', because in reality one can only saw two planks of 1 metre out of a plank of 2.5 metre. For instance, in the study of Verschaffel et al. (1994), only 17 % of all answers given by a group of 75 fifth-graders to the ten P-items of a collectively administered paper-and-pencil test, could be considered as 'realistic'.

Inadequate conceptions and beliefs about mathematical problem solving

Schoenfeld (1985) refers to beliefs or belief systems as "the individual's mathematical world view, that is, the perspective from which one approaches mathematics and mathematical tasks" (p. 45). This mathematical world view includes beliefs about mathematics, mathematical tasks, oneself as a doer or learner of mathematics, and the environment or context within which one is doing or learning mathematics.

Research has revealed that many weak pupils hold inadequate domain-related beliefs about mathematics and mathematics learning, teaching and problem solving such as the following:

- Mathematics problems have one and only one right answer.
- There is only one correct way to solve any mathematical problem; usually the rule the teacher had most recently demonstrated in the class.
- Ordinary pupils cannot solve mathematical problems by themselves.
- Solving a mathematical problem should not take more than five minutes.
- Being able to solve a mathematical problem is a mere question of luck.
- The mathematics learned in school has little or nothing to do with the real world.

There is evidence that domain-related beliefs exert a powerful influence on pupils' willingness to engage in mathematical problems, on the kind of activities they perform when confronted with a mathematical problem, on the kind of knowledge they are inclined to activate during their problem-solving attempts, and on the way they evaluate their success or failure to solve problems (De Corte et al., 1996; Lester et al., 1989; Schoenfeld, 1992).

Interaction between the different types of difficulties

So far, we have mentioned several types of difficulties which may be responsible for pupils' failure in solving mathematical application problems. It is obvious that these different categories of pupil difficulties - and their subcategories - may interact in varied and complex ways (see also De Corte et al., 1998). For example, pupils' failure to apply heuristics in a spontaneous and/or efficient way - as observed in the studies of De Bock et al. (1996), De Corte and Somers (1982) and Van Essen (1991) - is not necessarily caused by the fact that these pupils did not know the possibly relevant heuristics; another reason may be that they did not have access to the conceptual knowledge base required to implement these heuristics properly (Van Essen, 1991) or that they were not inclined to apply them because they were not aware of the problematic nature of the task (De Bock et al., 1998). This example provides evidence of a negative kind about the importance of the integrated acquisition of the different kinds of aptitudes discussed in the first section.

Instructional determinants of children's insufficient problem-solving aptitudes

One could ask: which factors are responsible for these insufficiencies in pupils' abilities to solve context-based mathematical application problems? According to many scholars, they are induced and shaped by several characteristics of the current practice and culture of mathematics education, and especially by the way in which teaching and learning wor(l)d problem solving are typically organized in the mathematics lessons (Burkhardt, 1994; Greer, 1993; Lester et al., 1989; Verschaffel & De Corte, 1997a).

Nature of the problems

A first important feature is the impoverished and stereotyped diet of problems currently used in the lessons in word problem solving. As explained and illustrated in the previous section, pupils frequently bypass the representation phase of the problem-solving process and move directly to a mathematical expression on the basis of syntactical, surface clues. Such an approach is directly elicited and reinforced by the stereotyped nature of word problems presented to the pupils, since it produces a high degree of superficial success. Indeed, whoever observes pupils and pupils in classroom and homework situations can find again and again how few textbook problems force pupils to do an in-depth semantic analysis. For instance, Schoenfeld (1992) reported that in a widely used elementary textbook series in the US, more than 90 % of the application problems could be correctly solved by the superficial key-word strategy. Similarly, Säljö and Wyndhamn (1987) noted that Swedish textbooks often contain headings that clearly and systematically spell out the nature of the tasks to be performed, so that pupils frequently know what arithmetic operation to select before they even have begun to read the problems.

The same holds for pupils' strong tendency toward suspending real-world knowledge and realistic sense-making from their problem-solving endeavours. In the current practice of school mathematics, pupils are repeatedly confronted with standard problems, in which the relationship between the problem context and the required calculation(s) is straightforward; non-standard problems inviting or even forcing them to take seriously the varied and problematic nature of mathematical modelling and to engage in rich and complex model-eliciting activities, are very rare. Or, as Freudenthal stated it (1991, p. 70), "... in the textbook context each problem has one and only one solution. There is no access for reality, with its multiply solvable and its unsolvable problems". Only when pupils are confronted with a significant number of problems in which there is access for reality, and - therefore - for multiple representations and solutions, will they begin to view the use of one or more arithmetic operations as only one out of a number of candidate models (Greer, 1993; Verschaffel et al., 1994). Or, as Nunes, Schliemann, and Carraher (1993, p. 148) put it: "If mathematics education is going to be realistic, problems will have to be sought that respect assumptions about life outside school".

Partly as a result of these criticisms against the nature of the problems used in traditional textbooks, textbook writers have started constructing problems with more problematic and more realistic characteristics. But in most textbooks there still is a preponderance of problems of trivial modelling assumptions which allow pupils to build wrong beliefs about mathematical problem solving and superficial strategies for handling them.

Instructional techniques

A second feature of the current practice of teaching and learning mathematical applications is the lack of an explicit and systematic attention to the development of valuable modelling and problem-solving skills and self-regulatory strategies, which have shown to be crucial aspects of expert problem solving. Classroom observations have revealed that lessons in mathematical word problem solving are traditionally of the type called 'exposition, examples, exercises' (Burkhardt, 1994; Schoenfeld, 1992), wherein pupils mainly work individually on a series of problems presented in textbooks or worksheets and asking for the application of a particular piece of mathematics (a concept, an algorithm, a formula, ...), after some explanation and demonstration by the teacher. It is clear that such lessons in word problem solving are not very inductive to the development of higher-order knowledge and skills in pupils. As recent research in the domain of teaching thinking and problem solving has convincingly shown, problem-solving expertise can only be acquired through engaging pupils in powerful teaching/learning environments focusing on the strategic aspects of problem solving. Among other things, such powerful environments are characterized by the extensive and systematic use of:

- Modelling to demonstrate how an expert selects and applies heuristic and metacognitive strategies.
- Several kinds of 'scaffolds' which help the pupils with the execution of some aspects or parts of the problem-solving process which they cannot carry out autonomously.
- Co-operative learning in small groups (Collins, Brown, & Newman, 1989; De Corte et al., 1996; Schoenfeld, 1992).

Classroom culture

A third aspect of school arithmetic that seems responsible for the unintended and undesirable pupil learning behaviour and learning outcomes described in the previous section, relates to the culture of the mathematics classroom (De Corte et al., 1996; Gravemeijer, 1994; Lampert, 1990; Schoenfeld, 1992). Recently, researchers have started to document how several subtle, 'invisible' aspects of the daily mathematics classroom rituals and practices can contribute to unwanted mathematical behaviour and learning outcomes, such as the development of superficial strategies for coping with word problems or the construction of erroneous conceptions and beliefs about what mathematics and mathematics learning and teaching all is about. According to these researchers, pupils' learning of mathematics occurs essentially through being immersed in a particular classroom culture, and what they learn as a result of their daily involvement with the rituals and the cultural practices does not always correspond to the explicitly intended learning goals. As a typical illustration of this approach, we recall Lamperts' (1990, p. 31) commentary explaining why so many pupils develop strange conceptions about what mathematics and especially mathematical problem solving all is about: "Commonly mathematics is associated with certainty, knowing it, with being able to get the right answer, quickly. These cultural assumptions are shaped by school experience, in which doing mathematics means following the rules laid down by the teacher; knowing mathematics means remembering and applying the correct rule when the teacher asks a question, and mathematical truth is determined when the answer is ratified by the teacher. Beliefs about how to do mathematics and what it means to know it are acquired through years of watching, listening, and practising". The reversed side of these analyses revealing how the classroom culture may negatively affect pupils' problem-solving behaviour, is that the culture of mathematics classroom can also be positively used as a vehicle for realizing (more) authentic learning experiences and learning outcomes with respect to beliefs about and strategies for applied mathematical problem solving, as will be demonstrated in the next section.

Teaching and learning realistic mathematical modelling and problem solving

In recent years, several scholars have begun to design and evaluate instructional environments aimed at the development in children of genuine mathematical modelling and problem-solving abilities, which embody the major ideas emerging from the recent theoretical analyses and empirical studies reviewed in the previous sections. Hereafter, three such intervention studies will be presented and discussed, namely (a) a study by Lester et al. (1989) at the middle school level designed to enhance the use of heuristics and metacognition in seventh-graders' mathematical problem solving, (b) a study by Verschaffel and De Corte (1997b) aimed at breaking upper elementary school children's tendency toward non-realistic mathematical modelling of application problems in school arithmetic, and (c) a study by Van Haneghan et al. (1992) about the possibilities of new information technologies like videodiscs to enhance complex applied mathematical problem solving in upper elementary school children. The description of these three intervention studies could be extended with many other examples of attempts at the research-based design of powerful learning environments for mathematical modelling and problem solving for upper elementary school children, such as the work of Lampert (1990), Siemon (1992), and Silver, Smith, Lane, Salmon-Cox, and Stein (1990) (see also the chapters by Nelissen en Van Luit, this book).

Teaching heuristics and metacognitive skills in seventh-grade pupils

During the eighties, Lester et al. (1989) pursued a major intervention study designed to study the effects on pupils' problem solving behaviour of an instructional environment that involved (a) practicing pupils in the use of a number of valuable heuristic strategies (= strategy training), (b) training pupils to be more aware of the strategies and procedures they use to solve problems (= awareness training), and (c) training pupils to monitor and evaluate their actions during problem solving (= self-regulation training). In other words, this instructional environment focused strongly on the heuristic and metacognitive aspects of expert problem solving.

The instructional environment designed by Lester et al. (1989) consisted of (a) a set of appropriate problems and tasks and (b) a series of lesson plans with teacher roles and activities. The selection of these instructional tasks and teaching activities was based on a theoretical framework of skilled problem solving, consisting of a series of heuristic or cognitive strategies (i.e., guess and check, look for a pattern, work backwards, draw a picture, make a table, simplify the problem) embedded in an overall metacognitive strategy consisting of four stages (i.e., orientation, organization, execution, verification).

The problems used in the programme were of two broad types: routine and non-routine problems. Routine problems were typical multi-step word problems wherein pupils had to translate a verbally described real-world(like) situation into a mathematical expression. Typical of this type of problem is the following: 'Laura and Beth started reading the same book on Monday. Laura read 19 pages a day and Beth read 4 pages a day. What pages was Beth on when Laura was on page 133?' The non-routine tasks, on the other hand, involved problems with superfluous and insufficient information, but also so-called 'process problems', for which there is no standard algorithm or computation available. Illustrative of this category of non-routine problems is the following: 'A caravan is stranded in the desert with a 6 day walk back to civilization. Each person in the caravan can carry a 4 day supply of food and water. A single person cannot carry enough food and water and would die. How many people must start out in order for one person to get help and for the others to get back to the caravan safely?' The different types of tasks - standard multi-step problems, multi-step problems with irrelevant and superfluous information, and process problems - were included because each seemed especially suited to exemplify the need for and provide practice with particular aspects of the cognitive/metacognitive framework described earlier.

The instruction focused on solving problems in small groups, in combination with whole-class discussions and individual assignments. During these classroom activities the teacher had to fulfill three different but closely related roles: (a) serve as external monitor during problem solving; (b) encourage discussion of behaviour considered important for the internalization of the heuristic and metacognitive skills, and (c) model good executive behaviour. Table 1 delineates the required teacher behaviours before, during and after the problem solving activity, and Table 2 presents a card with problem-solving tips to be used by the teacher and the pupils and reflecting the major aspects of the intended cognitive/metacognitive strategy.

The instructional programme was realized by one of the investigators to a regular-level and an advanced-level seventh-grade class about three days per week for a period of 12 weeks. The total actual instruction time devoted to the programme was about 15 hours spread over 12 weeks of instruction, averaging about 1/3 of the total mathematics classroom time during the instructional period.

Before, during, and after the instruction, Lester et al. (1989) employed a large set of assessment instruments including written tests, clinical interviews, observations of individual and pair problem-solving sessions, and videotapes of the classroom instruction.

The results on the written pretest and post-test, involving a set of new routine and non-routine word problems, revealed positive effects of the training on the pupils' problem-solving skills. Both the regular class and the advanced class realized a considerable overall gain in total score from pre- to post-test, but this gain was not as large as hoped. Interestingly, large interindividual differences were observed: while the scores of most pupils in both classes increased from pre- to post-test, there was also a significant number of pupils whose scores remained the same or even decreased. The results of the individual interviews revealed that among the four categories of activities in the cognitive/metacognitive framework, the category called 'orientation' had the most important effect on pupils' problem solving.

Finally, observations of the lessons and reflections by the researchers offered some insights in the strengths and weaknesses of the instructional environment, and led to several suggestions for improvement. First, it was observed that the teacher experienced serious trouble in maintaining his role of model, facilitator and monitor in the face of classroom reality, especially when pupils had difficulties with basic subject matters. Second, the overall quality of the thinking and the interaction in the small groups was not as high as expected. Third, it appeared that the time provided for the experimental instruction was too short and too fragmented to be fully effective.

In addition to these difficulties raised by the authors themselves, some other

Table 1 Teaching Actions for Problem Solving (adapted from: Lester et al., 1989, p. 26)

	Teaching Action	Purpose
		BEFORE
1	Read the problem to the class or have a pupil read the problem. Discuss words or phrases pupils may not understand.	Illustrate the importance of reading problems carefully and focus on words that have special interpretations in mathematics.
2	Use a whole-class discussion about understanding the problem. Use problem-specific comments and/or the Problem-Solving Guide[1].	Focus attention on important data in the problem and clarify parts of the problem.
3	(Optional) Use a whole-class discussion about possible solution strategies. Use the Problem-Solving Guide.	Elicit ideas for possible ways to solve the problem.
		DURING
4	Observe and question pupils to determine where they are in the problem-solving process.	Diagnose pupils' strengths and weaknesses related to problem solving.
5	Provide hints as needed.	Help pupils pass blockages in solving a problem.
6	Provide problem extensions as needed.	Challenge the early finishers to generalize their solution strategy to a similar problem.
7	Require pupils who obtain a solution to 'answer the question'.	Require pupils to look over their work and make sure it makes sense.
		AFTER
8	Show and discuss solutions using the Problem-Solving Guide as a basis for discussion.	Show and name different strategies used successfully to find a solution.
9	Relate the problem to previously solved problems and discuss or have pupils solve extensions of the problem.	Demonstrate that problem-solving strategies are not problem-specific and that they help pupils recognize different kinds of situations in which particular strategies may be useful.
10	Discuss special features of the problem, such as a picture accompanying the problem statement.	Show how the special features of a problem may influence how one thinks about a problem.

[1]For the problem-solving guide see Table 2

Table 2 Problem Solving Tips (adapted from: Lester et al., 1989, p. 197)

Understanding the problem

- Read the problem carefully; often you should read it two or more times.
- Be sure you understand what the question is asking; ask yourself: "Do I understand what I am trying to find?"
- If you aren't sure you understand the problem, draw a picture or diagram of the information.
- Write down all the important information and the question; these are called: 'What I know and What I want to find'.

Solving the problem

- Explore the problem to get a good 'feel' for what the problem is about.
- Don't do anything hard until you have tried easy ideas first; if easy things don't help, then you may need to do something more complicated.
- When you don't have any idea of what to do, try to make a good guess and then check it out with the important data.
- Use the strategies that you have learned; for example:

draw a picture	guess and check	look for a pattern
make a table	work backwards	simplify the problem

Getting an answer and evaluating it

- Be sure to check your work along the way, not just at the end; you may be able to avoid some unnecessary work by finding a mistake early.
- Be sure that you used all the important information.
- Write your answer in a complete sentence; this makes it easier to decide if the answer is reasonable.
- Ask yourself: "Does my answer make sense?"

problematic aspects of the experimental programme can be mentioned. First, according to Siemon (1992), the language used in Lester et al.'s (1989) cognitive/metacognitive problem-solving strategy (see Table 2) was too alienating and too unfamiliar for children of that age. Therefore, Siemon (1992) suggests another model - the so-called 'Ask-Think-Do' model - in which the different aspects of the problem solving process are represented in a more natural and more user-friendly way. Second, as the focus of the programme of Lester et al. (1989) was on the development of a set of heuristic and metacognitive skills for complex mathematical problem solving, the major criteria for selecting and designing tasks was the kind of heuristic and metacognitive skills likely to be tapped during the problem-solving process. Unfortunately, this resulted in a number of rather unrealistic, puzzle-like problems (such as the example routine problem about children reading a precise number of book pages every day given earlier), which may hinder rather than foster the development of a genuine disposition toward realistic mathematical modelling (Greer, 1993).

Teaching realistic mathematical modelling to fifth graders

Starting from the finding that many upper elementary school pupils tend to approach arithmetic word problems in a stereotyped and non-realistic way, Verschaffel and De Corte (1997b) carried out a teaching experiment in which they tried to 'break' this tendency among pupils and to develop in them a disposition towards (more) realistic mathematical modelling and problem solving. This was attempted by immersing them into a classroom culture that fosters pupils' awareness that representing and solving application problems involves more than figuring out the correct operation(s) to apply with the given numbers.

Three classes from the same school participated in the experiment: one experimental class of fifth-grade children, and two control classes of sixth-grade children. The pupils from the experimental class participated in an experimental programme on realistic modelling consisting of five Teaching/Learning Units (TLU) of about 2 1/2 hours each, spread over a period of about 2 1/2 weeks. During the experiment, the pupils from the two control classes followed the regular mathematics curriculum. The major characteristics of the experimental programme are the following.

First, the impoverished and stereotyped diet of standard word problems offered in traditional mathematics classrooms, was replaced by a set of non-routine problem situations that were especially designed to stimulate pupils to pay attention at the complexities involved in realistic mathematical modelling, and at distinguishing between realistic and stereotyped solutions of mathematical applications. Each TLU focused on one prototypical problematic topic of realistic modelling accessible to these children. The topic of the first TLU was: making appropriate use of real-world knowledge and realistic considerations when interpreting the outcome of a division problem involving a remainder. The opening problem involved a story about a regiment of 300 soldiers doing several military activities. Each part of the story was accompanied with a question which always asked for the same arithmetic operation (namely 300 : 8 = .) but required each time a different answer (respectively, '38', '37', '37.5' and '37 remainder 4'). In the four other TLUs pupils were confronted with: problem situations about the union or separation of sets with joint elements (TLU 2), problems wherein the result of adding or subtracting the two given numbers yields an answer that is 1 more or 1 less than the correct one (TLU 3), problems in which one has necessarily to take into account several relevant elements that are not explicitly nor immediately 'given' in the problem statement but that belong to one's common-sense knowledge base (TLU 4), and problem situations wherein one has to realize that solutions based on direct proportional reasoning are inappropriate (TLU 5).

Second, not only the nature of the problems but also the teaching methods differed considerably from traditional mathematics classroom practice. The opening problem of each TLU was solved in mixed-ability groups of three-four pupils. This group assignment was followed by a whole-class discussion, in which the answers, the solution processes and possible additional comments of the different groups were compared. Then, each group was given a set of four or five new problems, some with and some without the same underlying modelling difficulty as the opening problem. This group assignment was again followed by a whole-class discussion. Finally, each pupil was individually administered one problem that involved once again the topical modelling difficulty, and the reactions to this individual assignment were also discussed afterwards in a whole-class discussion.

Third, attempts were made to establish a new classroom culture by explicitly negotiating new socio-mathematical norms about the role of the teacher and the pupils in a (mathematics) classroom, and about what counts as a good mathematical problem, a good solution procedure, or a good response (see Gravemeijer, 1994).

Before this programme was implemented in the experimental class, the three groups were given the same pretest consisting of five standard items (S-items) and ten 'problematic' items (P-items) related to each of the five modelling difficulties involved in the programme. Some of these P-items had problem contexts which resembled those used during the intervention (= learning items), but others were built around contexts that were more different from those encountered during the training (= near-transfer items). Table 3 gives an example of a P-item from the pretest related to each of the five kinds of modelling difficulties involved in the programme.

At the end of the experimental course a parallel version of the pretest was administered in the three classes as a post-test. However, in one of the control classes the post-test was preceded by an introduction in which the pupils were explicitly warned that the test would contain several problems for which routine solutions based on adding, subtracting, multiplying or dividing the given numbers, are inap-

Table 3 Examples of P-items from the Pretest (Verschaffel & De Corte, 1997b)

1	228 tourists want to enjoy a panoramic view from the top of a high building. In the building there is only one elevator. The maximum capacity of the elevator is 24 persons. How many times must the elevator ascend to get all tourists on the top of the building?
2	At the end of the school year, 66 school children try to obtain their swimming diploma. To get this diploma one has to succeed in two tests: swimming 100 metre breaststroke in 2 minutes and treading water during one minute. 13 children do not succeed in the first test and 11 fail on the second one. How many children get their diploma?
3	This year the annual rock festival Torhout/Werchter was held for the 15th time. In what year was this festival held for the first time?
4	This flask is being filled from a tap at a constant rate. If the depth of the water is 4 cm after 10 seconds, how deep will it be after 30 seconds? (This problem was accompanied by a picture of a clearly *cone-shaped* flask)
5	A man wants to have a rope long enough to stretch between two poles 12 metres apart, but he has only pieces of rope 1,5 metres long. How many of these pieces would he need to tie together to stretch between the poles?

The number before each problem refers to the corresponding teaching/learning unit (TLU) propriate. One month after the post-test, the pupils from the experimental class were administered a retention test consisting again of five S-items and ten P-items, half of which were contextually different from but structurally similar to the ones used during training (= near-transfer items), whereas the other half involved completely new mathematical modelling difficulties (= far-transfer items). An example of such a far-transfer item is: 'Bruce and Alice go to the same school. Bruce lives at a distance of 17 kilometres from the school and Alice at 8 kilometres. How far do Bruce and Alice live from each other?'. Depending on the presence or absence of the utilization of context-related real-world knowledge in their problem solving endeavours, pupils' reactions to the P-items in the pretest, post-test and retention test were scored as a Realistic Reaction (RR) or a Non-realistic Reaction (NR), respectively.

The experimental programme had a positive effect on children's disposition toward realistic modelling and interpreting of arithmetic word problems: in the experimental class the overall percentage of RRs on the P-items increased significantly from 7% during the pretest to 51% during the post-test; in the two control classes the progress from pretest to post-test was non-significant (namely from 20 to 34% in the first control group and from 18 to 23% in the second control group). The relatively small and non-significant increase in percentage of RRs in the first control class (wherein the pupils were explicitly warned about the problematic nature of some of the items in the post-test) indicates that merely telling and illustrating that routine solutions for word problems are not always appropriate, is not enough to transform pupils from mindless and stereotyped task performers into more critical and more realistic problem solvers. Moreover, the positive effect of the experimental programme was not restricted to the five so-called learning items (from 9 to 60% RRs), but transferred also to the five near-transfer (from 6 to 41% RRs). Finally, the results on the retention test (i.e., 41% RRs) showed that the positive effect of the experimental programme did not disappear after the training had stopped, and provided some evidence of far-transfer effects of the programme (see Verschaffel & De Corte, 1997b, for a more detailed report of the results.)

Although the above-mentioned results warrant a positive conclusion about the feasibility of breaking pupils' tendency to approach arithmetic application problems in a stereotyped and non-realistic way, some caution is in order. First of all, the reported positive results are jeopardized by some methodological weaknesses of the teaching experiment, such as the small size of the experimental and control groups and the absence of a retention test in the control classes. Second, as in Lester et al.'s (1989) study, large interindividual differences in learning, transfer and retention effects were found. A detailed analysis of the data revealed that the programme was especially effective for the strongest pupils of the experimental class. Third, the videotapes of the lessons revealed a number of difficulties with respect to the design and the implementation of the instructional programme. In this respect, we mention that the programme was too heavily focused at making children aware of the imperfections of their stereotyped and routine-based approach toward mathematical problem solving; although the programme was indeed successful in stopping pupils' tendency toward stereotyped and non-realistic mathematical problem solving, it provided insufficient help in developing a set of valuable heuristic and metacognitive skills for handling such complex and ambiguous problem situations. The videotapes revealed also that several low-ability pupils were frequently not actively and productively involved in the small-group activities and the classroom discussions; this explains why they profited much less from the programme than their high- and medium-ability peers. Finally, the videotapes showed that instructional interventions aimed at the establishment of the intended new social and socio-mathematical norms occurred too rarely and too implicitly to be fully effective.

Using videodisc technology to anchor mathematical problem solving in realistic contexts

As Verschaffel and De Corte (1997b), Van Haneghan et al. (1992) started from the assumption that traditionally pupils' experiences with word problem solving consist mainly of choosing the arithmetic operation with the given numbers in the problem to figure out the correct answer, without paying attention to the realities of the context evoked by the problem statement; this is far removed from problem solving in the real world, in which posing and defining problems, model building, planning and decision making, and interpreting outcomes - always linked to the real-life context - are major activities. However, while the tasks used by Verschaffel and De Corte (1997b) were mostly 'non-routine' or 'problematic' variants of the standard word problems used in the traditional word problem solving lessons, Van Haneghan et al. (1992) applied videodisc technology to confront pupils with rich, authentic, and complex problem-solving spaces offering ample opportunities for problem posing, exploration, and discovery. However, these videodisc-based problem spaces are only one of the components that they believe to be important; other conditions include (a) the guidance provided by an expert teacher, who organizes the learning experience, who stimulates co-operative learning and discussion in small groups, and who explicitly addresses the culture of the classroom, and (b) the availability to children of rich and realistic sources of information.

One series of videodiscs for mathematics instruction in grade 5 and 6 developed by the Vanderbilt group is called The adventures of Jasper Woodburry. In the initial videodisc of this series a person named Jasper Woodburry takes a river trip to see an old cabin cruiser he is considering purchasing. Jasper and the cruiser's owner test-run the cruiser, after which Jasper decides to purchase the boat. As the boat's running lights are inoperative, Jasper must determine if he can get the boat to his home dock before sunset. Two major questions that form the basis of Jasper's decision are presented at the end of the disc: (a) does Jasper have enough time to return home before sunset, and (b) is there enough fuel in the boat's gas tank for the return trip?

The major design principles underlying the Jasper series and their functions are the following:

- Video-based presentation format: According to the Vanderbilt group, there are reasons to situate instruction in a video-based format, mainly because the medium allows a richer, more realistic, more dynamic presentation of information than textual material. At the same time the video-based format has some advances over real-life contexts, simply because these latter methods are not always practical, efficient, and well-structured and difficult to organize in a school situation.
- Narrative format: the presentation of the problem in the form of a story helps pupils to create a meaningful context.
- Generative structure: by having the pupils themselves generate the resolution of the story, their active involvement in the learning process is stimulated.
- Embedded data design: by having the information that will be relevant to the solution embedded in the story, pupils are enabled to take part in problem identification, problem formation and pattern recognition activities that traditional word problem solving does not allow.
- Problem complexity: by posing very complex mathematical problems - sometimes comprised of more than 15 interrelated steps or subproblems - pupils are provided with the opportunity to engage in a kind of sustained and applied mathematical thinking that traditional curricula rarely offer.

Initial studies with the Jasper series have produced encouraging results (Van Haneghan et al., 1992). A baseline study revealed that even above-average sixth-graders were very poor in their approach to complex application problems of the kind used in the Jasper series without instruction and mediation. According to the authors, this was not too surprising, as pupils rarely have the opportunity to engage in such complex problem formulation and problem solving activities.

However, a subsequent controlled teaching experiment showed that videodisc-based anchored instruction can substantially improve pupils' problem solving processes and skills. Participants belonged to a fifth-grade class of above-average pupils. The first day of the experiment the Jasper video was shown to all pupils and then they were pretested. After pretesting, pupils were assigned either to an experimental or a control group and both groups received three additional teaching sessions. During these sessions the experimental group engaged in problem analysis, problem detection, and solution planning to check Jasper's trip-planning decision, thereby intensively relying on videodisc/computer technology controlled by the instructor. In the control group traditional teaching methods were used to instruct pupils in solving traditional unrelated-context word problems. Following instruction, pupils received two post-tests - one consisting of traditional word problems and one about organizing information in the Jasper video for problem solving - and a transfer test that assessed pupils' abilities to identify, define, and solve problems similar to those posed in the Jasper series.

The results can be summarized as follows. On the traditional problems of the first post-test the experimental pupils performed as well as the control pupils. On the second post-test the experimental group showed significant gains from pretest to post-test, whereas control pupils showed no significant improvement. Moreover, the analysis of interview protocols relating to children's problem solving with respect to a video near-transfer problem showed significant transfer in the experimental group, but not in the control groups.

At the end of their report, the authors stress that their initial attempt to test the facilitating effects of the Jasper context is preliminary. Indeed, the study involved a small and selective group of pupils and was restricted to a couple of hours with one videodisc. Continued research is necessary both to validate the claims of anchored instruction and to see whether the encouraging findings can be replicated in other samples and with other anchors. Meanwhile, additional studies of the Cognition and Technology Group at Vanderbilt (1997) have confirmed the initial positive results of anchored instruction based on the Jasper series, have led to supplementing the original Jasper series with effective analog and extension problems and appropriate scaffolding tools, and have provided evidence for the practical applicability of this system of anchored instruction based on the Jasper series in real classroom situations. Another weakness of the study is the absence of a detailed description and account of the instructional environment in which the videodisc was embedded.

Conclusions and discussion

In this chapter we presented an expert model of solving mathematical application problems involving the integrative and interactive application of different categories of aptitudes during the distinct phases of the problem-solving process. Then this model was used to describe and analyse some well-documented research findings concerning elementary school pupils' difficulties with the solution of application problems. Afterwards we pointed to a number of characteristics of the current practice and culture of school mathematics that are responsible for these difficulties in pupils, namely: the artificial and stereotyped nature of the word problems, the lack of explicit and systematic teaching of the intended higher-order knowledge and skills, and the absence of a supportive classroom culture. Finally, three evaluation studies of experimental instructional programmes in mathematical modelling and problem solving in upper elementary school children, were exemplary presented. In these programmes a set of carefully designed application problems, a collection of highly interactive teaching methods and an attempt to change the socio-mathematical classroom norms were combined to create a teaching/learning environment that focused at the development of a mindful and a realistic approach toward mathematical modelling and problem solving. On the one hand, the results of these studies indicate that these interventions had a positive effect on pupils' mathematical modelling and problem solving behaviour. On the other hand, the positive effects obtained were not totally convincing. In addition to the critical comments mentioned at the end of each reported study, we provide a more general list of problematic aspects and unresolved questions, which need to be addressed in further research and developmental work.

First, all three intervention studies had a rather specific aim and scope. In Lester et al.'s (1989) study pupils were trained in the mindful and flexible use of a fixed set of cognitive and metacognitive strategies for dealing with non-routine word problems; Verschaffel and De Corte's (1997b) major goal was to break pupils' superficial and non-realistic approach to arithmetic word problems; Van Haneghan et al. (1992) investigated the possibilities of new information technologies to bring reality into the mathematics lessons. Taking into account the complex nature of the skill of applied mathematical problem solving and the multiform nature of children's difficulties in acquiring it, instructional environments aimed at the development of this skill will have to combine and integrate the most powerful aspects from these different instructional approaches.

Second, although all authors stress the necessity of appropriate interventions aimed at the establishment of a new classroom culture that fits with the aims and scope of their instructional programme, it seems that this pillar was not addressed in a sufficiently systematic way in these programmes. In all three cases, the teacher initiated some discussion, raised questions, made remarks, and gave feedback with a view to changing pupils' beliefs and to inducing new classroom norms (e.g., about what counts as a good problem, a good solution strategy, a correct answer and a proper explanation; or about what can be expected from the pupils and the teacher in a mathematics class). But it seems that most of these discussions, questions and comments were rather unsystematic and inconspicuous, and they elicited too rarely mindful thoughts and reflections among the pupils. Certainly, this aspect of the experimental programmes should be studied in greater detail in future research (Gravemeijer, 1994; Lampert, 1990; Schoenfeld, 1992).

Third, while small group work was applied all experimental programmes, the results suggest that some scepticism about the effectiveness of these co-operative learning groups is warranted, as several pupils were found to be rather passive during these small-group activities and - consequently - to profit relatively little from this kind of instruction. Probably, this was partly due to the fact that the pupils had little or no experience with co-operative learning in the mathematics class at the beginning of the programme. Another explanation may be the lack of specific arrangements aimed at guaranteeing the active and productive participation of all pupils. Some possibilities for improving the small-group activities toward that end deriving from these studies are:

- Structuring the interaction and the division of labour by alternating roles and accountabilities during the group work (e.g., a process regulator, an executor, a controller, a reporter...).
- Providing more systematic and more intentional coaching during group assignments, for instance, by asking questions such as "What are you doing?"

and "Why are you doing this?" (Good, Mulryan, & McCaslin, 1992).

Fourth, all three instructional programme occurred over a relatively small period of time. However, the available research on teaching and learning problem solving indicates that developing problem-solving expertise in a complex subject-matter domain as mathematics is very difficult to realize, partly because it often involves 'unlearning' inappropriate metacognitive beliefs and control behaviours developed through prior instruction (Schoenfeld, 1992), partly because of the complex interplay between the strategic and the content-related aspects of the problem-solving ability. If we want pupils' problem-solving abilities to increase more than in the studies reported above, they will need appropriate instruction over a more prolonged period of time.

Fifth, we remind that - although the three experimental programmes were planned and realized in collaboration with regular classroom teachers - the actual teaching was always done by one of the researchers. Future research should investigate how these new instructional materials and methods are implemented by regular classroom teachers in regular mathematics classrooms, taking into account the teachers' cognitions and beliefs as a mediating variable. Indeed, it has been convincingly demonstrated that many teachers have inappropriate conceptions, skills and attitudes with respect to (realistic) mathematical modelling and problem solving, which may negatively interfere with their teaching behaviour (see, for example, Gravemeijer, 1994; Verschaffel, De Corte, & Borghart, 1997). Therefore, the introduction of new teaching and learning materials aimed at enhancing pupils' mathematical problem solving and at connecting mathematics more to the experiential worlds of pupils, will have to be complemented with initiatives aimed at stimulating and supporting teachers to construct for themselves the proper concepts, skills and attitudes that are needed for realistic mathematical modelling and problem solving.

Sixth, the three experimental programmes presented above consisted of a series of teaching/learning units organized separately from the regular mathematics lessons. However, it should be obvious that we do not plea for the implementation of a special programme in the mathematics curriculum wherein pupils are trained in a systematic way in realistic mathematical modelling of rich and complex context problems, whereas the rest of the mathematics curriculum remains unchanged. On the contrary, in a more authentic and problem-oriented approach to mathematics education, as envisaged in the current reform documents relating to mathematics education (see, e.g., NCTM, 1989), the development of a disposition toward realistic mathematical modelling and problem solving should permeate the entire curriculum from the outset. Programmes like the one developed by Lester et al. (1989), Verschaffel and De Corte (1997b) and Van Haneghan et al. (1992) can be especially valuable during the first steps towards the implementation of such a new approach toward more authentic mathematical problem solving (see also: Tabberer, 1987).

Finally, there is the issue of transfer. Although all three intervention studies aim at enhancing transfer not only to other domains of the school curriculum but also to out-of-school settings, none of them yields convincing evidence of effects going beyond transfer of knowledge and thinking skills to problem situations that are (very) similar to those encountered in the programme. In the future, theoretical analyses and empirical research should provide a clearer picture of the conditions under which the concepts, skills and attitudes acquired in these programmes will effectively transfer to complex problems occurring in other parts of the mathematics curriculum, in other subject-matter domains and in out-of-school contexts.

References

Baruk, S. (1989). *Wie alt ist der Kapitän?* [How old is the captain?]. Stuttgart: Birkhäuser.

Burkhardt, H. (1994). Mathematical applications in school curriculum. In T. Husén & T.N. Postlethwaite (Eds), *The international encyclopaedia of education* (2nd ed., pp. 3621-3624). Oxford: Pergamon.

Carpenter, T.P., Lindquist, M.M., Matthews, W., & Silver, E.A. (1983). Results of the third NAEP mathematics assessment: Secondary school. *Mathematics Teacher, 76*, 652-659.

Cognition and Technology Group at Vanderbilt (1997). *The Jasper Project. Lessons in curriculum, instruction, assessment, and professional development.* Mahwah, NJ: Erlbaum.

Collins, A., Brown, A., & Newman, S.E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics. In L.B. Resnick (Ed.), *Knowing, learning and instruction. Essays in honour of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Erlbaum.

De Bock, D., Verschaffel, L., & Janssens, D. (1998). Solving problems involving length and area of similar plane figures and illusion of linearity: An inquiry of the difficulties of secundary school students. *Educational Studies in Mathematics, 35*, 65-83.

De Corte, E., Greer, B., & Verschaffel, L. (1996). Mathematics teaching and learning. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 491-549). New York: Macmillan.

De Corte, E., & Somers, R. (1982). Estimating the outcome of a task as a heuristic strategy in arithmetic problem solving: A teaching experiment with sixth-graders. *Human Learning, 1*, 105-121.

De Corte, E., Verschaffel, L., & Van Coillie, V. (1988). Influence of number size, problem structure and response mode on children's solutions of multiplication word problems. *Journal of Mathematical Behaviour, 7*, 197-216.

Freudenthal, H. (1991). *Revisiting mathematics education.* Dordrecht: Kluwer.

Gravemeijer, K.P.E. (1994). *Developing realistic mathematics education.* Utrecht: Freudenthal Institute.

Good, T.L., Mulryan, C., & McCaslin, M. (1992). Grouping for instruction in mathematics: a call for programmic research on small-group processes. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 165-196). New York: Macmillan.
Greer, B. (1992). Multiplication and division as models of situations. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 276-295). New York: Macmillan.
Greer, B. (1993). The modelling perspective on wor(l)d problems. *Journal of Mathematical Behaviour, 12*, 239-250.
Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal, 27*, 29-63.
Lester, F., Garofalo, J., & Kroll, D. (1989). *The role of metacognition in mathematical problem solving. A study of two grade seven classes.* Bloomington, IN: Mathematics Education Development, Indiana University.
National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: NCTM.
Nunes, T., Schliemann, A.D., & Carraher, D.W. (1993). *Street mathematics and school-mathematics.* Cambridge, MA: Cambridge University Press.
Radatz, H. (1983). Untersuchungen zum Lösen eingekleideter Aufgaben [Reaearch to the problem solving of context tasks]. *Journal für Mathematik-Didaktik, 4*, 205-217.
Säljö, R., & Wyndhamn, J. (1987). The formal setting as context for cognitive activities. An empirical study of arithmetic operations under conflicting premises for communication. *European Journal of Psychology of Education, 2*, 233-245.
Schoenfeld, A.H. (1985). *Mathematical problem solving.* Orlando, FL: Academic.
Schoenfeld, A. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-370). New York: Macmillan.
Siemon, D. (1992, August). *Enhancing metacognition in the mathematics classroom.* Paper presented at the Seventh International Conference on Mathematics Education. Quebec, Canada.
Silver, E.A., Smith, M., Lane, S., Salmon-Cox, L., & Stein, M. (1990). *QUASAR summary project.* Pittsburgh, PA: University of Pittsburgh.
Sowder, L. (1988). Children's solutions of story problems. *Journal of Mathematical Behaviour, 7*, 227-238.
Tabberer, R. (1987). *Study and information skills in schools.* Windsor, Berkshire: NFER-Nelson.

Van Essen, G. (1991). *Heuristics and arithmetic word problems*. Amsterdam: State University.

Van Haneghan, J., Barron, L., Young, M., Williams, S., Vye, N., & Bransford, J. (1992). The Jasper series. An experiment with new ways to enhance mathematical thinking. In D.H. Halpern (Ed.), *Enhancing thinking skills in the sciences and mathematics*. Hillsdale, NJ: Erlbaum.

Verschaffel, L., & De Corte, E. (1997a). Word problems. A vehicle for authentic mathematical understanding and problem solving in the primary school. In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective* (pp. 69-97). Hove, East Sussex, UK: Psychological Press.

Verschaffel, L., & De Corte, E. (1997b). Teaching realistic mathematical modelling and problem solving in the elementary school. A teaching experiment with fifth graders. *Journal for Research in Mathematics Education, 28,* 577-601.

Verschaffel, L., De Corte, E., & Borghart, I. (1997). Pre-service teachers conceptions and beliefs about the role of real-world knowledge in mathematical modelling of school word problems. *Learning and Instruction, 28,* 577-601.

Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modelling of school arithmetic word problems. *Learning and Instruction, 4,* 273-294.

10

Teaching mathematical thinking to children with special needs

J.E.H. Van Luit

Introduction

Over the past 20 years, increasing attention has been paid to the idea that schools should be less concerned with imparting information and more concerned with the kind of teaching which concentrates on the way children learn. At the same time, it has been more widely realized that childrens' thinking abilities are, in general, underestimated. Coles and Robinson (1991) show, in particular, that the critique and reinterpretation of Piagetian theory have enormously encouraged attempts to teach thinking directly, an approach which is reflected in the literature (e.g., Baron & Sternberg, 1987; Coles & Robinson, 1991; Halpern, 1992; Hamers & Overtoom, 1997). Before explaining how teaching thinking is instantiated in a special maths programme, the most important components of this approach should first be introduced. It is useful here to mention the types of thinking specified by Baron and Sternberg (1987), who divide thinking skills into three kinds:

- Executive processes which are used to plan, monitor and evaluate one's own thinking.
- Performance processes which are actually used to carry out thinking.
- Learning processes which are used to learn how to think.

Examples of executive processes would include identifying, and formulating a question, keeping the situation in mind and organising one's thoughts. Performance processes include seeing similarities and differences, deducing, and making value judgements, while asking and answering questions of clarification such as 'What do you mean by that?', and listening carefully to other people's ideas, are examples of learning processes (Coles & Robinson, 1991). In the second part of this chapter we will discuss the way a specific maths programme attempts to teach these different kinds of thinking. However, Baron and Sternberg (1987) claim that it is possible to teach thinking skills in separate courses for reading and for mathematics, as well as informing the entire curriculum. For the first option we have chosen teaching mathematical thinking skills in a specific programme, in the first place because the field of special education needs curriculum-based programmes and, in the second place, because of the results in earlier studies with this kind of programme (see Van Luit, 1994a).

Thinking is a process we have to learn. We build up our own personal frameworks of interpretation and understanding, and techniques of problem solving, through experience. A rich culture like ours offers many of these frameworks ready-made, in language, mathematics and science, for example; but each of us has to make these frameworks our own (Nisbet, 1991). An example of this idea in the field of mathematics is given by Hasemann (1994), who demonstrates experimentally that the similarity of external and internal representation makes the acquisition of knowledge more effective when new knowledge is presented to learners in the form of structured networks, or when they are encouraged to construct such networks themselves. There are skills and strategies in thinking which we develop from experience. Some people are quicker than others at acquiring these skills; but appropriate teaching can help all of us to improve our competence. However, educating poor students in thinking skills will be more difficult than better students.

There is a consensus among mathematics educators and researchers, that to learn mathematics means to construct mathematics. With the adoption of realistic mathematics education in primary schools (Schoenfeld, 1989; Siegler & Jenkins, 1989; Skemp, 1989; Treffers, 1987), children with disappointing achievements have also been given more attention in the last ten years. This is not surprising, because the weaker pupil, capable of only a limited contribution to the learning process (Geary, 1994), tends to drop out earlier than previously. Realistic mathematics education means learning to use several problem-solving strategies by top-down processing. This means that:

- Children discover adequate strategies themselves by solving a variety of mathematics problems. Furthermore, they discover the quickest problem-solving strategy for a given situation.
- All children are given brief instruction by the teacher and, subsequently, the (individual) solutions of a problem are discussed in small groups. The teacher does not provide solutions, but leads the discussion of different strategies suggested by the children themselves.

- Most mathematics problems for group discussions are presented in meaningful and rich contexts, using a variety of mutually linked models, schemes and graphs. In addition there is also time for individual exercise.

A realistic method thus provides interactive education involving the entire group. However, these group assignments alone allow poorer pupils in mathematics to master too little information. These pupils are often not sufficiently involved in the lesson to master mathematics (Wilkinson, Martino, & Camilli, 1994).

Theoretical orientation

'Cognitive mathematics' (Van Luit, 1994a) is an active field of research. A considerable body of results has accumulated, and several theories have been proposed regarding cognitive processes underlying childrens' simple mathematical problem-solving strategies (for reviews, see Ashcraft, 1992; Clark & Campbell, 1991; Lemaire, Barrett, Fayol, & Abdi, 1994; McCloskey, Harley, & Sokol, 1991; Rickard, Healy, & Bourne, 1994). Beyond the basic question "How do we do mathematics in our heads?" much effort has been devoted to investigating related issues, such as children's acquisition of mathematics knowledge and skills (e.g., Campbell & Graham, 1985; Siegler & Jenkins, 1989), and the nature of the problems children with learning or mathematics disabilities have when presented with mathematical tasks (e.g., Geary, 1990; Geary, Brown, & Samaranayake, 1991).

Only a few experiments (e.g., Geary et al., 1991) have examined specific programmes or treatments for children with severe mathematics disabilities. Questions like "How do children with special needs solve simple mathematics problems?" are the topic of these experiments. Nevertheless, questions like "Is mathematical thinking in children with special needs comparable with that of 'average' childrens' thinking" are not included in these experiments.

Current developmental models (e.g., Ashcraft, 1992; Fayol, 1990) suggest that counting-based knowledge of facts plays an important role in the development of young children's mental mathematics. Indeed, it can be considered as a prerequisite for mathematical thinking. Fundamental to any mental problem-solving is the child's ability to use simple mathematical facts adequately; that is, the ability to determine quickly and accurately the answers to such problems as 3x3. This basic knowledge must be automatic for children, to facilitate more difficult problems, such as 3x6, 3x13, and so on. Recent research (e.g., Ashcraft, 1987; Campbell, 1987; Miller & Paraedes, 1990; Rickard et al., 1994; Siegler, 1988) shows that, whereas children often consciously use counting knowledge, especially in the early acquisition of problem-solving skills (e.g., children often solve 3x6 by adding 6+6+6), the subsequent development of these skills leads toward retrieval of mathematical facts directly from memory.

In this experimental study, we investigate the efficacy of teaching children with special needs to think about different problem-solving strategies in multiplication and division, and to practice these strategies adequately. The study thus focuses on the one hand on seeking solutions, rather than merely memorizing procedures, and on the other hand on exploring patterns, rather than memorizing formulae (Schoenfeld, 1992). For remedial basic mathematics, Schoenfeld recommends trying to induce 'critical thinking' or 'analytical reasoning' skills. The difficulty of achieving this recommendation in the education of children with mathematical disabilities will be highlighted.

In schools for special education, especially for the learning disabled (LD) and the educable mentally retarded (EMR) children, the number of children with mathematics disabilities seems to be on the increase. These difficulties begin at an early stage in the child's school life. Many children in special schools are unable to learn the four basic operations of addition, subtraction, multiplication and division before leaving primary education at the age of about twelve years. The reason for these difficulties can be found in their lack of certain fundamental thinking skills needed for successful mathematical reasoning; for example, understanding causal relationship, recognizing and criticizing assumptions, analysing means-goals relationships, assessing degrees of likelihood and uncertainty, incorporating isolated data into a wider framework, and using analogies to solve problems (Halpern, 1992).

Since the seventies, there have been many studies aimed at investigating whether educable, either mentally retarded children or learning disabled children are better able to learn and/or memorize by teaching them the skills (executive processes, performance and learning processes) Baron and Sternberg (1987) mentioned above. It has been shown that education based on the use of open verbalizations by children about the choices they make in the solution procedure has a number of advantages. Children who are aware of the different possibilities for solving mathematical problems (partly) meet the requirements which Schoenfeld (1992) mentioned. For example: a child has to solve the problem 7x8. If (s)he does not know the answer automatically the child can use a strategy to solve this problem. One possibility would be 5x8 + 2x8= 40+16=56 and a less adequate strategy could be 1x8=8, 2x8=16, ..., 7x8=56.

Self-instruction is an important approach in teaching children these strategies. Training in self-instruction has been employed to stimulate the general behaviour required for academic success (see for a review Van Luit, 1989; Whitman, 1987). Schoenfeld (1991) notes that it is very important in thinking about mathematics to communicate mathematically, to express oneself using the language of mathematics. In mathematical education, children have to be asked to use mathematical terminology when they write and speak about their solving of a mathematics problem, i.e. to think (aloud) about possible solutions in problem-solving and the steps needed to complete such a solution, but also about the intermediate steps. For example: a child has to solve 16x8. First he must ask himself what to do. He can consider different problem-solving strategies, such as 10x8 + 6x8, 8x8 + 8x8, 2x8 + 2x8 + 2x8 + 2x8, etc. The child can help himself with this kind of thinking by asking himself questions. Before solving the problem the child asks himself questions like: "What should I do?", "Which strategies do I know?", "Which one is the easiest for me?", and "How do I have to do this?". Having chosen a given strategy, he thinks about the steps needed to complete the task, for example: "First I multiply 10x8 which makes 80, then I have to do 6x8 makes 48. And 80 together with 48 makes 128, so 16x8 is 128. Oh yes, that is correct. I see now, because I know by heart that 15x8 makes 120, so 16x8 is 8 more!".

The research in the field of education related to self-instruction training has been designed by Meichenbaum (1977). The goal of such a training is to teach children to think about the possible steps toward solving a problem before actually doing so. A child must ask himself what the (mathematical) problem demands from him. He has to think (flexibly) about the main aspects of the problem, on the one hand to remember what he already knows about this kind of problem and how this knowledge can facilitate the solution and, on the other hand, exactly what information about the problem he needs to solve the problem adequately. One way of facilitating the teaching of self-instruction, to enable children to consider possible strategies in problem-solving, is modelling. Through modelling, by the trainer/teacher, low ability children learn to think in problem-solving steps by asking themselves a series of questions about the nature of the problem. The main purpose of self-instruction training is to make children aware of their own responsibility to find a quick and accurate answer to mathematical problems.

On the basis of teaching children to use self-instruction in mathematical problem-solving we have developed the SRVV programme (Special Mathematics Programme for Multiplication and Division; Van Luit, Kaskens, & Van de Krol, 1993) for teaching children with mathematics disabilities to learn to think about problem-solving in multiplication and division. In this study, we examine whether this programme can influence children's ability in problem solving in this mathematical domain.

Method

Subjects

Sixty students participated in this study. These students were selected from two school systems providing special education to children, who are educable, either mentally retarded (EMR) or learning disabled (LD). All these children had severe mathematics disabilities. This group of 60 children was selected from 114 children whose performance on a standardized paper-and-pencil mathematics test fell below the level of 50 % correct. Further selection was performed by individual diagnostic research (see Van Luit, 1994b). The following functions of the mathematics operations are checked in each student: orientation (operation meant to explore the situation, as preparation for the implementation), implementation (the accomplishment of the intended operation result), control (of the result accomplished) and reflection (thinking critically about one's own operations).

The 60 children selected were unable to think mathematically in terms of understanding relations between multiplication and addition, division and subtraction, multiplication and division, and in understanding associations between different ways of problem solving. At the beginning of the training the average age of the 30 children in EMR schools was 12 years and 8 months (SD = 11 months) and the average age of the 30 children in the LD schools was 10 years and 10 months (SD = 10 months). Their mathematics level was comparable to that of third grade pupils in a primary school. Children in the EMR schools were thus retarded by about three and a half years and the children in the LD schools by about one and a half years.

Procedure

In this study we used a pretest, posttest control group design with follow-up. In each school system fifteen children in experimental groups were trained in working groups of five children each, with the SRVV programme. The other fifteen children in the control groups in each school system received a mathematics training in their own classroom making use of instruction procedures and materials of the standard curriculum with a great deal of individual help from their (remedial) teacher. The experimental groups received instruction from remedial teachers. During the experiment of sixteen weeks, the children did not work on any other mathematics programme nor did they receive any other mathematics instruction. The experiment took place in separate classrooms three times a week, 45 minutes each time. Each session consisted of the mathematics problems in accordance with the instructions given in the SRVV programme.

Materials

In the SRVV programme the childrens' problem solving within and between domains (especially multiplication and division) in mathematics was taken into account. For example, children with mathematics disabilities do not combine new information with already known information. Within the domain of multiplication they do not make the connection between an already known task (5x8 = 40), and a new task (6x8 = ?). In addition these children do not understand the connections which can be made between domains, for example, between multiplication and division (24:4 = 6 and 6x4 = 24). The purpose of the study therefore was to teach children with severe mathematics disabilities how to think mathematically by using adequate problem-solving strategies.

In the SRVV programme (Van Luit et al., 1993) the following goals for learning were designed to increase the child's:

- Understanding of multiplication as repeated additions.
- Understanding of number system and the premises of some problem solving strategies like reversibility (5x9 = 9x5), associations (9x7 = 10x7 - 1x7), and doubling (8x6 = 4x6 + 4x6).
- Importance of memorization of basic multiplication facts below 100 to facilitate the problem solving of more difficult multiplication problems.
- Understanding of division as repeated subtractions.
- Understanding and using as a control activity the connection between division and multiplication.
- Importance of memorization of all the division facts below 100 to facilitate the problem solving of more difficult division problems.
- Application of multiplication and division in real and imagined situations.

The SRVV programme comprises 23 lessons in multiplication and 19 lessons in division. The programme is made up of the following series. The multiplication lessons consist of eight lessons on the basic procedures (relations between addition and multiplication and, related to that, the association between multiplication and long addition, and introduction of the meaning of the tables), 11 lessons on the different problem-solving strategies and memorization of multiplication tables, and four lessons on specific problems with one number below 10 and one number between 10 and 20 (e.g., 8x17). The lessons in division consist of seven lessons on the basic division procedures, six lessons on the different problem solving strategies and memorization of division without remainders and the connection with multiplication, five lessons on division on the different problem-solving strategies and memorization of division with remainders, and one lesson with the denominator between ten and twenty (e.g., 72:12). After the last lesson in the series, there is a curriculum based test. If a child scores 80 % or more correct answers on a test at the end of each series and he shows adequate strategies in problem solving, he may go on to the following series.

Each of the series involves teaching and discussing thinking about new steps in problem solving related to specific tasks. A series always starts with an orientation phase, and then the task can be solved with the help of materials. After that the connection is made with mental problem solving. The child then has to learn to check the solution. Finally, this cycle ends with the control phase - shortening - automatization - generalization. The teacher tries to bring about this complex method of working with use of self-instruction. For children who need more assistance, suggestions or repetition of a problem solving strategy by their teacher turn into more specific help. Self-instruction included in the programme offers, for example, modelling. The programme includes explicit information about how to teach these children who need considerable help to address their lack of fundamental thinking skills, especially in the domain of mathematics.

The SRVV programme involves setting the opportunity for problem-solving strategy generation and use (e.g., Butterfield & Nelson, 1989). Children are given the opportunity to apply their own solution or strategy to problems. The teacher's task is to lead the discussion in the direction of the use of adequate strategies and to facilitate the discussion of the strategies put forward by the children. A child can use any problem-solving strategy he or she wishes, but the teacher assists the children in discussion and reflection about the choices made. The teacher presents training involving several strategies, which become increasingly shorter. For example an initial strategy for 8x7 could be: 1x7 + 1x7 + 1x7 + 1x7 + 1x7 + 1x7 + 1x7 + 1x7. Some children may find this is too long a strategy and use: 2x7 + 2x7 + 2x7 + 2x7 or 4x7 + 4x7, or 5x7 + 3x7, for example. The teacher ensures that each child understands the different strategies and encourages them to use the most efficient strategy. The main goal is to help children to think about an adequate strategy with the use of simple multiplication and division facts in solving more complex problems. For more specific information about the SRVV programme see Van Luit (1994b) and Van Luit et al. (1993).

Results

Treatment effects

Mean scores and standard deviations for pretest, posttest and follow-up maths scores, and effect sizes are presented in Table 1. These data show that both the experimental and the control groups improved with instruction, although, the degree of improvement differed substantially. The effect size for all children in the experimental groups in comparison with the control groups is large on the posttest (2.10) as well as on the follow-up test (2.04). The differences between pretest and posttest scores show that the LD experimental group ($t(1,14)=18.57$, $p<.001$) and EMR experimental group ($t(1,14)=21.19$, $p<.001$) improved significantly. However, the LD and EMR control groups pre-post results were also significant ($t(1,14)=6.41$, $p<.001$) and ($t(1,14)=8.56$, $p<.001$), respectively.

Differences between groups were tested with ANOVA. Between group results on the pretest, differences between experimental and control groups show that the LD groups ($F(1,28)=.236$, $p>.10$) and the EMR groups ($F(1,28)=.155$, $p>.10$) have comparable scores. These results show that the procedures used for matching the two EMR and the two LD groups yielded samples with similar pretest scores. The analyses of the experimental and control groups' posttest results show that the experimental LD group achieve better scores after the training period then the LD control group ($F(1,28) = 33,847$, $p<.001$). The LD between groups' mean score differences expressed as an effect size is 2.36. These results indicate that the SRVV programme had considerable influence on the mathematics performance of the experimental group. The results are similar for the EMR groups. The experimental EMR group also show higher scores on the posttest than the EMR control group ($F(1,28)=54,878$, $p<.001$). The effect size (2.99) for the EMR children shows that these children in the experimental group achieved substantially larger scores on the

Table 1 Pretest and Posttest Means, Standard Deviations, and Effect Sizes for 40 Multiplication and Division Problems

Group	N	Pre-Test		Post-Test		Follow-up		Pre-Post	Pre-Follow-up
		Mean	SD	Mean	SD	Mean	SD	Effect*	Effect*
EMR & LD									
Exp.	30	11.4	6.2	31.9	5.4	31.8	5.1	2.10	2.04
Control	30	12.2	6.6	18.0	8.4	17.9	9.1		
EMR									
Exp.	15	7.3	4.5	28.3	4.4	28.5	4.7	2.99	3.15
Control	15	8.0	4.8	13.5	6.4	13.4	7.0		
LD									
Exp.	15	15.4	5.1	35.6	3.5	35.1	2.6	2.36	2.13
Control	15	16.3	5.4	22.5	8.0	22.3	9.0		

* Effect Size (d) = $(M_1-M_2)/Sp$ $[Sp^2=(N_1-1).S_1^2+(N_2-1).S_2^2/N_1+N_2-2]$

posttest measures than the EMR children in the control group, supporting the effectiveness of the SRVV programme.

The differences in scores on the pretest measures between children in EMR and LD schools is striking. The LD and EMR groups (experimental and control together) differed significantly at the pretest ($F(1,59=42,142$, $p<.001$) as well as posttest ($F(1,59)=12.092$, $p<.001$). The children with learning disabilities answered approximately 40% of the mathematics problems on the pretest correctly while the EMR children only answered 16% of the problems correctly. At the posttest these percentages were 62% and 54% respectively. The eight measurements (see Table 2) obtained during the training period showed that the experimental LD children children made a gradual and consistent progression but that the experimental EMR children did not show much progress until the third month. The performance of the control groups showed a slight increase during the treatment interval.

Follow-up

The results of the three month follow-up show no differences in comparison with the posttest scores obtained at the conclusion of the training period (see Table 1). The differences between posttest and follow-up are non-significant for all four roups. The small decrease for the total experimental group (mean score from 31.9

Table 2 Percentage Correct During the Training Period at the End of Two Week Period of Training

	Weeks 2 - 16							
Group	2	4	6	8	10	12	14	16
EMR								
Experimental	12	15	21	26	46	63	73	75
Control	16	23	24	27	29	33	37	38
LD								
Experimental	41	55	61	74	81	90	92	92
Control	43	48	51	55	56	58	61	60

to 31.8) is statistically non-significant ($t(1,29)=.20$, $p>.10$). The students in the total control groups also show a statistically non-significant ($t(1,29)=.50$, $p>.10$) difference between posttest and follow-up test scores (mean score from 18.0 to 17.9).

This means that the progress of the students in the experimental groups is stable. After three months, including six weeks summer holidays and return to their regular classrooms, the students of the experimental group in the LD schools have maintained their high scores. Further progress was not expected for this group because of ceiling effects. The students of the experimental group in the EMR schools maintained the posttest level reached.

Generalization

Prior to starting the programme, the trainers assessed each students' individual skill level. They then taught the children to use specific mathematics materials and self-instruction on the most difficult type of multiplication and division problems mastered. The mastery levels were considerably easier than the types of mathematics problems the pupils had been working on in their own classrooms. Therefore the mathematics material was of a lower level than usually practised in the classrooms. Many children, especially those in LD schools, found this practice childish, because the students overestimated their level of knowledge.

The trainers of the experimental groups used the first two weeks of the four month programme to acclimatize the children to the newly formed groups. After this period, the pupils were trained in the specific components of the treatment. For many children, the new mathematics programme and the self-instruction procedure were very demanding, especially for the children from EMR schools. For example

Problem:	In a little street there are two rows of 7 houses. How many houses are there in the street?
Trainer:	"Can you solve this problem?"
Bianca:	"Yes, that makes ... eh ... fourteen."
Trainer:	"How did you do that Bianca?"
Bianca:	"Just easy, I know that two times seven makes fourteen."
Trainer:	"Can you show me that little street with the help of these blocks or can you draw the little street for me?"
Bianca:	... "How do you mean?"
Trainer:	"Can you solve the problem by laying down the blocks or by drawing the houses?"
Bianca:	She lays down one group of two blocks and separately one group of seven blocks and she is saying "Two times seven makes fourteen." She laughs softly when he says "That's not right."
Trainer:	"How can you do it in another way?" "What does the little street looks like?"
Bianca:	"I don't know."
Trainer:	"It doesn't matter. We will learn that in the coming weeks."

Figure 1 Example of a solution by Bianca before the training

verbalizations of the students were often found to be limited (see illustration in Figure 1).

After about six weeks of training, a change became apparent. Most children became aware of the purpose of the training and they slowly gained insight into the goals and problem-solving procedures of the programme. Some children quickly acquired a new procedure and came to use adequate problem-solving strategies on their own. Others consistently had difficulties in discovery similar rules between mathematics problems, the adequate use of problem-solving strategies, and so on. In general, most children, especially the children from LD schools, did acquire different strategies as a result of discussion and/or instruction. An example of the use of such a strategy is shown in Figure 2. In this figure Bianca can solve problems adequately by the end of the training.

The children were examined for far generalization (Brooks & Dansereau, 1987; Butterfield & Nelson, 1989) by evaluating video recordings made during specific test sessions. Three independent observers evaluated the videos (inter-rater reliability was 96.7%). To examine far generalization at the end of the training period the observers analysed the childrens' problem-solving. The students were given mathematics problems of a type that was substantially more difficult than the type of problems they were previously trained in. The observers categorized

Problem:	In the box are 8 cans of oil. How many cans there are in 9 boxes?
Trainer:	"Can you solve this problem?"
Bianca:	"Yes that's ... seventy-two cans."
Trainer:	"How did you do that Bianca?"
Bianca:	"Well first I did ten times eight that makes eighty cans and after that I subtract one times eight. So that makes eighty minus eight is seventy-two."
Trainer:	"Can you show it to me with help of the blocks or can you draw it for me?"
Bianca:	"Sure!" She lays down ten ten-sticks, puts a piece of paper on all last two blocks of these sticks and then puts away one ten-stick. "Nine times eight makes seventy-two."
Trainer:	"Very well done Bianca!"

Figure 2 Example of a solution by Bianca at the end of the training

problem-solving into three classes: effective, potentially effective and ineffective. An effective solution means a short route problem-solving strategy with a correct answer; a potentially effective solution means a long solution route with about 50% probability of a correct answer; and an ineffective problem-solving strategy means different solution routes without an acceptable strategy and mostly without correct answers. Table 3 shows the division of the problem-solving strategies of the groups of children in these three classes.

Table 3 shows the far generalization of the groups of children in EMR and LD schools in their use of problem-solving strategies. One hundred percent of the experimental children in LD schools used an effective or potentially effective way

Table 3 The Methods of Problem-Solving on Far Generalization Tasks

		Ways of problem-solving		
Group	N	Effective	Potentially Effective	Non-effective
EMR				
Experimental	15	1	3	11
Control	15	0	0	15
LD				
Experimental	15	12	3	0
Control	15	4	4	7

of problem-solving on the far generalization tasks. Only 27% of the experimental children in EMR schools were also able to solve unfamiliar tasks in an effective or potentially effective way. For the children in control, LD, and EMR schools, these percentages were 53 and zero, respectively. The achievement of the children in the experimental groups appears to be more effective than the achievement of the children in the control groups. Teaching mathematics to children in LD schools using the SRVV programme seemed to encourage the spontaneous application of problem-solving to unfamiliar mathematics tasks. Only a few children in EMR schools were also capable of solving these tasks in a similar manner. Far generalization thus not only depends on the programme but also on the cognitive capacity of the children. This finding is consistent with findings of Rickard et al. (1994) and Van Luit (1994a).

Conclusions

The results of the study suggest that the SRVV mathematics training programme can be effectively employed in teaching problem-solving strategies in multiplication and division to children in EMR and LD schools who are poor at mathematics. The findings are consistent with other research results, e.g. Leon and Pepe (1983), Naglieri and Gottling (1995), Van Luit (1989), and Whitman (1987). The effectiveness of this training can be explained in terms of successful integration of parts of the mathematics curriculum, thinking about different ways of problem-solving and self-instruction. The children in the experimental groups perform significantly better on mathematics tasks compared with the children of the control groups. They became capable of thinking about the best possible problem-solving strategy for a specific mathematics task. These results suggest that the children in LD and EMR schools benefited from the specific training programme (Van Luit et al., 1993). However, only the children in LD schools are capable of making far generalizations adequately after the experimental training.

These results, like those of Perry (1991), suggest that children in LD schools may improve in mathematics if they are provided with the underlying principles and given the opportunity to generate their own procedures for solving mathematical problems, and that this may result in generalization of knowledge. In this experiment, generalization was possible for most of the children in the LD schools but not for the mentally disabled children. Geary (1994), noting the complexity of the difficulties faced by these children, remarks that the most likely consequences of these difficulties is the lack of transfer. Day and Hall (1988) and Campione, Brown, and Ferrara (1982) also note that EMR children only show modest near transfer if the transfer training is of sufficient duration. Campione, Brown, Ferrara, Jones, and Steinberg (1985) suggest that children with EMR need considerable help before they can understand and solve near transfer problems. One important reason for the lack of transfer is that children who show transfer are also good at strategy use (Pressley, Snyder, & Cariglia-Bull, 1987). The literature about children with EMR show that these children are poor at strategy use; it is therefore understandable that we do not find transfer in children with EMR.

We conclude that the SRVV programme of strategy instruction, applied in the teaching of certain mathematics domains, may have positive influences on the achievement of children with special needs because adequate problem-solving strategies were facilitated. In general, the results of this study suggest that the programme may provide a method for regular and remedial teachers to help children with disabilities improve their 'maths thinking' performance. Further research with programmes for other domains that include self-instruction of strategy use should be considered, especially for children with severe mathematics disabilities.

References

Ashcraft, M.H. (1987). Children's knowledge of simple mathematics: A developmental model and simulation. In J. Bisanz, C.J. Brainerd, & R. Kail (Eds.), *Formal methods in developmental psychology* (pp. 302-338). New York: Springer-Verlag.

Ashcraft, M.H. (1992). Cognitive mathematics: A review of data and theory. *Cognition, 44*, 75-106.

Baron, J.B., & Sternberg, R.J. (Eds.) (1987). *Teaching thinking skills: Theory and practice.* New York: Freeman.

Brooks, L.W., & Dansereau, D.F. (1987). Transfer of information: An instructional perspective. In S.M. Cormier & J.D. Hagman (Eds.), *Transfer of learning: Contemporary research and applications* (pp. 121-151). San Diego: Academic.

Butterfield, E.C., & Nelson, G.D. (1989). Theory and practice of teaching for transfer. *Educational Technological Research and Development, 37*, 5-38.

Campbell, J.I.D. (1987). Production, verification, and priming of multiplication facts. *Memory & Cognition, 15*, 348-364.

Campbell, J.I.D., & Graham, D.J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Canadian Journal of Psychology, 39*, 338-366.

Campione, J.C., & Brown, A.L. (1990). Guided learning and transfer: Implications for approaches to assessment. In N. Frederiksen, R. Glaser, A. Lesgold, & M.G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 141-172). Hillsdale, NJ: Erlbaum.

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). Cambridge, MA: Cambridge University Press.

Campione, J.C., Brown, A.L., Ferrara, R.A., Jones, R.S., & Steinberg, E. (1985). Breakdowns in flexible use of information: Intelligence-related differences in transfer following equivalent learning performance. *Intelligence, 9*, 297-315.

Clark, J.M., & Campbell, J.I.D. (1991). Integrated versus modular theories of number skills and acalculia. *Brain and Cognition, 17*, 204-239.

Coles, M.J., & Robinson, W.D. (1991). Teaching thinking: What is it? Is it possible? In M.J. Coles & W.D. Robinson (Eds.), *Teaching thinking* (pp. 9-24). London: Bristol Classical Press.

Day, J.D., & Hall, L.K. (1988). Intelligence-related differences in learning and transfer and enhancement of transfer among mentally retarded persons. *American Journal on Mental Retardation, 93*, 125-137.

Fayol, M. (1990). *L'enfant et le nombre* [The child and the number]. Neuchatel: Delachaux & Niestle.

Geary, D.C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology, 49*, 363-383.

Geary, D.C. (1994). *Childrens' mathematical development. Research and practical applications.* Washington, DC: APA.

Geary, D.C., Brown, S.C., & Samaranayake, V.A. (1991). Cognitive addition: A short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Developmental Psychology, 27*, 787-797.

Halpern, D.F. (1992). A cognitive approach to improving thinking skills in the sciences and mathematics. In D.F. Halpern (Ed.), *Enhancing thinking skills in the sciences and mathematics* (pp. 1-14). Hillsdale, NJ: Erlbaum.

Hamers, J.H.M. & Overtoom, M.T. (Eds.) (1997). *Teaching thinking in Europe.* Utrecht: Sardes.

Hasemann, K. (1994). Exploration of individual learning processes in mathematics. In J.E.H. Van Luit (Ed.), *Research on learning and instruction of mathematics in kindergarten and primary school* (pp. 181-200). Doetinchem: Graviant.

Lemaire, P., Barrett, S.E., Fayol, M., & Abdi, H. (1994). Automatic activation of addition and multiplication facts in elementary school children. *Journal of Experimental Child Psychology, 57*, 224-258.

Leon, J.A., & Pepe, H.J. (1983). Self-instructional training: Cognitive behaviour modification for remediating arithmetic deficits. *Exceptional Children, 50*, 54-60.

McCloskey, M., Harley, W., & Sokol, S.M. (1991). Models of arithmetic fact retrieval: An evaluation in light of findings from normal and brain-damaged subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 377-397

Meichenbaum, D. (1977). *Cognitive-behavior modification: An integrative approach.* New York: Plenum.

Miller, K.F., & Paraedes, D.R. (1990). Starting to add worse: Effects of learning to multiply on childrens' addition. *Cognition, 37*, 213-242.

Naglieri, J.A., & Gottling, S.H. (1995). A cognitive education approach to math instruction for the learning disabled: An individual study. *Psychological Reports, 76*, 1343-1354.

Nisbet, J. (1991). Projects, theories and methods: The international scene. In M.J. Coles & W.D. Robinson (Eds.), *Teaching thinking* (pp. 25-34). London: Bristol Classical Press.

Perry, M. (1991). Learning and transfer: Instructional conditions and conceptual change. *Cognitive Development, 6*, 449-468.

Pressley, M., Snyder, B.L., & Cariglia-Bull, T. (1987). How can good strategy use be taught to children? Evaluation of six alternative approaches. In S.M. Cormier & J.D. Hagman (Eds.), *Transfer of learning. Contemporary research and applications* (pp. 81-120). San Diego: Academic.

Rickard, T.C., Healy, A.F., & Bourne, L.E. (1994). On the cognitive structure of basic arithmetic skills: Operation, order, and symbol transfer effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1139-1153.

Schoenfeld, A.H. (1989). Teaching mathematical thinking and problem solving. In L.B. Resnick & B.L. Klopfer (Eds.), *Toward the thinking curriculum. Current cognitive research.* Washington, DC: ASCD.

Schoenfeld, A.H. (1991). What's all the fuss about mathematical problem solving? In A.L.Costa (Ed.), *Developing minds* (pp. 144-146). Alexandria, VA: ASCD.

Schoenfeld, A.H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-370). New York: MacMillan.

Siegler, R.S. (1988). Strategy choice procedures and the development of multiplication skills. *Journal of Experimental Psychology: General, 117*, 258-275.

Siegler, R.S., & Jenkins, E.A. (1989). *How children discover new strategies.* Hillsdale, NJ: Erlbaum.

Skemp, R.R. (1989). *Mathematics in primary school.* London: Routledge.

Treffers, A. (1987). *Three dimensions. A model of goal and theory description in primary mathematics education: Wiskobas.* Dordrecht: Reidel.

Van Luit, J.E.H. (1989). Self-instructional training for remediating arithmetic deficits in EMR- and LD-children. In J.J. Dumont & H. Nakken (Eds.) *Learning disabilities. Vol. 2: Cognitive, social and remedial aspects* (pp. 139-148). Amsterdam: Swets & Zeitlinger.

Van Luit, J.E.H. (1994a). The effectiveness of structural and realistic arithmetic curricula in children with special needs. *European Journal of Special Needs Education, 9*, 16-26.

Van Luit, J.E.H. (1994b). A method to diagnose arithmetic disabilities. In J.E.H. Van Luit (Ed.), *Research on learning and instruction of mathematics in kindergarten and primary school* (pp 35-41). Doetinchem: Graviant.

Van Luit, J.E.H., Kaskens, J., & Van de Krol, R. (1993). *Speciaal Rekenhulpprogramma Vermenigvuldigen en Verdelen* [Special Mathematics Programme for Multiplication and Division]. Doetinchem: Graviant.

Whitman, T.L. (1987). Self-instruction, individual differences, and mental retardation. *American Journal of Mental Deficiency, 92*, 213-223.

Wilkinson, L.C., Martino, A., & Camilli, G. (1994). Groups that work: Social factors in elementary students' mathematics problem solving. In J.E.H. Van Luit (Ed.), *Research on learning and instruction of mathematics in kindergarten and primary school* (pp. 75-104). Doetinchem: Graviant.

11

The use of connectives in different textual genres: A developmental analysis in a 'thinking skills' perspective

L. Chanquoy

Introduction

This research[1] aims at carefully analysing children's writing skills while dealing with different types of text. To reach this objective, some specific surface marks, the connectives, are investigated in different genres of text written by 10 and 13 year olds. As it seems impossible to study all the writing activity, it is necessary to make some choices. In this study of writing in different genres, the analysis of connectives seems to fit our objective, that is to try to answer the following question: Do children manage differently their writing, as a function of the type of text they are asked to write down? The study of connectives could provide important information about writing skills and strategies that children as young as 10 years old are able to use.

1

A part of this work has been presented in Nijmegen (The Netherlands) for the 6th European Conference for Research on Learning and Instruction (August, 1995). The author would like to thank Pierre Coirier (Université de Poitiers, France) and Gert Rijlaarsdam (University of Amsterdam, The Netherlands) for their helpful comments on a previous version of this paper

To study these writing skills, connectives have been considered as appropriate surface indicators of deep writing processes. Indeed, connectives are a good way to analyse how children convert the multiple dimensions of their thought into one linear dimension necessary for writing. In addition, this could provide information about how children acquire the ability of 'juggling' (Flower & Hayes, 1980) with the very complex processes of writing and how these cognitive processes develop and become automatized. In order to do so, thirty six children successively wrote a narrative, a description and an argumentation concerning similar topics. An analysis of interclausal connectives was led to study the children's ability to use specific connectives according to specific genres. Writing on different genres or different topics is not considered to require the same amount of skills (Chanquoy, 1991). Indeed, writing an argumentative text is always considered to be more difficult than writing a narrative, even for expert writers. These different degrees of difficulty, if they have an intrinsic value, can be analysed via certain particular surface textual marks. Connectives seem to be one of the more appropriate means to study children's developmental abilities to produce different types of text (Costermans & Fayol, 1997).

Before presenting this experiment and its main results, the first part of this chapter focuses on the writing process and the acquisition of connectives, and their relationships with more general thinking skills. Indeed, writing requires a number of specific thinking skills and we postulate that the use of connectives can be viewed as one of them. Then in the last part of this chapter, the results of the experiment are discussed in relation with thinking and writing skills. Thus, the experiment presented here has an indirect educational perspective. It can be viewed as an empirical study aimed at providing concrete results for further educational purposes into connectives and textual genres.

Theoretical framework

General presentation

Writing is difficult and requires complex cognitive work. The subject must find and organize information from long term memory and context in order to elaborate a text. Language, and particularly writing skills, can be viewed as a part of a more global set of general thinking skills and considered as a part of problem solving abilities (Parrila, 1995). Writing is considered as subordinate to learning and thinking (Boscolo, 1995; Galbraith, 1992), and a great number of researchers now see writing as a major tool for encouraging the development of thinking skills (Hartley, 1991; Kellogg, 1993). However, the relationship between thought and language is debatable because it remains difficult to clearly define these two concepts (Gould, 1980). Sternberg (1987) distinguishes three types of thinking skills processes: (a) executory processes, (b) non-executory task processes, and (c) non-executory learning processes. Boekaerts and Simons (in Hamers & Overtoom,

1997) find globally the same types of thinking skills and emphasize on the importance of metacognition, too.

Writing is viewed as part of an executory process, since it concerns planning and translating ideas into a text, evaluating the whole text, and monitoring these subprocesses (see Hayes (1996) or Kellogg (1996) writing models). Concerning education, writing is viewed as a non-executory learning and a non-executory task process.

Writing is different from speaking (Levelt, 1989); it needs a very formal learning at school and the acquisition of specific knowledge. Speaking can be considered as 'a natural behaviour', while writing is highly artificial and must be learned at school. Indeed, writing can be considered as "one of the most complex skills taught in school" (Zimmerman & Risemberg, 1997, p. 97). In addition, writing cannot be considered as a simple transformation of an oral code into a written code. Writing is often described as a very complex activity that is difficult to study.

The main problem of researching writing is that it is quite impossible to globally analyse writing activities. Thus, it seems necessary to focus on one aspect of writing. Generally, research on writing is divided into two main fields, depending on different methodologies: research using 'off-line' methods and those using 'on-line' methods. Both are, of course, productive and interesting. The first method focuses on text, that is, on the written products, and the second method analyses writers' behaviours, through more complex methodologies (for a review, Kowal & O'Connell, 1987). On-line methods are more difficult to use with children, because they are difficult and/or very unnatural. In this work, an off-line method has been used to analyse children's writing.

The writing process

During these past years, several conceptions of language production have been elaborated. Some works, that analyse more particularly oral language are based on linguistic or psycholinguistic theories (Levelt, 1989). Other works, that study written language, have a more psychological and/or educational approach, and have provided information on written processes (Bereiter & Scardamalia, 1987; Scardamalia & Bereiter, 1987). At the same time, the emergence of studies managed in real time (on-line studies), and centered on the writer's behaviour has led to real progress in the knowledge of writing processes and to the generation of models (Hayes, 1996; Hayes & Flower, 1980; Kellogg, 1996).

Some years before these models, writing was conceptualized in terms of three successive stages: (a) a prewriting stage, with planning activities, (b) a writing stage, during which the writer is composing a draft, and (c) a rewriting stage which involves editing and revising the draft. Thus writing was considered as a linear sequence of stages. However, more recent observations of the process of writing, using on-line methods, like thinking-aloud protocols, have shown that the writer seldom proceeds sequentially but recursively. The famous model elaborated by Hayes and Flower (1980) shows writing as composed of three major components: (a) the task environment (the rhetorical problem, the text produced so far, the writing tools and the external sources of information available while writing), (b) the writer's long term memory (involving the writer's knowledge about the topic, the audience and the possible plans), and (c) the writing process. This process includes three sub-processes, controlled by a cognitive monitor:

- A planning process with two main steps: the generation of relevant information from long term memory, and the organization of the retrieved ideas, in relation with the goals of the production.
- A translating process, necessary to translate the ideas into text, combining two components: lexicalization and linearization (Fayol & Schneuwly, 1988). The lexicalization is the selection, in linguistic paradigms, of lexical items, corresponding to the retrieved information; the linearization is the translation of ideas into syntactically correct sequences.
- A reviewing process that allows the writer to evaluate and revise the text produced so far.

Thus, during writing, when the relevant ideas have been retrieved, they must be organized according to the task constraints, in order to translate them into a text. At this level, differences between children and adults, and/or between novices and experts in writing, are obvious. According to Scardamalia and Bereiter (1987), there are two main ways to organize information, depending on the writer's expertise (or development) level:

- The first, 'knowledge telling strategy', which characterises children who would directly formulate the retrieved information, in order to produce utterances. This strategy is a direct transformation of ideas into words, managed step by step.
- The second, 'knowledge transforming strategy', which only appears in adults or experienced writers, consists in building and elaborating information, as a function of the topic, and as it relates to the reader.

According to these models of writing, it seems that "novice writers write without thinking, while experts write as problem-solver" (Boscolo, 1995, p. 353).

In conclusion, a comparison between thinking skills and writing can be proposed. Baron and Sternberg (1987) divide thinking skills into three main processes: (a) executive processes aimed at planning, monitoring and evaluating thinking; (b) performance processes involved in carrying out thinking, and (c) learning processes that help to learn how to think. These three thinking processes can easily be combined with writing processes: executive thinking processes could correspond to the planning process as previously defined (Hayes & Flower, 1980); the revising process could be viewed as a part of performance thinking processes: it needs the writer to act as a critical reader of his/her own text; finally, learning processes are intrinsically combined with all the writing processes: in learning by writing, and in writing to learn how to write better, children discover how to write, how to improve their writing both while writing and by taking into account teacher's remarks.

Different textual genres

Very early on, children appear able, in conversation, to describe an object or a situation, to tell a story, or to convince somebody to accept a demand. However, the writing of a complex text confronts the subject with considerable difficulties, and the control of different genres of text, such as description, narrative, or argumentation, appears much later in their development (Pappas, 1993). Although subjects demonstrate a relative facility in the production of narratives at ten years of age (Fayol, 1985), argumentation remains difficult even for older subjects (Coirier & Golder, 1993). In addition, the writer must take into account several parameters: the purpose, the context of the production (the time, the place and the eventual reader), and the theme of her/his production. The consideration of these parameters supports the management of different operations on two levels: (a) a pre-linguistic level, where the writer elaborates a representation of the reference theme (macro-structural planning) and a hierarchical organization of ideas from which the final structure of the text will result (super-structural planning; see Levelt, 1989), and (b) a linguistic level of lexicalization and linearization (micro-structural organization (Fayol & Schneuwly, 1988; Schneuwly, 1988; Schneuwly, Rosat, & Dolz, 1989). These operations are notably expressed, in the textual structure, by the appearance of modalization, cohesion and connection-segmentation markers (De Weck & Schneuwly, 1994; Hedberg & Fink, 1996).

These different management operations present specific difficulties depending on textual genres. The analysis of discursive models consists in finding specific morphological organizations for each textual genre (Adam, 1992). The text appears as an organization of interconnected clauses in sequences of different linking modes. These sequences and modes are characterized by types of textual configuration. A comparison between narrative and description is interesting because they are opposite in terms of organization of the reference area: narrative is a sequential representation of a temporal series of organized events, whereas description is a spatial representation of a set of elements (Chanquoy & Fayol, 1995).

The acquisition of the narrative is today well-understood, thanks notably to Fayol's (1985) work. Researcher has confirmed the hypothesis of a narrative schema, that guides the management of narrative structures (Mandler & Johnson, 1977; Stein & Glenn, 1982). This has led to postulate the existence, at a cognitive level, of a narrative 'model' with limited abstract categories (setting, initial event, goal, path-to-the-goal, result, end), that are linked together by precise arrangement rules (Kintsch & Van Dijk, 1978). The difficulty for the subject consists in the passage from this cognitive-semantic representation to the text. Fayol (1985) demonstrated the double evolution of subjects' skills. Concerning content organization, the youngest children were just able to juxtapose independent events (texts written as 'news announcements'), while the oldest children were able to elaborate one or more episodes of events organized according to the narrative schema. Fayol also observed a progressive control of the intra-textual functioning: a decrease in the marks usually found in oral conversations and an increase in the management of the temporal aspects of the different episodes and their articulation. Thus, the progressive control of narrative operations appears to be linked to, and facilitated by the presence of a pre-linguistic representation as the schema. Writing a narrative text seems to depend on the ability to transfer a mental organisation into a linear text. This is a highly complex skill, progressively acquired.

Description, though much less studied, does not seem to present a framework comparable with the narrative schema; it is a matter of linearly developing a set of clauses whose organization is originally neither causal, nor chronological (Adam, 1992). The subject seems to possess two types of strategy to write a descriptive text: (a) to produce an enumeration, from the spatial contiguity principle of elements, by using enumerative organizers characterized by juxtaposition, or (b) to construct a text plan by borrowing a specific order from spatial or temporal systems (Chanquoy & Costeplane, 1995).

Although the acquisition of textuality appears facilitated in the case of narratives by the presence of a cognitive representation, according to the narrative schema, it seems, in the case of descriptions, to depend on the subject's textual and linguistic abilities, and on a familiarity, notably attained at school, with more or less conventional descriptive models. It can be considered that descriptive models seem more 'intuitive' than narrative ones; in addition, at school, children cope much better with narratives than with descriptions.

In contrast, argumentation presents specific difficulties (Coirier, 1996) in which the writer must construct a discourse designed to modify the reader's representation on a given theme (Coirier & Golder, 1993). Two essential processes underlie the production of argumentations: a supporting process that bases an assertion on one or several reasons (Coirier & Golder, 1993; Golder, 1992), and a negotiation process that has as its purpose to bring the reader accepting these reasons (Coirier, 1996). The fundamental structure used in the argumentative discourse is the relationship 'stated position/foundation of this position'. By contrast with narratives, there are not many works dealing with an eventual 'argumentative schema' that could be used while writing an argumentation. Nevertheless, argumentative texts, collected by Schneuwly (1988) from 10 to 14 year olds, underline the child's belated capacity to write this kind of text and to take into account a viewpoint different from his/her own. The acquisition of control of argumentations seems, from the studies reviewed, never completely mastered, even by adults.

Connectives

During learning to write, children must learn how to manage a system of linguistic units, sometimes very different from oral (Perera, 1986). Among these units, connectives seem to be particularly interesting. Indeed, the use of connectives in texts could provide information about children's progressive mastery of their thinking abilities, via writing. Studying textual connectives could help to better understand writing processes and skills that children use while writing about different genres of text. Connectives are words or short phrases that join clauses or simple sentences (Millis, Graesser, & Haberlandt, 1993).

The production of the different types of text, previously presented, seems facilitated by the presence of an a priori logically organized representation, and seems hindered when the elaboration of the textual structure depends only on the logical, linguistic and textual capacities of the subject, who must manage all these operations (Chanquoy & Coirier, 1997). The analysis of variations in connectives, considered as traces of planning and textualization, may provide information on these operations (Chanquoy, 1996; Chanquoy & Coirier, 1997). These marks indicate, in textual surface, the degree and nature of interclausal connections between ideas or groups of ideas referring to states or events belonging to the 'mental model' (Chanquoy, 1991; Chanquoy & Fayol, 1995). Indeed, during writing, it is necessary to 'linearize' ideas, that rarely have a sequential organization: relationships between elements separated in the textual surface must be explicit, and, in the same way, the absence of relationships between two close elements must be indicated (Costermans & Fayol, 1997). The writer may thus have to link elements distanced in the cognitive representation, or, on the contrary, to separate conceptually unconnected elements (Fayol & Schneuwly, 1988). Connectives may be considered as clues to these operations. Their essential function is to provide textual cohesion and thematic progression (Shapiro & Hudson, 1997). Schneuwly (1988) and Schneuwly et al. (1989) considered them as indicators, in the textual surface, of three main operations (the last two from a micro-structural level):

- Beaconing operations, from which text structuring results. Macro-structural beaconing operations correspond to sequential organization procedures of the reference representation; super-structural beaconing operations correspond to the marking of phases in the text plan, and to taking into account of production parameters.
- Packaging operations that preferentially use coordinating and subordinating conjunctions to connect clausal units (syntactic level).
- Linkage operations that provide the continuity of the enunciative undertaking, and that especially appear in children's oral productions.

Thus, writing a text to describe, relate or argue mobilizes different logical, linguistic, and textual operations. One of the preferential indicators in the textual surface is the system of connectives (Chanquoy & Costeplane, 1995), which indicates two types of operation:

- At an interclausal or micro-structural level, they provide the continuity of the statement, or clearly indicate the nature of the relationship between ideas of the mental reference representation.

- At a textual, macro- or super-structural level, they may either contribute to the cohesion of articulated clausal sequences, by indicating, once again, the nature of this relationship, or they may serve to mark the different phases of the text plan (Chanquoy & Costeplane, 1995).

Therefore, the analysis of variations in connectives, considered as traces of planning and translating (see Hayes & Flower's model, 1980), can provide information on the writing processes involved in different types of text. Children must progressively learn how to correctly use these interclausal marks, which are already known and used, or at least partly, in oral productions. This leads both to a new apprenticeship and to the adjustment of an already existing system, so as to use it differently and efficiently in writing (Chanquoy, 1991).

In conclusion, the different writing processes above mentioned are not easy to analyse: most of them take place in the subject's mind, and are therefore difficult to identify and to clarify. Indeed, the main difficulties found in analysing writing are due to the fact that the most part of the writer's behaviour is not directly observable. Only the finished product (the text) can be easily studied. Consequently, it is important to study written productions, as the main manifestation of cognitive and linguistic activities.

The main goals of the research reported here, which was focused on developmental thinking skills related to writing abilities in the intermediate grades, were threefold.

Firstly, to determine which developmental skills might contribute to writing acquisition, we analysed the development of management capacities in written production operations and their progressive automatization. More accurate thinking skills, and an increase in the number of available connectives should lead to a reorganization of the connective system. We expected a decrease in the connectives maintaining the enunciative framework (e.g., mainly subordinating conjunctions such as: 'and', 'then', 'after', Chanquoy & Costeplane, 1995), and an increase in the connectives demonstrating a greater power of planning in writing. At the same time, texts written by older children would be longer than those written by the youngest ones, demonstrating a greater management of the planning and translating processes.

Secondly, to determine whether these developmental skills may contribute to the use of specific connectives depending on textual genres, three types of text were written by children: a description, a narrative, and an argumentation. From a typological point of view, we expected important differences in the nature and the frequency of connectives depending on these types of text. Many studies have shown that these parameters are determined by the textual structure in which connectives are found. In the case of narrative, where the referent is chronologically or causally organized, subjects seem to benefit from a cognitive representation (i.e., the canonical schema) that leads to the appearance, in the textual surface, of specific markers based on the relationships between cognitive categories (Fayol, 1985). Descriptions have no such cognitive framework and the spatially (or not) organized referent may lead to different descriptive routes (Adam, 1992). To elaborate a descriptive configuration, the writer must alone manage the sequential organization of the text. In argumentation, where the textual framework is entirely constructed and managed by the writer, the referent corresponds to the taking into account and the integration of different opinions. Children may thus use very different connectives depending on text types:

- In narrative, connectives specific to the different episodic moments would appear (for example, 'but' or 'suddenly' could be used, when an obstacle rises), while juxtapositions of 'news announcements' (linked by 'and') would decrease (Fayol, 1985).
- In description, connectives linking organized descriptive sequences (for example: foreground and background) would appear, while enumeration markers (and, then, after) would decrease (Adam, 1992).
- In argumentation, specific logical connectives linking hierarchical argumentative sequences would appear (first, while, therefore), while enumeration markers would decrease (Chanquoy, 1996; Chanquoy & Coirier, 1997).

Writers may use very differentiated logical connectives in argumentation, while using temporal connectives more frequently in narratives, and spatial connectives in descriptions. In addition, the complex management of the different operations for descriptions and argumentations would lead to more connectives than in narratives.

Thirdly, to precisely determine which skills developed, we analysed the variations in textual structuring, by studying the proportion of linearity and textual structuralization marks among the connectives used. The precise analysis of interclausal connectives, considered as surface cues of mental processes, may provide information about the different skills used by children while dealing with different textual genres and about the development of these skills.

Method

Participants

Eighteen fifth-graders (9 girls and 9 boys; mean age: 10.8 yr, ranging from 10.4 to 11.1), and eighteen 8th-graders (11 girls and 7 boys; mean age: 13.6 yr, ranging from 12.9 to 14) took part in the experiment. All children were French native speakers and came from a village school near Montpellier.

Materials

For the description, a picture depicting a scene was presented to the children: a character in a boat alongside a river bank on which a house stood surrounded by trees. This picture had a high number of details. For the narrative, a comic strip composed of six pictures was presented. The strip corresponded to the canonical form of the narrative schema (framework: a child is playing on the beach; initial event: a dog appears; goal: the child approaches the dog to play; path-to-the-goal: game between dog and child, the castle is destroyed; result: the child chases the dog away; end: the child, now alone again, is crying). For the argumentation, a picture depicting a child at the seaside was presented. The child can see a fishing boat and a sailing boat in the distance. The scene thus refers to the alternatives proposed to the children during the writing of this text.

Procedure

All subjects participated in the three text sessions, with each session separated by an interval of three weeks. The text sessions were always presented in the same order: description, narrative, and argumentation. Results from a pre-test session for all genres of text showed the necessity to train children in argumentative writing. The argumentation session was thus preceded by a preparatory session for all children, in which different argumentative topics were orally discussed in the classroom.

During the experimental sessions, each subject received two sheets of paper: one with the picture associated to the required text, the other describing the instructions and limiting the production to approximately twenty lines. Then, children were asked to read the instructions closely, to ask questions if necessary, and to write down a text about one of the proposed topics. This was done three times. Each time, the instructions asked the children to carefully look at the picture(s) and to describe what they saw, to tell the story for or to argue to convince, one friend who was not able to see the picture(s).

Data collection

Two dependent variables were analysed: the textual length, estimated in inter-clauses (i.e., the number of clauses - 1), considered to give a good insight into the cognitive difficulty to produce in one genre (Levy & Ransdell, 1996), and the mean proportion of interclausal connectives (i.e., the total number of interclausal connectives divided by the total number of interclauses). These data were analysed in a 2x3 analysis of variance: (A: children's age (or school level) with 2 modalities [10 y.o. (5th grade) and 13 y.o. (8th grade)]) x (T: type of text with 3 modalities:

Table 1 Mean Length of the Different Types of Text, According to the Two Grades

Text Type	School Level		
	5th Grade	8th Grade	Mean
Description	11.4	24.5	18.0
Narrative	11.8	19.9	15.9
Argumentation	21.4	23.1	22.2
Mean	14.9	22.5	

description, narrative, and argumentation), with repeated measures on the last factor.

Results

Textual length

The mean length for each type of text and for each age is presented in Table 1. Ten year-old children coarsely wrote texts significantly shorter than the 13 year-olds (14.88 vs 22.50; $F(1,34)=46.478$, $p<.001$). More surprisingly, the length significantly varied according to the types of text. The narratives (15.9) were shorter than the descriptions (18.0), which were in turn, shorter than the argumentations (22.2; $F(2,68)= 21.866$, $p<.001$). The significant interaction between ages and types of text ($F(2,68)=17.012$, $p<.001$) brought to light global differences. Indeed, for the younger children, descriptions and narratives had similar lengths (respectively: 11.4 and 11.8), whereas argumentations were significantly longer (21.4; $F(2,34)= 33.014$, $p<.001$). Training children on a new genre had led to an increase of the length for argumentative texts.

Conversely, in the 8th-grade, the differences were significant but less marked: the 8th-graders wrote descriptions (24.5) and argumentations (23.1) that were significantly longer than narratives (19.9; $F(2,34)=5.638$, $p<.01$). Table 1 emphasizes the particular case of argumentation. Although the descriptive and narrative texts produced by the older children were approximately twice as long as those produced by the younger group, the mean length of the argumentative texts was relatively stable. An a posteriori comparison opposing argumentative texts to descriptive and narrative texts was, in this respect, significant ($F(1,68)=39.134$, $p<.001$).

Interclausal connectives

The mean proportions of connectives, according to the school levels and the text types are shown in Table 2.

Both 5th- and 8th-graders used similar proportions of connectives (.55 and .52;

Table 2 Mean Proportion of Connectives, in the Three Types of Text, According to the Two Grades

Text Type	School Level		
	5th Grade	8th Grade	Mean
Description	.61	.51	.56
Narrative	.40	.43	.42
Argumentation	.64	.63	.63
Mean	.55	.52	

$F(1,34) < 1$, NS). However, the proportion of interclausal connectives significantly varied according to textual genres ($F(2,68)=15.375$, $p<.001$): Connectives were more numerous in argumentations (.63) and in descriptions (.56) than in narratives (.42). The two factors did not interact ($F(2,68)=1.475$, NS), this distribution was the same whatever the age.

Results showed that, not surprisingly, the 8th-graders produced longer texts than the 5th-graders, but did not use more connectives. The stability of connectives did not entirely fit our predictions, but was consistent with previous research results with younger children (see Chanquoy, 1991). Two hypotheses may be advanced to explain this phenomenon. Instead of assuming a 'drop of performance' in the 8th-graders, it seems that a better control of writing skills (i.e., planning and text generating) led them to use less connection markers. Step-by-step management of written production (i.e., the use of knowledge telling strategy), requiring frequent marking of the nature of relationships among clauses or sentences, would be progressively replaced by a more rigorous and planned management (i.e., the use of knowledge transforming strategy), that could allow to organize the text in global units (Scardamalia & Bereiter, 1987).

The acquisition of more per formant writing abilities led to the use of more specific modes of linkage, more intra-clausal clues (for example, different verbal tenses, anaphoras, etc.), and therefore fewer interclausal connectives to indicate the textual progression. This evolution is based on a double process: (a) the appearance of specific connectives and a change of the mode of text functioning of already used connectives, and (b) a more accurate 'perception' of writing that led children to progressively manage globally their production, while developing more efficient writing skills. Another explanation, that was not tested here, is the interference of a relatively classic rule, confirmed by studies about the system of punctuation and connectives, according to which more punctuation leads to fewer connectives (see Chanquoy, 1991). If this hypothesis is correct, the stability of connective proportions could be due to the acquisition of a more extensive and more specialized punctuation system expressing a better structuring of the text into specific thematic units. Only a comparative study of the acquisition and functioning of these two systems would allow an answer to this question.

The results about the different genres of text were also surprising. They showed that the shortest texts were narratives, followed by descriptions, and then argumentations. The proportion of connectives followed the same progression. The hypothesis predicted that the production of a narrative would be less difficult because of the presence of a precociously acquired narrative schema. This schema would facilitate writing by decreasing the cognitive load associated with the writing process. Conversely, the absence of such a schema for descriptive and argumentative texts, and the need for structural and thematic management of these texts, led to the expectation of shorter texts. But the results showed that narratives were the shorter texts, with the weakest proportion of interclausal marks.

Among the factors that may have influenced these results, pictorial characteristics occupy an important position. For the description, the wish to spare children from several descriptive paths, ending in different textual configurations, led to choosing a picture with several richly detailed potential perspectives. The precise examination of the texts revealed that the majority of children engaged in a quasi-exhaustive enumeration of these details. This may have resulted in a bias, leading both to a lengthening of the descriptions and to the use of numerous connectives to link the different successively enumerated elements. In the case of argumentation, the freedom in the choice of arguments was enriched by the proposals emitted in the preparatory session. Each child was thus able to call on a large number of arguments, from different sources, that she/he had to articulate with connectives. By contrast, in the case of narratives, the comic strip comprised few details and emphasized successive event episodes. It seems that the presence of the narrative schema may have had a restrictive effect: the children went to the main plot of the story and used connectives only to link the most important episodes. In the second part of this study, we will try to answer some of the questions raised by these general results. Two ways of textual structuring have been analysed.

Analysis of linearity and textual structuration connectives

To explain variations in the frequency of interclausal connectives, the texts were analysed by classifying each connective according to its function in the textual structuring. Two broad types of connectives were chosen: linearity markers, and macro- and/or super-structural markers, called textual structuring markers (Schneuwly, 1988). The linearity connectives made obvious an 'oral mode' of writing, with a step-by-step management of writing, using close interclausal links, whereas the macro- or super-structural connectives reflect a 'written mode' of writing, with a more global management, using textual (or macro-) connections, instead of micro-connections. They thus represent two ways of thinking about writing. The mean proportions of these two categories of connectives were estimated by computing the number of interclausal connectives divided by the number of interclauses. Two analyses of variance were performed, using the same experimental design as before.

The 5th-graders tended to use more linearity markers than the older children

Table 3 Mean Proportions of Linearity (Lin.) and Textual Structuration (T.S.) Connectives

	5th Grade		8th Grade	
	Lin.	T.S.	Lin.	T.S.
Description	.60	.01	.45	.06
Narrative	.33	.07	.35	.08
Argumentation	.51	.13	.42	.21
Mean Proportion	.48	.07	.41	.12

(.48 vs .41; $F(1,34)=3.171$, $p<.09$). Descriptive texts had more linearity markers than, respectively, argumentative and narrative texts (.52, .46 vs .34; $F(2,68)= 12.994$, $p<.001$). The interaction between age and type of text tended toward significance ($F(2,68)=2.753$, $p<.07$); it indicated that the linearity connectives decreased with age for argumentative and descriptive texts, and was constant for narratives. In addition, it showed that textual differences were only significant for the younger children (see Table 3). A complementary analysis of variance was performed for textual structuring markers and revealed that 8th-graders' texts contained significantly more textual structuring markers than the texts produced by the 5th-graders (.07 vs .12; $F(1,34)= 6.561$, $p<.02$). Argumentative texts (.17) exhibited more structuring connectives than narrative (.08) and descriptive texts (.04; $F(2,68)= 31.864$, $p<.001$). The two factors did not interact ($F(2,68)= 1.529$, NS).

The results showed that 13-year-olds produced texts in which the proportion of linearity markers, indicating the linearization process, was slightly lower than in texts produced by 10-year-olds. This effect was mainly due to differences between descriptions and argumentations. Conversely, the proportion of structuring markers, indicators of a more powerful planning, was greater in all the texts written by 8th-graders. These results confirmed the hypothesis that variations in connective proportions would be accompanied by an internal restructuration in their use. Here, the overall stability of connective proportions seemed to cover two phenomena: (a) a slight inferiority of linearity connectives, and (b), in a complementary manner, a more marked superiority of structuring connectives in 8th-graders' texts. The step-by-step management of written production, typical in younger children, seemed to be replaced during the 8th grade by a more controlled management in which the planning process became more important and more powerful.

Theoretically, this process would be based on two types of operations that the writer must accomplish: (a) to sequentially elaborate thematic units of information during the activation and organization of the textual representation in memory (macro-structure), and (b) to organize these units according to a plan of the final text (super-structure). The writer is thus able to use available connectives to indicate beaconing operations in textual surface. However, the mean proportion of structuring connectives was relatively weak whereas the proportion of linearity connectives was still very high, even in older children's texts. Twenty six out of the 54 texts produced by 5th-graders did not contain any structuring connectives (67%), whereas the proportion in 8th-graders was only 15 out of 54 texts (29%). Thus for example, 5th-graders' narratives were often limited to enumerations or juxtapositions of states or events, connected by coordinating conjunctions (for example: and, then or after). Six out of the 18 5th-graders' narratives corresponded to 'news announcements' (33%), and the remaining 12 had a canonical framework (67%). For 8th-graders, only three texts (17%) were considered as 'news announcements', the other texts were well-structured, and followed the narrative schema. The younger children seemed to build their texts mainly to ensure enunciative continuity (as they did orally), provided by linkage operations (such as the connective 'and'), and were less prone to use textual and thematic continuity for indicating the nature of relationships that link their clauses. The textual framework was limited, in most cases, to an enumeration or to a state or event juxtaposition, leading to the frequent use of connectives linking or separating the states or the events, but without introducing a more global textual hier- archization. In narratives, the mean proportions of structuring connectives was constant for the two ages. This finding confirmed one of our hypotheses: it seems that the existence of a canonical structure could explain this stability.

Descriptions and argumentations developed in a similar manner, as there was an increase in structuring connectives and a decrease in linearity connectives for both, but in expectively different proportions. These results confirmed the influence of the organization in memory of structures associated to types of text. Descriptions contained the highest proportion of linearity connectives, and the lowest proportion of structuring connectives. These children had no a priori sequential structure, and simply enumerated or juxtaposed the elements of the picture, without any hierarchy or global organization. In addition, they used many connectives to indicate the passage from one element to another. The differences in textual structures produced by the two grades also indicated particular difficulties in writing descriptions. The 5th-graders' descriptions were not at all structured, and this was barely the case for the 8th-graders. Twelve out of 18 8th-graders' descriptions showed sequential organization (67%). The only descriptive text was a narrative one, imagined from the picture, by a 5th-grader who clearly illustrated the difficulty of description skill: instead of writing a poor description, this child chose to write in a familiar genre, and produced a narrative that followed an already acquired textual structure.

Finally, the argumentations contained few linearity connectives, but had the highest proportion of structuring connectives. The subjects, on their own to elaborate a textual framework, were concerned about sequentially structuring their arguments. The ability to structure arguments was clearly greater in 8th-graders. Using the classification of Coirier and Golder (1993), argumentative texts could be organized as follows:

- For 5th-graders, only one text juxtaposed unsupported stands (6%), four texts presented a minimal argumentative structure (22%), nine texts had a more elaborate structure (50%), and four texts integrated restrictions and counter-arguments (22%).
- For 8th-graders, all the texts presented supported argumentation sequences, one text had a minimal structure (6%), fourteen texts were more elaborate (78%), but only three texts demonstrated complex argumentative abilities (16%).

These results showed that the writing and thinking process involved in creating an argumentative text remained a very difficult exercise even for 14-year-old. Depending on the text type, the children demonstrated overall and, in varying degrees corresponding to their school level, progressive textual control. The acquisition of a better management of planning and textualization processes led to a restructuration and a diversification in the use of connectives. It would be therefore interesting to analyse more 'qualitatively' the connectives, in order to closely examine the main categories of connectives used by children according to their school level, and whether this diversification led to a specialization in their use.

Categories of connectives

This last study was designed to compare the nature and frequency of connectives, by taking into account four categories of marks (Chanquoy, 1991; Schneuwly, 1988) and as a category of connective all its own, spatial connectives (e.g., in front of, on the left, near), temporal connectives (e.g., after, then, while), and non-temporal connectives (e.g., but, also, and logical or argumentative connectives) connectives. Table 4 summarizes the numbers and the percentages of connectives in descriptions, narratives and argumentations.

'And' is mainly used in narratives, representing a bit less than half of the connectives. Conversely, 'and' represented only a quarter to a third of the connectives in argumentations, and was rarer in descriptions. Spatial connectives were exclusively found in descriptions, which also contained non-temporal connectives, but no temporal connectives. This is remarkable insofar as it indicates an already high degree of specialization in the use of the different categories of connectives, for different types of text. The temporal connectives remained rare in argumentations and, even though their rate was higher in narratives, their percentage were weak (between 10% and 18%). Non-temporal connectives were the most frequent category of connectives, except for narrative texts. They represented between a third and a half of the connectives used in narratives and descriptions, and two-thirds in argumentations. Two main categories of connectives were used for each type of text: spatial and non-temporal for descriptions; 'and' and non-temporal for narratives; logical 'and' and for argumentations. Temporal connectives were rarely used; it seemed that ' and' mainly indicated the chronological progression of texts.

Globally, the textual framework was the main organizer of the discourse. The presence of a narrative schema in memory underlaid the textual elaboration, and the written story appeared as a linear succession of events whose temporal relationships were specified mainly with 'and' and temporal connectives. Thus, the writer's planning seemed considerably lightened. It could be limited to a global structuring of the text, and to the useful specifications of event relationships with non-temporal connectives (mainly but and suddenly). From the 5th- to 8th-grade, the decrease of temporal connectives and the increase of non-temporal connectives indicated a greater precision in the signalling of breaks between episodes. Nevertheless, the high percentages of 'and' in the two grades should be analysed. In descriptions, connectives allowed textual progression by linking the successive elements of the picture. Writing a description without a schema resulted in

Text type	Connectives	10 y.o.	13 y.o.
Description	And	20 (16%)	36 (17%)
	Spatial Conn.	48 (39%)	96 (43%)
	Non-temporal Conn.	54 (44%)	88 (38%)
Narrative	And	41 (48%)	61 (41%)
	Temporal Conn.	15 (17%)	15 (10%)
	Non-temporal Conn.	30 (35%)	72 (49%)
Argumentation	And	62 (25%)	84 (32%)
	Temporal Conn.	24 (10%)	12 (5%)
	Non-temporal Conn.	158 (65%)	163 (63%)

children using different strategies that led to different textual configurations. Spatial (in front of, behind, etc.) and non-temporal connectives (relative pronouns and subordinating conjunctions) dominated on both levels. In argumentations, the subjects' assertions had no a priori logical organization. The writer had to select his/her arguments while taking into account the possible objections, and then to organize them into a coherent and logical text. The role of connectives was thus to specify the links among arguments. Children mainly used non-temporal (because, but, relative pronouns and subordinating conjunctions) and argumentative connectives (in addition, thus, conversely, etc.).

This analysis can be summarized in two aspects concerning the functioning modes of connectives. First, children were able to use specific connectives according to the different types of text. The younger children used 'and' and temporal connectives to give order to their texts, and specific connectives (i.e., spatial connectives in descriptions and logical connectives in argumentations) to explain links between two successive clauses (Chanquoy, 1991; Schneuwly, 1988). Second, the organization of textual frameworks seemed to have varying levels of difficulty. When the textual framework was based on a logical organization in memory (for example, the canonical schema), connectives were rare. Conversely, when such a framework did not seem to exist, connectives were numerous, as in descriptions and in argumentations.

Concerning the functioning modes of connectives according to the two grades, two levels were distinguished (see also Bereiter & Scardamalia, 1987): the first level corresponded to a step-by-step management of written productions, in which linkage and packaging operations dominated. The textual continuity was maintained by and, and by coordinating connectives (but, also, after, etc.). The multifunctionality of these connectives allowed them to appear in all genres. The most frequent organization of these texts was reduced to an enumeration or to a juxtaposition of elements in descriptions, to a news announcement in narratives, and to unsupported and unarticulated stands in argumentations. Gradually, with the development of thinking, of thinking about writing, and of writing skills, a second and to unsupported and unarticulated stands in argumentations. Gradually, with the development of thinking, of thinking about writing, and of writing skills, a second level appeared which was noticeable in 5th-graders but more frequent in 8th-graders. This level demonstrated precise macro-structural competences in textual structuring, marked by (a) the appearance of descriptive sequences, where children were able to choose a perspective to write the description; (b) the appearance of narrative sequences, with the construction of coherent episodes; and (c) the appearance of argumentative sequences, with supporting relationships. In some rare texts, this structuring competence was super-structural; a plan of text comprising successive articulated phases was superimposed on thematic sequences. The nature of connections between the different sequences and their hierarchical importance were specified. These modifications in the functioning modes of the different categories of connectives seemed to indicate a better control in planning operations, expressed by a greater complexity in connection operations, and thus more efficient writing skills. Nevertheless, most of the texts written by the 8th-graders were still not very well constructed and illustrated the difficulty that they had, particularly in writing descriptions and argumentations. These two last genres need to be learned in a more efficient way; thus instead of always asking children to write narrative texts, it seems useful to suggest teachers to help children to develop more accurate thinking skills about writing in different genres, especially for such complex texts as argumentations and descriptions.

Discussion

The objective of this study was to examine the variations in connectives used in three textual genres, in relation with the development of writing abilities and thinking processes. The analysis of variations in connectives, considered as traces of planning and textualization, have provided information, according to the writer's level, about the different writing processes (Costermans & Fayol, 1997). Concerning variations in text lengths and the connectives' proportions, the results partially invalidated our hypotheses. Although it is a commonplace to note that the texts produced by the 5th-graders were shorter than those of 8th-graders, it is surprising to note that narratives were the shortest texts. Two hypotheses were advanced: one concerning a possible bias due to the pictures, and the other concerning an obviousness for children of a simple narrative structure, due to the narrative schema, and confirmed by our results for children as young as 10 years of age. Conversely, descriptions and argumentations were longer and more complex texts, with many spatial or argumentative connectives, that indicated, at least for the older children, their relative ability to produce these texts.

A closer analysis of text structuring levels confirmed our hypotheses, and demonstrated the following:

- Childrens' narratives were well-structured texts, for both groups, with a high number of specific connectives (but and represented half of the connectives) that appeared in preferential places of the textual framework. This can be explained by the strong cognitive structure of narratives, with a chronological and linear organization that helped children during writing. In this experiment, they had just to 'describe' the events related in each picture, in a chronological order.

- Descriptions posed greater difficulties. Reduced to enumerations in younger children's texts, they were equally ill-structured in older children's productions. Descriptions had neither a chronological nor a linear organization, and children mainly used spatial and non-temporal connectives to specify the spatial arrangement of the different elements of the picture.
- Argumentations were characterized by a slight difference in the functioning modes of connectives between the two levels. From the 5th grade, children demonstrated some argumentative abilities, but this competence in the 8th-graders, although somewhat more developed, was still only partially evident. Argumentations have been well-known to be difficult, without any a priori cognitive schema or a priori rules that could help children in the 'building' of their texts. In addition, even though there were pictures for the three types of text, the picture accompanying the argumentative text might probably not be directly usable for children. This picture might have given some ideas to the children, but without providing help in the textual elaboration.

Overall, our results were in agreement with other works. For example, concerning the nature and distribution of connectives according to text types, connectives were not randomly used, but varied according to the text (Chanquoy, 1991), and their diversification was accompanied by a relative specialization, and by a reorganization of their functions (Schneuwly, 1988). The results of this study showed that the use of connectives depended on: (a) the state of development of these marks, (b) the textual genres in which they appeared, (c) the production topics, and (d) the writing levels of skills (Chanquoy & Fayol, 1995). The usual opposition between oral and written productions, concerning the use of connectives, does not seem to be justified here. The general acquisition of connectives seems to occur concurrently in oral language as in written language, and their use largely depends on production contexts (genre, topic, and eventual reader). The choice among different connectives is not random, but is a function of some linguistic and non linguistic factors about the situation being spoken of or described, even for young children (Chanquoy, 1996).

The findings of the present experiment should prove particularly interesting for those concerned with instruction. These results highlight the variable and adaptable nature of young writers' composing processes (Zecker, 1996), and also the types of text as an important and influential factor on these processes and on the acquisition of writing skills. However, as said previously, the writing process is not easy to analyse because it takes place in the subject's mind. The single text study is not sufficient, but must be followed or led with analyses of the writer's behaviour during writing.

Further research is needed to study how children can use information provided about different types of text, including more qualitative research allowing a comparison, child per child, of the three text types. Thus, it could be possible to see if children use very different categories of connectives in each text. The next step could be to develop a pedagogical tool to help children to write about different genres (Charney & Carlson, 1995) and to increase their knowledge about the different meanings shared by the same connectives. The knowledge of particular forms of discourse is also important for writing coherent texts (Bereiter & Scardamalia, 1987). Thus, the results of this study could have some implications for writing instruction: first, teachers should try to develop interest in different written genres (Albin, Benton, & Khramtsova, 1996), instead of 'concentrating' writing on one or two types of text; second, it seems important to develop children's knowledge about different writing topics, so that they can be aware of the specific conventions, organization and rules that control the different genres; third, it could be interesting to focus on one aspect of the written texts (as here, for example, on connectives) to provide children a way to analyse the structuring of their own productions and to learn how to think about writing. Children must become aware that writing is not simply a transformation of an oral output into a written output, using conventional arbitrary signs, but that writing must be thought of differently, organized and deeply analysed to be improved (see also Bakunas, 1996).

To conclude, even if this aspect has not been taken into account in this work, it is important to help children to take care of the potential reader, to consider the audience, especially for argumentative texts, where the reader, as the person to convince, is very important. As Sperling (1996) said: 'Learning to write means, to a large degree, learning to anticipate that (and how) one's words will be read' (p. 54).

It is necessary to help children to develop metacognitive skills in their own writing, to help them to think better and to learn more efficiently (Hamers & Overtoom, 1997). Children need to acquire both procedural and declarative knowledge about writing. A deeper reflection about writing could lead to better thinking skills (Hartley, 1991). Indeed, "writing tends to promote greater self-reflection and the taking of broader perspectives than does oral expression." (Wade, 1995, p. 24). Similarly, it seems that the development of such a complex cognitive skill as writing may need a stage of explicit guided monitoring before children will be able to guide their writing by a more autonomous self-monitoring.

Once again, much more research is needed. Some studies relating the types of exercises done in school to children's skills and performances would be very fruitful. Particularly, these kinds of studies could be useful during primary school to track learning problems. Long-term interventions do have effects on the development of cognitive abilities. Writing and thinking are main parts of these abilities. And finally, they could allow to show different individual learning patterns, in order to propose adapted apprenticeships for children with different thinking and writing skills. More emphasis on each individual pattern and on the possibility of cognitive progress through training and practice must be envisaged. In addition, the social and cultural context in which writing, and, more generally, literacy are acquired must be examined to improve the quality of instructional programs in writing. To do so, it is necessary to build a link between fundamental and applied research. These two kinds of research are, still today, too far from each other. "Writing can be fruitfully viewed as the prototypical thinking task" (Kellogg, 1993, p. 3).

References

Adam, J.M. (1992). *Eléments de linguistique textuelle* [Elements of textual linguistics]. Bruxelles: Mardaga.

Albin, M.L., Benton, S.L., & Khramtsova, I. (1996). Individual differences in interest and narrative writing. *Contemporary Educational Psychology, 21,* 305-324.

Bakunas, B. (1996). Promoting idea production by novice writers through the use of discourse-related prompts. *Applied Psycholinguistics, 17,* 385-400.

Baron, J.B., & Sternberg, R.J. (Eds.) (1987). *Teaching thinking skills. Theory and practice*. New York: Freeman.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition.* Hillsdale, NJ: Erlbaum.

Boscolo, P. (1995). The cognitive approach to writing and writing instruction: A contribution to a critical appraisal. *Current Psychology of Cognition, 14,* 343-366.

Chanquoy, L. (1991). *Ponctuation et connecteurs: Acquisition et fonctionnement. Etudes comparatives chez l'enfant et l'adulte (méthodes off-line et on-line)* [Punctuation marks and connectives: Acquisition and functioning. Comparative studies in children and adults (off-line and on-line methods)]. Dijon: Université de Bourgogne.

Chanquoy, L. (1996, October). *Connectives and argumentative texts. A developmental study.* Paper presented at the First International Workshop on Argumentative Texts. Barcelona, Spain.

Chanquoy, L., & Coirier, P. (1997, August). *Connectives in argumentative texts.* Paper presented at the Seventh European Conference for Research on Learning and Instruction. Athens, Greece.

Chanquoy, L., & Costeplane, J. (1995, August). *The use of connectives in three types of text written by children.* Paper presented at the Sixth European Conference for Research on Learning and Instruction. Nijmegen, The Netherlands.

Chanquoy, L., & Fayol, M. (1995). Analyse de l'évolution de l'utilisation de la ponctuation et des connecteurs dans deux types de texte [Analysis of the evolution in the use of punctuation marks and connectives in two textual genres]. *Enfance, 2,* 227-241.

Charney, D.H., & Carlson, R.A. (1995). Learning to write in a genre: What student writers take from model texts. *Research in the Teaching of English, 29,* 88-125.

Coirier, P. (1996). Composing argumentative texts: Cognitive and/or textual complexity. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Current trends in writing research: Theories, models and methodology* (pp. 135-152). Amsterdam: Amsterdam University Press.

Coirier, P., & Golder, C. (1993). Writing argumentative text: A developmental study of the acquisition of supporting structures. *European Journal of Psychology of Education, 8,* 169-181.

Costermans, J., & Fayol, M. (1997). *Processing interclausal relationships. Studies in the production and comprehension of text.* Mahwah, NJ: Erlbaum.

De Weck, G., & Schneuwly, B. (1994). Anaphoric procedures in four text types written by children. *Discourse Processes, 17,* 465-477.

Fayol, M. (1985). *Le récit et sa construction. Une approche de psychologie cognitive* [The beginnings of narratives. A cognitive perspective]. Neuchâtel: Delachaux & Niestlé.

Fayol, M., & Schneuwly, B. (1988). La mise en texte et ses problèmes [Translating and its problems] In J.L. Chiss, J.P. Laurent, J.C. Meyer, H. Romian, & B. Schneuwly (Eds), *Apprendre/enseigner à produire des textes écrits* (pp. 43-61). Bruxelles: De Boek.

Flower, L.S., & Hayes, J.R. (1980). The dynamics of composing. Making plans and juggling constraints. In L. Gregg & E. Steinberg (Eds.), *Cognitive processes in writing* (pp. 31-50). Hillsdale NJ.: Erlbaum

Galbraith, D. (1992). Conditions for discovery through writing. *Instructional Science, 21,* 45-72.

Golder, C. (1992). Production of elaborated argumentative discourse: The role of cooperativeness. *European Journal of Psychology of Education, 7,* 51-59.

Gould, J.D. (1980). Experiments on composing letters: Some facts, some myths and some observations. In L.W. Gregg & E.R. Steinberg (Eds), *Cognitive processes in writing* (pp. 98-127). Hillsdale, NJ: Erlbaum.

Hamers, J.H.M., & Overtoom, M. (1997). Theory. In J.H.M. Hamers & M. Overtoom (Eds.), *Inventory of European programmes for teaching thinking* (pp. 13-29). Utrecht: Sardes.

Hartley, J. (1991). Psychology: Writing and computers. A review of research. *Visible Language, 25,* 339-375.

Hayes, J.R. (1996). A new framework for understanding cognition and affect in writing. In C.M. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 1-28). Mahwah, NJ: Erlbaum.

Hayes, J.R., & Flower, L.S. (1980). Identifying the organisation of the writing process. In L.W. Gregg & E.R. Steinberg (Eds), *Cognitive processes in writing* (pp. 3-30). Hillsdale, NJ: Erlbaum.

Hedberg, N.L., & Fink, R.J. (1996). Cohesive harmony in the written stories of elementary children. *Reading and Writing: An Interdisciplinary Journal, 8,* 73-86.

Kellogg, R.T. (1993). Observations on the psychology of thinking and writing. *Composition Studies, 21,* 3-41.

Kellogg, R.T. (1996). A model of working memory in writing. In C.M. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 57-72). Mahwah, NJ: Erlbaum.

Kintch, W., & Van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363-394.

Kowal, S., & O'Connell, D.C. (1987). Writing as language behavior: Myths, models, methods. In A. Matsuhashi (Ed.), *Writing in real time* (pp. 108-131). Norwood, NJ: Ablex.

Levelt, W.J.M. (1989). *Speaking. From intention to articulation.* Cambridge, MA: MIT.

Levy, C.M., & Ransdell, S. (1996). Writing signatures. In C.M. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 149-161). Mahwah, NJ: Erlbaum.

Mandler, J.M., & Johnson, N.S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9,* 111-151.

Millis, K.K., Graesser, A.C., & Haberlandt, K. (1993). The impact of connectives on the memory for expository texts. *Applied Cognitive Psychology, 7,* 317-339.

Pappas, C.C. (1993). Is narrative 'primary'? Some insights from kindergarteners' pretend readings of stories and information books. *Journal of Reading Behavior, 25,* 97-129.

Parrila, R.K. (1995). Vygotskian views on language and planning in children. *School Psychology International, 16,* 167-183.

Perera, K. (1986). Language acquisition and writing. In P. Flechter & M. Garman (Eds.), *Language acquisition. Studies in first language development* (pp. 494-520). Cambridge, MA: Cambridge University Press.

Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In R. Rosenberg (Ed.), *Advances in applied psycholinguistics* (Vol. 2, pp. 142-175). Cambridge, MA: Cambridge University Press.

Schneuwly, B. (1988). *Le langage écrit chez l'enfant* [The written language in children]. Neuchâtel: Delachaux et Niestlé.

Schneuwly, B., Rosat, M.C., & Dolz, J. (1989). Les organisateurs textuels dans quatre types de textes écrits. Etude chez des élèves de dix, douze et quatorze ans [Textual organizers in four genre of written texts. A study with 10, 12, and 14 year olds children]. *Langue Française, 81,* 40-58.

Shapiro, L.R., & Hudson, J.A. (1997). Coherence and cohesion in children's stories. In J. Costermans & M. Fayol (Eds.), *Processing interclausal relationships. Studies in the production and comprehension of text* (pp. 23-48). Mahwah, NJ: Erlbaum.

Sperling, M. (1996). Revisiting the writing-speaking connection: Challenges for research on writing and writing instruction. *Review of Educational Research, 66,* 53-86.

Stein, N.L., & Glenn, C.G. (1982). Children's concept of time: The development of a story schema. In W.J. Friedman (Ed.), *The developmental psychology of time* (pp. 187-205). New York: Academic.

Sternberg, R.J. (1987). Questions and answers about the nature and teaching of thinking skills. In J.B. Baron & R.J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp 13-36). New York: Freeman.

Wade, C. (1995). Using writing to develop and assess critical thinking. *Teaching of Psychology, 22,* 24-28.

Zecker, L.B. (1996). Early development in written language: Children's emergent knowledge of genre-specific characteristics. *Reading and Writing: An Interdisciplinary Journal, 8,* 5-25.

Zimmerman, B.J., & Risemberg, R. (1997). Research for the future. Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology, 22,* 73-101.

12

Thinking skills in reading and text studying

H. Van Oostendorp & M. Elshout-Mohr

Introduction

Reading and text studying are complicated activities and many thinking skills are involved in comprehending the meaning of texts and in processing text information according to the situation and purposes at hand. Relevant thinking skills have been identified at various levels. In this chapter we distinguish between the basic level, the strategic level and the higher order level.

Basic level skills, for instance, have been investigated by Hunt, Lunneborg, and Lewis (1975) who showed that encoding speed, e.g. time to retrieve a name as well as time to retrieve semantic or phonological information associated with a name, is related to reading ability (see also Perfetti & Lesgold, 1977). Another example of a basic skill comes from Kintsch and Vipond (1979), who identified another basic ability which correlates strongly with reading performance (see also Daneman & Carpenter, 1983). Kintsch and Vipond found that the greater the functional capacity of a reader's working memory, the easier it is to comprehend references more than a few sentences back. Reading performance is then improved because time consuming and error-prone searches in long-term memory can be avoided.

Strategic skills have been described by Johnston and Afflerbach (1985), among others. They described a number of strategies that are employed by proficient readers when a text is difficult. For instance, students make deliberate use of linguistic information such as lexical repetitions and signal words and use contextual information, while focusing on main points. They ask themselves for instance: What will the teacher find important in this text? (Schellings & Van Hout-Wolters, 1995). Other examples will be given in this chapter.

Baker and Brown (1984) and Palincsar and Brown (1984) focused on higher order, metacognitive, skills such as determining the reading goal, monitoring whether the goals are reached, and deciding what to do next when goals are not met. To give a concrete example: preparing an essay exam demands goal setting (the goal might be to understand information and to memorize definitions and other main points), monitoring (the monitoring strategy might consist of summarizing and reviewing, which are means to diagnose one's own ability to retrieve information from memory), and taking decisions (for instance to proceed reading or to reread parts of the text). The proficient orchestration of the various procedures and strategies is often called metacognition, or self regulation.

It is far beyond the scope of this chapter to review all skills involved in reading and text studying, and this is certainly not what we intend to do. Our plan is to focus on a sample of difficulties that frequently occur in reading and studying informative texts, and on thinking skills that might be helpful to deal with those difficulties. Three categories of difficulties that are identified in literature and that we addressed in our own research, will be discussed. First, some interesting difficulties have been identified in the 'use of prior knowledge' during the process of constructing a textbase (a representation of the content of the particular text) and a situation model (a representation of some aspect or portion of the real world as suggested by text information, see van Dijk & Kintsch, 1983). Second, there is a category of difficulties that do come forth when new text information is conflicting with already available information. In this case readers must detect the conflict, and use adequate strategies to update or restructure 'older' representations. Third, there are difficulties that arise when a person, a student, is required to study textual information about a topic that he or she is not familiar with. In this case the obvious drawback is the relative shortage of relevant prior knowledge. The student has then to handle the information in a selective and goal directed manner, the actual reading process is difficult and the student must try to integrate the ideas that are presented in the text in the personal knowledge base.

The three categories of difficulties, concerning the use of prior knowledge, updating and restructuring prior knowledge, and constructing new knowledge structures, are discussed in later sections. In each section we first describe how these difficulties have been highlighted from the perspective of experimental reading research by manipulation of text materials, reading tasks, reading abilities of subjects and so on. Subsequently we discuss these difficulties from the perspective of thinking skills. It has been indicated by several authors that reading can be viewed as a higher order skill (Just & Carpenter, 1980; Van Dijk & Kintsch, 1983), because research on reading revealed that complex thinking skills, like inductive reasoning and problem solving, are required throughout the reading process. The idea that it might be wise to incorporate such skills in educational reading programs has been suggested (e.g. Resnick, 1987) and put to the proof with some success (Palincsar & Brown, 1984). We hope to contribute to this line of research by focussing attention on a few categories of difficulties that frequently occur during reading processes and the specific thinking skills that enable strong and weak readers to sometimes overcome these difficulties. Increasing the specificity of the links between difficulties and thinking skills will not only enhance insight in the required skills, but can also facilitate accurate description of lower order skills that help readers to reach automaticity and higher order skills that are needed to organize and monitor appropriate appliance of the necessary skills. How the difficulties that we selected for elaboration in this chapter are related to the broader reading process, is discussed in the next section, in which the general framework is exposed.

Framework

The general framework for this chapter is presented in Figure 1. This framework is partly based on ideas of Scardamalia and Bereiter (1991).

An essential idea in discourse comprehension and learning from text is that constructing a coherent text representation, and ultimately an integrated mental model or situation model, in the form of a connected network of (semantic) propositions, is pivotal. The more connected the network, internally and externally to prior knowledge, the better the understanding and transfer. However, the construction of such a coherent text representation or textbase is only possible when relevant prior knowledge is available, and also when it is activated (or accessed) from long-term semantic memory or episodic memory, when needed (Scardamalia & Bereiter, 1991). We assume that most of the knowledge is organized in the form of situation models or mental models (Johnson-Laird, 1983; Van Dijk & Kintsch, 1983). Having available and activated them at the right moment is an essential condition to be able to construct a coherent text representation (this process is indicated by arrow (1) pointing to the right). Having constructed such a representation may lead to a new situation model (arrow (2) pointing to the left). This knowledge may be of an episodic or of a more permanent, general character. The difficulties and skills involved in this process of using prior knowledge (arrow (1)) will be discussed first, in the following section, labelled 'Use of prior knowledge'.

The two other situations that we wish to address in this chapter are not explicitly present in Scardamalia and Bereiter's framework. Both situations occur very frequently, however, not only in experimental-educational settings but also in more real life settings. In all of these the reader has to make use of prior knowledge (the

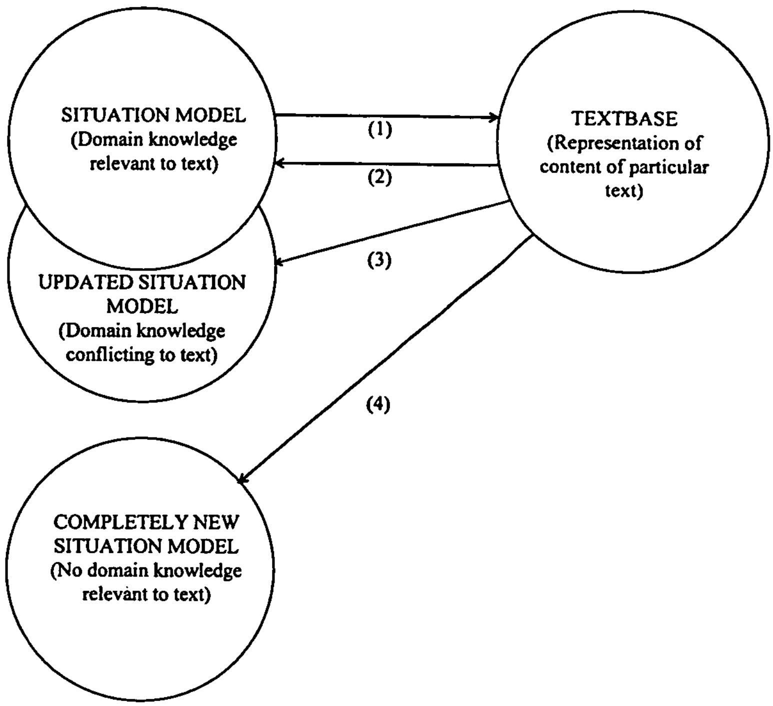


Figure 1 Interaction processes between situation model construction and textbase construction

situation model) to construct a coherent text representation.

The first situation that we want to discuss concerns the updating of situation models. Frequently, information is presented that conflicts in some way to existing models. The reader has in this case relevant prior knowledge, but the text contains conflicting information. After having constructed a coherent textbase representation - after having 'comprehended' as far as possible the text - readers need to change the already existing situation model, to update that model and bring it in correspondence to the new data, presented in the current text (or discourse). This updating process (arrow (3) pointing to the left) has some specific characteristics and difficulties which will be discussed in the section 'Updating and restructuring prior knowledge'.

The last type of situation that we want to discuss in this chapter occurs when students process study-texts and have to build a completely new situation model (arrow (4) pointing to the left). In that case the student has general semantic or world knowledge available (Kintsch & Franzke, 1995), but he or she has no appropriate situation models or detailed schemata available. On the contrary, the reader has to build up these new situation models. The shortage of relevant prior knowledge will, of course, result in the construction of a textbase that is at first incoherent and incomplete. Consequently, also the new models constructed will be incomplete. Only by extra efforts will these inadequacies of the mental representation be repaired. The difficulties and skills involved will be discussed in the section 'How to proceed when relevant knowledge is not yet represented in the situation model'. The framework we described and represented in Figure 1 will be followed in the next sections.

Each section contains a table which presents an overview of the kind of 'difficulties' and 'skills' involved in the process under consideration. Where possible, we refer to training studies and give examples of how these skills might be learned or developed. Throughout the chapter we will use the three levels, basic, strategic and higher order level to present and categorise skills. The relationship between these levels is the following (see also Elshout-Mohr, 1992; Pressley, Snyder, & Cariglia-Bull, 1986). Readers use strategies. When readers confront a text, they identify elements of the situation that are similar to previous encounters, and they use these similar elements to select an appropriate strategy. Good readers are familiar with many strategies, both general and more specific to domains in which the individual is an expert. They have to decide when, where, why and how to use them and they must switch strategies when all is not going well. This demands all sorts of monitoring and regulating skills that we subsume under the label 'higher order skills'.

The strategies used are acquired. With time and experience some aspects of strategy skill are automatized. An example might be 'processing referential information within the limits of one's working memory'. Such automatized components of strategies are subsumed under the label 'basic skills'. They may be closely connected to individual differences in capacity (e.g. capacity of working memory) and to differences in the knowledge base (many strategies can only be executed by people who possess a lot of knowledge). A consequence of this conceptualisation of the interrelations between skills at the basic, strategic and higher order level is that the distinctions are flexible in the following sense. A skill that is 'strategic' at one moment of time in a novice reader (and therefore under command of higher order skills), may become automatized when the reader becomes more proficient. It then becomes a basic skill that is not under direct command of higher order skills. In the same manner metacognitive (higher order) skills, like comprehension monitoring, may become incorporated in the strategic skills of proficient learners. In general, one could say that experience has the effect that new higher order skills are developed, former higher skills are incorporated in strategic skills and former strategic skills become automatized basic skills that demand no conscious attention of the reader (see also Nelson, 1996). The skills that will be discussed in this chapter are cognitive skills. We take it for granted that these skills are rightly referred to as 'thinking skills', although some are more clearly formulated as such (like 'inductive reasoning') than others (like 'making inferences').

Use of prior knowledge

Goal of using prior knowledge to build an appropriate textbase

Prior knowledge, represented in the situation models of a person, is essential for a person's understanding of textual information and building a coherent text representation (the process corresponding to the arrow (1) pointing to the right in Figure 1). In this section we discuss what skills are needed to make proficient use of available prior knowledge during reading and text studying. First, a reader must be able to activate components of prior knowledge that are relevant for the reading task at hand and these components must be put to use to make sense of the text. Second, to construct a coherent text representation, the reader must draw inferences and tie together different meaning aspects. Third, the reader must take care that the coherence of the textbase is not superficially based on linguistic cues. It has to be linked to conceptual and episodic prior knowledge. Otherwise it is impossible to detect errors in the information and to use the information productively, for instance to answer higher order questions. To achieve this requires monitoring, regulating and checking the quality of the textbase under construction.

Almost every theory on discourse understanding acknowledges the idea that activating knowledge and drawing inferences play an important role in constructing a propositional semantic structure, in particular when one wants to construct a coherent representation (see for instance Schnotz & Ballstaedt, 1996; Van Dijk & Kintsch, 1983). However, at the same time, these theories tend to underestimate the complexity of the regulating and monitoring processes that are required. For theoretical purposes it is important to investigate these processes under various conditions, but for practical reasons this is important also. Deficiencies in the monitoring and regulation processes may lead to superficial text processing and lack of flexibility in regard to variations in reading tasks and reading conditions. Proficient readers display a high flexibility in processing, and they accommodate the completeness of processing to circumstances, such as the nature of text and information and their own task conception. To identify the skills that are involved in the use of prior knowledge during self regulated reading, we shall first discuss a number of relevant concepts and studies describing characteristic difficulties readers have in putting knowledge in use during reading. Table 1 provides an overview of the difficulties which we will discuss and also presents a list of skills relevant for the use of prior knowledge.

Processes, variables and difficulties concerning the use of prior knowledge

A first relevant concept in understanding how readers put prior knowledge to use in constructing an adequate textbase, is the concept of 'semantic relatedness', that is the overlap in semantic features of concepts. Semantic relatedness is one of the material factors influencing the activating and monitoring process of constructing an episodic semantic representation. Erickson and Mattson (1981) presented subjects with questions as 'How many animals of each kind did Moses take on the

Ark?'. The majority of subjects answered this question without noticing that the name, Moses, is not correct. Subjects who showed evidence in a posttest that they perfectly knew that it was Noah (not Moses) who took animals on the Ark, failed to activate this knowledge. Or, more precisely, they failed to activate the information completely enough to detect the use of the erroneous, but semantically related, name. This phenomenon, found under varying circumstances, is called the Moses-illusion. It also occurs for instance when the subjects' task is to verify sentences such as 'Moses took two animals of each kind on the Ark', True or False? (Reder & Kusbit, 1991; Van Oostendorp & De Mul, 1990; Van Oostendorp & Kok, 1990). On the basis of these studies, among others, we hypothesized that readers continuously monitor the semantic cohesion of a text representation under construction, and regulate further processing on the basis of a comparison of the perceived cohesion to some internal comprehension criterion. During the initial processing of a sentence the perceived cohesion is primarily dependent on the semantic relatedness between involved concepts (Van Oostendorp, 1994).

Seemingly easy textual information can, thus, foster incomplete processing, and not-noticing errors, thereby creating characteristic difficulties. This conclusion can also be drawn from a study by McNamara, Kintsch, Songer, and Kintsch (1996) in which they investigated the completeness of processing textual information under various conditions. The same information, about a biological topic (hart disease), was presented to subjects in a high coherent and a low coherent version. The level of coherence was manipulated by 'linguistic cues', like supplying pronoun and synonym referents, adding macro propositions and headings, and adding sentence connectives and descriptive elaborations to connect familiar and unfamiliar terms. The subjects in the experiment differed in background knowledge about the biological topics. The results were the following: The coherent version led to better recall among both high and low-knowledge readers. Performance on inference and problem-solving questions was also better among low-knowledge readers who read the coherent text. However, the opposite was the case for high-knowledge readers who performed better after reading the less coherent text version. Kintsch and Kintsch (1996, p. 523) draw the following conclusion: "It appears that readers who possess adequate knowledge form a better situation model representation when they must draw these connective inferences on their own. This is because readers must draw on their knowledge to generate the inferences and in so doing construct a more elaborated mental model that is well linked with concepts in the personal knowledge base." This conclusion refers primarily to the situation model that is constructed by the readers, but it does also suggest that the quality of the readers' inference processes was influenced by linguistic coherence of the text. A text can be so 'easy' that a person is tempted to construct a coherent textbase without complete processing and optimal use of his or her prior knowledge. The reader has an 'illusion of knowing' simply because the text is coherent linguistically (Glenberg, Wilkinson, & Epstein, 1982). The risk involved in this illusion is that content-errors in the text are overlooked and that the understanding is superficial.

Table 1 Using Prior Knowledge to Construct a Coherent Textbase

Goal: use prior knowledge to build an appropriate textbase
Condition: relevant prior knowledge is available (in the reader's mental model)

Subtasks
- activating relevant components of the mental model
- monitoring, regulating and checking the quality of the textbase

Difficulties
- incomplete processing and overlooking of errors by seemingly easy information, e.g. semantic relatedness
- incomplete processing by inappropriate goals or task conception
- linguistic coherence of the text facilitates the construction of a textbase and elicits superficial processing of the information
- activated prior knowledge is not always relevant to the specific context

Skills
- Basic level
  . activating context relevant features of concepts
  . generating context relevant inferences and elaborations
- Strategic level
  . focussing attention on context-relevant features of concepts
  . focussing attention on coherence-contributing attributes, concepts, and elaborations
  . developing a task conception that fosters 'careful' reading
- Higher-order level
  . accommodating one's reading strategy and speed to different types of texts conditions
  . keeping one's eyes open when a text is surprisingly easy: checking for incongruencies
  . monitoring the quality of the textbase (avoidance of superficiality)

The quality of the monitoring and regulating processes depends on the text, but also on the reader's proficiency and goals. Proficient readers slow down their processing if they suspect that a seemingly easy text may contain errors and they re-read passages when the text is incoherent (Johnston & Afflerbach, 1985). Another important factor is the task-conception of the reader, a conception that may be instigated externally or internally. It has been shown experimentally that the explicit external instruction 'to read carefully' leads to a high degree of inferencing and integrating of information belonging to 'scriptal' knowledge of readers (Schank & Abelson, 1977). In one experiment Van Oostendorp (1991) instructed subjects to read carefully a text about a train-journey. The readers then activated more concepts belonging to a train script (which are not mentioned in the text such as tickets, bookstalls, seats, and so on) than subjects who read the text 'normally'. Apparently reading goals influence knowledge activation. Effects of reading goals which are internally set are highly similar to effects of goals that are externally set. Readers who habitually set a high comprehension criterion show the same performance, regarding inferencing and integrating, as readers who respond to the external instruction to read carefully.

To conclude this overview, we turn to individual differences that do not concern the regulation and monitoring aspects of prior knowledge activation, but the activation process as such. What actually happens when a reader activates prior conceptual knowledge while reading a sentence, is somewhat different for high and low proficient readers. It has been found in several studies that proficient readers activate and actualise meaning aspects which are more relevant in the context than the meaning aspects activated by weak readers (Merrill, Sperber, & McCauley, 1981). Van Oostendorp and Anbeek (1983) presented readers of 11-12 years context sentences on a computer screen such as 'The jeweller carved the glass with a diamond'. Immediately after reading this sentence a question was presented in which subjects had to judge the importance of an attribute of a concept from the context sentence. Examples are the questions 'Diamond is sharp. Important or not-important here?' (sharpness is in this context a relevant attribute) or, opposed to this, 'Diamond is precious. Important or not-important here?' (precious is in this context an irrelevant attribute). The speed of making the judgements was registrated. The results showed that proficient readers (i.e. readers who scored high on the ISI reading comprehension test) answered the questions with relevant attributes much faster than weak readers, while the questions with irrelevant attributes showed no significant differences between proficient readers and weak readers. This research indicated an important difference between proficient and weak readers. Proficient readers are able to elicit aspects of prior knowledge that are specifically relevant to the text at hand (e.g. the sentence) while less proficient readers seem to activate prior knowledge less selectively.

A similar phenomenon has been observed in the inference and elaboration processes of proficient and weak readers. Both proficient and weak readers make inferences and generate elaborations trying to make the textbase coherent. Stein, Bransford, Franks, Owings, Vye, and McCraw (1982) showed, however, that weak readers make elaborations that may be semantically correct, but not highly relevant. Skilled readers, on the other hand, generate elaborations that are not only semantically correct, but also highly relevant. That is, they are contributing to the coherence of the representation. Subjects received sentences as 'The strong man helped the woman' which they had to continue in such a way that they could understand the described situation. Skilled readers generated relevant continuations such as 'carry heavy packages', while weak readers generated low relevant elaborations as 'cross the street' (the latter does not explain 'strong'). Knowledge activated by readers, in particular weak readers, is not always relevant to the specific context. In follow-up studies Franks, Vye, Auble, Mezinksi, Perfetto, Bransford, Stein, and Littlefield (1982) and Stein et al. (1982) did some training experiments in which they tried to teach readers the thinking skills involved. These will be discussed in the following paragraph.

Using prior knowledge: thinking skills and training

Table 1 presents a list of skills that are relevant for readers' proficiency in the use of prior knowledge. This list is divided in basic skills, strategic skills and higher order skills. These skills can be viewed as reading skills, but we intend these skills to represent general skills that are integrated in the reading task. General thinking skills, like activating relevant aspects of information, generating relevant inferences, focussing attention, developing an appropriate task conception, and seeking accuracy, cannot be developed in isolation of specific tasks or content, like reading a text about Noah's ark. Such skills, however, have the distinct quality that they can be integrated in all sorts of reading tasks and in tasks that are quite different from reading as well, like doing mathematics or handling political problems. Whether it is feasible to develop the skills which are mentioned in Table 1 is but partly known. Some training studies that are relevant for developing these general skills within the field of reading are mentioned.

A study in which a basic skill did indeed improve was performed by Beck, Perfetti, and McKeown (1982). In this study students learned the meaning of new words in a training program that made exceptionally high demands on precision, accessibility and flexibility of word meaning knowledge. The skills relevant here to text comprehension are being able to activate context-relevant features of concepts, and in the first place being able to focus attention to make these features available. The results showed that this training had a positive effect on text comprehension and more specifically on the comprehension of parts of the text that were not processed deeply by 'normally trained' students.

Franks et al. (1982) and Stein et al. (1982) did some training experiments in which weak readers learned to focus on the relevance of concepts in a text to the coherence of the text representation. In the given example ('The strong man helped the woman') subjects were encouraged, for instance, to find a reason why the word 'strong' is mentioned. Both studies resulted in positive effects on understanding and recall of text information, also when more complex texts were used. Apparently the skill to generate context-relevant inferences and elaborations, and to focus attention on coherence-contributing attributes fostered text understanding. Further research is needed to find out whether readers' adoption of the strategic skill (focussing on the relevance of concepts) results in the long run in an improvement of the basic skill (unmediated activation of relevant features of concepts).

We present here three examples of training studies that are relevant to higher order skills. The first one is a series of studies done by Palincsar and Brown (1984). They focused on strategic and, particularly, on higher order skills. They made children (7-years to 11-12 years old) familiar with a procedure to regulate the quality of their own reading by the use of four strategies: posing questions, summarising, predicting what comes next in a text and clarifying texts for clarity, internal consistency, or compatibility with known facts. The higher order and strategic skills were trained by a so-called 'reciprocal teaching' method. In this method the teacher first models the four skills, and gradually the guidance and support provided by the teacher lessens, and is taken over by the children themselves. Their results showed that their method was very effective and also showed transfer to text understanding in other domains (Brown & Palincsar, 1989).

Beck, McKeown, Worthy, Sindora, and Kucan (1996) studied the effect of asking readers a new type of questions about the texts that they are reading. In 'Questioning the author' students are asked to think about the author's intended message, to evaluate how it was conveyed and whether the author was successful. Thinking about why the author chose a particular piece of information and evaluating how it contributed, challenged students to activate their own prior knowledge and improved students' active participation.

A third example of successful training in higher order skills, is a program developed by Paris and others, the so-called 'Informed strategies for learning' program (Paris, Cross, & Lipson, 1984). The program's aim was to teach reading and learning skills, how and why they work, and when they have to be applied. The training included training in strategies to grasp the meaning of a text by activating prior knowledge, summarising, and so on. Moreover, the students were trained to monitor and control the reading process - in other words, to develop a task conception that fosters careful reading -, and to make use of strategies like rereading, posing questions, checking for incongruencies, etc. The skill to accommodate reading strategy to circumstances was an explicit part of the training. Paris and Oka (1986) showed that also this program had positive effects. The scope of the program was of course much broader than just 'improving readers' use of prior knowledge in constructing a coherent textbase. It showed however that strategic skills and higher order skills, like regulating and monitoring the comprehension process, can be trained and that such training can result in substantial improvement.

Updating and restructuring prior knowledge

The goal of updating and restructuring an old situation model

Often the primary goal of reading texts is to understand something new, for instance understanding the continuously changing world and to accommodate our established ideas to changes that happened in this world. Or, when it concerns studying texts, to understand a theory or body of facts which deviate from a point of view taken earlier. In these cases readers have to update their old situation models or to restructure them. We define updating here as filling in new values (new data) in variables that make up existing schemata or situation models. Restructuring requires a more drastic change. The existing schemata or models have then to be restructured completely and replaced by new ones.

A recurring finding of studies on updating situation models in the context of reading and studying, is that updating is relatively easy when it involves merely elaborating or adding some new components to the model. However, updating is often much more difficult under conditions in which new knowledge
has a predecessor in the reader's memory, even if this predecessing mental model is constructed just before. Updating then implies transforming this already available knowledge and this is found to be a difficult and skill demanding process. In terms of the preceding framework of Scardamalia and Bereiter (1991) we focus here on the modification of an existing situation model by new text information (arrow (3) pointing to the left in Figure 1). The existing situation model may already exist in semantic memory and be constructed long before, or it can be constructed just before, even during reading the same text.

The questions that are addressed concern the following aspects of the tasks: How and when do readers decide whether they should modify the model, or should hold on to the old model? and How do they actually modify the model? Table 2 provides an overview of the difficulties and skills which we will discuss in the next subsections.

Processes, variables and difficulties in updating and restructuring prior knowledge

In several studies it has been found that the existing memory representation was not effectively updated after reading new, correcting information (Johnson & Seifert, 1994; Van Oostendorp & Bonebakker, 1996, in press; Wilkes & Leatherbarrow, 1988). Readers often keep on making inferences on the basis of old information, that should have been corrected. Van Oostendorp and Bonebakker (in press), for instance, presented subjects (university students) with a story on a fire in a warehouse. In the experimental condition a correction was inserted, which denied a fact reported earlier. Subjects in the experimental condition read a sentence as 'inflammable materials were carelessly stored in a side room'. Later in the story they read that 'the side room happened to be empty'. Subjects in the control condition read first a neutral, irrelevant sentence as 'Both the fire brigade and police were involved with the investigation', and after that (exactly as in the experimental condition) 'the side room happened to be empty'. The conclusion was that the influence of old, obsolete information in the experimental condition was not fully neutralized by the new, discrediting information. Answers on inference questions (e.g. 'What was the cause of the explosion?', or 'For what reason could an insurance company here refuse a claim?') were frequently based on the old information, even by subjects who were aware of the fact that certain information was discredited. In these experiments, even the explicit instruction that information might be corrected did not lead to a better updating. Readers continued to use the misinformation, and kept making inferences based on the corrected information.

A recent study by Campanario and Van Oostendorp (in press) showed that a very strong manipulation, like a clear-cut statement that the given information was

Table 2 Updating and Restructuring Prior Knowledge

Goal: updating situation model, that is building new, related model
Condition: having related situation model available

Subtasks
- noticing the need to update/ restructure
- integrating old and new information

Difficulties
- correction endangers coherence of existing mental models
- integration demands much effort and working memory
- the need to update/ restructure is often not strong enough
- quality of existing model

Skills
- Basic level
  . functional working-memory capacity
- Strategic level
  . mobilising prior knowledge
  . detecting misfits and gaps in current knowledge
  . deep processing new information
  . paying attention to the conflict between old and new
  . weighing and investigating pro and contra's of various models
  . using substrategies to integrate old and new information
- Higher-order level
  . hold the epistemological belief that restructuring is worthwhile
  . monitoring and regulating the direction of attention (to textbase construction or updating existing situation models)

incorrect, is needed to make subjects use the new information instead of the old information. An additional finding was that correction of incorrect information is facilitated when the text not only corrected the information (about the inflammable materials), but provided an alternative causal explanation of the event (the explosion). In the example it was helpful to present the information 'that there were indications that the fire had been started deliberately', through insertion of a sentence as 'Firefighters have found evidence of gasoline-soaked rags'. The additional information makes a strong alternative mental model available for the reader, a mental model which enables the reader to integrate all isolated information units without further need for the discredited information. This finding suggests that one source of difficulty for readers is that they are reluctant to correct information when rejection of the information will not contribute to the coherence of an existing situation model. A factor that is relevant here, besides the explicitness of the correction, is the saliency of the original incorrect information. In the same study by Campanario and Van Oostendorp (in press) it was also shown that strengthening the old information, by inserting text information that indirectly reinforces it, lessens the degree of updating.

Limitations in functional working-memory capacity are a second source of difficulty in updating processes (Kintsch & Vipond, 1979). Such limitations can urge a reader to give priority to those pieces of information, whether old or new, that are easiest to use in further processing. The discarded, 'suppressed', information has then less chance to be used and to influence the final representation (Mannes, 1994). Another factor causing difficulties in updating are 'epistemological beliefs', like the belief that the integration of ideas underlying a text, is important to understanding (Kardash & Scholes, 1996; Rukavina & Daneman, 1996; Schommer, 1990).

The examples of difficulties in updating that were given until now are relatively simple, and have a restricted scope. Chinn and Brewer (1993) describe more complex cases in which the existing knowledge that must be restructured consists of deeply entrenched ideas and conceptual knowledge. In these cases learners are found to use complex higher order strategies to determine whether (and, if so, how) their existing conceptual model needs to be changed in order to achieve a successful coordination of their existing model and incoming data. Chinn and Brewer (1993) distinguish the following strategies that learners are observed to use to handle new, deviating ('anomalous') information: (a) ignore anomalous, new information, (b) reject new information, (c) peripheral theory change; that is, a partial peripheral change but preserving the core of an old model, (d) reinterpret the new information while retaining the old model, (e) exclude the new information from the domain of the old model, (f) hold the new information in abeyance, that is, some kind of compartmentalization, and finally (g) accept the new information and restructure the old model. The variety of strategies that are used to postpone or reject modification of the existing model is remarkable. Only the last strategy involves accepting the new data and changing the existing model in an appropriate way. An interesting additional finding that is reported by Chinn and Brewer concerns the quality of the information processing strategies that guide evaluation of the incoming contradicting information. Changes of existing models are more likely to occur when the incoming information is processed deeply. Deep processing, in turn, can be enhanced by fostering personal involvement in the issue the text is about (Tesser & Shaffer, 1990).

Other studies show that the quality of the already available situation model is a relevant factor also. Peeck, Van den Bosch, and Kreupeling (1982) did the following experiment. Learners (students in primary education, 11-12 years) were instructed to mobilise their knowledge of foxes prior to studying a text on a fictional kind of fox. The text contained information that was consistent or inconsistent with the learner's existing knowledge of foxes. In comparison with a nonmobilisation control group, the learners were found to remember the same amount of consistent information but significantly more inconsistent information. This result fits in nicely with a suggestion by Kintsch (1980; see also Berlyne, 1965) that "it is change, incongruity, surprise that leads to new learning (..). Misfits between the apperceptive mass and new information (..) provide the right conditions for learning, which is now conceived of as a correction or addition to existing knowledge structures" (pp. 92-93). The mobilisation phase facilitates the detection of misfits and knowledge gaps by making the learner more aware of his or her initial existing knowledge structure. In a related study Van Oostendorp (1996) found that updating of existing knowledge with new, incongruent information was performed better for subjects who had a more precise and accessible situation model than for subjects with a weaker situation model. In this study subjects were presented with two related texts about the situation in Somalia, one after the other. The second text contained transformations or incongruencies of facts mentioned in the first text. Subjects who had initially constructed a strong situation model and also knew the related information in the first text - assessed by means of an inference judgment task presented directly after reading the first text - judged inferences concerning the transformations in the second text better and faster than subjects with a originally weak model. To summarise: having available a precise model, being aware of incongruencies, misfits and gaps in one's own knowledge and deep processing new information are factors that promote finding the right balance between the two processes: textbase construction and updating one's situation models.

Studies that investigated the actual processes that occurred during updating and restructuring of the mental representation show that updating is often difficult (Van Oostendorp, in press). The extent to which updating is realised is affected by text characteristics. A relevant factor is the distance in the textbase between the old corrected and new correcting information. Further, in the research of Garcia-Arista, Campanario and Otero (in press) it has been shown that the monitoring and updating strategies were dependent on the context (setting) in which the comprehension task was carried out. In that study readers received the same text containing contradictions, either in a 'scientific' or in a 'newspaper' setting. Readers detected more contradictions in the 'scientific' setting, and regulated their comprehension better than in the newspaper setting. This is probably in line with the observation that credibility of the news source (Chinn & Brewer, 1993) is influential on updating processes. More generally, there seems to be a regulation mechanism that is weighing evidence pro and con to certain situation model, and according to the outcomes, readers do hold on or restructure their existing model.

Updating existing situation models: Thinking skills and training

Table 2 presents a list of skills that are relevant for readers' proficiency in updating existing situation models. Training studies that aim at improving the proficiency in updating or restructuring situation models are scarce. They are highly relevant however, both from a theoretical and a more practical point of view. Chinn and Brewer (1993) developed a framework to help people to create clear-cut conceptual conflicts and to solve these conflicts in a constructive manner. The main components of the suggested strategies, on the readers's side, are involving the 'characteristics of prior knowledge' and the 'processing strategies' a learner is employing. Several strategies mentioned in Table 2 are incorporated in this framework.

How a reader responds to 'anomalous' information is dependent on characteristics of prior knowledge. Particularly important here is the 'entrenchment' of related conceptions and beliefs in the cognitive structure of a learner. The more deeply embedded in a network of other beliefs, the harder a belief is to change. In attempting to reduce the entrenchment of prior beliefs, models etc, it is necessary to figure out precisely why that belief is entrenched, that is, to identify the crucial components. A change in belief and a concomitant change in knowledge - that is, an adequate updating - is difficult without addressing these components (Vosniadou, 1992).

As mentioned above, the processing strategy employed by the learner is also determining the amount and quality of updating. The processing of presented sources may include mental processes as paying carefully attention to the old and new information, deep processing new information and attempting to understand the new model, elaborating the relationships between the new information and competing models, and considering the fullest available range of evidence pro and contra (Nickerson, 1991). Deep processing of information might be promoted by enhancing self-involvement of learners (Tesser & Schaffer, 1990) and letting them to produce self-explanations, that is, to instruct learners to justify their reasoning. Several studies showed that this last instructional strategy is very successful in knowledge building (Chi, deLeeuw, Chiu, & LaVancher, 1994; Coté & Goldman, in press; Perfetti, Rouet, & Britt, in press)

On a higher-order level of skills, going back to the framework Scardamalia and Bereiter (1991) sketched, it is important that the process back and forth between the domain knowledge (already established situation models) and the episodic textbase is monitored effectively. Exclusive concentration on 'situation model checking' maximizes gains in new knowledge at the cost of opportunities to reach significant modification of one's understanding as a result of text studying. This one-sided approach leads to what Bazerman (1985) calls 'selective evaluation'. It is often found in the reading of experts (e.g. physicists): They tend to ignore textual information that is different from their existing schemas and pay much more attention to what can be easily assimilated. Opposed to this, exclusive concentration on the construction of a textbase may inhibit integration with prior knowledge and result in weak subsequent appliance of the text information to real world situations or problems (see Voss, Blais, Means, Greene, & Ahwesh, 1989). Thus a higher-order level skill consisting of a fine-tuned balance of both processes has to be developed.

How to proceed when relevant knowledge is not yet represented in the situation model

Goal to construct a completely new situation model

It often occurs, in particular in an educational context, that a person has to process a large amount of text about a new topic. An example of this type of situation is a student who studies an introductory textbook. The student has to deal with a lot of new terms and subjects that cannot contribute to text understanding, because the relevant knowledge is not yet available. In the situation that we want to address the person's goal is to understand the text and to use the new information to increase his or her body of knowledge. In this combined reading and knowledge acquisition task the reader wants to construct a textbase (see arrow (1) in figure 1) and build new mental models (see arrow (4) in figure 1). Although this situation occurs very frequently, many specific questions about the process are but partly answered.

Three questions are discussed in this chapter. The first, and best researched question, concerns main point selection. Considering the complexity of the task, the reader has to be selective in deciding which concepts and relations are so important that they should be processed (and incorporated in a mental model) with priority. How is this selection task performed by proficient and less proficient students? The second question pertains to the reading process. How does the reader cope with the overload of new information that he or she cannot properly understand and how is this information integrated into a coherent textbase? The third question concerns the construction of new mental models by text studying. It is well known that the construction of a new model, for instance of a theoretical concept, demands a variety of processes, like distinguishing examples and counter-examples. Is it possible for students to realise the required activities during text studying? Do students conceive the construction tasks as problematical and what are the skills involved in solving or circumventing the problem? Table 3 provides an overview of the difficulties and skills which we will discuss in the next subsections.

Processes, variables, and difficulties in reading a text under conditions of minimal prior knowledge

Several studies that focus on the selection of main points in a text show that lack of prior knowledge is indeed an important factor. Dee-Lucas and Larkin (1986) compared strategies of novices and experts who were invited to select important sentences in scientific texts and to rate the importance of text sentences. Novices, like experts, considered definitions to be more important than facts and made use of the hierarchical structure of the text. Novices, however, tended to overgeneralise their own, self-imposed rules regarding the importance of information type and passage organisation. For instance, they attended to unimportant definitions in scientific texts at the expense of more important information of other types.

Novices, in other words, were highly dependent on external cues. Brown, Day,

Table 3 Building New Situation Models by Studying

Goal: building new situation models
Condition: a shortage of relevant prior knowledge

Subtasks
- selecting main points
- reading a text that conveyes an overload of new information
- constructing new components of the mental model

Difficulties
- main point selection has to be based mainly on external cues
- a cyclic studying process is needed to handle the overload of new information and to work one's way up to deeper understanding
- integrating new information into the existing knowledge structure involves the use of schemas ('models')

Skills
- Basic level
  . functional capacity of working memory
  . decontextualisation skill
- Strategic level
  . using thinking skills like inductive reasoning, comparing, categorising and abstracting
  . using study skills, like concept-mapping and schematising
  . using goal-directed tactics, like 'taking different perspectives' and 'listing to cram it all in'
  . resisting 'easy' approaches that promote bias and preservation of misconcepti-ons
  . using schemata ('models') to integrate new information
- Higher-order level
  . using complex study methods, like SQ3R, to orchestrate the use of thinking and study skills and to relate learning goal to local criteria
  . balancing processes directed at construction of a text base and building new mental models
  . assessing the need for additional (social) resources

and Jones (1983) found, in a study on note taking (writing down main points), that novice students often use a simple strategy, called the 'copy-delete' strategy: Students copy 'important sentences' (like definitions) and ignore the remaining text. Experts, who have relevant prior knowledge, and students who have developed good study habits use more complex strategies, based on drawing inferences, restructuring and integration. Schellings (1995) and Schellings and Van Hout-Wolters (1995) studied differences between novices (students in secondary education) and experts (biology teachers) in underlining main sentences in instructional texts, and they compared the main point selection of students and their teachers. They found large differences in selected main points, both between groups and within groups. Interviews with students revealed that they used three different approaches: 'educational' (the student tries to select the points which are probably considered important by the teacher), 'linguistic' (the student uses linguistic cues) and 'personal interest' (the student selects what he or she finds personally interesting). Correspondence between students and teachers, in the selected main points, was highest for students who reported the use of the educational approach. A second study by the same authors, showed that proficient students were able to adjust their selection strategies to external conditions and instructions, while weak students were less flexible.

To summarise: In the absence of appropriate prior knowledge, readers develop their own selection rules (like 'definitions are important'), use linguistic information, try to predict what the teacher will consider important, and take into account their own ideas about what is interesting. Proficient students are less susceptible to overgeneralisation than weak students and adjust their selection strategies more flexibly to external circumstances.

The second issue concerns the reading process. How does reading take place when a text contains so much new information that it is very difficult to perform the usual activities, like making inferences to foster understanding and chunking information into appropriate models or schemata? It is stated in the theory of Van Dijk and Kintsch (1983) that integrating new information in an adequately organised macrostructure, is one important way to reduce the overload that threatens reading processes. The macrostructure constructed during reading can guide the processing of subsequent information (Aarnoutse, 1982; Guindon & Kintsch, 1982). When a student reads the same text a number of times, this might lead to a gradual growth of understanding. Neisser (1976) described a 'comprehension cycle' that students might use to bootstrap their way up: Parts of the new information, however small, are used to enlarge prior knowledge. This enlargement influences subsequent processing, leading again to a knowledge modification, and so on. Depending on the number of cycles, comprehension can be deeper or more superficial. According to this view, the quality of the new mental models is influenced by the number of cycles, and thus by the students' skill to phase the studying processes, to monitor their own level of understanding, and to diagnose which passages to reread.

Whether a person is able and willing to transform a 'reading' process into a cyclic 'studying' process, probably depends on external conditions and task conception. One rather fundamental difficulty is, however, that many students find it difficult to decide whether the optimal learning result is reached yet. During text studying it is far from clear what criteria, like 'understanding' and 'integration in the personal knowledge base', mean (Entwistle, 1997). It is even less clear which degree of understanding and integration is attainable at all for a novice who enters a new domain by studying an introductory text (Elshout-Mohr & Van Daalen-Kapteijns, 1985). In a study on the acquisition of new conceptual knowledge by students, Elshout-Mohr and Van Daalen-Kapteijns (1990) focused on the way students regulated their text studying processes. Analysis of protocols of students who were 'thinking aloud' during text studying showed that students steer the process by self-imposed learning goals, but also by self-imposed local learning criteria. A 'learning goal' was defined as an objective that is to some extent abstracted from the text at hand (decontextualised). Examples of learning goals are 'to be able to use newly acquired conceptual knowledge to tell something about an unfamiliar aspect of a phenomenon or object' and 'to be able to explain the relation between concepts'. A 'local learning criterion' was defined as an objective that is strictly related to the text at hand. Examples of local learning criteria are 'to be able to name the four aspects on page 16' and 'to be able to give an overview of the correspondences between the graph on the upper half of page 17 and the associated text on the lower half of the page'. Students differed in the role that they attributed to local learning goals and criteria. Some students considered the text studying task fulfilled as soon as the local learning criteria were achieved. These students often worked very hard, but the effect was limited because they restricted their attention to information that was text-bound and not connected to already available personal knowledge. Other students were primarily interested in achievement of learning goals. Those students combined top-down processing of the text, guided by learning goals, and bottom-up processing guided by local learning criteria. They often experienced tension between what they wanted to achieve (the learning goals) and the local textual information. A student for instance studied a text about 'frustration'. She wanted to find out 'What is frustration?', but she was tempted to switch over to the question 'When does frustration occur?' simply because the answer to the latter question was more salient in the text than the answer to the former question. Solving mis-matches between personal learning goals and the content of the text at hand, is one of the major problems of studying text about new topics. In actual practice the majority of students is pragmatic: when in doubt about goals and criteria, they just memorise main points. This provides at least the 'illusion of knowing', not only for the students but also for those teachers who use reproductive assessment procedures.

Under circumstances like the ones that we discussed just now, textbase construction is viewed as subordinated to constructing new mental models. The balance between the two construction processes can also be very different. In that case the student is primarily focused on text comprehension and likely to forget that the studying process is not yet finished with the construction of a coherent textbase. To bring the process to an end, the student must see to the model building part of the studying process (arrow (4) in figure 1). One possibility is to memorise parts of the textbase and incorporate those in long term memory. An often preferred alternative requires that one or more new situation models be built by incorporating new information in appropriate conceptual schemata. To accomplish this the student must have such schemata ('models') available. Van Daalen-Kapteijns and Elshout-Mohr (1981) investigated the use of models in a study on the acquisition of the meaning of new words. High and low verbal university students were instructed to infer the meanings of neologisms from contexts. An example of one of the neologisms used was 'kolper', meaning 'a window that transmits little light because of something in front of it'. An example of one of the five context sentences for this neologism was: 'I was afraid that the room might have kolpers, but when I went and saw it turned out that plenty of sunlight came into it'. All students tried to construct a coherent textbase for each context, but to construct new word meanings (verbal units that could be incorporated in semantic memory) they had to go one step further. This step involved the use of 'models' or word meaning schemata. In this respect high and low verbal students differed. High verbal subjects used models that were better structured and more appropriate to decontextualise and assemble aspects of the new word's meaning than the models of low verbal subjects. This finding was replicated with younger students, aged 11-12 (Van Daalen-Kapteijns, Elshout-Mohr, De Glopper, & Schouten-van Parreren, 1997). It is highly similar to the finding, discussed earlier, that updating of existing knowledge is in general performed better by subjects who have a more precise and accessible situation model than for subjects with a weaker model.

Building new situation models by studying: Thinking skills and training

Table 3 presents a list of skills that are relevant for readers' proficiency in building new situation models. Student learning is a domain in which many studies have been performed. Reviews are presented by Bereiter and Bird (1985), Dansereau (1985), Pressley, Borkowski, and Schneider (1989), and Weinstein and Van Mater Stone (1996). Trainers and researchers have described and tested many strategic and higher order skills to handle textual information in the absence of adequate prior knowledge.

Many of the strategic skills are thinking skills, like the inductive reasoning skill that is relevant for any reading process. However, text studying demands additional thinking skills like comparing, categorising and abstracting. Such skills are needed to make optimal use of external cues, to organise information and diminish cognitive load. To help students to use the skills in a coordinated manner, trainers designed 'study methods' in which those skills are incorporated. Much recommended and frequently used are methods that involve spatial strategies, like concept mapping and schematising (Holley, Dansereau, McDonald, Garland, & Collins, 1979). Special tools to facilitate studying specialised texts, like for instance scientific research reports, have been developed also (Elshout-Mohr, 1983; Novak & Gowin, 1984). The meticulous use of study methods does not diminish students' dependence on 'external cues' but promotes flexibility. Being aware that the same text can be approached on several levels and from different perspectives is one of the characteristic of proficient students. Johnston and Afflerbach (1985) analysed the tactics of proficient readers who were invited to read very difficult texts on unfamiliar subjects, and they observed for instance the following tactics:

- Listing to 'cram it all in'.
- Refocusing on a different level of text.

- Scanning to find the source of difficulty.
- Consolidation or reviewing to 'firm it up'.

Not all strategies used by experts are recommendable, however. One strategy worth mentioning here is the strategy to concentrate processing on information that is recognised as vaguely familiar and to skip the rest (Bazerman, 1985). It will be clear that this strategy reduces overload at a very high cost. The reader risks overlooking misconceptions, and the strategy does help to extend existing mental models, but not to build new ones. An overview of textbook study strategies is given by Cavely and Orlando (1991). They conclude that "most study strategies are effective, but no study strategy is appropriate for all students in all study situations." (p. 155).

Basic skills and higher order skills are important factors in solving the difficulties that we discussed. Two relevant basic skills are the functional capacity of working-memory and decontextualisation skill. Whitney, Ritchie, and Clark (1991) compared readers with low and high working-memory capacity who were reading difficult passages. They found that readers with a low working-memory capacity made very specific inferences and that many of these inferences turned to be wrong later in the passages. Opposed to this, high working-memory capacity readers used more general inferences that left the interpretation more open ended. By keeping their options open, and remembering these options, high span readers were able to make their decision later in the passage, when they were more likely to be correct. We may conclude from this finding that the (cyclic) way to deeper understanding (and building a coherent text base) is substantially less tiresome for high-span readers who can delay choices until they have a good chance to be correct, than for low-span readers who make early choices which are often wrong.

A similar conclusion follows from the study by Van Daalen-Kapteijns and Elshout-Mohr (1981) in which word meaning acquisition by students with high and low verbal ability scores (on a vocabulary test) were compared. As we mentioned earlier all students tried to construct a coherent text base for each context, but the high verbal students were better able to decontextualise word meaning aspects; these aspects were subsequently tested and accumulated over various sentences into one coherent mental model of the word's meaning. The low verbal students were less able to decontextualise word meaning aspects and this made it more difficult for them to test and accumulate their findings. Low proficiency on basic skills made the complex task even more complicated for low verbal students.

Higher order skills are of course essential to orchestrate all processes involved and to maintain an adequate balance between textbase construction and building new situation models. To teach students to steer these processes effectively, several higher order study strategies have been developed. One such strategy is the so called SQ3R-strategy. It directs students to perform the following five steps: Survey the material and convert the subheadings in the text into Questions; then start Reading, Reciting and Reviewing the textual information. For students who apply this method, the 'questions' serve as 'learning goals' that enable them to combine a top-down (goal-driven) approach with a bottom-up (text-driven) approach.

Studies on the effects of strategy training, like the ones that were mentioned in the beginning of the subsection, emphasise that the best prospects are in long term training programs in which both strategic skills and higher order skills are incorporated. An additional idea is that classroom culture should be transformed as well to create a social and motivational climate that is optimal for learning how to study texts. This latter idea is eloquently elaborated by Goldman (1997). She argued that research about learning from text is mostly grounded on 'individualistic models of cognitive activity', and she proposed to research text studying as it might occur 'in classroom learning environments that support thinking and collaborative activity'. In the settings that Goldman had in mind, multiple sources of information play an important role in the learning process. These multiple sources may include texts and study texts, but also peers and knowledgeable others. We tend to agree that students who participate in such settings are likely to gain useful experience with different goals of text studying, with different levels of text understanding, with the use of various types of resources that can promote progress when text comprehension is faltering, and with the use of multiple sources of information to test and consolidate newly acquired situation models. These settings might, for one thing, help solve the difficulty we mentioned earlier, that it often occurs that a single study text provides insufficient information to build and test a new situation model. This approach of learning from text brings to the fore new skills that might be important for readers, like communicative skills and skill in the assessment of 'resource needs' before or during the text studying process (Silverman, 1995). A discussion of these skills is, however, beyond the scope of this chapter.

Discussion and conclusions

We started this chapter with the statement that reading is a complicated process. Now, at the end we are even more convinced that this judgement is fair. A great number of relevant activities have been distinguished such as activating prior knowledge, and each of these activities generated its own specific difficulties. By consequence, a proficient reader or learner has to master a lot of skills. The most typical skills are probably located at the level of strategic skills. It is probably true that readers should have available a great number of strategies or strategic skills for solving the complicated task of reading and studying. Just because of this, readers should also have available higher order skills - in order to monitor and regulate the strategies - as well as basic skills - in order to reduce cognitive overload.

We think that it is useful to analyse all these different components of reading and studying, to identify all the different difficulties and to examine what thinking skills are needed in different situations. On the basis of this analysis it is possible to design training programs focused on the specific circumstances. Garner (1990) underscores the importance of the specific setting of training programs. To promote transfer possibilities of learned skills to settings where skills have to be practiced, one has to analyse the specific circumstances (see also Elen, 1992).

Though we distinguished a great number of activities, it does not imply that there is no overlap in the processes involved. For instance, mobilising prior knowledge is needed to understand a text (arrow (1) in figure 1), but it is also needed in the situation of updating (arrow (3)) or in the situation of selecting main points which is relevant to acquiring new knowledge (arrow (4)). The same holds for, e.g., working memory capacity. The same 'building blocks' emerge in different situations though in different relationships to each other, making up different complex skills.

We focused in this chapter on three aspects of reading and text studying and treated them separately, because these three aspects occur very frequently and play each a central role in the reading and study process:

- Using prior knowledge is central to reading and comprehending of informative texts as well as of narrative texts. The degree of activation of prior knowledge is crucial. Too much activation of prior knowledge, too many inferences and elaborations provide the danger of activating irrelevant ideas, hindering to grasp the central message of the text. Too little activation leads to not-detecting inconsistencies and to an understanding of the text that is too sloppy and not connected to prior knowledge.
- Updating is a crucial activity in all situations where representations have to be replaced by new ones. In these cases readers should be able to decide whether replacement is desirable, and if so, to what degree. An additional difficulty here is that the task of updating of an old representation is very different from building a coherent textbase. Consequently, there are two competing activities (or processes), and learners have to find an adequate balance between the two in terms of time and effort.
- Building new knowledge on the basis of text information is a central activity in all situations where the goal of reading is the acquisition of knowledge. It is crucial here that readers are able to select information on the basis of linguistic characteristics and goals and that they understand the selected information and build new situation models. This task puts, in general, heavy demands on readers, especially when it concerns a text where readers have relatively low prior knowledge.

We have described a number of programs focused on teaching to read, based on theoretical notions we distinguished, e.g., that of Brown and colleagues. That is, however, not the end of the story. A number of (ecologically) important ways of reading have still to be implemented in educational programs. We mention here one: the reading and understanding of multiple, related texts (e.g. newspaper articles) or documents from different sources (Goldman, 1997; Perfetti et al., in press). It will be clear that here particularly updating processes are crucial.

There is still need to study the training of skills. For example, what is the optimal order of learning different (basic, strategic, or higher-order) skills? It is not difficult to raise a number of other interesting and also important research questions. To avoid misunderstanding, we do not wish to claim that all difficulties we distinguished have to be trained, or that all skills we distinguished have to become subject of formal training. For a great part they describe characteristic difficulties readers face with understanding and learning new text materials. Most readers will automatically acquire the required skills after practice in school or at home. For some readers, however, the difficulties encountered remain problematic, and they don't acquire the strategies to cope with these difficulties. For these cases it is useful to know exactly what the nature of the difficulties is and how the deficits have to be trained.

Finally, we saw that skills such as mobilising prior knowledge are important to reading and text studying, but prior knowledge also has to be available. Our concluding remark is that fostering the acquisition of well-organised and person-involved knowledge is at least as important as the training of skills.

References

Aarnoutse, C.A.J. (1982). *Aspecten van begrijpend lezen in het vierde leerjaar van het gewoon lager onderwijs* [Aspects of reading comprehension in the fourth year of elementary school]. Nijmegen: Berkhout.

Baker, L., & Brown, A.L. (1984). Metacognitive skills in reading. In P.D. Pearson (Ed.). *Handbook of reading research* (pp. 353-395). London: Longman.

Bazerman, C. (1985). Physicists reading physics: Schema-laden purposes and purpose-laden schema. *Written Communication, 2*, 3-23.

Beck, I.L., McKeown, M.G., Worthy, J., Sindora, C., & Kucan, L. (1996). Questioning the author: A year-long implementation to engage students with text. *Elementary School Journal, 96*, 385-414.

Beck, I.L., Perfetti, C.A., & McKeown, M.G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74*, 506- 521.

Bereiter, C. (1990). Aspects of educational learning theory. *Review of Educational Research, 60*, 603-624.

Bereiter, C., & Bird, M. (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. *Cognition and Instruction, 2*, 131-156.

Berlyne, D.E. (1965). *Structure and direction in thinking*. New York: Wiley.

Brown, A.l., Day, J.D., & Jones, R.S. (1983). The development of plans for summarizing texts. *Child Development, 54*, 968- 979.

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

Campanario, J., & Van Oostendorp, H. (in press). Updating mental representations when reading scientific text. In J.F. Rouet, J. Levonen & A. Biardeau (Eds.). *Proceedings of the using complex information systems conference*. Poitiers.

Cavely, D.C., & Orlando, V.P. (1991). Textbook study strategies. In Flippo, R.F. & Caverly, D.C. (Eds.), *Teaching reading and study strategies at the college level* (pp. 86-166). Newark: International Reading Association.

Chi, M.T.H., deLeeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439-477.

Chinn, C.A., & Brewer, W.F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63,* 1-49.

Coté, N., & Goldman, S.R. (in press). Building representations of informational text: Evidence from children's think-aloud protocols. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representations during reading*. Mahwah, NJ: Erlbaum.

Daneman, M. & Carpenter, P.A. (1983). Individual differences in working memory and reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9,* 561-584.

Dansereau, D.F. (1985). Learning strategy research. In S.C.J. Segal & R. Glaser (Eds.), *Thinking and learning skills* (Vol. 1, pp. 209-241). Hillsdale, NJ: Erlbaum.

Dee-Lucas, D., & Larkin, J.H. (1986). Novice strategies for processing scientific texts. *Discourse Processes, 9,* 329-354.

Elen, J. (1992). *Toward prescriptions in instructional design: A theoretical and empirical approach.* Leuven: Katholieke Universiteit.

Elshout-Mohr, M. (1983). Communicatiepatronen: een hulpmiddel bij het bestuderen van studieteksten [Communication models: A tool in studying texts]. *Tijdschrift voor Taalbeheersing, 5,* 1-18.

Elshout-Mohr, M. (1992). Metacognitie van lerenden in onderwijsleerprocessen [Metacognition of students in teaching-learning processes]. *Tijdschrift voor Onderwijs Research, 17,* 273-290.

Elshout-Mohr, M., & Van Daalen-Kapteijns, M.M. (1985). Het leren van begrippen, in het bijzonder in het eerste stadium van het hoger onderwijs [Concept learning in the first phase of higher education]. *Pedagogische Studiën, 62,* 459-470.

Elshout-Mohr, M., & Van Daalen-Kapteijns, M.M. (1990). Self-imposed local learning criteria in studying text. In H. Mandl, E. De Corte, N. Bennett, & H.F. Friedrich (Eds.), *Learning and instruction* (Vol. 2.2, pp. 253-271). Oxford: Pergamon Press.

Entwistle, N. (1997, August). *Metacognitive and strategic awareness of learning processes and understanding*. Paper presented at the Sevenyh European Conference for Research on Learning and Instruction. Athens, Greece.

Erickson, T.D., & Mattson, M.E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior, 20*, 540-551.

Flavell, J.H. (1979). Metacognition and cognitive monitoring. A new area of cognitive-developmental inquiry. *American Psychologist, 34,* 906-911.

Franks, J.J., Vye, N.J., Auble, P.M., Mezinski, K.J., Perfetto, G.A., Bransford, J.D., Stein, B.S., & Littlefield, J. (1982). Learning from explicit versus implicit texts. *Journal of Experimental Psychology: General, 3*, 416-422.

Garcia-Arista, E., Campanario, J.M., & Otero, J. (in press). Influence of subject matter setting on comprehension monitoring. *European Journal of Psychology of Education.*

Garner, R. (1990). When children and adults do not use learning strategies: Toward a theory of settings. *Review of Educational Research, 60*, 517-529.

Glenberg, A.M., Wilkinson, A.C., & Epstein, W. (1982). The illusion of knowing: Failure in self-assessment of comprehension. *Memory and Cognition, 10*, 597-602.

Goldman, S.R. (1997). Learning from text: Reflections on 20 years of research and suggestions for new directions of inquiry. *Discourse Processes, 23*, 357-398.

Guindon, R., & Kintsch, W. (1982). *Priming macrostructures*. Boulder, CO: University of Colorado.

Holley, C.D., Dansereau, D.F., McDonald, B.A., Garland, J.C., & Collins, K.W. (1979). Evaluation of a hierarchical mapping technique as an aid to prose processing. *Contemporary Educational Psychology, 4*, 227-237.

Hunt, E., Lunneborg, C., & Lewis, J. (1975). What does it mean to be high verbal? *Cognitive Psychology, 7,* 194-227.

Johnson, H.M., & Seifert, C.M. (1994). Sources of the continued effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 1420-1436.

Johnson-Laird, P.N. (1983). *Mental models*. New York: Cambridge University Press.

Johnston, P., & Afflerbach, P. (1985). The process of constructing main ideas from text. *Cognition and Instruction, 2*, 207- 232.

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

Kardash, C.M., & Scholes, R.J. (1996). Effects of preexisting beliefs, epistomological beliefs, and need for cognition on interpretation of controversial issues. *Journal of Educational Psychology, 88*, 260-271.

Kintsch, E., & Kintsch, W. (1996). Learning from text. In E. De Corte & F.E. Weinert (Eds.), *International encyclopedia of developmental and instructional psychology* (pp. 519-525). Oxford: Pergamon.

Kintsch, W. (1980). Learning from text, levels of comprehension, or: Why anyone would read a story anyway. *Poetics, 9,* 87-98.

Kintsch, W., & Franzke, M. (1995). The role of background knowledge in the recall of a news story. In R.F. Lorch & E.J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 177-196). Hillsdale, NJ: Erlbaum.

Kintsch, W., & Vipond, D. (1979). Reading comprehension and readability in educational practice and psychological theory. In L.G. Nilsson (Ed.), *Perspectives on memory research* (pp. 329-365). Hillsdale, NJ: Erlbaum.

Mannes, S. (1994). Strategic processing of text. *Journal of Educational Psychology, 86,* 377-388.

McNamara, D.S., Kintsch, E., Songer, N.B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14,* 1-43.

Merrill, E.C., Sperber, R.D., & McCauley, C. (1981). Differences in semantic encoding as a function of reading comprehension skill. *Memory and Cognition, 11,* 12-21.

Neisser, U. (1976). *Cognition and reality. Principles and implications of cognitive psychology.* San Francisco: Freeman.

Nelson, T.O. (1996). Consciousness and metacognition. *American Psychologist, 51,* 102-117.

Nickerson, R.S. (1991). Modes and models of informal reasoning: A commentary. In J.F. Voss, D.N. Perkins, & J.W. Segal (Eds.), *Informal reasoning and education* (pp. 291-309). Hilsdale, NJ: Erlbaum.

Novak, J., & Gowin, D.B. (1984). *Learning how to learn.* Cambridge, MA: Cambridge University Press.

Palincsar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1,* 117-175.

Paris, S.G., Cross, D.R., & Lipson, M.Y. (1984). Informed strategies for learning; A program to improve children's awareness and comprehension. *Journal of Educational Psychology, 76,* 1239-1252.

Paris, S.G., & Oka, E.R. (1986). Children's reading strategies, metacognition and motivation. *Developmental Review, 6,* 25-56.

Peeck, J., Van den Bosch, A.B., & Kreupeling, W.J. (1982). The effect of mobilizing prior knowledge on learning from text. *Journal of Educational Psychology, 74,* 771-777.

Perfetti, C.A., & Lesgold, A.M. (1977). Discourse comprehension and sources of individual differences. In M. Just & P.A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 265-287). Hillsdale, NJ: Erlbaum.

Perfetti, C.A., Rouet, J.F., & Britt, M.A. (in press). Toward a theory of documents. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representations during reading.* Mahwah, NJ: Erlbaum.

Pressley, M., Borkowski, J.G., & Schneider, W. (1989). Good information processing: What is it and how education can promote it. *International Journal of Educational Research, 13*, 857-867.

Pressley, M., Snyder, B.L., & Cariglia-Bull, T. (1986). How can good strategy use be taught to children? Evaluation of six alternative approaches. In S.M. Cormier & J.D. Hagman (Eds.), *Transfer and learning* (pp. 81-119). New York: Academic Press.

Reder, L., & Kusbit, G.W. (1991). Locus of the Moses Illusion: Imperfect encoding, retrieval or match? *Journal of Memory and Language, 30*, 385-406.

Resnick, L.B. (1987). *Education and learning to think.* Washington, DC: National Academy Press.

Rukavina, I., & Daneman, M. (1996). Integration and its effect on acquiring knowledge about competing scientific theories from text. *Journal of Educational Psychology, 88*, 272-287.

Scardamalia, M., & Bereiter, C. (1991). Literate expertise. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise.* New York: Cambridge University Press.

Schank, R.C., & Abelson, R. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Erlbaum.

Schellings, G.L.M. (1995). *An educational approach to selecting main point in texts.* Amsterdam: University of Amsterdam.

Schellings, G.L.M., & Van Hout-Wolters, B.H.A.M. (1995). Main points in an instructional text, as identified by students and by their teachers. *Reading Research Quarterly, 30*, 742-756.

Schnotz, W., & Ballstaedt, S.-P. (1996). Comprehension: teaching and assessing. In E. De Corte & F.E. Weinert (Eds.), *International encyclopedia of developmental and instructional psychology* (pp. 492-498). Oxford: Pergamon.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology, 82*, 498-504.

Silverman, B. (1995). Computer supported collaborative learning. *Computers & Education, 25*, 81-91.

Stein, B.S., Bransford, J.D., Franks, J.J., Owings, R.A., Vye, N.J., & McGraw, W. (1982). Differences in the precision of self-generated elaboration. *Journal of Experimental Psychology: General, 4*, 399-405.

Tesser, A., & Shaffer, D.R. (1990). Attitudes and attitude change. *Annual Review of Psychology, 41*, 479-523.

Van Daalen-Kapteijns, M.M., & Elshout-Mohr, M. (1981). The acquisition of word meanings as a cognitive learning process. *Journal of Verbal Learning and Verbal Behavior, 20*, 386-399.

Van Daalen-Kapteijns, M.M., Elshout-Mohr, M., De Glopper, K., & Schouten-Van Parreren, C. (1997). *Natuurlijke strategieën bij het afleiden van woordbetekenissen uit context* [Natural strategies in inferring word meaning from context]. Amsterdam: SCO-Kohnstamm Institute.

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

Van Oostendorp, H. (1991). Inferences and integrations made by readers of script-based texts. *Journal of Research in Reading, 14,* 3-21.

Van Oostendorp, H. (1994). Text processing in terms of semantic cohesion monitoring. In H. Van Oostendorp & R.A. Zwaan (Eds.), *Naturalistic text comprehension.* Norwood, NJ: Ablex.

Van Oostendorp, H. (1996). Updating situation models derived from newspapers articles. *Medienpsychologie, 8,* 21-33.

Van Oostendorp, H. (in press). Holding on to established viewpoints during processing news reports. In W. van Peer & S. Chatman (Eds.), *Narrative perspective: Cognition and emotion.* Buffalo, NY: Suny University Press.

Van Oostendorp, H., & Anbeek, G.H.J. (1983). *Tekstverwerking: Evaluatie van samenhang en verschillen tussen lezers* [Text Processing: Evaluation of coherence and differences between readers]. Utrecht: Utrecht University.

Van Oostendorp, H., & Bonebakker, C. (1996). Het vasthouden aan incorrecte informatie bij het verwerken van nieuwsberichten [Holding on to incorrect information during processing news reports]. *Tijdschrift voor Communicatie-wetenschap, 24,* 57-74.

Van Oostendorp, H., & Bonebakker, C. (in press). Difficulties in updating mental representations during reading news reports. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representations during reading.* Mahwah, NJ: Erlbaum.

Van Oostendorp, H., & De Mul, S. (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica, 74,* 35-46.

Van Oostendorp, H., & Kok, I. (1990). Failing to notice errors in sentences. *Language and Cognitive Processes, 5,* 105-113.

Vosniadou, S. (1992). Knowledge acquisition and conceptual change. *Applied Psychology: An International Review, 41,* 347-357.

Voss, J.F., Blais, J. Means, M.L., Greene, T.R., & Ahwesh, E. (1989). Informal reasoning and subject matter knowledge in the solving of economics problems by naive and novice individuals. In L.B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 217-249). Hillsdale, NJ: Erlbaum.

Weinstein, C.E., & Van Mater Stone, G. (1996). Learning strategies and learning to learn. In E. De Corte & F.E. Weinert (Eds.), *International encyclopedia of developmental and instructional psychology* (pp. 419-423). Oxford: Pergamon.

Whitney, P., Ritchie, B.G., & Clark, M.B. (1991). Working-memory capacity and the use of elaborative inferences in text comprehension. *Discourse Processes, 14,* 133-145.

Wilkes, A.L., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *Quarterly Journal of Experimental Psychology, 40A,* 361-387.

13

Evaluating the effectiveness of cognitive programmes: Some methodological issues

W. Hager

Introduction

This volume as well as many other recent publications show that there is a growing interest in various interventions intending to foster cognitive skills, strategies, abilities, performances, and/or competencies. Any intervention programme aiming at a cognitive domain can be called a 'cognitive programme'. This name can also be given to any programme which aims at other goals but seeks to reach them by means of changing cognitive processes as is the case with mental training for athletes (cf. Orlick, 1986). All these programmes "should be based on a theory of intellectual performance specifying mental processes" (Sternberg, 1983, p. 6; see also Brown, Bransford, Ferrara, & Campione, 1983). But theories 'specifying mental processes' belong to the domain of basic research such as general or differential psychology, and they usually refer to phenomena, events, and processes and to their precise description and especially they try to explain them. As Bunge (1967) has pointed out, such a theory only can (and of course should) give a general framework for constructing programmes and for further theoretical considerations; it should 'inspire' authors of programmes. Generally they do not contain a component that tells us how to act if we want to change processes or phenomena. But since intervention programmes aim at changes they must also contain rules of action which tell us how to act in order to achieve these changes. This is the main reason why it is not possible to deduce programmes logically from general theories, but they can be created by adding general rules of action and personal knowledge including 'extra-scientific' experience to the inspiring theory (Herrmann, 1984). In the end we get 'systems of technological rules' according to Bunge (1967), that is, special rules of action. Thus, authors like Bunge (1967) and Cronbach (1983) distinguish basic from technological research, and Herrmann (1984) adds the activities of psychologists who work practically without doing any research. Technological research in the first place aims at providing the latter colleagues tools and knowledge so that they can be more efficient in their profession.

Considering cognitive programmes as '(systems of) technological rules', they have to meet with general criteria quite different from those for a psychological theory referring to basic phenomena such as mental processes: They must be "effective, reliable, without negative side effects, easily applicable, and economic" (cf. Herrmann, 1984, p. 28). Even if the theory inspiring them is imprecise, incomplete, contradictory, or even lacking, this does not affect technological rules as long as they turn out to be effective in practical circumstances. This also means that the statement that a programme is well-founded theoretically does not imply its effectiveness.

For this reason the basic question concerning programmes refers to whether its intentions, goals, or objectives are attained. To answer it, the effectiveness of a programme has to be evaluated and demonstrated empirically. Although the basic concepts of evaluation research have been known for some time, they have not always been taken into account when evaluating cognitive or other programmes, and especially some of their consequences for designing an evaluation study are disregarded of in many instances. For this reason these consequences are discussed in some detail. The point of view taken in this discussion and underlying my arguments is strictly hypothesis testing, where 'hypothesis' mostly refers to 'effectiveness hypotheses'. Usually, however, these hypotheses cannot be derived from a theory; so where do testable hypotheses come from? The programme authors' hopes and assertions that their programme is effective (if they were to think otherwise, they would not have developed a programme) can (and should) be framed as an hypothesis to be examined by means of scientific methods (see also Popper, 1992). Taking the view of the programme authors, the most obvious hypotheses concern the (global) effectiveness of their programmes. This does not preclude that other colleagues are of different opinions which will, of course, lead to different hypotheses, but this fact does not invalidate the subsequent considerations.

Types of programmes, goals of programmes, and criterion variables

There are three basic questions the authors of a particular programme should attend to: "(a) what to teach, (b) how to teach it, and (c) how to adapt what and how to individual differences, that is, who is being taught" (Brown et al., 1983, p. 126). The second and third questions by Brown et al. (1983) lead to distinguishing a number of components of cognitive programmes such as the problems or tasks, the instructional methods, and their appropriateness for or their adaptability to participants of different abilities and/or age; with a few exceptions, these subsystems of technological rules cannot be presented or discussed in this chapter.

The first question addressed by Brown et al. (1983) refers to the goals of a programme. On a general level, programmes can aim at improving performances, at improving competencies or abilities or aptitudes (which includes some kind of transfer), and at accelerating development (see for another classification Hanson, 1984, p. 396). These general aims may refer to particular tasks, to one or more skills, to declarative and/or procedural knowledge and to strategies applicable to the successful solving of different problems, that is, they may be narrow and specific or broad and general: Programmes to improve reading comprehension or inductive reasoning may be called rather specific programmes, whereas programmes to improve broader abilities or aptitudes such as intelligence, thinking, or reasoning may be called (more) general programmes. The most ambitious (broad) goal consists in accelerating development, and only few programmes explicitly aim at this goal, for example the programme CASE by Adey, Shayer, and Yates (1989). When choosing a goal, there should be good reason to believe that it is attainable for the persons for whom the intervention is designed; theories and empirical findings especially from the domain of developmental psychology should provide answers to this important question. For example, the operational efficiency of the short-term memory span seems to be one ability which develops, but cannot be improved by interventions (cf. Case, Kurland, & Goldberg, 1988). It is beyond the scope of this chapter to address various other possible goals (see, e.g., Brown et al., 1983; Datta, 1986; Friedrich & Mandl, 1992; Hanson, 1984) or to discuss their adequacy (see, e.g., Nickerson, 1987; Paul, 1990).

Given a multitude of different goals the question to be addressed next refers to the criterion variables which are appropriate with respect to these goals. General claims to use 'multiple criteria' or using global measures of 'success' may lead astray as long as it is unclear in which way these multiple or global measures relate to the target programme and its goals (see Bransford, Stein, Arbitman-Smith, & Vye, 1985; Campione & Armbruster, 1984; Dansereau, 1985; Hager, 1995). But which criteria should be used?

Training, coaching, transfer, and competence

The most modest claim to be ascribed to a particular programme is that it enhances a chosen performance as a result of teaching specific knowledge or specific strategies. Especially in the United States, many programmes exist whose goals consist in 'coaching' for a certain test or a particular exam or the like. To this end, participants practice tasks which are very similar, parallel, or even identical to those of the test, and they do so shortly before they have to perform the test. Usually nothing is done to enable transfer to situations other than this particular test or exam.

Although the exact definition of 'coaching' is a matter of some debate, "the essential point of agreement is that special preparation and coaching usually refer to efforts to raise examinees' scores on a particular target examination" (Haney, 1990, p. 123; see also Anastasi, 1981). According to Pike (1978), such scores can be influenced through three different means: (a) the true score component can be changed, that is, the skills and abilities underlying the score and probably assessed by the criterion test are improved; (b) the primary test-specific component is improved, that is the "examinees' familiarity with the format, conventions, and particular types of tasks encountered in the target test" is improved (Haney, 1990, p. 123); and (c) the secondary test-specific component is improved, that is the "examinees' levels of confidence and efficiency in test taking" is improved (Haney, 1990, p. 123). The two test-specific components defined by Pike (1978) can be changed by practice and coaching, as various reviews show (e.g., Kulik, Bangert-Drowns, & Kulik, 1984; Sackett, Burris, & Ryan, 1989), but there is evidence that these changes are of short duration and that they 'may be of little use in life activities' (Anastasi, 1981, p. 1089).

On the other hand, cognitive programmes usually aim at changing the true score component addressed by Pike (1978), that is, at altering the skills and abilities underlying better performance; their scope is broader than mere coaching and it includes some kind of transfer. If the term 'competence' is used to mean dispositions to act which are relatively stable over time, one can say that cognitive training try to improve competencies, referring to skills, abilities, aptitudes, knowledge, strategies, and so on; these improvements must last for a while. Thus. programmes which change competencies, which enable transfer, and whose effects last for some time can be called 'training programmes' and distinguished from 'coaching programmes' (see Hasselhorn, 1995; Stankov, 1991). Both types of programmes have in common that they rely on practice, which is repeated with similar tasks, and both of them constitute active interventions. But they have different kinds of objectives the attainment of which has to be assessed by different criterion variables.

The basic problem in assessment, however, lies in the fact that changes of true scores as well as changes in the test-specific components can only be assessed by performances based on observable scores or variables (which reflect true scores as well as situation- and person-specific 'errors'). There are two main ways of dealing with this problem. First, to choose observable measures that are appropriate with respect to the goals, and second, to choose experimental designs and comparisons which are appropriate with respect to the goals and the questions referring to them. Any design referred to consists of at least two experimental groups.

Based on these considerations there are two main possibilities to empirically separate training effects from mere coaching effects. Firstly: Criterion tasks should be chosen which are solvable only if some (near) transfer in the sense given by Royer (1979) takes place, that is, they should be similar to the tasks of the programmes, but neither parallel nor identical to them. Different theories of transfers and of the cognitive processes underlying them (see, e.g., Adams, 1989; Adey et al., 1989; Brown, 1978; Cormier & Hagman, 1987; Perkins & Salomon, 1989; Royer, 1979; Salomon & Perkins, 1989; Singley & Anderson, 1989; VanLehn, 1996) seem to agree that some kind of similarity between the tasks or problems of the programme and the criterion tasks or problems is essential for all types of transfer (with the possible exception of 'unspecific transfer'; Royer, 1979, p. 54-55). There is no consent, however, what 'similarity' exactly means (Vosniadou & Ortony, 1989), but as a provisional guideline it often may suffice to choose criterion measures which differ from the programme's tasks in the format or surface structure (Palincsar & Brown, 1984). If a programme shows its effectiveness not only with respect to identical, but also with respect to similar tasks, this may be taken as the first evidence that it is a training rather than a coaching programme. Secondly: Coaching effects are presumed to be of short duration, whereas effects of a training should improve competencies, which means that the beneficial effects in performance should last for some time (maintenance of effects). Therefore, an additional possibility of separating both kinds of effects consists in assessing them some weeks after the end of the intervention (delayed posttest) or to give it immediately after completion of the programmes and to repeat it some weeks or months later (follow-up). The relevant comparisons refer either to the delayed posttests or to the changes within each programme between the immediate posttest and the follow-up (Hasselhorn, 1995). If the programme shows beneficial effects in these comparisons, this is a further evidence that it is a training rather than a coaching programme.

Since the various kinds of transfer or generalization have to be practised, too, it is necessary to add particular components to a programme which furnish links between the problems of the programme and those of the criterion tasks and links to situations or tasks quite dissimilar to those of the intervention context, that is, to some 'real-world behaviour'. Although such components are often lacking, many authors think them to be most important (Adams, 1989; Adey & Shayer, 1993; Belmont & Butterfield, 1977; Bransford, Sherwood, & Sturdevant, 1987; Brown, 1978; Friedrich & Mandl, 1992; Pressley, Snyder, & Cariglia-Bull, 1987; Sternberg, 1983). The same holds true for special monitoring or metacognitive strategies which enable, among other aspects, detection of similarities between apparently 'dissimilar' problems and tasks.

Narrow or specific and broad or general goals and their assessment
But there may be other problems when considering certain objectives and their relations to criterion measures. The broader the goals of the programme the more likely it will be to find appropriate criterion variables. For instance, if one wants to foster general intelligence by the programmes reviewed by Sternberg (1984), many diagnostic tests may be appropriate to assess the programmes' effectiveness. The narrower its scope, however, the more difficult it may be to find criterion variables that suit the programme's objectives. Thus, if a particular programme intends to improve inductive reasoning ability, that is, to develop rules from similarities and dissimilarities between objects and/or relations (Klauer & Phye, 1994), it might be rather difficult to find a test which enables assessment of this ability. Since inductive reasoning plays a central role in fluid intelligence, one might use a test such as the Culture Fair Test by Cattell to evaluate the programme's effectiveness, and with respect to this test the programme proves effective. But as fluid intelligence encompasses not only inductive reasoning, but also further abilities (Horn, 1985), it may be questioned as to whether this effectiveness is indeed based on improved inductive reasoning or on improvements of other abilities or both. Although improvements in fluid intelligence would be positive side effects (see below) they leave the question unanswered whether the programme reached its particular goal(s). Looked upon from another perspective, the problem addressed is one of construct validity in the sense of Cook, Campbell, and Peracchio (1990): Construct validity is violated if the (observable) criterion variables do not enable proper assessment of a programme's (non-observable) goals.

Proximal (near) and distal (far) goals and cognitive acceleration
Another distinction concerning the goals must be made with training programmes, since very often two further kinds of goals can be identified. 'Proximal (or near) goals' are directly intended by the programme and refer to a certain cognitive function such as, e.g., perception or reasoning. But 'the ultimate (or distal) goal of any strategy training consists in successfully improving strategy use in real learning and problem solving situations in school, university, and at work' (Friedrich & Mandl, 1992; see also, e.g., Adams, 1989; Sternberg & Bhana, 1986). 'Distal (or far) goals' can only be achieved by means of (far) transfer.

Besides a follow-up some or a long time after completion of the programme assessments of distal goals usually require criterion variables different from those appropriate for the proximal goals. That means, for example, that measures of success in school or in a profession are appropriate for the respective distal goals, but usually not for proximal goals. To give an example: The proximal goal of the perceptual programme by Frostig, Horne, and Miller (1972) is the improvement of visual perception. Achievement of this goal should be assessed by appropriate criterion variables referring to visual perception, where 'appropriate' does not necessarily mean the test developed by Frostig, Maslow, Lefever, and Whittlesey

(1964), which consists of problems either very similar or even parallel to those of the programme. Since (visual) perception is a basic ability, improving it may lead to improvements in other cognitive domains such as general intelligence. If one interprets transfer to general intelligence as another proximal goal, an appropriate test of intelligence should be administered in addition to perceptual tests. If, in addition, one seeks to assess the programme's transfer effects on other abilities or cognitive domains this can also be done using appropriate criterion variables. If enhanced intelligence is viewed as a predictor of better school performance and if this constitutes a distal goal, school reports or exams are appropriate with respect to this goal.

The same holds true when a particular programme is said to have positive effects on certain cognitive areas, but not on others. According to Schneider, Visé, Reimers, and Blaesser (1994) the German version of the meta-linguistic programme by Lundberg, Frost, and Peterson (1988) should have positive effects on phonological awareness, but no effects on intelligence, memory capacity, and on information processing speed. To test the respective hypotheses empirically, criterion variables have to be chosen which are related to the cognitive areas mentioned.

If a programme is designed for particular long-term effects (e.g., Byrne & Fielding-Barnsley, 1995) or for cognitive acceleration (Adey et al., 1989; Kuhn, 1974), at least one long-term follow-up one or more years after the programme's completion should be applied to assess whether the programme has resulted in cognitive acceleration, which includes far transfer from one cognitive domain to another (Adey & Shayer, 1993). The criterion tasks should be chosen accordingly, that is, they should enable assessment of the progress in various cognitive areas. According to the proposal by Hasselhorn (1995), the relevant comparisons refer to the pretest, if this has been given, and the long-term follow-up measures. The nature of the (statistical) comparisons will be discussed below.

To sum up: Whatever goals a programme is developed or intended to achieve: It is of primary importance to assess whether it attains these goals or not, whether it shows the intended effects.

Effects and side effects

Besides the intended effects, there may be unintended or side effects which are partly positive and partly negative. Positive side effects of a cognitive programme to improve reasoning processes may consist in a more reflexive general strategy of problem solving or in improved metacognitive skills or in a more positive self-concept (cf. also Hanson, 1984). Negative side effects may arise if a well established strategy is to be replaced by a more versatile one, and the process of replacing may cause a temporary drop in performance (cf. Lohman, 1986; Snow, 1977). Negative side effects may also occur if the programme is boring or not attractive to the participants.

Although these side effects are legitimate goals of evaluations, their systematic investigation should not be the first step in evaluating programmes, especially for children, if additional criterion variables are necessary to assess them. Younger children can concentrate only for a very short period (about 15 to 25 minutes at most). Thus there are narrow limits for active data gathering with children, if one wants results which are not mere numbers, but are interpretable with respect to one's hypotheses and research questions. The first evaluations of a programme should focus on the intended goals. During these first evaluations 'informal' attention should be directed at possible negative side effects, which often become apparent during the intervention, and the programme should be changed accordingly.

In some cases, the occurrence of positive side effects instead of the intended effects is interpreted as documenting the programme's effectiveness, and its goals are reformulated to meet with the positive side effects. This is reason enough to demand a clear separation between both kinds of effects.

Effects of programmes and statistical measures of effect sizes

Evaluation of programmes in the domain of educational psychology has a long tradition in considering (statistical) effect sizes (ES) such as standardized mean distances (d) or correlations (r) in addition to tests of significance or instead of them. Often, researchers claim to have discovered large effects which have remained insignificant, and they base their conclusions on effect sizes instead of tests of significance. In other instances, the same (or other) researchers state that their statistical effects are only small or moderate, but statistically significant, and in this case, statistical significance is given more weight than the effect sizes. Without discussing any details, I take the position that effect sizes generally should supplement statistical tests, but not replace them (see also APA, 1994). A statistical test tells us whether there is a significant effect, and an additional computation tells us how large this effect is. In agreement with the rationale underlying statistical tests and the usual interpretation of nonsignificant results, insignificant effect sizes should be interpreted as random deviations from the 'true' value ES = 0 despite their magnitudes (cf. Dush, Hirt, & Schroeder, 1989, p. 101), whereas significant effects can be interpreted according to their magnitudes taking into account random variations. This convention, although not in accordance with some meta-analytic procedures, avoids conflicting decisions of the kind addressed above.

The problem addressed becomes even more serious when one performs a non-parametric test (such as a U-test) and subsequently computes a parametric effect size such as d. This is an appropriate and 'empirically meaningful' measure (in the sense of Suppes & Zinnes, 1963) for some parametric, but not for non-parametric tests: Its size changes under admissible transformations of the original data, which preserve their rank order, but the size of d does not change under transformations which are admissible for data at interval scale level which is needed for the usual interpretations of results achieved by parametric tests such as t or F tests. Thus, the ES measure should conform to the statistical tests one performs, especially as power analysis always refers to a criterion ES which is closely associated with the test to be performed (Cohen, 1988; Hager, 1995).

It is well-known that the results of statistical tests depend heavily on sample sizes, and on the sizes of the effects. But although effect sizes are independent of sample sizes, they are not independent of other factors, especially they are not attributes of programmes. Instead, their magnitudes depend, among other factors, on the nature of the particular programme (special vs. more general), the duration of its application, its execution (with one or more participants at a time), the age and the initial status (abilities and aptitudes) of the participants, the quality of the intervention (especially the expertise of the persons who administer the pro gramme), the kind of criterion variables in relation to the programme's tasks (identical, similar, dissimilar, different; see above), and last but not least, on random variations. Therefore, it would seem inappropriate, if not unwise to compare these measures across evaluation studies which are not similar enough to make these comparisons useful, despite what is done in many meta-analyses. If, for instance, a particular programme is evaluated twice under conditions which differ mainly with respect to the criterion tasks chosen (parallel vs. similar in the sense that transfer is necessary to solve them successfully), greater statistical effects can be expected for the parallel tasks than for the similar tasks, all other things being about equal.

The question of the magnitude of effect sizes to be expected with particular programmes or kinds of evaluations will be addressed below. The next paragraph deals with some kinds of evaluations.

Types and conducts of evaluation with some references to effect sizes

Some short considerations concerning experimental designs and experimental validity

As has been indicated above, programmes most often are evaluated using a randomized or non-randomized pretest-posttest design with at least two experimental groups to enable comparisons between the groups. Only rarely the effectiveness of a programme is assessed by simply comparing pretest to posttests scores only (e.g., Klauer & Phye, 1994), a basically invalid design. In many instances, the designs are supplemented by one or more follow-ups which serve various objectives, and more complex experimental designs have to be used in order to answer special questions (e.g., Adey & Shayer, 1993; Byrne & Fielding-Barnsley, 1995; Miller & Dyer, 1975). Because of space limitations I restrict discussion to the basic layout, as there is one general rule for choosing a design: 'Select the design which is appropriate with respect to your hypotheses and which allows of control of factors either pretending or masking beneficial effects of the programme, and the 'intervention-bound' factors (see below) deserve special attention in any case.' Although randomization is important for internal validity (Cook et al., 1990), even non-randomized or quasi-experimental studies can have high internal validity. This is the case when empirical data can only or at least best be interpreted by one particular and well-founded hypothesis (or theory), whereas apparently 'plausible' rival explanations are far less well-founded. The simple assertion that there are rival explanations for the data is not sufficient for arguing that internal validity is lowered or lacking. With respect to the follow-ups potential threats to validity such as history (Cook et al., 1990) sometimes cannot be ruled out by randomization: It may be possible that activities between the several assessments are more similar to one programme than to another, thus leading to repeated practice of certain aspects of this programme. This, in turn, might cause better performances for this programme in the next follow-up which are not due to, say, accelerated development, as one might conclude. For further details concerning the validity of evaluation studies the reader is referred to Cook et al. (1990), Cronbach (1983), and Hager (1995).

Although the pretest is not necessary if sample sizes are large enough to ensure randomization to be successful, often it gives first information concerning the abilities of the participants. In some programmes, instructional methods or other components differ with respect to various levels of abilities, which means that if no other sources are available a pretest is necessary to choose the appropriate methods and/or problems. The fact that a pretest is given does not, by the way, lower the so-called 'external validity' because of pretest-sensitization effects, since usually programmes are chosen on the basis of prior assessments as to the participants' suitability for the particular programme (see also Cronbach, 1983). Moreover, pretest data can (and should be) correlated with posttest and follow-up data to get some information of their sizes and especially of their algebraic sign: If pairwise correlations are near zero or even negative this is a clear evidence that something went wrong with the intervention. Either there is no regularity between the data of two times of measurement, or high scores at time T_j are paired with low scores at time $T_{j'}$ ($j < j'$; $j = 1, 2, ..., J - 1$) and vice versa. Both data patterns indicate a lack of systematic training effects, despite possible differences in means as expected. These pairwise correlations between different times of assessments should be computed and reported routinely as well as the intercorrelations between the criterion variables assessed. Whether pretest data are used for random blocking or analysed by means of an analysis of covariance or by a repeated measures analysis of variance or another technique deemed appropriate for such comparisons between different occasions and programmes and for the kind of data to be expected may depend on the time necessary to analyse the pretest data and on the assumptions one is willing to accept concerning the various techniques of data analyses.

Types of evaluation: Process and outcome evaluation, global and analytical evaluation

The classifications of kinds of evaluations in the present part of this chapter are mainly based on the considerations presented amongst others by Baron (1987), Nickerson et al. (1985), and by Scriven (1967, 1991), but the particular classification scheme is proposed by the present author. 'Type of evaluations' mainly refers to different objectives of the evaluations, whereas the classification 'conduct of evaluations' primarily refers to the way empirical evaluations can be performed. This distinction is more or less arbitrary because of the many relations which exist between the various evaluation activities. The most important aspects of the first classification are given little attention here, as they seem to be well-known. Since a particular distinction between different conducts of evaluations bears certain consequences which usually are not attended to, I shall discuss these consequences in some detail (see Cronbach, 1983, for a different theoretical framework for evaluations, which partly is discussed in Cook et al., 1990).

Process evaluation (Scriven, 1991) is directed to changes during the intervention, whereas outcome evaluation focuses on the results of the intervention, often with respect to the baseline before the beginning of the intervention. Both types of evaluations are complementary, but most often researchers restrict their activities to evaluating outcomes, since process evaluations are time and cost consuming and very often are not possible with children who can concentrate and cooperate only for a short time (see above). This restriction seems often justified since the effectiveness of a particular programme is expected to be best after the complete intervention. Sometimes, however, process evaluations may reveal that the intended effects show up prior to the end of a particular programme, and this fact should be taken as a clue to modify the programme accordingly (see Scriven, 1991, for further benefits of process evaluations).

Global evaluation means that a single 'score' is allocated to a programme, that is, its effectiveness is evaluated as a whole (Scriven, 1991). This type of evaluation answers the most important (technological) question as to whether a programme is effective or not. In contrast, analytical evaluations refer to questions concerning the potential causes underlying a programme's effectiveness or its failure. They concern parts or components of the programme and their contributions to the overall effectiveness; this kind of questions resembles those of basic research rather than those typical for technological research. By definition, such a component evaluation (Scriven, 1991) cannot be a global evaluation. On the other hand, evaluations of a programme's dimensions or (expected) effects and to the (positive and negative) side effects can either be achieved by analytical or by global evaluations. Evaluations of a programme's dimensions or effects by an analytical or by a global evaluation may be characterized by aspects belonging to the basic as well as to the technological area. It is basic since the question is directed towards the effects underlying the programme's effectiveness or its causes. It is technological since any programme may be said to have been constructed to give raise to certain effects which lead to its effectiveness. As long as there is no reasonable doubt concerning the close relation between effectiveness and the effects underlying it, this type of investigation does not seem most urgent.

Besides evaluating the programme as a whole (global evaluations) or parts of it (analytical including component evaluations) other features extraneous to the programme itself may be focussed on such as the situations under which the programme is administered (e.g., one-to-one intervention, intervention in small groups or in school classes) or level of ability of the participants or characteristics of the trainers, and so on. Research questions like these can be put to the whole programme and/or to its components. Some of the relations between various evaluative activities are summarized in Figure 1.

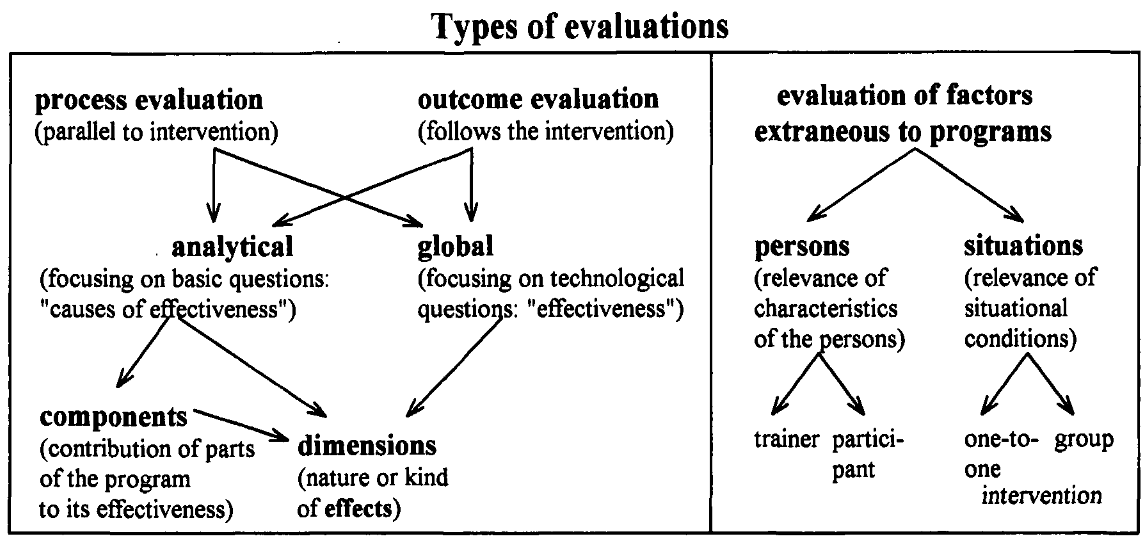
Conduct of evaluations (1): Formative and summative evaluations

Within the category 'conduct of evaluations', basically two paradigms of evaluations can be distinguished according to Scriven (1967): formative vs. summative evaluations and non-comparative vs. comparative evaluations. The present paragraph deals with the first distinction.

During a programme's construction it should be evaluated by means of formative evaluations to improve it or its components. This is largely an informal activity which encompasses theoretical as well as empirical analyses, where empirical data do not necessarily rest on rigorous methods of design and analysis (Nickerson, Perkins, & Smith, 1985; Scriven, 1991). Formative evaluations are mostly done as 'internal evaluations' by the programme authors themselves. After the (provisional) completion of the programme it is subject to various kinds of summative evaluations (Scriven, 1991, p. 340) "for the benefit of some external audience and decision-maker". Summative evaluations should also or mainly be performed as 'external evaluations' by researchers not affiliated with the programme authors (Scriven, 1991; Sternberg & Bhana, 1986). Process, outcome, global and analytical as well as component evaluations can be framed either as formative or as summative evaluations (see Figure 1).

Conduct of evaluations (2): Comparative and non-comparative evaluations

Basically, the distinction between comparative and non-comparative evaluations refers to the type of experimental group to whom a programme is compared. There are mainly three comparison groups: (a) a 'no-treatment' or 'waiting' group who do not experience a particular intervention, but only take part in the tests administered to the experimental group; (b) a group to whom a programme is given with goals quite different from those of the experimental programme to be evaluated, and (c) a group to whom another programme is given which has the same goals as the target programme. If comparison groups fall in the categories (a) and (b), the evaluation is non-comparative; if it belongs to the third category, the



Conducts of evaluations				
formative evaluation	+	+	+	(„Evaluations“ of extraneous factors usually do not fit into the categories to the left, although they can be subject to meta-evaluations)
summative evaluation	+	+	+	
non-comparative ev.[a]	+	+	+	
comparative ev.[a]	+	+	+	
meta-evaluation	+	+	+	+ + + +

Notes. [a] Non-comparative and comparative evaluations can be connected with formative and with summative evaluations. Besides the relations explicitly addressed there may be many more. +: The respective type of evaluation can be combined with a particular type of conduct. - See Scriven (1991) for the types and conducts of evaluation which have been modified with respect to evaluations of factors extraneous to a given program and with respect to dimensions or effects of programs. - Besides the classifications addressed, all evaluations can be performed either by the program authors (internal evaluation) or by scientists not affiliated with the program authors (external evaluation).

Figure 1: Types and conducts of evaluations of cognitive programs

evaluation is comparative. The aims of non-comparative evaluations consist in assessing a programme's effectiveness as such: 'Is the programme effective?' ('programme-bound effects'), whereas the aim of comparative evaluations is to answer questions referring to the relative effectiveness of at least two programmes with equal or very similar goals: 'Is the new programme more or less effective than a comparable one with the same goals?'

Whenever an intervention is given, there will be 'intervention-bound factors' like the special attention directed to the participants or motivational factors inherent in the particular intervention situation (Ball, 1990; Snow & Swanson, 1992) which cannot be ruled out by randomization. These factors usually lead to enhanced performances which often are (falsely) interpreted as programme-bound effects; sometimes, however, they may conceal programme-bound effects. In order to separate intervention-bound from programme-bound effects, it is usually demanded that there should be an intervention with the comparison group, too (e.g., Sternberg & Bhana, 1986). Let us consider the possible comparison groups in some detail.

Non-comparative evaluations, plausible hypotheses, and effect sizes

In non-comparative evaluations the programme's 'pure' effectiveness usually should be assessed against a control programme which does not have the same objective(s) as the target programme. The effects of the latter should even not be achieved by means of transfer, but if this occurs, the effects should be incidental and considerably smaller than the corresponding effects of the programme to be evaluated. Moreover, the second programme must take place under conditions which are identical or very similar to those of the target programme (e.g., equal lengths, equal group sizes during intervention), and the same holds true for its demands concerning concentration, level of effort, its attractivity, and so on. Most often, an available programme has to be changed to meet with these demands, that is, to serve its function as a control programme. Although it may be difficult sometimes to find a programme that does not aim at the target programme's objectives even by means of transfer, the tailoring itself poses no problem at all since a control programme is not subject to evaluation, but only serves to make the intervention-bound factors as similar as possible to those of the programme to be evaluated. Therefore, in non-comparative evaluations it is sufficient to use only those criterion variables which are relevant for the target programme.

In rare instances, however, the programme is imbedded in a holistic theoretical conception where the particular intervention setting is interpreted as part or component of the programme; this is the case with the programme in visual perception by Frostig et al. (1972). Here, the programme's effectiveness should be assessed using a comparison group without any intervention at all ('waiting group'), since intervention-bound factors are part of the programme and need not be controlled separately.

The preceding discussion of the meaning of intervention-bound factors mainly refers to programmes given to small groups of persons or to a single participant taken from their regular classes or courses. Sometimes it is possible to make a programme part of regular lessons, whereas other classes or courses are given regular lessons without the particular programme (see for an instance of this Schneider et al., 1994). In cases like these there is no reason to expect different intervention-bound factors in the experimental groups, and only the occurrence of such factors in one, but not in the other experimental group lowers the validity of the investigation. Administering a programme to entire school classes or courses often precludes strict randomization, but this would not necessarily lower the study's validity (see above).

Taking the view of the programme authors, the most obvious hypothesis in a non-comparative evaluation states that the programme leads to greater improvements than in a comparison group, who experience an intervention with objectives quite different from those of the programme to be evaluated. It thus refers to the programme's 'pure' or programme-bound effectiveness. Usually, the effectiveness of a programme is inferred if comparisons between the two experimental groups turn out to be significant. As long as these comparisons are analysed statistically by a single (say, F) test, their interpretation may not be unambiguous, especially when the changes in the experimental groups point in opposite directions. To avoid this ambiguity, one should first test the predictions that there will be improvements with the target programme and that there are no changes or improvements in the control group, too. These predictions mean that certain disordinal interactions with respect to the factor 'time of assessment' may occur without precluding unequivocal interpretations as to the effectiveness of the programme to be evaluated. If the results of these tests are as predicted the changes are compared across both experimental groups, preferably by focussed (univariate) tests on directional statistitical hypotheses since we expect greater improvements for the programme of interest. The programme of interest should only be called 'effective' if this test, too, comes out as predicted. In addition, the outcome of the test should be accompanied by an effect size 'not too small'. If the preliminary tests within each programme show that there are no improvements for the programme to be evaluated, a further comparison is not necessary.

Although we do not expect changes to point in different directions, they sometimes occur (see, e.g., Slack & Porter, 1980). Among other interpretations, this result may cast doubts on the appropriateness of the control programme, its execution, and so on, and it makes clear interpretation regarding the effectiveness of the target programme more difficult, if not impossible; and especially the size of its (statistical) effects may be overestimated. In instances like this it may be best to suspend decision on the effectiveness hypothesis (see for more comprehensive discussion of possible results within and between the groups and their possible meanings Hager, 1995; cf. also Brown et al., 1983; Campione & Armbruster, 1984; Cook et al., 1990). A direct comparison between the programmes or a direct test of the corresponding hypothesis should also not be performed for those criterion variables whose correlations between pretest and posttest or posttest and follow-up and so on are negative or about zero, as this may be a hint for irregularities during the intervention and/or for the differential appropriateness of a programme with respect to the participants. As particular a priori effectiveness hypotheses are to be tested statistically the main data analysis should rest on focussed univariate tests instead global multivariate ones (Hager, 1995). The intercorrelations between the criterion variables do not threat the validity of the univariate tests, but introduce an unavoidable dependency between the decisions on effectiveness hypotheses referring to the several criterion variables. These intercorrelations should be reported routinely (see above).

But effects of what magnitude should be expected in non-comparative evaluations? Here, effect sizes (ES) tell us how large the 'pure' (statistical) effect of the programme is, if the comparison group is a control group in the sense given above. The demands or expectations range from Bloom's (1984) 'two-sigma criterion' (a standardized mean difference of d = 2.0) under ideal conditions (one-to-one tutoring with mastery learning) to d = 0.5 through d = 1.0 for quite successful programmes administered to larger groups (Friedrich & Mandl, 1992, pp. 38-39). Many authors refer to the conventions proposed by Cohen (1988) according to whom an effect is large if d = 0.8, medium if d = 0.5, and small if d = 0.20. However, Cohen (1969, 1988, p. 25) restricts his proposals to situations where no other information is available. But despite the multitude of factors influencing their magnitudes (see above), much is known about effect sizes in the domain of evaluating particular programmes so that adhering to Cohen's conventions maybe greatly misleading (see also Howell, 1997), if the particular conditions which gave raise to an effect of a certain magnitude are neglected.

Because of these factors influencing effect sizes, it seems impossible to give any guidelines except the general statement that in non-comparative evaluations the statistical effects, at least shortly after the intervention, should be 'large' especially as the effects of programmes tend to vanish if it is of short duration and if nothing is done to maintain them (Zigler, 1979). Despite all difficulties, measures of ES should be considered and reported to give an impression of how large the effects are which arise with a particular programme under particular conditions of intervention with particular participants, and so on. These effects can then be used as a basis for planning one's own evaluation study of the same programme, as a criterion value of the ES to be detected has to be specified in advance to determine the power $1-\beta$ of one's tests (cf. Cohen, 1988). Any important dissimilarity between the former and the new study can be taken as a reason to change the ES to be detected accordingly. In other cases the only realistic way is to take the given (or expected) sample size, to prespecify a conventional level of α, and some values of β (or of the power function) and to determine how large the ES has to be 'in the populations' under these conditions. Prespecifications like these are arbitrary, but choosing sample sizes and some conventional values of α is arbitrary as well and disregards the power considerations (see for a more thorough discussion of these matters and some further proposals Hager, 1995).

Comparative evaluations, plausible hypotheses, and effect sizes

When a new programme has shown its effectiveness under various circumstances in some non-comparative evaluations it may and should be compared to other programmes which have the same or at least very similar objective(s) as the new programme and which for this reason are competitors or rivals to the new programme. This type of evaluation usually is called comparative evaluation, and its focus lies in comparing the relative effectiveness as a basis for evaluating the relative efficiency (see below).

Usually competing or rival programmes will differ in many ways to achieve the same objectives: They will offer different tasks, different instructions and/or strategies for working on these tasks, they may rely on different instructional methods, their lengths as well as their particular intervention context may differ, and so on. As these differences are part of the programmes they must not be changed in a comparative evaluation, since any change will also alter the programme. In their original versions the programmes are competitors to each other; as soon as one of the programmes is changed, as is often done with the aim of controlling variables which must not be controlled in comparative evaluations,

Table 1 Distinctions Between the Most Important Comparison Groups with Evaluations of Cognitive Programmes and Their Classification with Respect to Comparative and Non-Comparative Evaluations

Characteristics of the treatments and of the intervention settings in the comparison groups	Goals of the comparison programme with respect to the goals of the programme to be evaluated	
	Different	Same or very similar
Different	no treatment (programme)	rival programme
Same or very similar	control programme	'quasi-rival' programme
Conduct of evaluation	non-comparative	comparative

Notes The characteristics of the comparison groups including the particular intervention settings are compared to the programme to be evaluated. - The tasks and problems as well as the instructional methods of all comparison programmes usually differ from those of the programme of interest. In comparative evaluations the intervention settings of the programmes to be compared may not be different, whereas in non-comparative evaluations these settings must be very similar. Here, amongst others, the cognitive demands, the duration, and the attractivity of the control programme must be similar to the target programme. - See text for further explanations it should not be called a rival programme, but it may serve as a 'quasi-rival' to the other programme - a fourth possible comparison group. One of the problems with this type of programme is that its effectiveness usually has not yet been assessed, since the change of the original programme can affect its effectiveness. Whenever there is (additional) interest in knowing the non-comparative effectiveness of a programme it is advisable to additionally administer a further programme which serves as a control with respect to the programme in question. In these instances, a comparative evaluation is combined with a non-comparative one. But even if such control groups are not introduced, a comparison between two rival programmes is reasonable. Since a comparative evaluation can only answer comparative questions, any result is interpretable and valuable: The programmes are either equally effective (or ineffective), or one of them outperforms the other. The inclusion of one or two control group(s), however, enables to distinguish between 'equally effective' and 'equally ineffective'.

In comparative evaluations, it is necessary to consider the criterion variables relevant to both programmes and their objectives, respectively. Although the programmes have the same goals (non-observable or 'theoretical' level), the programme authors may differ with respect to the observable or empirical criterion variables they think appropriate for their programme's goals. In order to avoid biases for or against the competitive programmes criterion variables must be considered which are relevant to either programme (see Hager, 1995, for the details).

The kind of statement in a comparative evaluation refers to the relative effectiveness of (at least two) rival programmes which have the same (or very similar) goal(s). Basically, there are two effectiveness hypotheses the examination of which demands a comparative evaluation. The first hypothesis claims that a new programme is more effective than a rival one (superiority hypothesis). The second hypothesis represents a more modest claim and says that the new programme is not inferior to the competitor (non-inferiority hypothesis), that is, it is equally effective as or more effective than the other programme. Under both hypotheses, it can be predicted that the programmes lead to improvements from pretest to posttest. But even if improvements only show up with one programme, the intended comparative statement can be made: Under the circumstances given, this programme is more effective if the comparison between the experimental groups comes out significant. If, however, there are no improvements for both programmes this may indicate inappropriate execution, inappropriate side conditions, and the like. In case of this data pattern the programmes should not be compared.

The superiority hypothesis is supported if the comparison between both programmes leads to a statistically significant result in favour of the programme expected to be superior. If this test comes out insignificant, the only statement permitted is that the programmes are equally effective or equally ineffective. If there is interest in distinguishing between 'equally effective' and 'equally ineffective', either proper control groups should be added to the design (see above), or the improvements within the programmes should be compared to those in other investigations under very similar circumstances and with the same criterion variables. These comparisons may be performed using meta-analytic techniques (comparisons of effect sizes) such as those outlined by Hedges and Olkin (1985).

In contrast, the non-inferiority hypothesis claims the new programme not to be inferior to its competitor. In this instance, statistical power requires special attention, since the comparison between both programmes should remain insignificant in order to lent support to the hypothesis. Statistical significance is not always required for positive statements about the effectiveness of programmes. A more thorough discussion of predictions and their testing in comparative evaluations can be found in Hager (1995). Both hypotheses, by the way, should not be tested for those variables for which correlations between pairs of measurement occasions are near zero or even negative (see above).

In comparative evaluations it is no threat to experimental validity if the rival programmes are applied by different persons who are 'experts' for their programme. This expertise in administering a programme, then, may be interpreted to be another part of the programme itself. But as far as possible the experts for one of the rival programmes should also administer the respective control programme if this part of the design.

Measures of effect sizes should be considered in comparative evaluations, too. Using d, its correct interpretation refers to differences in effectiveness of the two programmes, not to their effectiveness as such. This fact is important, since it means that the discussion about the magnitude of effect sizes given above is not applicable to comparative evaluations. No one would reasonably expect two rival programmes, aiming at the same goals, to differ by two or even by one standard deviation(s). Unfortunately, this makes it even harder to give any general guidelines concerning the magnitude of effect sizes, and the reader is referred to the options given above.

Comparative and non-comparative evaluations: Some further details

The distinction between comparative and non-comparative evaluations is essential, although there are evaluations that do not seem to fit into this classification. For example, Miller and Dyer (1975) considered four preschool programmes and tried to assess their 'dimensions and effects'. They did not direct much attention to the programmes' specific objectives, but chose criterion variables they thought to be relevant in assessing the programmes' (differential) effects. This, however, is not a (hypotheses-guided) question of effectiveness with special attention to the programmes' particular goals, but one that arises when research is directed to certain questions without first considering hypotheses as tentative answers to these questions, a view contrary to the one taken in this chapter. As the objectives of the programmes overlapped only partly they cannot be classified as control programmes to each other nor are they competitors as a whole. Maybe they would best be classified as rivals or 'quasi-rivals' with respect to some goals and criterion variables and as control programmes with respect to others. From the view taken here, an investigation such as the large scale one done by Miller and Dyer (1975) suffers from identifying the criterion variables for which there should be large differences among all or some programmes and those criterion measures for which small or no differences at all are to be expected without necessarily leading to negative statements concerning the programmes' (non-comparative) effectiveness. But this study clearly shows that the important differentiation between two conducts of evaluations does not necessarily refer to a study as a whole, but instead may refer to the criterion variables in a particular study: For some variables the evaluation is non-comparative and for others comparative. If, for instance, a programme aiming at enhancement of (visual) perception and (inductive) reasoning is compared to a programme claiming to foster (inductive) reasoning alone, the study is non-comparative with respect to criterion measures of (visual) perception, but comparative with respect to measures of reasoning abilities.

Effectiveness, efficiency, some criteria of effectiveness, and meta-evaluation

The empirical evaluation of the effectiveness of any programme is a conditio sine qua non. Effectiveness refers to the basic question as to whether a programme shows beneficial effects in the intended cognitive domain and with the participants it has been designed for. Another question in evaluations refers to the programme's efficiency as compared to the efficiency of 'similar' or rival programmes (see Rossi & Freeman, 1993). Are the benefits and their maintenance of a particular programme worth the material and immaterial costs, the effort, especially if compared to the benefits of similar programmes? Is it easy to apply and is it 'robust' with respect to different trainers? Are its effects reliable? In contrast to other questions, those referring to efficiency are not completely answerable by empirical data, since efficiency encompasses effectiveness and other factors such as economy, attractivity for the participants, applicability to groups, and so on. Instead of dealing with efficiency more systematically, a list of general criteria of effectiveness for cognitive training programmes will be proposed, based on the considerations presented above and in the respective literature (e.g., Adams, 1989; Baron, 1987; Belmont & Butterfield, 1977; Bransford, Franks, Vye, & Sherwood, 1989; Palincsar & Brown, 1984; Sternberg, 1983). The specific application of these general criteria to a particular programme, however, cannot be outlined because of the multitude of factors influencing, say, effect sizes or maintenance of effects or amount of transfer: Questions like these seem to have to be answered for each programme separately, but with reference to programmes which are similar or comparable to the one under study. All comparisons mentioned should refer to the usual ones between groups as well as to the analysis of single cases within each experimental group.

The assessment of the criteria require a complete research programme, and partly they may be understood as 'operationalizations' of the general criteria for any system of technological rules given above, i.e. that it must be "effective, reliable, without negative side effects, easily applicable, and economic" (cf. Herrmann, 1984, p. 28): (a) assessment of the programme's proximal goals and intended effects with special attention to the distinction between coaching and training programmes and the persons the programme has been designed for (non-comparative evaluation; comparison between posttest measures of performance or between the changes from pretest to posttest measures; 'material' transfer to similar tasks); (b) significant statistical effects of 'sufficient' sizes when comparing the target programme to a control programme (same comparisons as before); (c) assessment of the specifity of effects with respect to the programme's goals to exclude mere intervention-bound effects (same comparisons as before, but including criterion measures sensitive to intervention-bound factors); (d) assessment of plausible positive and negative side effects (same comparisons as before, but choice of 'plausible positive' criterion measures in addition to or instead of the goal-oriented measures); (e) assessment of maintenance of effects which by definition is evidence for changes in competencies (comparisons of changes between posttest and follow-up measures or between follow-up measures according to Hasselhorn, 1995); (f) assessment of far and 'temporal' transfer to situations and problems outside the intervention context (comparisons as before; appropriate situations and criterion measures for far transfer); (g) assessment of the reliability of the effects by replications under 'similar' and under varying conditions (non-comparative evaluations and same comparisons as before); (h) assessment of 'robustness' of the programme's effects with respect to the trainers (focus lies on different trainers rather than on different programmes); (i) successful comparisons with rival programmes with special attention to efficiency, if there are any rivals (comparative evaluations; choice of criterion measures as outlined above); (j) assessment of distal goals (comparative or non-comparative evaluation; appropriate choice of criterion variables); (k) assessment of acceleration of development or of cumulative effects as a further possible (but usually not necessary) and very ambitious goal or effect the assessment of which is obligatory for programmes designed for acceleration of development (comparisons between 'delayed' follow-ups some years later or comparisons between the changes from pretest to 'delayed' follow-ups according to Hasselhorn, 1995). Especially criteria (j) and (k) require a rather long time, and they should not prevent authors to present their programmes to the public before these criteria are met.

When all or at least a good deal of these evaluation activities have taken place it is advisable to re-analyse and review them, that is, (l) to perform meta-evaluations (Cook & Gruder, 1978), which, according to Scriven (1991, p. 229), are "the professional imperative of evaluation". Meta-evaluations may focus on certain aspects of a programme and its evaluations or may try to be more or less comprehensive, but usually they have in common to give an impression of the programme's effectiveness, its advantages, and its possible shortcomings. They encompass reanalysing the primary studies as to their validity and categorizing them accordingly. Other or additional classification schemes may focus on the hypotheses (mainly concerning effectiveness) being examined in the primary studies or to reconstruct them from what the researchers had in fact done. Whatever classification schemes seem appropriate with respect to the aim(s) of a meta-evaluation, it is essential to carefully distinguish between comparative and non-comparative evaluations (see above). Reviews of a programme's effectiveness across various situations, for persons of different age, and for different criterion measures may be a topic of meta-evaluations (see for further aspects for reviews Hager, 1996). It is desirable that meta-evaluations are performed mainly by external evaluators, and they seem especially important since (nearly) any empirical investigation can be designed to maximize the probability of results being as one expects or hopes them to be - willingly or not (cf. Datta, 1986). Administering a programme to one person at a time and comparing the results to the pretest-posttest data of a no-treatment group is an example where the design of the study leaves only very little room for an answer different from 'the programme is effective' (but probably because of intervention-bound factors). The opposite is also possible: Using an effective rival or 'quasi-rival' programme as a comparison programme in a 'non-comparative' evaluation will maximize the probability of the statement 'The programme is not effective', since the 'desired' statistically significant effects will not show if the programmes are about equally effective.

Some concluding remarks

Many researchers seem not to favour the hypothesis testing approach advocated in this chapter, but hypotheses specify exactly what information you are interested in the first place, and the design and the methods can then be chosen accordingly (see Benson & Michael, 1990; Hager, 1995). Any question directed to programmes and their effectiveness can be tentatively answered by testable hypotheses. Testing hypotheses formulated in advance should not prevent the researcher from inspection of the data to get further information which may be interesting in itself and which may generate further testable hypotheses. For an endeavour to be scientific it must be required that as many as possible of the necessary decisions, criteria, and/or rules are disclosed and justified, if possible. To formulate hypotheses is one way to make one's inevitable 'biases' and 'presuppositions' (Kuhn, 1970; Popper, 1992) explicit, and empirically testable. In the case of evaluating programmes, the best choice seems to be the programmes' goals and the intended comparisons either to a control programme or to a rival programme. The replacement of hypotheses by the attitude "Let the data speak for themselves usually translates to I have not thought that far ahead" (Howell, 1997, p. 218). And:

If no hypotheses have been formulated in advance, no study can be designed appropriately (Hager, 1995), and the data will remain silent, or their interpretation will be arbitrary.

As soon as basic distinctions in the domain of programme evaluation and especially their consequences for empirical research are attended to, some conflicting evidence concerning cognitive programmes can be resolved, although others will remain unresolved: Refined methodologies and methods will never lead to 'miracle cures', but those considered in this chapter are no 'snake oil remedies' either (cf. the title of the article by Sternberg & Bhana, 1986). The classifications presented have proved valuable in my own empirical evaluations of some cognitive programmes, but in turn they have made it difficult to adequately interpret the results of various evaluations, in which the categories presented in this chapter had not been taken into account, despite the fact that the basic distinctions have been known for many years (see above). Therefore, the main goal of this chapter lies in directing the readers' attention to these distinctions and some of their most important consequences so that evaluations of cognitive programmes can be designed the results of which are more easily interpretable than has been the case all too often in the past.

References

Adams, M. J. (1989). Thinking skills curricula: Their promise and progress. *Educational Psychologist, 24,* 25-77.

Adey, P., & Shayer, M. (1993). An exploration of long-term far-transfer effects following an extended intervention programme in the high school science curriculum. *Cognition and Instruction, 11,* 1-29.

Adey, P., Shayer, M., & Yates, C. (1989). *Thinking science: The curriculum materials of the Cognitive Acceleration through Science Education (CASE) project.* London: Macmillan.

American Psychological Association (APA) (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.

Anastasi, A. (1981). Coaching, test sophistication, and developed abilities. *American Psychologist, 36,* 1086-1093.

Ball, S. (1990). Unintended effects in educational research. In H.J. Walberg & G. D. Haertel (Eds.), *The international encyclopaedia of educational evaluation* (pp. 139-142). Oxford: Pergamon.

Baron, J.B. (1987). Evaluating thinking skills in the classroom. In J.B. Baron & R.J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 221-247). New York: Freeman.

Belmont, J.M., & Butterfield, E.C. (1977). The instructional approach to developmental cognitive research. In R.V. Kail & J.W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 437-481). Hillsdale, NJ: Erlbaum.

Benson, J., & Michael, W.B. (1990). Basic principles of design and analysis in evaluation research. A twenty year perspective on evaluation study design. In H.J. Walberg & G.D. Haertel (Eds.), *The international encyclopaedia of educational evaluation* (pp. 545-553). Oxford: Pergamon.

Bloom, B.S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*, 6, 4-16.

Bransford, J.D., Franks, J.J., Vye, N.J., & Sherwood, R.L. (1989). New approaches ot instruction: Why wisdom can't be told. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 470-497). Cambridge: Cambridge University Press.

Bransford, J.D., Sherwood, R.D., & Sturdevant, T. (1987). Teaching thinking and problem solving. In J.B. Baron & R.J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 162-181). New York: Freeman.

Bransford, J.D., Stein, B.S., Arbitman-Smith, R., & Vye, N. (1985). Three approaches to improving thinking and learning skills. In J.W. Segal, F. Chipman, & R. Glaser (Eds.), *Thinking and learning skills: Relating instruction to basic research* (Vol. 1, pp. 133-200). Hillsdale, NJ: Erlbaum.

Brown, A.L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 1, pp. 77-165). Hillsdale, NJ: Erlbaum.

Brown, A.L., Bransford, J.D., Ferrara, R.A., & Campione, J.C. (1983). Learning, remembering, and understanding. In P.H. Mussen (Ed.), *Handbook of child psychology* (Vol. 3, pp. 77-166). New York: Wiley.

Bunge, M. (1967). *The search for truth* (Vol. 2). Berlin: Springer.

Byrne, B., & Fielding-Barnsley, R. (1995). Evaluation of a programme to teach phonemic awareness to young children: A 2- and 3-year follow-up and a new preschool trial. *Journal of Educational Psychology, 87*, 488-503.

Campione, J.C., & Armbruster, B. (1984). An analysis of the outcomes and implications of intervention research. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 287-304). Hillsdale, NJ: Erlbaum.

Case, R., Kurland, D.M., & Goldberg, J. (1988). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386-404.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cook, T.D., Campbell, D.T., & Peracchio, L. (1990). Quasi-experimentation. In M. D. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, 2nd ed., pp. 491-576). Palo Alto, CA: Consulting Psychologists Press.

Cook, T.D., & Gruder, C.L. (1978). Meta-evaluation research. *Evaluation Quarterly, 2,* 5-51.

Cormier, S.M., & Hagman, J.D. (1987). *Transfer of learning. Contemporary research and applications.* San Diego, CA: Academic.

Cronbach, L.J. (1983). *Designing evaluations of educational and social programmes.* San Francisco, CA: Jossey-Bass.

Dansereau, D.F. (1985). Learning strategy research. In J.W. Segal, S.F. Chipman, & R. Glaser (Eds.), *Relating instruction to research* (Vol. 1, pp. 209-239). Hillsdale, NJ: Erlbaum.

Datta, L.-E. (1986). Benefits without gains: The paradox of the cognitive effects of early childhood programmes and implications for policy. In M. Schwebel & C.A. Maher (Eds.), *Facilitating cognitive development. International perspectives, programmes, and practices* (pp. 103-126). New York: Haworth.

Dush, D.M., Hirt, M.L., & Schroeder, H.E. (1989). Self-statement in the treatment of child behaviour disorders: A meta-analysis. *Psychological Bulletin, 106,* 97-106.

Friedrich, H.F., & Mandl, H. (1992). Lern- und Denkstrategien - ein Problemaufriß [Learning strategies and thinking strategies]. In H. Mandl & H.F. Friedrich (Eds.), *Lern- und Denkstrategien. Analyse und Intervention* (pp. 3-54). Göttingen: Hogrefe.

Frostig, M., Horne, D., & Miller, A.M. (1972). *The developmental programme in visual perception.* Chicago, IL: Follett.

Frostig, M., Maslow, P., Lefever, D.W., & Whittlesey, J.R.B. (1964). *The Marianne Frostig developmental test of visual perception.* Palo Alto, CA: Consulting Psychologists Press.

Hager, W. (1995). Planung und Durchführung der Evaluation von kognitiven Förderprogrammemen [Planning and forthgoing of evaluation from cognitive programmes]. In W. Hager (Ed.), *Programmeme zur Förderung des Denkens bei Kindern. Konstruktion, Evaluation und Metaevaluation* (pp. 100-206). Göttingen: Hogrefe.

Hager, W. (1996). Kriterien der Effektivität von Trainingsprogrammemen [Criteria for evaluation of training programmes]. In E. Witruk & G. Friedrich (Eds.), *Pädagogische Psychologie im Streit um ein neues Selbstverständnis* (pp. 259-267). Landau: Empirische Pädagogik.

Haney, W. (1990). Coaching for tests and examinations. In H.J. Walberg & G.D. Haertel (Eds.), *The international encyclopedia of educational evaluation* (pp. 122-124). Oxford: Pergamon.

Hanson, M.J. (1984). The effects of early intervention. In M.J. Hanson (Ed.), *A typical infant development* (pp. 385-406). Baltimore: University Park Press.

Hasselhorn, M. (1995). Kognitive Trainings: Grundlagen, Begrifflichkeiten und Desiderate [Cognitive training: Backgrounds, understanding, and opinions]. In W. Hager (Ed.), *Programmeme zur Förderung des Denkens bei Kindern. Konstruktion, Evaluation und Metaevaluation* (pp. 14-40). Göttingen: Hogrefe.

Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Herrmann, T. (1984). Methoden als Problemlösungsmittel [Methods for problem solving]. In E. Roth (Ed.), *Sozialwissenschaftliche Methoden* (pp. 18-46). Munich: Oldenbourg.

Horn, J.L. (1985). Remodeling old models of intelligence. In B.B. Wolman (Ed.), *Handbook of intelligence. Theories, measurements, and applications* (pp. 267-300). New York: Wiley.

Howell, D.C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Duxbury.

Klauer, K.J., & Phye, G.D. (1994). *Cognitive training for children*. Seattle: Hogrefe & Huber.

Kuhn, D. (1974). Inducing development experimentally: Comments an a research paradigm. *Developmental Psychology, 10,* 590-600.

Kuhn, T.S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago, IL: University of Chicago Press.

Kulik, J.A., Bangert-Drowns, R.L., & Kulik, C.-L.C. (1984). Effectiveness of coaching for aptitude tests. *Psychological Bulletin, 95*, 179-188.

Lohman, D.F. (1986). Predicting mathemathanic effects in the teaching of higher-order thinking skills. *Educational Psychologist, 21,* 191-208.

Lundberg, J., Frost, J., & Peterson, O.P. (1988). Effects of an intensive programme for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23,* 261-284.

Miller, L.B., & Dyer, J.L. (1975). Four preschool programmes: Their dimensions and effects. *Monographs of the Society for Research in Child Development, 40,* no. 162.

Nickerson, R.S. (1987). Why teach thinking? In J.B. Baron & R.J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 27-37). New York: Freeman.

Nickerson, R.S., Perkins, D.N., & Smith, E.E. (1985). *The teaching of thinking*. Hillsdale, NJ: Erlbaum.

Orlick, T. (1986). *Psyching for sports: Mental training for athletes*. Champaign, IL: Human Kinetics.

Palincsar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1,* 117-175.

Paul, R.W. (1990). Critical and reflective thinking: A philosophical perspective. In B.F. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 445-494). Hillsdale, NJ: Erlbaum.

Perkins, D.N., & Salomon, G. (1989). Are cognitive skills context bound? *Educational Researcher, 18,* 1, 16-25.

Pike, L.W. (1978). *Short-term instruction, testwiseness, and the Scholastic Aptitude Test: A literature review with recommendations.* Princeton, NJ: Edcuational Testing Service.

Popper, K.R. (1992). *The logic of scientific discovery* (4th ed.). London: Routledge.

Pressley, M., Snyder, B.L., & Cariglia-Bull, T. (1987). How can good strategy use be taught to children? Evaluation of six alternative approaches. In S.M. Cormier & J.D. Hagman (Eds.), *Transfer of learning. Contemporary research and applications* (pp. 81-120). San Diego, CA: Academic Press.

Rossi, P.H., & Freeman, H.E. (1993). *Evaluation. A systematic approach* (5th ed.). Newbury Park, CA: Sage.

Royer, J.M. (1979). Theories of transfer of learning. *Educational Psychologist, 14,* 53-69.

Sackett, P.R., Burris, L.R., & Ryan, A.M. (1989). Coaching and practice effects in personnel selection. In C.L. Cooper & I. Robertson (Eds.), *International review of industrial and organizational psychology 1989* (pp. 145-183). New York: Wiley.

Salomon, G., & Perkins, D.N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist, 24,* 113-142.

Schneider, W., Visé, M., Reimers, P., & Blaesser, B. (1994). Auswirkungen eines Trainings der sprachlichen Bewußtheit auf den Schriftspracherwerb in der Schule [Explanation of a training in knowledge in speaking on education in spelling in school]. *Zeitschrift für Pädagogische Psychologie, 8,* 177-188.

Scriven, M. (1967). The methodology of evaluation. In R.E. Stake (Ed.), *Perspectives on curriculum evaluation* (pp. 39-83). Chicago, IL: Rand McNally.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.

Singley, M., & Anderson, J.R. (1989). *The transfer of cognitive skill.* Cambridge, MA: Harvard University Press.

Slack, W.V., & Porter, D. (1980). The Scholastic Aptitude Test: A critical appraisal. *Harvard Educational Review, 50,* 154-175.

Snow, R.E. (1977). Research on aptitudes: A research report. In L.S. Shulman (Ed.), *Review of research in education* (Vol. 4, pp. 50-105). Ithasca, IL: Peacock.

Snow, R.E., & Swanson, J. (1992). Instructional psychology: Aptitude, adaptation, and assessment. *Annual Review of Psychology, 43,* 583-626.

Stankov, L. (1991). The effects of training and practice on human abilities. In H.A.H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 97-117). Hillsdale, NJ: Erlbaum.

Sternberg, R.J. (1983). Criteria for intellectual skills training. *Educational Researcher, 12,* 2, 6-12.

Sternberg, R.J. (1984). How can we teach intelligence? *Educational Leadership, 42,* 1, 38-48.

Sternberg, R.J., & Bhana, K. (1986). Synthesis of research on the effectiveness of intellectual skills programmes: Snake-oil remedies or miracle cures? *Educational Leadership, 44,* 2, 60-67.

Suppes, P., & Zinnes, J.L. (1963). Basic measurement theory. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 1-76). New York: Wiley.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology, 47,* 513-539.

Vosniadou, S., & Ortony, A. (Eds.). (1989). *Similarity and analogical reasoning.* Cambridge: Cambridge University Press.

Zigler, E. (1979). Project Head Start: Success or failure? In E. Zigler & J. Valentine (Eds.), *Project Head Start: A legacy of the war on poverty* (pp. 495-507). New York: Free.

14

Converging aims and diverging means of teaching thinking: An epilogue

B. Csapó & J.H.M. Hamers

Introduction

Education, one of the most ancient human enterprises, aims at cultivating children's minds. However, throughout the long history of organized education, there have been different views of what this aim really means and how it can best be achieved. In the second half of this century, there has been a growing emphasis on those aspects of cultivating the mind that enable learners to manage, process, organize and apply the information they acquire. Teaching thinking is one of the umbrella-keywords under which the research into this problem seems to find its place, but there are a number and classical or recently emerging areas of research that are also aimed at the same target. When we attempt to review the main research trends in this field and the related areas, all efforts initially appear to point in the same direction. If they deal with knowledge, procedural components are emphasized, and the quality of students' knowledge is characterized by how well is it organized, and how easily it can be mobilized in new situations. If students' school learning is studied, meaningfulness and understanding are central issues; students' potential to learn is to be improved or their learning strategies are to be developed. This apparent convergence of the broad aims of research and development diminishes if the particular research projects are considered. Although their theoretical frameworks are becoming more consistent, the particular solutions they propose for the general problems studied are quite different.

This book mirrors these two faces of current research and programme development. The research projects that are the bases of the chapters seem to be aimed at the same target, and, as we have already discussed in the first chapter of this volume, they tend to turn to the same pool of philosophical and theoretical sources. There are several links between the projects presented in this book, but at present they rather seem to be mosaics or fragments of a larger picture than parts of a consistent research paradigm. Diversity and consistency are not the only issues that the editors of this book faced when selecting and organizing the chapters, but the dilemma they pose is characteristic of the whole field of research as well. Thus, it is hard to synthesize the results of the book without outlining the broader problems of the area. Therefore, as we draw a final conclusion from the work presented in this book, we outline a strategy of managing this diversity as well. We address general issues related to the present and future of research regarding teaching thinking research which is illustrated in the chapters of this book. Diversity of ideas, approaches and practical solutions is of great value, especially if the actual field of research is in a phase of rapid development. However, the particular projects may become easily isolated if they do not support each other and if they lack the links that organize the concepts used within certain projects into consistent conceptual networks. One of the challenges that will influence the next few years of research regarding teaching thinking may be how a healthy and fruitful balance can be found between the benefits and drawbacks of diversity and consistency.

Diversity: Advantages and drawbacks

There are many advantages of diversity, many rooted in the premise that: different approaches to the same problem can be cross-fertilizing when they are contrasted and discussed in the same framework. Interpreting particular contradictions often requires broadening the field of vision and shaping a conceptual framework. Researchers in the field have already encountered the diversity of research regarding teaching thinking, and (beyond the scheme we propose in the first chapter) have attempted to classify the approaches several times. For example, Nickerson, Perkins and Smith (1985), who reviewed around thirty projects carried out in the United States until the mid 1990s, classified them into five main categories. Jones and Idol (1990) used six dimensions of thinking, and Nickerson (1988) listed seven aspects of teaching thinking. However, these attempts not only indicate that the number of approaches - as well as their diversity - is growing, but even attempts to synthesize the results show impressive variety.

Most of these attempts aim at classifying, categorizing and synthesizing United States projects, and one of the best known collections of teaching thinking programmes (Costa, 1991) also presented programmes that have already been implemented in the United States. When Hamers and Overtoom (1997) published a comparable inventory of European programmes for teaching thinking, this added another dimension of variety to the already enormous complexity of the field: they presented an inventory of 42 projects in a consistent format; in doing this, they introduced programmes that were devised and implemented in different countries, in different cultures and educational settings, and in different languages. McGuinness and Nisbet (1991) also published about the European research on teaching thinking. Their volume was the first one that presented enough information about some programmes to prompt the intention of implementing them somewhere else or at least using the experiences of other programme developers. There have already been some well-known attempts to transfer teaching thinking projects from one culture to another (for example, the Venezuelan projects are the best known ones, see Dominguez, 1985; Sánchez, 1987), but in Europe it is different. Because of linguistic isolation (and also because of the former ideological and political divisions of the continent) a number of original ideas, theoretical concepts and practical methods have appeared and were nurtured in several countries that may fertilize works in other research communities. However, such an enterprise also prompts the questions: how and how far can programmes devised and implemented in one culture be used in other countries? Under what conditions can a programme that is successful in one educational setting be used in another country with approximately the same efficiency? How can experiences from one culture be transferred into another, or putting it in a different way, how far can research results be replicated under different circumstances?

Our present book continues to pose such dilemmas, although there are some historical parallels that we should learn from. Language and cultural issues are not new at all. Some terms introduced by German psychologists, like 'Gestalt' have become known not only to psychologists around the world, but are also now part of standard English (as well as many other languages), while Piaget's 'structure d'ensemble' is known only by the specialists in the field, although it is also known that this French term is the one that best expresses Piaget's original concept. It is also known that 'activity' has a rather different connotation than its Russian original, and the problems caused by inappropriate translations of Piaget's early works into other languages have also been broadly discussed in the literature (for some current instances see Adey & Shayer, 1994). Despite all these controversies, it is undeniable that the works of German psychologists, Vygotsky and his followers, and the Geneva school have become the common knowledge base of the present English-speaking research community. Or to cite a more recent example, Dutch psychologists benefited greatly by drawing from the works of another generation of Russian psychologists (see references in the first chapter and other chapters by Dutch authors). Benefiting from this type of linguistic and cultural diversity is definitely slowing down since English is becoming the dominant language of scientific communication. Maybe there are ideas and terms presented in this book that probably lose their original Finnish, Dutch or Hungarian flavour when translated into English, but today's researchers, even if English is not their first language, keep in mind 'how would it sound in English' when choosing terms for developing conceptual frameworks.

As for the linguistic aspects of the difficulties of the synthesis, it is not the ideas originally expressed in different languages that cause the main complication. A major part of the problem lies within the English terminology itself. The proliferation of terms used in the field of teaching thinking seems to be accelerating, and quite often they are applied inconsistently and so increase the complexity of problems unnecessarily. In some cases, some words are used as synonyms and are varied to make the presentation of ideas more attractive; in other cases, the selection of terms indicates sharp differences of theoretical orientations.

At least four groups of terms can be distinguished. (a) The first group refers to the mental process itself. The most frequently used words are: thinking, reasoning, cognition, and information processing. Although some authors strictly distinguish them and prefer or avoid one or several of them, these words are quite frequently used as synonyms to name the mental processes studied in general. (b) In the second group, there are terms that refer to the dispositions or attributes that lie behind the mental processes. The words typically used are: skills (specific, general, higher order thinking, reasoning), abilities (mental, cognitive, specific, general), mind, intelligence (general, fluid, crystallized, practical), procedural knowledge, cognitive strategies, operations, structures, operational structures, competence or even aptitude. The choice of term in this case is fairly characteristic for the author's approach and theoretical position. The programmes (including those presented in this book) aim at improving these dispositions or attributes. (c) The terms in the third group are used to name the change itself, or the process that results in the desired change of the disposition or attributes. A large number of terms can be identified, for example, teaching, developing, improving, training, instructing, educating, modifying, fostering, enhancing, increasing, stimulating, accelerating, remedying. The choice of this type of term is usually determined by the theoretical paradigm the authors identify with, by the type of expression selected from the second group or by the target population of a specific programme, (e.g., teaching or developing is more often used in the normal population while training or remedying fits better to children in special education, and expressions like fostering or stimulating are more often used for naming the programmes devised for exceptional children). (d) Finally, the terms that describe or name the specific programmes are usually the combination of words listed in the previous groups. Their structure is frequently 'doing something with something'. In the vast body of research literature almost every combination can be found. Besides the most common combinations, e.g., teaching thinking or improving cognitive abilities, a large number of original or unusual terms also appear, for example 'teaching intelligence' (Blagg, 1991), 'cognitive instruction' (Jones & Idol, 1990), or the really unusual 'training of intellectual aptitude' (Snow, 1982). Some combinations refer to specific areas of intervention, e.g., improving operational abilities, fostering inductive reasoning, while others name whole research paradigms or orientations. For example, an international association was organized around 'cognitive education', and a journal has been published (see also Scheinin and Mehtäläinen, this volume).

The chapters of this volume are no exceptions to this trend: they also use a variety of expressions to name the object of their study. Although there are trademark-like associations of terms in this book that are also associated with certain types of research or research communities (e.g., Experiential Structuralism introduced by Efklides and Demetriou; Cognitive Acceleration through Science Education, CASE, trademark of the research by Adey and Shayer; Scheinin's and Mehtäläinen's Formal Aims of Cognitive Education, FACE; or Csapó's Operational Enrichment, OE) in elaborating the theoretical frameworks, a wider consistency can be observed. The usage of terms and expressions usually does not go beyond the Piagetian, Vygotskyan, constructivist, information processing and psychometric terminology. This consistency of terms helps to bridge the differences between the specific theoretical foundation and practical implementation of the research programmes. The chapters still do not use a well-defined terminology but there are a number that overlap making the conceptual frameworks 'translatable'.

Cognitive research and/or programme development

There is one more dimension to the variety of teaching thinking projects that, quite often, is characterized by the theoretical or practical orientation of the researchers or programme developers, but in fact the sources of the differences are deeper than that. Actually, the orientation of researchers and thus the outcome of their work is determined by a number of different factors, and among these, professional considerations that are derived from different philosophical ideas also play an important role.

Those who are closer to the positivist view of the development of sciences and who tend to share the values best expressed by natural scientists doing basic research, believe in the step-by-step accumulation of scientific knowledge in the field of human cognition and in studying cognitive development as well. In this view, devising and testing programmes for improving thinking skills is part of an empirical research process and the results contribute to the growing body of knowledge about the development of thinking. They follow the classical principle that says: "If you want to understand it, try to change it." Research being done in this vein requires sophisticated theories, consistent effort, careful design and good coordination of work done by different researchers at different places. Controlling other's findings, replicating experiments, synthesizing the results via the quantitative processes of meta-analysis or conceptual analyses are all broadly accepted processes of this paradigm. Then the ever-changing and permanently tested body of knowledge can be utilized to solve the actual problems, which in our case are the design of curricula, courses or other programmes for teaching thinking.

On the other hand, the applicability or at least the universal validity of this strategy in the research into human cognition, or more precisely in the study of teaching thinking, is often challenged. The enormous complexity of the problem can be the first objection. Too many variables need to be taken into account so the models or theories that describe the whole phenomena would be hopelessly complicated. Therefore, for a longer period of time, developing training programmes for the practical situation has come to dominate teaching thinking. Programme developers, in general, have not paid too much attention to the theoretical foundations, or used particular or ad-hoc theories. This works resembled to medical or pharmaceutical research, or engineering. There has been a need to solve problems arising in practice: processes or treatments have to be applied even if their scientific bases are not fully understood. Particular technologies were tested in practice, and if the treatment worked, no one bothered about the theories. As Hager (this volume) also points out, technologies or technological rules may work well in practice even if their theoretical foundations are weak. Many such programmes have been developed and tested (see Costa, 1991), including Feuerstein's Instrumental Enrichment (IE), de Bono's CoRT programmes, and Lipman's Philosophy for Children. They have become known worldwide and have been adapted in several countries (for a review of these programmes, see for example Blagg, 1991). These training programmes consist of certain instruments, tasks, teaching materials, specifically organized sessions of teaching or longer courses. If these 'pre-packaged' programmes are properly applied, as described in their manuals, they will probably have the desired effects.

However, if the training involves complex processes, application to new circumstances always requires some adaptation as well. If the programme is theoretically not well established, modifications cannot be theoretically understood either. During such modifications one can question, how far these programmes preserve their own identity? How does their adaptation to new circumstances influence their efficiency? And if the working of a programme is not understood well, how can it be systematically improved? And if it is not adapted, or the technological rules are too rigid, how can an efficient implementation be expected? These dilemmas can be illustrated by the status of one of the most well known programme, Feuerstein's IE. As Adey and Shayer (1994) describe, IE has become 'fossilized' in its original form, and although it was promising at the beginning, it is hardly efficient when applied in an environment that differs from the designers' original settings. Some broadly known unsuccessful adaptations of IE indicate the problem (see for example Blagg's (1991) work in Britain). Thus, developing training programmes without a strong research background can only provide a solution to particular problems.

The need for consistent research has already been expressed in the literature and several of the authors of this book are among those who are working on bridging the gap between research and practice. The programmes they have devised for improving a particular or a wider area of thinking are embedded in a broader and long-term research agenda. In this way, the underlying and supporting research makes generalization and integration of the research results possible. For example, the work of Adey (this volume) in the CASE project relates science education and teaching thinking. The neo-Piagetian background connects his research to that of Efklides' group (this volume; also see Demetriou, Shayer, & Efklides, 1992). Klauer (this volume) has devised programmes for improving inductive reasoning; furthermore, these programmes have already been applied in other countries and adapted to other languages (Hamers, De Koning, & Sijtsma, 1998; Klauer & Phye, 1995; Klauer, Resing, & Slenders, 1996). On the one hand, the theoretical foundations (e.g., the precise definition of the concept of inductive reasoning) and the broader research background to his work (see Klauer, 1993) are strong enough that the original concept can be applied to a number of new situations. New directions, like the classroom applications, have grown out of this work, and the results can be synthesized via meta-analysis (Klauer, this volume). Some concepts, borrowed from the psychometric or individual differences tradition form a link between Efklides' and Klauer's work, as well as that of Adey, Csapó, and Scheinin and Mehtäläinin. Efklides's ability approach provides a common basis to Scheinin's work and Klauer's topic, while inductive reasoning also belongs to the main theme of classical intelligence research. However, each keeps a certain distance from the often discredited concept of intelligence and they share the views of Carroll (1993) who also prefers to speak about cognitive abilities.

Other chapters in this volume are embedded in the research conducted within the information processing paradigm, like Van Oostendorp's and Elshout-Mohr's work on text comprehension, and Chanquoy's work on text production. Describing change of knowledge as a constructive process, even if the sources of these changes are texts and not the reality itself, relates the analysis of Van Oostendorp and Elshout-Mohr to the chapters where constructivist views are more directly expressed (e.g., De Koning & Hamers; Nelissen).

Direct or indirect school application of the research forms a common basis for all chapters, regardless of whether the researchers approach the classroom application from the direction of other research areas or whether the problems they are dealing with have arisen immediately from the school practice. Reading (De Koning & Hamers; Van Oostendorp & Elshout-Mohr), writing (Chanquoy) and mathematics (Nelissen, Van Luit, Verschaffel) has formed the core of education since medieval times. If science (Adey, Csapó, Klauer) and grammar (Csapó, Klauer) are added, almost the whole range of school subjects is covered. However, despite all these links and overlapping, further efforts are needed to improve the consistency of research and to ensure the results can be more easily replicated, and the findings more comparable and controllable.

At the present state of cognitive research, it would be unwise to deny the necessity, relevance or importance of designing programmes for teaching thinking. It is obvious that one of the ultimate benefits of cognitive research is embodied in the form of practically applicable programmes. On the other hand, again taking into account the present state of research into teaching thinking, we would strongly argue for emphasizing the importance of coordinated research efforts that serve to accumulate widely applicable knowledge for educational practice. After reviewing the chapters of this book, we may conclude that the more efficient accumulation of knowledge in the field of teaching thinking requires a new research agenda.

Outline of an agenda for future research

Teaching thinking and related areas have belonged to the main line of educational research for the past decades. Especially in the late 1970s and early 1980s interest was focused around these topics: a large number of international conferences were organized, and books were published about this theme. Then in the late 1980s, for several reasons, the intensity of interest decreased somewhat, and other issues became dominant. Since then, a number of changes have taken place, of which the trends in globalization and the revolutionary developments in information technology have had the most visible impacts on society. These changes influence schools and school education (main sites of educational research) as well as the conditions and possibilities for educational research. For example, improved communication and accessibility have had an impact on the organization of research. In the last section of this epilogue, we list some areas where we expect developments may help the synthesis of results (some of these issues have already been discussed in more detail in Csapó, 1997b).

Theoretical frames

Today's educational research draws from a number of different sources. Brain research and neuroscience offer new insight into the biological foundations of cognition and their results have already been suggested for educational application (see for example Jensen, 1988). More or less abstract models of cognitive science (e.g., models of parallel distributed processes) are also often considered as possible sources of educational innovation. However, we may question how broad a field educational researchers have to observe for resources, or how many new ideas educational theories can accommodate without becoming helplessly eclectic, complex and inapplicable.

Thus, first a proper level for educational theories should be found, including theories that can accommodate frameworks for teaching thinking. An appropriate level of abstraction and generalization is needed for such theories that are embedded in the general conceptual frames of cognitive sciences, that are firmly grounded in practice as well, that are neither very abstract nor too simple. A number of recent publications have already called for such a new framework or paradigm, most of them suggesting the renewal of instructional psychology through the adaptation of the results of cognitive science (e.g., Glaser, 1991). A new developmentally valid instructional psychology, a cognitive educational psychology or a 'cognitive pedagogy' (Csapó, 1992) may be close to the desired theoretical framework. Some believe that the time for a cognitive revolution in education has not arrived yet (Ohlsson, 1990), others argue that the outline of the new paradigm is already apparent (Vosniadou, 1996). We tend to agree with those who believe that such a new paradigm may appear in the near future, but it will not come without the concentrated and conscious efforts of the interested research community.

Consistent terminology

Needless to say, a firm, clear and unambiguous terminology is a precondition of any theory-building. However, psychological and educational concepts are not easy to define, and even if well known definitions exist, their interpretations may be changing continuously. Productive conceptual developments should not be stopped or limited, but there is a need to control redundancy. Physicists already have a solution for such problems: when they define a basic concept or dimension, a process of its measurement is part of the definition. In psychology, psychometricians have followed this method while behaviorists have attempted to overcome 'word magic' via carefully operationalizing their concepts. The cognitive sciences also offer firm ground for defining basic terms. The problems of fuzzy and ever changing concepts cannot be solved completely, but they require continuous attention. Developments of achieving consensus in the usage of terms can be stabilized by synthetic reviews, encyclopedic collections and dictionaries. Eysenck's (1990) work on a closely related field may be an example for such efforts.

Taxonomy: Mapping the mind

What are we going to change when we teach thinking or develop students' cognitive abilities? There are almost as many answers to this question as there are different theoretical frameworks. Based on the Piagetian, factor-analytic or more recent cognitive studies, there are a large number of lists, taxonomies, systems or models of thinking skills, cognitive abilities and other similar constructs. However, even if there are empirically underpinned structures available, like Carroll's factor analytic model, they require further interpretation (see, for example, Spearritt (1996) on Carroll's work) and empirical research if we are to attempt to apply them in educational contexts. Among others, the educational or practical relevance of the identified skills or abilities should also be examined. For example, inductive and deductive reasoning are often analysed in parallel in theoretical models or in cognitive studies, but their importance in real-life cognitive processes seems different. Research into teaching thinking has to involve systematic mapping of the mind in educational contexts. Demetriou's and Efklides' (1994) research into the structure of the mind is an example of how this can be done.

Development and modifiability

Describing static structures would scarcely be useful for educational applications. If researchers intend to devise programmes to stimulate development, we have to know how development takes place without the special stimulating processes: not only the structures but also their development should be described. Figures of developmental trends or exact developmental curves of the target skills or abilities would be of great help to programme developers. These developmental curves serve as base lines or points of references for intervention studies. " ... one needs evidence that a change in children's development has been achieved. For this there must also exist normative data against which the effect can be shown ..." as Shayer (1992, p. 108) put it. Systematic and comparable measurements of developmental trends in basic skills or abilities provide firm ground to estimate at what age, and how, interventions would result in the best effects (for example, for the development of inductive reasoning, see Csapó, 1997a). The ultimate goal of intervention studies is to determine if development can be modified (e.g., stimulated, accelerated). To make the picture more complicated, we have to take into account that modifiability of skills is age-dependent, so intervention studies should deal with more than one age group. Therefore, it is the modifiability of thinking skills that intervention studies examine (or should examine) and a systematic description of their modifiability would be one of the best ways to integrate research results.

Methodological standards

In the past decade the methodology of intervention studies has developed a great deal (Hager, this volume). For example, because publishing effects sizes has become standard practice, results are comparable and the synthesis of results is easier. However, if we want to compare and integrate the results into a larger picture, there are still a number of problems in finding standards. For example, 'doses', or units of treatment in the interventions, type and length of training sessions, and length of the whole training have to be taken into account when the results of intervention studies are compared. One of the main difficulties of doing meta-analysis studies is the lack of standards or measures of experimental treatments (Goossens, 1992).

Human thinking is one of the most complex phenomena researchers have ever studied and stimulating its development is the most ambitious goal of education. Describing the structure of thinking skills, their interdependence, their relevance, their development and their modifiability as a function of age is the major aim of research. In the near future, the success of research in this field will depend largely on how the complexity of problems is managed and whether researchers find ways of coordinating their efforts. The chapters of this book show that it is necessary, it is possible, and also that there is still much to be done.

References

Adey, P., & Shayer, M. (1994). *Really rising standards. Cognitive intervention and academic achievement.* London: Routledge.

Blagg, N. (1991). *Can we teach intelligence? A comprehensive evaluation of Feuerstein's instrumental enrichment program.* Hillsdale, NJ: Erlbaum.

Carroll, J.B. (1993). *Human cognitive abilities. A survey of factor-analytic studies.* Cambridge: Cambridge University Press.

Costa, A.L. (Ed.) (1991). *Developing minds. A resource book for teaching thinking.* Alexandria, VA: ASCD.

Csapó, B. (1992). *Kognitiv pedagógia* [Cognitive pedagogy]. Budapest: Akadémiai Kiadó.

Csapó, B. (1997a). Development of inductive reasoning: Cross-sectional measurements in an educational context. *International Journal of Behavioral Development, 20,* 609-626.

Csapó, B. (1997b, August). *Improving cognitive abilities in schools: Constraints and promises.* Paper presented at the Seventh European Conference for Research on Learning and Instruction. Athens, Greece.

Demetriou, A., & Efklides, A. (1994). Structure, development and dynamics of mind: A meta-Piagetian theory. In A. Efklides & A. Demetriou (Eds.), *Intelligence, mind and reasoning. Structure and development* (pp. 75-109). Amsterdam: North-Holland.

Demetriou, A., Shayer, M., & Efklides, A. (1992). *Neo-Piagetian theories of cognitive development. Implications and applications for education.* London: Routledge.

Dominguez, J. (1985). The development of human intelligence: The Venezuelan case. In J.W. Segal, S.F. Chipman, & R. Glaser (Eds.), *Thinking and learning skills* (Vol. 1, pp. 529-537). Hillsdale, NJ: Erlbaum.

Eysenck, M.W. (Ed.) (1990). *The Blackwell dictionary of cognitive psychology*. Oxford: Blackwell.

Glaser, R. (1991). The maturing of the relationship between the science of learning and cognition and educational practice. *Learning and Instruction, 1*, 103-118.

Goossens, L. (1992). Training scientific reasoning in children and adolescents. A critical commentary and quantitative integration. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development. Implications and applications for education* (pp. 160-179). London: Routledge.

Hamers, J.H.M., De Koning, E., & Sijtsma, K. (1998). Inductive reasoning in third grade: Intervention promises and constraints. *Contemporary Educational Psychology, 23*, 132-148.

Hamers, J.H.M., & Overtoom, M.Th. (Eds.) (1997). *Teaching thinking in Europe. Inventory of European programmes*. Utrecht: Sardes.

Jensen, E.J. (1988). *Teaching with the brain in mind*. Alexandria, VA: ASCD.

Jones, B.F., & Idol, L. (1990). *Dimensions of thinking and cognitive instruction*. Hillsdale, NJ: Erlbaum.

Klauer, K.J. (Ed.) (1993). *Kognitives Training* [Cognitive training]. Göttingen: Hogrefe.

Klauer, K.J., & Phye, G.P. (1995). *Cognitive training for children*. Seattle: Hogrefe & Huber.

Klauer, K.J., Resing, W.C.M., & Slenders, A.P.A.C. (1996). *Cognitieve training voor kinderen* [Cognitive training for children]. Göttingen: Hogrefe.

McGuinness, C. & Nisbet, J. (1991). Teaching thinking in Europe. *British Journal of Educational Psychology, 61*, 174-186.

Nickerson, R.S. (1988). On improving thinking through instruction. *Review of Research in Education, 15*, 3-57.

Nickerson, R.S., Perkins, D., & Smith, E.E. (1985). *The teaching of thinking*. Hillsdale, NJ: Erlbaum.

Ohlsson, S. (1990). Cognitive science and instruction. Why the revolution is not here (yet). In H. Mandl, S.N. Bennett, & H.F. Friedrich (Eds.), *Learning and instruction. European research in an international context* (Vol. 2.1, pp.561-600). Oxford: Pergamon Press.

Sánchez, M. (1987). Teaching thinking processes. In D.N. Perkins, J. Lochhead, & J. Bishop (Eds.), *Thinking: The second international conference* (pp. 413-430). Hillsdale, NJ: Erlbaum.

Shayer, M. (1992). Problems and issues in intervention studies. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development. Implications and applications for education* (pp. 107-121). London: Routledge.

Snow, R.E. (1982). The training of intellectual aptitude. In D.K. Detterman & R.J. Sternberg (Eds.), *How and how much can intelligence be increased?* (pp. 1-37). Norwood: Ablex Publishing Corporation.

Spearritt, D. (1996). Carroll's model of cognitive abilities: Educational implications. *International Journal of Educational Research, 25,* 107-198.

Vosniadou, S. (1996). Towards a revised cognitive psychology for new advances in learning and instruction. *Learning and Instruction, 6,* 95-109.

Contributors

P.S. Adey, School of Education, King's College, London, United Kingdom

L. Chanquoy, University Paul Valéry, Montpellier, France

B. Csapó, Department of Education, Szeged University, Hungary

E. De Koning, Department of Education, Utrecht University, the Netherlands

A. Efklides, School of Psychology, Aristotle University of Thessaloniki, Greece

M. Elshout-Mohr, SCO-Kohnstamm Institute and Institute of Teaching and Learning, University of Amsterdam, the Netherlands

W. Hager, Institute of Psychology, Göttingen, Germany

J.H.M. Hamers, Department of Education, Utrecht University, the Netherlands

K.J. Klauer, Department of Educational Science, Technical University, Aachen, Germany

J. Mehtäläinen, Department of Education, University of Jyväskylä, Finland

J.M.C. Nelissen, School Advising Center, Utrecht, the Netherlands

P.M. Scheinin, Department of Education, University of Helsinki, Finland

J.E.H. Van Luit, Department of Education, Utrecht University, the Netherlands

H. Van Oostendorp, Department of Psychonomics, Utrecht University, the Netherlands

L. Verschaffel, Center for Instructional Psychology and Technology, University of Leuven, Belgium

Subject Index

V

W

Z

CONTEXTS OF LEARNING
Classrooms, Schools and Society

1, *Education for All.* Robert E. Slavin
1996. ISBN 90 265 1472 7 (hardback)
ISBN 90 265 1473 5 (paperback)

2, *The Road to Improvement: Reflections on School Effectiveness.* Peter Mortimore
1998. ISBN 90 265 1525 1(hardback)
ISBN 90 265 1526 X (paperback)

3, *Organizational Learning in Schools.* Edited by Kenneth Leithwood and Karen Seashore Louis
1999. ISBN 90 265 1539 1 (hardback)
ISBN 90 265 1540 5 (paperback)

4, *Teaching and Learning Thinking Skills.* Edited by J.M.H. Hamers, J.E.H. van Luit and B. Csapó
1999. ISBN 90 265 1539 1 (hardback)

5, *Managing Schools Towards High Performance: Linking school management theory to the school effectiveness knowledge base.* Edited by Adrie J. Visscher
1999. ISBN 90 265 1546 4 (hardback)

The main goal of this volume is to provide an up-to-date grounded review of the psychological and educational literature on thinking, problem solving and teaching thinking. Problem solving in mathematics, reading comprehension, text production and in our everyday lives are the areas addressed in this volume.

This volume covers most of the topics that traditionally form the foundation of the field, and includes several major approaches to development and education: the Piagetian and neo-Piagetian, information processing, learning, contextual and psychometric. Each of the chapters deal with one or two theoretical approaches to thinking, problem solving and/or teaching thinking in programs, and most of the chapters include theoretical considerations, research and program proposals. This book can be read as an introduction to the study of thinking and problem solving in several areas. Due to the diversity of approaches, an introduction and concluding chapter are included to help tie the chapters together.

ISBN 90-265-1545-6

9 789026 515453

SWETS & ZEITLINGER PUBLISHERS