

Teaching Inductive Reasoning in Primary Education

Els de Koning

Leiden University, Leiden, The Netherlands

Jo H. M. Hamers

Utrecht University, Utrecht, The Netherlands

Klaas Sijtsma

Tilburg University, Tilburg, The Netherlands

and

Adri Vermeer

Utrecht University, Utrecht, The Netherlands

This study deals with smoothing the transition from developmentally oriented education in the preparatory phase to curriculum-oriented education from the third grade onward. The emphasis on inductive reasoning that already exists in the preparatory phase was extended to the third and the fourth grades by implementing two classroom-teaching programs. Theories on inductive reasoning provided the foundation for the operationalization of the tasks that were used and the reasoning processes that were taught. The tasks and processes were embedded in a three-phase procedure for teaching, which was based on the development of metacognitive knowledge. The materials chosen for teaching inductive reasoning were adapted to the developmental level and educational needs of the pupils: in Grade 3, visual material (building blocks and pictures) was used, and in Grade 4 verbal material (texts) was used.

Results demonstrated that the teachers were able to apply the programs as intended, although they needed support to shift their attention from the reasoning product to the reasoning process. They also experienced difficulties in implementing the role swap between the teacher and the pupils in the third phase of the teaching procedure. The pupils who were administered the programs significantly outperformed pupils from the control condition on near-transfer tests. Doubly taught pupils (i.e., in Grade 3 and Grade 4) significantly outperformed singly taught pupils on a far-transfer test, and singly taught pupils significantly outperformed pupils from the control condition. The learning effects of both programs indicated that pupils' better performance was due to their increased inductive reasoning ability. Suggestions are

Address correspondence and reprint requests to Jo H. M. Hamers, Department of Special Education, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, The Netherlands. E-mail: J.Hamers@fss.uu.nl.



provided for improving the teaching programs and for future research on developmentally appropriate and cognitively based change in education. © 2002 Elsevier Science (USA)

This article reports on a 2-year-long longitudinal study on teaching of inductive reasoning in Grades 3 and 4 of primary education. This study extended the present emphasis on inductive reasoning in the preparatory phase to the third and the fourth grades by implementing two teaching programs.

Induction is the formulation of general laws from particular cases, for example, "all observed cats are black, therefore all cats are black." Obviously, inductive reasoning does not necessarily results in valid laws. In this study, the process of induction is of interest, not the validity of the laws that are produced by the process. The fundamental role of the process of inductive reasoning for child development and its domain-independent applicability has been widely recognized (Holland, Holyoak, Nisbett, & Thagard, 1986; McShane, 1991; Snow, Kyllonen, & Marshalek, 1984; Sternberg, 1998; Sternberg & Gardner, 1983).

In most western countries, the preparatory phase of primary education closely follows and stimulates the *development* of children. That is, educating pupils aims at supporting their gradual understanding of the world. For example, education in the preparatory phase links with age-appropriate mental representations (e.g., Piaget, 1970) and age-appropriate mental processing capacity (e.g., Halford, 1993). Furthermore, the natural developmental mechanisms are supported, for example, by the equilibration process (Adey, 1999), causal reasoning (Goswami, 1998), and automating of cognitive processing for freeing memory space (Case, 1974). Examples of activities in the preparatory phase of primary education are classifying and seriating of blocks with respect to color, form, and number and discovering relationships such as action–reaction or cause–effect with lotto cards. These activities require pupils to order their environment by connecting units using the definition of laws. This is inductive reasoning.

From the third grade of primary education onward, such general reasoning activities are no longer part of the lessons. Instead, education becomes focused on reading, writing, and math, which are considered to be the main requirements for participation in our western society. This *curriculum-oriented* education is led by internal structures of specific domain knowledge (e.g., the syntax of the language and the metric system) and by the domain-tied cognitive processes that underlie efficient learning (e.g., distinguishing nouns from verbs and counting). Research on learning and information processing provides methods as to how knowledge could best be sequenced and what instruction procedures would work well.

The current transition from *developmentally oriented* education in the preparatory phase to curriculum oriented education from the third grade onward

is rather abrupt. There are two main reasons for wanting to bridge this gap. First, our increasingly more complex society demands people to handle huge amounts of information that becomes dated quickly (Hamers, De Koning, & Sijtsma, 1998). Hence, pupils should not only be taught considerable amounts of knowledge and skills for reading, writing, and math, but they should also be equipped with general reasoning skills to order information processing. This enables them, for example, to connect bits of information into meaningful representations and to distinguish relevant and irrelevant information for a particular task.

Second, reading, writing, and math performances are dependent on general reasoning skills (De Koning & Hamers, 1999). For example, text comprehension is a constructive process that involves making inferences and integrating information from separate words and sentences. Basically, this process is inductive. At present, it is assumed that all pupils enter the third grade with sufficient inductive reasoning skills for good performances on, for example, reading comprehension. Many studies showed that low-socioeconomic (SES) pupils do not reach standards set for reading and math performances and that they have lower scores on inductive reasoning tests than their classroom peers from middle-class backgrounds (Demery, 2000; Hamers et al., 1998; Hessels & Hamers, 1999; Janssen, Van der Schoot, Hemker, & Verhelst, 1999; Leseman & De Jong, 2001; Ma & Klinger, 2000; Stipek, 2001). De Koning (2000) showed that in a sample of 945 pupils the scores of low-SES pupils on an inductive reasoning test (Standard Progressive Matrices; Raven, 1958) were significantly lower than the scores of pupils from higher SES levels. An explanation is that these low-SES pupils lack experience with the verbal routines, such as making verbal comparisons and identifying attributes, typical of parent-child interactions in middle-class homes (Yuill & Oakhill, 1991). In the following, we show that these verbal routines are essentially inductive.

There have been many attempts to smooth the transition from developmentally oriented to curriculum-oriented education. On the one hand, in the preparatory phase activities have been implemented that underlie later reading, writing, and math. On the other hand, from the third grade onward, "adaptive education" aims at adapting the curriculum to the individual development of pupils. Implementing "adaptive education" requires major adaptations of the curriculum, the instruction, and the classroom organization. Up to now, this "cross-fertilization" between both phases in education resulted mainly in the addition of (domain-tied) third-grade activities to the preparatory phase. This is also the focus of many intervention programs meant to prevent academic backwardness of low-SES pupils. In contrast, this study aims at extending the stimulation of the development of inductive reasoning in the preparatory phase to the third and fourth grades by embedding special teaching programs in this part of the curriculum. It was expected

that inductive reasoning could be taught in regular classroom settings and then transferred to regular school-curriculum domains, independent of the material used for teaching.

Operationalization of Inductive Reasoning Tasks

Klauer (1989) defined inductive reasoning as the systematic and analytic comparison of objects aimed at discovering regularity in apparent chaos and irregularity in apparent order. Regularities and irregularities at the nominal level are recognized through comparing the attributes of elements, and comparisons at the ordinal level and the ratio level involve relationships among elements. As comparison processes can be aimed at finding similarities, dissimilarities, or both, attribute tasks and relationship tasks can both deal with any of these three modes. This results in the definition of six kinds of domain-independent inductive reasoning tasks (De Koning, Sijtsma, & Hamers, in press-a, -b). This means that, for example, classification and seriation of basic building blocks are considered to be based on the same reasoning processes as classification and seriation of sentences in a text, although the mental codes used for representing visual and verbal material are different. Common to these codes is that there is a network that links units of codes in categories and relationships among categories. This linking corresponds to classification, seriation, and finding analogies typically of tasks used for measuring inductive reasoning (e.g., Sternberg & Gardner, 1983).

For teaching inductive reasoning it was assumed that visual material would best facilitate third-grade pupils' learning as it links up with their preference for perceptual information processing. Also, the majority of the material regularly used in preparatory education demands perceptual processing. As development proceeds, semantic representations of conceptual categories and relationships among categories come along with ability to verbally process information. Therefore, verbal material (texts) was used for teaching inductive reasoning in the fourth grade. In addition, this material links up with the huge difficulties disadvantaged pupils encounter in comprehensive reading from the very start in the fourth grade onward. Figure 1 presents four examples of some relative easy tasks that were used in the first phase in either of the two teaching programs. In the upper half of the figure, the tasks require pupils to discover similarities and dissimilarities among attributes of objects (visual material) or sentences (verbal material). The bottom half of the figure shows examples of tasks that demand pupils to compare objects or sentences with respect to their relationships.

The domain independence of inductive reasoning expects pupils to transfer reasoning between visual and verbal material. However, transfer between domains does not occur easily (e.g., Brown, Bransford, Ferrara, & Campione, 1983). Many researchers agree now on the importance of incorporating meta-cognitive knowledge into teaching for transfer (e.g., Campione, Brown, & Ferrara, 1982). Several intervention studies explicitly aimed at enhancing


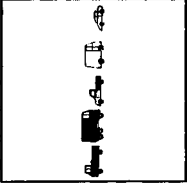
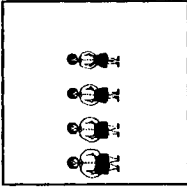

		Dissimilarity	
Similarity	Dissimilarity		
Group Tasks (Nominal Level)	Visual Material (Third grade program)	Verbal Material (Fourth Grade Program)	Visual Material (Third Grade program)
Categories based on <i>attributes</i> of objects/sentences		The teacher brings in sweets. And she decorated the classroom. She wears new clothes. Then John enters the classroom. <i>What does he think?</i> <i>Think of a title.</i>	
	<i>Does it belong together?</i>		<i>What doesn't belong?</i>
Row-Tasks (Ordinal Level)	Visual Material (Third grade program)	Verbal Material (Fourth Grade Program)	Visual Material (Third Grade program)
Series based on <i>relationships</i> among objects/sentences		Uncle Tom made a tree-hut. John climbs into the hut. Uncle comes in too. John hears the cracking of a huge branch. Uncle Tom looks down. <i>What happens next?</i> <input type="checkbox"/> Uncle Tom falls through the floor of the hut <input type="checkbox"/> The branch breaks. <input type="checkbox"/> John gets sticking plaster. <i>Think of a title.</i>	
	<i>Who's next?</i>		<i>What seems wrong here?</i> <i>Think of a title.</i>

FIG. 1. Examples of (easy) tasks of the visual and verbal programs that require students to compare either attributes or relationships in order to discover similarities or dissimilarities.

metacognitive knowledge showed higher performances in, for example, writing (Scardamalia & Bereiter, 1986), reading comprehension (Palinscar & Brown, 1984), and math (De Corte, Verschaffel, & Op 'T Eynde, 2000).

Operationalization of a Teaching Procedure

For operationalizing the teaching procedure of the third and the fourth grade teaching programs we adopted Boekaerts's (1997) model of self-regulated learning, which describes the development of metacognitive knowledge. The model consists of three levels of prior knowledge that can be used to teach for transfer. The process of transfer is on a *low level* (i.e., low road transfer: Salomon & Perkins, 1989) if it involves a simple application of conceptual and procedural knowledge in the same domain used for teaching the knowledge. The domain-specific subset of pupils' knowledge will be relatively easily activated when pupils are confronted with a learning task that clearly resembles the key features of the task previously used to teach the knowledge. The key features refer to surface attributes of the tasks, such as perceptual similarities (Baron & Sternberg, 1987). The knowledge can bridge only a small gap between two tasks; hence this low-level process of transfer results in higher scores on near-transfer tasks but not on far-transfer tasks.

The process of transfer is on a *middle level* if it refers to the use of strategic knowledge. This kind of knowledge relates to general cognitive strategies, such as elementary deductive and inductive reasoning rules and means–end analyses. It is necessary to explicitly direct instruction and practice at inducing transfer of these strategies to new domains (Ennis, 1990). This middle level process of transfer can bridge a wider gap between two tasks, which results in higher scores on near-transfer tasks and near-far transfer tasks, but not on far-transfer tasks.

The *highest* process level of transfer (i.e., high road transfer: Salomon & Perkins, 1989) is connected to decontextualized knowledge, such as scientific thinking schemes, main ideas, and final learning goals. The transfer relies on mindful application of the schemes' encompassing metacognitive strategies. These strategies sequence the information processing from "orientation" regarding the task at hand to the "planning" of actions that are "monitored" and "evaluated" with respect to predefined criteria. This high level process of transfer can result in high scores on near-transfer tasks, near-far transfer tasks, and far-transfer tasks.

In our study, the three levels of transfer were connected to a three-phased teaching and learning sequence of inductive reasoning processes. The measurement periods were before and after the teaching programs and involved tests comprising of near-transfer and far-transfer tasks. The section on instruments provides examples of these tasks.

Operationalization of Inductive Reasoning Processes

For operationalizing inductive reasoning processes of the third- and the fourth-grade teaching programs we followed Nisbett (1993) in defining top-

down processes and Sternberg and Gardner (1983), who defined the bottom-up processes of inductive reasoning. *Top-down* processing is important in induction given its probabilistic character: In contrast to computers, people make use of top-down shortcuts (Holyoak & Nisbett, 1988). For adults, Nisbett (1993) described two domain-independent inductive reasoning schemes that are based on two statistical rules, the "rule of large numbers" and the "rule of regression." Basically, these rules deal with the certainty with which the induced general laws sufficiently cover reality. De Koning and Hamers (1999) assumed that for pupils to develop these top-down inductive reasoning schemes, they could benefit from an *awareness* of how knowledge parts are connected. Following Nisbett's (1993) statistical terminology, for pupils two domain-independent schemes were described: a "group structure scheme", expressed on a nominal scale, and a "row structure scheme", expressed on an ordinal or rational scale (see also Halford, 1993). The group structure scheme directs classification processes, and the row structure scheme directs seriation processes (De Koning & Hamers, 1999).

Sternberg and Gardner (1983) investigated the *bottom-up* processes of inductive reasoning by analyzing how people solved analogies. For classroom teaching, the bottom-up processes in both group and row structure tasks were translated as follows: *search* for the relevant attributes or relationships, *compare* similarities or dissimilarities of attributes or relationships, *solve* the problem on the basis of comparison, and *control* the solution (De Koning & Hamers, 1999; Hamers, et al., 1998). See Fig. 2 for a schematic overview.

Integration of Tasks, Processes, and Teaching

In this study, the inductive reasoning tasks, the inductive reasoning top-down and bottom-up processes, and the three-phase procedure for teaching

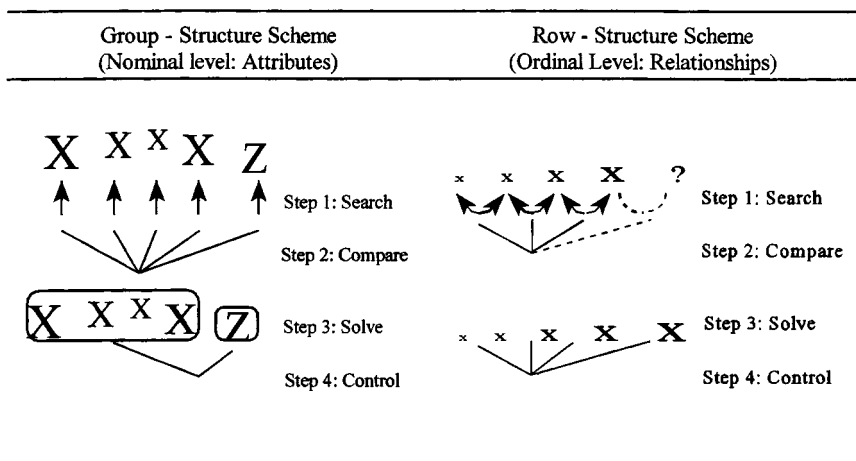


FIG. 2. Group and row structure schemes and bottom-up reasoning processes.

transfer were integrated and implemented into two programs. The program using visual material was a new classroom version of Klauer's *Program Inductive Reasoning I* (De Koning & Hamers, 1995) and was applied to the third grade. The *Program Inductive Reasoning in Reading Comprehension* (De Koning & Hamers, 1996) consisted of verbal material and was applied to the fourth grade. In both programs, teaching and learning were embedded in a classroom dialog in which the pupils and the teacher explicated exactly how they reasoned inductively. The use of dialog is rooted in Vygotsky's view that children's development moves from the social, intermental plane on which collaboration with knowledgeable other people is central to the psychological, intramental plane (Wertsch & Kanner, 1992). The programs introduce the reasoning processes using easy-to-perceive (visual program) and easy-to-understand (verbal program) materials. Hence, each pupil possessed the same prior inductive reasoning experiences when more abstract and unfamiliar contexts were introduced into the programs' tasks. More specifically, the visual material changed from building blocks to pictures of objects and to pictures of situations and abstract figures; the verbal material changed from short texts with everyday context to longer texts with everyday context and to longer texts with unfamiliar context (see Fig. 3). During the program, the variation in tasks demanding different inductive reasoning strategies increased. A more detailed description of the programs can be found under Method.

The three instruction phases aim at the integration of the top-down processes (i.e., the group structure scheme and the row structure scheme) and the bottom-up processes (i.e., searching, comparing, solving, and checking). During the three instruction phases, the responsibility for appropriate reasoning shifts from the teacher to the pupil (see Fig. 3). In the *introductory* phase, the teacher provides the pupils with maximum assistance, or even models the use of the top-down processes and the bottom-up processes, such that pupils become able to use parts of them. In the *practice and reflection* phase, the teacher guides the pupils such that they become aware of both kinds of processes. Before problem solving, the teacher activates the pupils' prior knowledge by reminding them of similar problems solved earlier. During problem solving, the teacher questions the pupils about their reasoning processes. After problem solving, the pupils explicate their reasoning processes and justify their reasoning results. Furthermore, the teacher enhances the pupils' reflection on the inductive reasoning processes by stimulating them to classify the tasks into group or row structure tasks and by explicitly repeating the pupils' application of the processes search, compare, solve, and check. Finally, the teacher provides the pupils with an adequate vocabulary to describe their thinking processes. In general, in this phase the teacher and the pupils share responsibility for the learning process. In the *reciprocal* phase, the teacher supports the integration of both the top-down processes and the four bottom-up processes such that pupils learn to regulate their

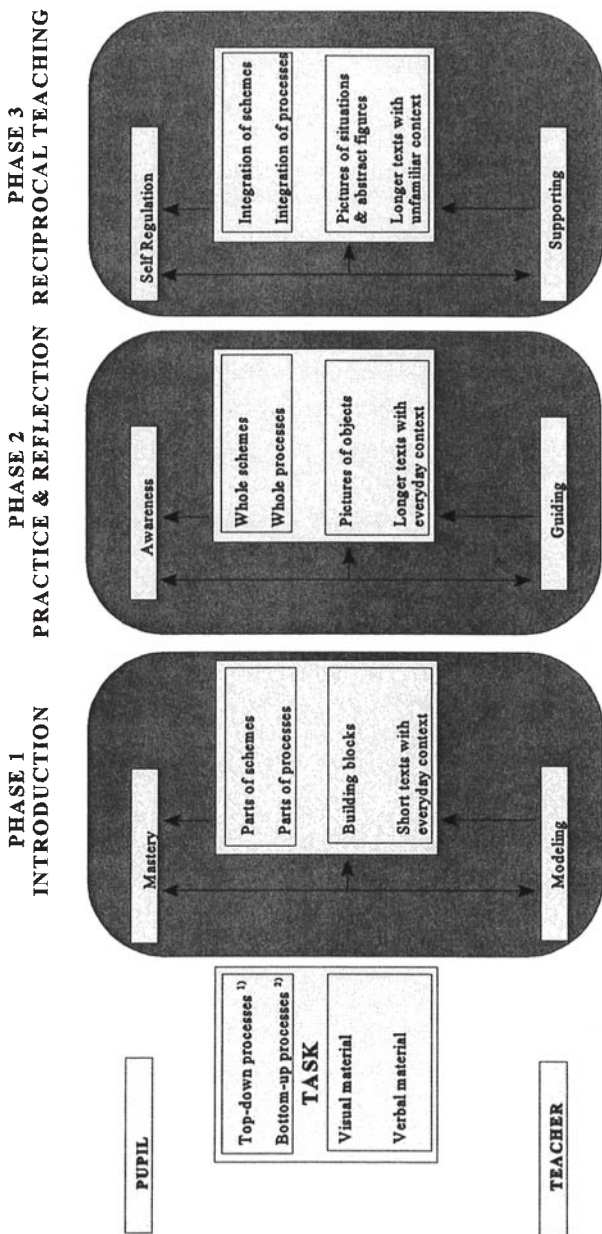


FIG. 3. A model for teaching and learning inductive reasoning.

1), Top-down processes: group and row schemes.

2), Bottom-up processes: searching, comparing, solving, and checking.

inductive reasoning. That is, they learn to orientate themselves on what kind of reasoning structure (i.e., group or row structure) they are dealing with and to monitor and evaluate the steps taken in the solution process. Furthermore, in this phase the pupils take over the role of the teacher so that it becomes clear when their self-regulation is insufficient.

In this study, first it was investigated whether teachers were able to apply the teaching programs to their own class. Second, it was investigated whether the pupils reached higher scores on transfer tests after finishing a program than the control pupils who followed the regular curriculum. The near-transfer tests measured learning effects on tasks that closely resembled the tasks from the programs. The far-transfer test consisted of tasks not trained in the programs. The tests were administered immediately after finishing the programs and once more after 3 months. Third, it was investigated whether pupils who followed both teaching programs in succession had higher scores on the transfer tests than pupils who followed only one teaching program.

METHOD

Design

A three-group pretest/posttest/follow-up test control group design was used (see Fig. 4). Conditions 1, 2, and 3 were experimental conditions, and Condition 4 was a control condition. Pupils in Condition 1 followed two teaching phases. In the first phase, the *Program Inductive Reasoning 1* (De Koning & Hamers, 1995) with visual material was applied to Grade 3, and in the second phase the *Program Inductive Reasoning in Reading Comprehension* (De Koning & Hamers, 1996) with verbal material was used in Grade 4. Pupils in Condition 2 were only taught with visual material in the first phase in Grade 3, and pupils in the third Condition were only taught with verbal material in the second phase in Grade 4. In the control Condition (C4), the pupils were not taught by means of any of the two programs.

Phase 1: Grade Three				Phase 2: Grade Four			
Sept.- Dec.	Jan.-April	May- June	Sept.- Oct.	Oct.-Jan.	Feb.- March	June- July	
Meas. 1	Program Ind. Reas. (Visual Tasks)	Meas. 2	Meas. 3	Program Ind. Reas. In Read. Comprehension (Verbal Tasks)	Meas. 4	Meas. 5	
C1	X	O	X	X	O	X	X
C2	X	O	X	X		X	X
C3	X		X	X	O	X	X
C4	X		X	X		X	X

FIG. 4. Design of the study with four conditions (C1 through C4), two teaching phases (Grade 3 and Grade 4), and five measurements (1 through 5).

Tests were administered during five measurement periods, before both teaching phases, immediately after, and 3 months later. Each condition comprised four classes (with four teachers) selected from eight different schools, thus preventing confounding between school and condition effects. For analyzing the near-transfer effect of the first teaching phase in Grade 3, Conditions 1 and 2 were compared to Conditions 3 and 4. The near-transfer effect of the second teaching phase in Grade 4 was evaluated by comparing Conditions 3 and 4. The Conditions 1, 2, 3, and 4 were compared for analyzing far-transfer effects.

Population and Sample

The population consisted of Dutch schools with more than 80% low-SES pupils. For these pupils their limited experiences in inductive reasoning at home are a risk for lower academic achievements. The majority of the pupils came from ethnic minorities (particularly from Turkey, Morocco, and the Netherlands Antilles). The inclusion criterion "school" was operationalized by means of a "school score," which reflects the number of pupils visiting the school, each pupil weighted by his or her SES score. Based on a higher school score, a school receives additional financial support from the national government. The weights are 1.25 for Dutch working-class children, 1.40 for bargee children not living with their parents, 1.90 for children having at least one non-Dutch parent, and 1.00 for all other children. If the school population consisted of more than 80% pupils with a weight of 1.90 it was a candidate for inclusion in the sample. From a list of 31 schools complying with this criterion, eight schools were randomly selected. The mean pupil SES weights did not differ across the conditions ($F = 1.123$, $df = 3$, $p = .34$). The mean age of the pupils was 85.73 months and the standard deviation was 6.09. The mean ages did not differ across the conditions ($F = .183$, $df = 3$, $p = .91$). The mean (standardized) score on the Standard Progressive Matrices (Raven, 1958) was comparable to the score of pupils with SES score of 1.90 in another sample that contained 945 pupils in total. In that sample the scores of the pupils with SES score of 1.90 differed significantly from the scores of pupils with SES scores of 1.25 and 1.00. Table 1 shows the numbers of pupils involved in this study. For testing the differential learning effects of the conditions, only pupils who took all the tests were included in the analyses.

Materials for Teaching and Learning Inductive Reasoning: Two Teaching Programs

The *Program Inductive Reasoning 1* consists of visual material: building blocks and large pictures that can be attached to the blackboard. The program comprises 20 lessons, each containing six tasks. In the first 4 lessons only generalization tasks (similarity of attributes) and seriation tasks (similarity of relationships) were discussed. From lesson five onward discrimination

TABLE 1
Frequencies of Pupils in the Conditions and the Two Teaching Phases

	Phase 1 + Phase 2	N	Phase 1	N	Phase 2	N
	C1	98	C1 + C2	189		
	C2	91				
	C3	61	C3 + C4	142	C3	61
	C4	<u>81</u>			C4	<u>81</u>
Total Conditions		331		<u>331</u>		142
Pupils from C3 and C4 who left after Phase 1				<u>11</u>		
Total Phase 1				342		
Pupils new in Phase 2: extra in Condition 3						14
Pupils new in Phase 2: extra in Condition 4						<u>28</u>
Total Phase 2						184

tasks (dissimilarity of attributes), disturbed seriation tasks (dissimilarity of relationships), cross-classification (similarity and dissimilarity of attributes), and system formation tasks (similarity and dissimilarity of relationships) were gradually introduced. First similar kinds of tasks were clustered together and then, in the last 2 lessons, each of the six kinds of tasks had to be solved.

The relevant attributes and relationships that had to be discovered were hidden to a higher degree as the lessons progressed. In the introductory phase, building blocks with simple shapes were used. By manipulating the blocks it was possible to actually show the results of various classifications on the basis of different attributes or seriations on the basis of different relations. In the practice and reflection phase, pictures of objects from everyday life were used. An example is a generalization task that required the pupils to find the common attribute in pictures of a helicopter, a butterfly, and a kite. In this phase, the manipulation part of the solution procedure was mental: the teacher and the pupils had to verbalize their thinking. In the reciprocal teaching phase, pictures of realistic situations and abstract (i.e., meaningless) material were used. An example of a cross-classification task asked pupils to place a sugar pot in a cupboard of which the content was organized with respect to two attributes that had to be discovered. The first attribute referred to the material the objects in the cupboard were made of (e.g., porcelain or not porcelain). The second attribute referred to the function of the objects (e.g., used for drinking or not used for drinking). To emphasize the importance of the group and row structures, during each lesson the group tasks remained visible on the right-hand side of the blackboard and the row tasks on the left-hand side.

The *Program Inductive Reasoning in Reading Comprehension* comprised 16 lessons and consisted of tasks composed of texts. The problem solving procedure was divided into a reading part and a thinking part (De Koning & Hamers, 1999). In the *reading part*, the structure of each sentence in the text was analyzed into a noun phrase (who is it about and what is it about) and a verb phrase (what is said about it). To reduce memory load, for each text a scheme was used in which the structure of the sentences and the text was visualized. That is, on the blackboard for every sentence the “who” part was written down and for the “what” part a picture was attached. Also the pupils had their own scheme and pictures. In this way, the pupils were shown that it is necessary to consciously convert the abstract verbal symbols (i.e., the written text) into semantic meaningful images (Yuill & Oakhill, 1991).

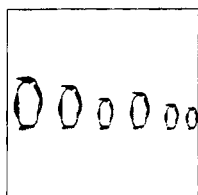
The thinking part was directed at the discovery of similarities and dissimilarities in attributes and relationships between text parts. The aim was to induce how the sentences in the text were ordered or why they were incorrectly ordered. Ordered texts consisted of sentences that dealt with the same theme (generalization task: similarity of attributes) or that had a logical sequence (seriation: similarity of relationships). In unordered texts, that is, texts with gaps or inconsistencies, there were sentences which did not deal with the main theme (discrimination task: dissimilarity of attributes) or in which the sequence was disturbed by an abnormal relationship (disturbed seriation task: dissimilarity of relationships).

In the introductory phase, the four task types were operationalized in separate texts. The group structure tasks (similarity and dissimilarity of attributes) and the row structure tasks (similarity and dissimilarity of relationships) were combined in the practice and reflection phase. Finally, in the reciprocal teaching phase, the four task types were combined into one long text. After solving the tasks pupils had to compose an appropriate title for the text. The texts changed from everyday content to one that was more abstract and unknown. The texts in the first half of the program involved situation sketches, which linked up to the pupils’ knowledge representations. The second half of the program dealt with expository texts. It was assumed that many pupils had insufficient knowledge of all the words in the text and of the theme of the text. By analogy with the build-up of the visual program, the relevant attributes and relationships that had to be discovered were thus hidden to a higher degree as the lessons progressed.

Instruments

Test for Inductive Reasoning (TIR). The near-transfer learning effects of the *Program Inductive Reasoning* with visual tasks were measured by means of the Test for Inductive Reasoning (TIR), which was specially designed for this research project (De Koning et al., in press-a, -b). The TIR contains six types of inductive reasoning problems that reflect the six types of tasks of the program. The six types of tasks are pictured both in daily life objects

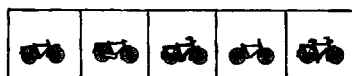
Row-Tasks

Grade 3 Teaching Program Task

What's wrong?

Test Item

(Test for Inductive Reasoning: near-transfer)



What's wrong?

Grade 4 Teaching Program Task

John goes to his chicken.
He looks for an egg in the chicken coop
John gives the egg to his uncle.
Uncle fries the chicken in the pan.
John likes his egg.

*What seems wrong here?
Think of a title.*

Test Item

(Reading Comprehension: near-transfer)

I give the flowers to mum.
I gather a bunch of flowers for mum.
There are lots of flowers in the field.
Mum puts the flowers in a vase.

What should be the last sentence?

FIG. 5. Examples of tasks in the teaching programs and test items in the near-transfer tests.

and in meaningless abstract shapes. The test consists of a pretest that is administered before the program and a posttest for measuring the learning effects after the program. The pupils' scores on the pretest and the posttest were transferred to a common scale comprising the estimates of their latent inductive reasoning ability. The pretest and posttest consist of 25 and 30 multiple-choice questions, respectively. Figure 5 contains an example of a teaching task and a test item that require pupils to focus on the dissimilarity of relationships among objects.

Tests for reading comprehension. Tests for reading comprehension (Verhoeven, 1992) were used to measure the near-transfer learning effects of the program using verbal tasks and the far-transfer effects of the program using visual tasks. The tests have satisfactory reliabilities and are widely used in Dutch primary education for comparing the reading comprehension achievement of individual pupils or groups of pupils. Two kinds of tests were used that closely resembled the learning aims of the program. Similar to the TIR, the pretest and posttest scores were transformed to a common scale of latent comprehensive reading ability. The first kind of tests measured whether pupils were able to construct a meaning from sentences with or without pictures and from combinations of sentences (short stories). The pretest consisted of

27 multiple-choice questions, and the posttest consisted of 30 questions. The second kind of tests measured whether pupils were able to construct correct inferences about persons, time, space, actions, and events figuring in short stories. Both pretest and posttest consist of 24 multiple-choice questions. Figure 5 contains an example of a teaching task and a test item that require pupils to focus on the dissimilarity of relationships among sentences.

Raven's Standard Progressive Matrices. The far-transfer learning effects of both the visual and the verbal programs were measured with the Standard Progressive Matrices (SPM; Raven, 1958). Much research confirmed that the SPM-Raven test is a valid and reliable measure of inductive reasoning (Carpenter, Just, & Shell, 1990) and that the scores are fairly independent from ones native tongue. The test consists of 60 items divided into five parts (called sets A through E). Each of the 60 pages, one item per page, is divided into two half-pages. On the upper half, a matrix of figures is depicted containing a missing element. This element has to be detected among the six alternatives (sets A and B) or eight alternatives (sets C through E) printed at the bottom of the page. As the SPM-Raven contains tasks of various difficulty, Bereiter and Scardamalia (1979) quantified 48 of the 60 SPM-Raven items in terms of Mental Demand (MD), which they defined as increasing from level 1 to level 5 (Fig. 6). For correctly solving the easiest items (MD = 1) all that is necessary is to retain one stimulus feature in mind while inspecting adjacent elements. Furthermore, apart from inductive reasoning processes Gestalt effects might also be part of the problem-solving procedure. Items with mental demands at levels 2, 3, 4, and 5 can only be solved by inductive reasoning. The items are progressively more complex because they require a mental coordination of an increasing amount of discovered features in memory. It was expected that the scores of pupils in this study would increase from MD2 items through MD4 items and that MD5 items would be too difficult (see also Case, 1974).

Other Tests

It was expected that the pupils of the low-SES sample would differ in their understanding of the Dutch language. For separating the learning effects due to *Program Inductive Reasoning* with visual material and the verbal

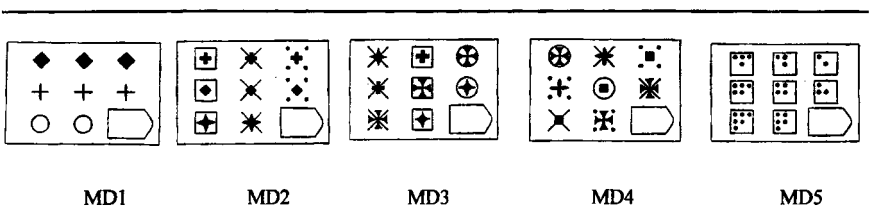


FIG. 6. Examples of SPM-Raven items with different levels of mental demands (MD).

communication skills, a Vocabulary Test (Verhoeven, 1996) and a Test for Listening Comprehension (CITO, 1995) were administered. The Vocabulary Test consisted of four pictures per item. The pupils had to indicate a picture that fit the description read by the teachers. The test consisted of 50 items. The Listening Comprehension Test consisted of 44 statements and short stories the teacher read out loud. The pupils had to indicate the picture that matched the teacher's reading.

The scores on reading comprehension are influenced by the ability to adequately decode the written symbols. Pupils in Grade 4 may differ remarkably in their decoding ability. To separate the learning effects due to the *Program Inductive Reasoning* with verbal material and the decoding ability, a Technical Reading Test (Verhoeven, 1995) was administered. The test consisted of a card on which 150 words are printed. Pupils had to read as many words as possible in 1 min. The score was the number of words the pupil was able to read correctly.

All tests have satisfactory reliabilities and are widely used in Dutch primary education for comparing the achievements of individual pupils or groups of pupils. Apart from the Technical Reading Test that had to be administered individually, all tests could be applied in classroom sessions. The administration of the SPM-Raven and the TIR took 45 and 60 min, respectively. The Reading Comprehension Tests, the Vocabulary Test, and the Listening Comprehension Test took approximately 90 min each.

Observations

The lessons of the teaching programs were observed and rated using 17 statements with Likert scales ranging from 1 to 5, each indicating to what degree the statements characterized the lesson correctly. The statements referred to the key features of the lesson, such as the classroom dialog, the three-phase instructions, and attention given to the inductive reasoning processes. One observer (first author) attended part of the lessons of all teachers. Six other observers (graduate students) each closely monitored a few teachers during most of the program lessons. The observers compared their observations of particular teachers. The observation forms were filled out during or immediately after the lesson and discussed with the teacher. Figure 7 provides an overview of the time schedule of the study, the conditions, and the instruments used.

Procedure

The teachers were provided with two kinds of information. First, the objectives underlying the three-phase teaching method were described in general, rather than abstract, terms. Second, the teachers received a manual with detailed examples of the typical tasks and dialogs to be used preferably in each separate phase. In a few sessions teachers were supported to connect the two kinds of information. Special attention was given to the classroom dialog.

Phase 1: Grade Three				Phase 2: Grade Four			
Sept.-Dec.	Jan.-April	May-June	Sept.-Oct.	Oct.-Jan.	Feb.-March	June-July	
Meas. 1	Program Ind. Reas. (Visual Tasks)	Meas. 2	Meas. 3	Program Ind. Reas. In Read. Comp. (Verbal Tasks)	Meas. 4	Meas. 5	
C1	X	O	X	O	X	X	X
C2	X	O	X		X	X	X
C3	X		X	O	X	X	X
C4	X		X		X	X	X
TIR (pretest)	Observations	TIR (posttest)	TIR (f-up test)	Observations	Read. Comp. (posttests)	Read. Comp. (f-up tests)	
SPM Raven		SPM Raven	SPM Raven		SPM Raven	SPM Raven	SPM Raven
Listening Comp. Vocabulary							
Technical Reading							

FIG. 7. Design of the study with four conditions (C1 through C4), two teaching phases (Grade 3 and Grade 4), five measurements (1 through 5), and the instruments used.

It was emphasized in particular that correct solutions without scheme-based and process-based justifications were not considered to be accurate answers.

The *Program Inductive Reasoning* using visual material was applied to Grade 3 after the midterm break from January until April/May. The *Program Inductive Reasoning* using verbal material in Grade 4 started in October before the midterm break and lasted until February/March. To prevent the pupils in Condition 3 from doubling their learning experience, all tasks that resembled the tasks in the *Program Inductive Reasoning* with verbal material were deleted from the regular curriculum.

Data Analysis Methods

Analysis of covariance (ANCOVA) was done with factor Condition with four levels (see Fig. 7). The pretest scores on the instruments that were used as dependent variables were incorporated in the analyses as covariates. For testing the learning effects of *Program Inductive Reasoning* using visual material on the TIR, Conditions 1 and 2 were compared with Conditions 3 and 4. The differences between the means of the scores on the Vocabulary Test and the scores on the Listening Comprehension Test were tested. In case of significant differences the scores on these instruments were incorporated as covariates. Comparison of Condition 3 and Condition 4 tested the learning effects on the tests for reading comprehension of *Program Inductive Reasoning* using verbal material. The difference between the means of the scores on the Technical Reading Test was tested. If the conditions differed significantly, the scores were incorporated as covariates. The differential learning effects among the four conditions on the SPM-Raven and the tests for reading comprehension were tested by means of ANCOVA with a repeated-measure procedure.

We anticipated that the singly taught pupils would outperform pupils from the control condition on near-transfer tests (Grade 3: Condition 1 and 2 versus Conditions 3 and 4; Grade 4: Condition 3 versus Condition 4). Furthermore, it was expected that doubly taught pupils (i.e., pupils in Condition 1 who followed the third- and fourth-grade programs) would outperform singly taught pupils on the far-transfer test (i.e., pupils in Conditions 2 and 3) and that these singly taught pupils would significantly outperform pupils from the control condition (i.e., pupils in Condition 4).

RESULTS

Teaching Inductive Reasoning: Classroom Observations

Without exception the lessons were very demanding for the teachers. First, teachers had to be assisted in bringing about a process-focused attitude while weakening their emphasis on the product of reasoning. For example, they had to be taught how to ask their pupils questions like “Why do you think this answer is correct” or “What other correct solutions do you suggest”

instead of the more usual “Right, this is correct.” In using the fourth-grade program in particular it was difficult for teachers to distract from the content and to concentrate on the underlying reasoning because the content resembled the regular reading comprehension curriculum. Second, planned three-phase instructions were not an integral part of daily educational practice. Although the teachers weakened their scaffolding in the process of learning, it proved to be difficult to fully implement the role swap in the reciprocal phase. By providing feedback after the lessons and, if needed, assistance during the lessons, all teachers gradually managed to implement the main requirements of the programs.

Learning Inductive Reasoning: Near-Transfer

Phase 1: Learning effects of Program Inductive Reasoning with Visual Material. The means, standard errors, and the 95% confidence intervals of the scores on the instruments that were applied to the third grade are displayed in Table 2. The mean sum scores on the vocabulary test and the listening comprehension test did not differ significantly between the experimental conditions (1 and 2) and the control conditions (3 and 4). Hence, these test scores were not incorporated as covariates into the next analyses of the scores on the TIR. ANCOVAs showed that the experimental conditions significantly outperformed the control conditions on the TIR posttest and the follow-up test. The corrected effect size on the follow-up test was .48.

Phase 2: Learning effects of Program Inductive Reasoning with Verbal Material. The means, standard errors, and 95% confidence intervals of the scores on the instruments that were applied to the fourth grade are displayed in Table 3. As the mean scores on the Test for Technical Reading differed significantly between Experimental Condition 3 and Control Condition 4, they were included as a covariate in the next ANCOVAs. With respect to the Test for Meaning Construction, ANCOVAs showed no significant differences between the conditions on the posttest, but the conditions differed on the follow-up test. The corrected effect size on the follow-up test was .58. For the Test for Inferences Construction, the ANCOVAs showed significant differences between the conditions on the posttest and on the follow-up test. The corrected effect size on the follow-up test was .43.

Learning Inductive Reasoning in Phase 1 and Phase 2: Far-Transfer

The means, standard errors, and 95% confidence intervals of the scores on the SPM-Raven are displayed in Table 4. An ANCOVA with repeated measures revealed a significant effect of Condition ($F = 18.311$, $df = 3$, $p = .000$). A multivariate analysis of variance (MANOVA) showed that the scores on the second through the fifth administration differed significantly among the four conditions.

In Fig. 8 the mean SPM-Raven scores on all items (total scores) and subgroups of items with different Mental Demands are displayed. Pairwise com-

TABLE 2

Mean, Standard Error, and 95% Confidence Intervals of the Scores on the Test for Listening Comprehension and the Vocabulary Test and Person Parameter Estimates on the Test for Inductive Reasoning in the Experimental Conditions and the Control Conditions on the Pretests, Posttests, and Follow-Up Tests in Phase 1

	Condition	N	Mean	Standard error	95% confidence interval		F test	
					Lower bound	Upper bound		
Test for listening comprehension								
Exp. Cond.	1 + 2	150	30.46	.50	29.48	31.44	$p = .556$	
Contr. Cond.	3 + 4	109	30.01	.58	28.86	31.16		
Vocabulary test								
Exp. Cond.	1 + 2	150	25.38	.53	24.34	26.42	$p = .482$	
Contr. Cond.	3 + 4	109	24.81	.62	23.59	26.03		
Test for inductive reasoning								
Pretest								
Exp. Cond.	1 + 2	150	.473	.03	.42	.53	$p = .000$	
Contr. Cond.	3 + 4	109	.437	.03	.38	.50		
Posttest								
Exp. Cond.	1 + 2	150	.787	.03	.73	.84		
Contr. Cond.	3 + 4	109	.567	.03	.51	.63		
Follow-up test								
Exp. Cond.	1 + 2	150	.885	.03	.83	.94	$p = .000$	
Contr. Cond.	3 + 4	109	.695	.03	.63	.76		

parisons of the *total scores* of the conditions for each administration were done to check (at a 5% level) whether the doubly taught pupils in Condition 1 outperformed the singly taught pupils in the Conditions 2 and 3 and whether the singly taught pupils outperformed the pupils in control Condition 4. In the second and third administration, Condition 1 significantly outperformed Conditions 3 and 4. Condition 2 significantly outperformed control Conditions 3 and 4 in the second administration and Condition 4 in the third administration. In the fourth and fifth administrations Condition 1 significantly outperformed the other conditions. Conditions 2 and 3 did not differ significantly and they both outperformed the control Condition 4. Pairwise comparisons across all administrations revealed that doubly taught pupils in Condition 1 differed significantly from the other conditions and that singly taught pupils in Conditions 2 and 3 did not differ from one another, but that

TABLE 3

Mean, Standard Error, and 95% Confidence Intervals of the Scores on the Test for Technical Reading and Person Parameter Estimates on the Tests for Reading Comprehension in the Experimental Condition and the Control Condition on the Pretests, Posttests, and Follow-Up Tests in Phase 2

	Condition	N	Mean	Standard error	95% confidence interval		F test	
					Lower bound	Upper bound		
Test for technical reading								
Exp. Cond.	3	54	40.38	2.66	35.11	45.64	$p = .027$	
Contr. Cond.	4	75	32.56	2.67	28.08	37.05		
Reading comprehension: meaning construction								
Pretest								
Exp. Cond.	3	54	92.61	.99	90.66	94.57	$p = .131$	
Contr. Cond.	4	75	91.75	.84	90.09	93.41		
Posttest								
Exp. Cond.	3	54	95.83	.83	94.20	97.47		
Contr. Cond.	4	75	93.75	.70	92.36	95.13		
Follow-up test								
Exp. Cond.	3	54	97.28	.83	95.63	98.93	$p = .006$	
Contr. Cond.	4	75	94.31	.71	92.91	95.71		
Reading comprehension: inferences construction								
Pretest								
Exp. Cond.	3	54	97.39	1.77	93.90	100.88	$p = .009$	
Contr. Cond.	4	75	96.47	1.50	93.51	99.43		
Posttest								
Exp. Cond.	3	54	103.43	1.72	100.02	106.83		
Contr. Cond.	4	75	97.84	1.46	94.96	100.73		
Follow-up test								
Exp. Cond.	3	54	107.82	1.87	104.12	111.51	$p = .006$	
Contr. Cond.	4	75	101.16	1.58	98.03	104.30		

they differed from the pupils in control Condition 4. These results confirmed our expectations.

Compared to the first administration, the corrected effect size on the fifth administration revealed small to strong effect sizes: Conditions 1–2: $d = .41$; Conditions 1–3: $d = .81$; Conditions 1–4: $d = 1.22$; Conditions 2–4: $d = .73$; and Conditions 3–4: $d = .30$ (Cohen, 1988). The program with the visual material, which was applied to Condition 2, caused better learning effects on the SPM-Raven than the program with the verbal material (Condition 3). It was expected that irrespective of the material used, the tasks and the teaching sequence in both programs would result in pupils' development of metacognitive knowledge about the top-down and bottom-up inductive reasoning processes. The effect sizes suggested that the visual program was superior.

TABLE 4

Mean, Standard Error, and 95% Confidence Intervals of the Scores of Pupils on the First Through the Fifth Administrations of the SPM Raven in the Four Conditions

Administration	Condition	N	Mean	Standard error	95% confidence interval		F test
					Lower bound	Upper bound	
1	1	70	20.54	.88	18.82	22.27	
	2	70	18.83	.88	17.10	20.55	
	3	43	23.09	1.12	20.89	25.29	
	4	54	21.87	1.00	19.91	23.84	
2	1	70	29.67	.95	27.79	31.55	
	2	70	26.99	.95	25.11	28.86	
	3	43	27.58	1.22	25.19	29.98	
	4	54	25.59	1.09	23.45	27.73	
3	1	70	30.54	1.03	28.51	32.58	$p = .00$
	2	70	28.01	1.03	25.98	30.05	
	3	43	29.63	1.32	27.03	32.23	
	4	54	27.30	1.18	24.98	29.61	
4	1	70	35.64	.93	33.81	37.48	$p = .002$
	2	70	30.13	.93	28.29	31.97	
	3	43	31.95	1.20	29.61	34.30	
	4	54	27.82	1.06	25.72	29.91	
5	1	70	36.96	.90	35.19	38.72	$p = .000$
	2	70	32.31	.90	30.55	34.08	
	3	43	33.47	1.14	31.21	35.72	
	4	54	29.83	1.02	27.83	31.84	
							$p = .000$

Pairwise comparisons of the scores on *subsets* of SPM-Raven items of the conditions in the fifth administration showed that the scores on the relatively easy MD1 items showed a ceiling effect and, thus, they did not differ (see Fig. 8). Notice that apart from inductive reasoning Gestalt effects probably influence the solving procedure as well. The scores on the relatively difficult MD5 items showed floor effects and thus did not differ. As expected, the scores on the MD2 items, MD3 items, and MD4 items decreased, which reflected the increasing demand on inductive reasoning. The graphs of the scores on these items showed common patterns across conditions. The fifth administration showed a significant effect of condition (MD2, MD3, and MD4: $p = .000$). Pairwise comparisons of the conditions showed that doubly taught pupils in Condition 1 significantly outperformed the other conditions and that pupils in the singly taught Conditions 2 and 3 did not differ. The scores of pupils in the control Condition 4 were significantly lower than the

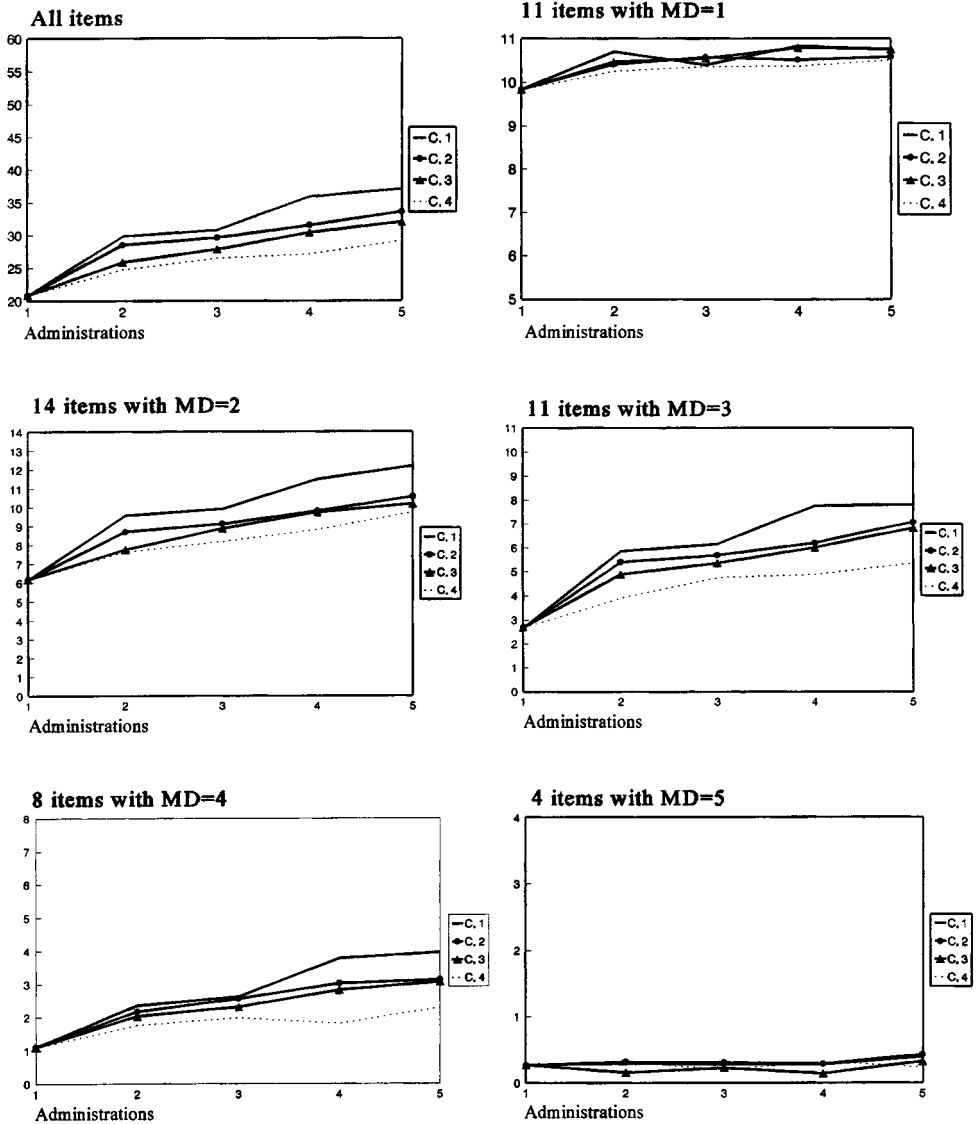


FIG. 8. Mean scores on the SPM-Raven of Condition 1 through Condition 4 (C1-C4). All items and subsets of items with different levels of mental demand (MD).

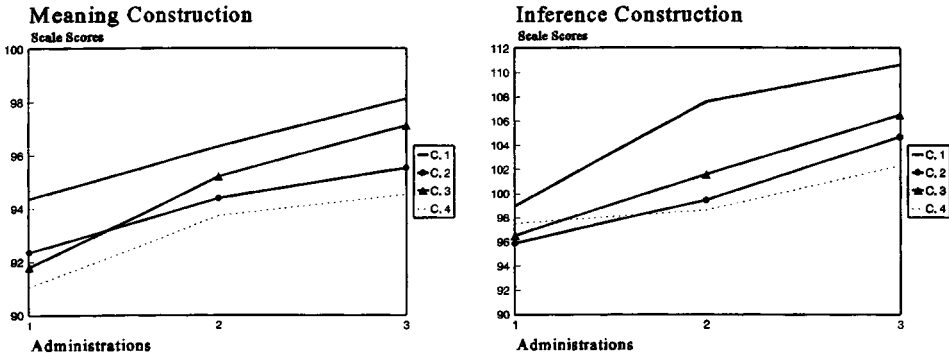


FIG. 9. Mean scores on the test for meaning construction and the test for inference construction of Condition 1 through Condition 4 (C1–C4).

scores of pupils in each of the other conditions, except for the scores with a MD of 2. In Condition 3, the scores on these MD2 items did not significantly differ from the scores in Condition 4. These results confirmed the expectations regarding the differential learning effects among the conditions and they indicated that the learning effects on the SPM-Raven were due to an increased inductive reasoning ability.

Figure 9 shows the scores on the Test for Meaning Construction and the Test for Inference Construction that were administered in Grade 4. The scores on the Test for Technical Reading were used as a covariate. ANCOVA with repeated measures of the scores on the Test for Meaning Construction revealed a significant effect of condition ($F = 6.253$, $df = 3$, $p = .000$). A MANOVA showed that the scores differed significantly among the four conditions in the three administrations (administration 1 through 3: $F = 3.589$, $df = 3$, $p = .014$; $F = 3.876$, $df = 3$, $p = .010$; and $F = 6.577$, $df = 3$, $p = .000$, respectively).

For administration 1, which came after the visual teaching phase in Grade 3, pairwise comparisons showed that the significant differences were due to significantly higher scores of doubly taught pupils in Condition 1 compared to the scores in the other three conditions. For administration 2, which came after the verbal teaching phase in Grade 4, the difference between the conditions was caused by doubly taught pupils in Condition 1 who significantly outperformed both the singly taught pupils in Condition 2 who followed the visual program and the pupils in control Condition 4. Pupils in Condition 3 who followed the verbal program did not differ significantly from Condition 1 or from Conditions 2 and 4. For administration 3, doubly taught pupils in Condition 1 again outperformed both the singly taught pupils in Condition 2 who followed the visual program and the pupils in the control Condition 4, but not the pupils in Condition 3 who followed the verbal program. The scores of these pupils differed significantly from Conditions 2 and 4. Pair-

wise comparisons across all administrations revealed that doubly taught pupils in Condition 1 differed significantly from the other conditions and that singly taught pupils who followed the verbal program in Condition 3 differed from the control Condition 4. Singly taught pupils in Condition 2 who followed the visual program did not differ significantly from control Condition 4.

An ANCOVA with repeated measurements of the scores on the Test for Inference Construction revealed a significant effect of condition ($F = 5.999$, $df = 3$, $p = .001$) (see Fig. 9). A MANOVA showed that the four conditions did not differ significantly in the first administration ($F = .953$, $df = 3$, $p = .415$), but that they differed in the second and the third administrations ($F = 9.200$, $df = 3$, $p = .000$; and $F = 5.476$, $df = 3$, $p = .001$, respectively).

For administration 1, which came after the visual teaching phase in Grade 3, pairwise comparisons showed that none of the conditions differed from one another. After the verbal teaching phase in Grade 4, the significant differences in the second administration were due to doubly taught pupils in Condition 1 who outperformed pupils in the other three conditions. Also in the third administration the significant difference was caused by Condition 1, which outperformed the other conditions. Furthermore, in this administration the singly taught pupils in Condition 3 who followed the verbal program outperformed the pupils in the control Condition 4. Pairwise comparisons across all administrations revealed that doubly taught pupils in Condition 1 differed significantly from pupils in the other conditions.

It can be concluded that a single application of the program with *visual* material did not result into a far-transfer to reading comprehension and that the single application of the program with *verbal* tasks did result in a near-transfer to reading comprehension. Successive application of both programs was superior to a single application of the program with verbal tasks. The development of decontextualized inductive reasoning top-down schemes and bottom-up steps during the verbal program in Grade 4 was probably facilitated by the visual program that was applied to Grade 3.

CONCLUSIONS AND DISCUSSION

Our research aimed at smoothing the transition from developmentally oriented education in the preparatory phase to curriculum-oriented education from the third grade onward. Therefore, we extended the emphasis on inductive reasoning in the preparatory phase to the third and the fourth grades by implementing two teaching programs that were applied by the teachers to their own classrooms.

All teachers needed support to shift their attention from the reasoning product (correct or incorrect answer) to the reasoning process (how was the answer found) in order to enhance the pupils' metacognitive development of the inductive top-down and bottom-up reasoning processes. Generally, the teachers were able to apply the programs as intended. However, the role

swap between teacher and pupil that should have been accomplished in the third and last instruction phases of both programs was difficult to implement. One explanation is that the pupils were not competent enough in Dutch to verbally reflect on their thinking processes. Another explanation is that teachers and pupils were not familiar with this kind of instruction because exchanging roles is not an integrated part of daily educational practice in the Netherlands.

The results showed that the taught pupils significantly outperformed pupils from the control condition on near-transfer tests. Doubly taught pupils (i.e., in the third and the fourth grades) significantly outperformed singly taught pupils on the far-transfer test, and singly taught pupils significantly outperformed pupils from the control condition. Pupils who followed the program with visual material showed greater learning effects than pupils who were taught using the program with verbal material. As both programs had the same objectives, probably the common visual mode of the third-grade program and the far-transfer test induced better learning results. To precisely focus on the inductive reasoning processes, the far-transfer test was divided into groups of tasks that differed in their demand on inductive reasoning. It was concluded that the learning effects of both programs indicated that pupils' better performance was due to their increased ability of inductive reasoning. This was corroborated by the facilitating influence of the visual program on the verbal program.

Contrary to Hager and Hasselhorn (1998) and other research on the original Program Inductive Reasoning 1 (Klauer, 1989), in our study equal importance was attached to the operationalization of inductive reasoning tasks, on the one hand, and the types and teaching sequence of the inductive reasoning processes, on the other hand. Our study corroborates with the view that learning is not considered a passive, knowledge-consuming, and externally directed process, but an active, constructive, and self-directed process (Bednar, Cunningham, Duffy, & Perry, 1991).

In this respect we follow Vermunt and Verloop (1999) in making a plea for conducting research that aims at the development of theories in which teaching, learning, and knowledge are intertwined. They integrated various teaching activities and learning strategies and showed possible interactions in terms of congruence and friction. Whether there is congruence or friction depends on the extent the teacher's and pupils' regulation of the learning process complement one another. In this study, the sequence of three instruction phases of modeling, practicing, reflection, and, finally, reciprocal teaching using role swapping, corresponds to three pupil-teacher interactions characterized as congruent. Also Elshout-Mohr, Van Hout-Wolters, and Broekkamp (1999) emphasized the usefulness of implementing learning phases, which aim at gradually transferring the responsibility from the teacher to the pupils. Furthermore, Elshout et al. (1999) suggested that training for transfer of knowledge should be prolonged and take place under vary-

ing conditions using different materials. Our study complies with these criteria as the design permitted comparison of different training materials (i.e., visual and verbal material) and duration (doubly and singly taught pupils).

Nevertheless, some critical remarks concerning this study and suggestions for future research can be made. Although the integration into two programs of theory-based operationalizations of tasks and processes and the developmentally based teaching sequence and choice of material seemed successful, it is not possible to conclude unequivocally which particular aspect of the programs was helpful. For instance, our results support the hypothesis that the underlying mechanism that produced the learning effects was metacognitive, but we did not prove it beyond doubt because we did not compare inductive reasoning programs with and without metacognitive elements. However, we expected that a transfer between two different contexts (i.e., pictures and texts) would be possible only if pupils developed conscious knowledge of the inductive reasoning top-down schemes and the bottom-up steps. This is supported by other research on inductive reasoning that showed that transfer between tasks of the visual and verbal modes is difficult to accomplish when the focus on metacognition in the instruction is less salient (see, for example, Resing & Roth-van der Werf, 1997).

In future research new observation tools and instruments could be designed for collecting more information about the dialogs that underlie the learning effects. Also more detailed analyses of the strategies that pupils use (e.g., thinking-aloud protocols) while solving inductive reasoning tasks could provide more information about the development of the decontextualization of the reasoning schemes and steps. Furthermore, instruments could be added for measuring the transfer levels after each of the three teaching phases.

A challenging question is how the role swap in the third instruction phase could be brought about more easily. We assume that an integration of the exploratory talk rules (Wegerif, Mercer, & Dawes, 1999) in the programs could improve the role swap. Wegerif et al. (1999) taught 8- and 9-year-old pupils to use exploratory talk and found learning effects on the SPM-Raven. Explorative talk is characterized by a critical and constructive engagement in which statements and suggestions are offered for joint consideration. Mercer (1996) stated that "in exploratory talk knowledge is made more publicly accountable and *reasoning is more visible*" (p. 369). Furthermore, self-regulation, which is reflected in a proper role swap, depends on cognitive variables and on affective variables (Boekaerts, 2000). Including affective variables should therefore broaden the cognitive focus of the model in Fig. 3.

How can future research on inductive reasoning contribute to the development of theories that connect development, teaching, and learning? The Cognition and Technology Group at Vanderbilt (1996) designed a framework for research. It suggested explicitly connecting scientific research to educational practice by increasing-the sample size from a few small training groups, through teaching individual classrooms and connected sets of classrooms,

to teaching whole schools. Furthermore, it suggested using the model of constructive learning, in which learning is considered an active, constructive, and self-directed process. Our study was carried out in connected classrooms and fit into this model of learning. Our future research could be directed at extending the sample size such that the school as a whole would be covered. For example, the programs for teaching inductive reasoning could be extended to the second and the fifth grades of primary education. For the second grade, it will be sufficient to reorganize the learning materials already available. In the fifth grade, it would be possible to embed the programs' principles into, for example, the domain of writing. After analyzing and constructing the meaning of texts in the fourth grade, it would be a logical next step to teach pupils to actively produce their own texts. This kind of research would reveal whether the integration of inductive reasoning tasks, processes, and teaching is suitable for pupils younger and older than those who participated in this study. By using analogous exercises in various contexts the inductive reasoning development will be continuously supported during primary education.

We hypothesize that there might be transfer effects on science education as well. Scientific reasoning involves generalizing rules from measurement results or everyday experiences, hypothesis formation and hypothesis testing, grouping of facts using two-dimensional or multidimensional tables, and concept formation and concept development. These are activities typical for inductive reasoning (Csapó, 1999). Science education starts, at least in the Netherlands, by the end of primary education and gets increasingly more important in secondary education. It would be interesting to investigate the presumed beneficial effects on science in primary and secondary education of pupils who have learned to consciously reflect on their inductive reasoning in primary school. We suggest that by focusing on the development of inductive reasoning it will be possible to define a line of thinking on how we best could design our learning materials and instruct our pupils during their whole course of primary and secondary education such that they become efficient and critical learners.

REFERENCES

- Adey, P. S. (1999). Thinking science: Science as a gateway to general thinking ability. In J. H. M. Hamers, J. E. H. van Luit, & B. Csapó (Eds.), *Teaching and learning thinking skills* (pp. 63–81). Lisse: Swets & Zeitlinger.
- Baron, J. B., & Sternberg, R. J. (1987). *Teaching thinking skills: Theory and practice*. New York: Freeman.
- Bednar, A. K., Cunningham, D., Duffy, T., & Perry, J. D. (1991). Theory into practice: How do we link. In G. J. Anglin (Ed.), *Instructional technology: Past, present, and future* (pp. 88–101). Englewood, CO: Libraries Unlimited.
- Bereiter, C., & Scardamalia, M. (1979). Pascual-Leone's M construct as a link between cognitive-developmental and psychometric concepts of intelligence. *Intelligence*, *3*, 41–63.

- Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and pupils. *Learning and Instruction*, 7(2), 161–186.
- Boekaerts, M., & Niemivirta, M. (2000). Self-regulated learning: Finding a balance between learning goals and ego-protective goals. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 417–450). San Diego: Academic Press.
- Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology: Cognitive development* (Vol. 3, pp. 77–166). New York: Wiley.
- 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, UK: Cambridge Univ. Press.
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97(3), 404–431.
- Case, R. (1974). Structures and strictures: Some functional limitations on the course of cognitive growth. *Cognitive Psychology*, 6, 544–573.
- Ceci, S. J. (1996). On intelligence. In *A bioecological treatise on intellectual development*. Cambridge, MA: Harvard Univ. Press.
- CITO (1995). *Luistertoets: Handleiding [Listening Comprehension Test: Manual]*. Arnhem: CITO.
- Cognition and Technology Group at Vanderbilt. (1996). Looking at technology in context: A framework for understanding technology and education research. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 807–840). New York: Macmillan.
- Csapó, B. (1999). Improving thinking through the content of teaching. In J. H. M. Hamers, J. E. H. Van Luit, & B. Csapó (Eds.), *Teaching and learning thinking skills* (pp. 11–36). Lisse: Swets & Zeitlinger.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- De Corte, E., Verschaffel, L., & Op 't Eynde, P. (2000). Self-regulation: A characteristic and a goal of mathematics education. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 687–726). San Diego: Academic Press.
- De Koning, E. (2000). *Project Inductief redeneren in begrijpend lezen: Verslag voor betrokken basisscholen [Project on inductive reasoning in reading comprehension: Report for schools that participated]*. Utrecht: Utrecht Univ. Press.
- De Koning, E., & Hamers, J. H. M. (1995). *Programma Inductief Redeneren 1 [Programme Inductive Reasoning I]*. Utrecht: Utrecht Univ. ISOR.
- De Koning, E., & Hamers, J. H. M. (1996). *Programma Inductief Redeneren en Begrijpend Lezen: Experimentele versie [Programme Inductive Reasoning in Reading Comprehension: Experimental version]*. Utrecht: Utrecht Univ. ISOR.
- De Koning, E., & Hamers, J. H. M. (1999). Teaching inductive reasoning: Theoretical background and educational implications. In J. H. M. Hamers, J. E. H. van Luit, & B. Csapó (Eds.), *Teaching and learning thinking skills* (pp. 156–188). Lisse: Swets & Zeitlinger.
- De Koning, E., Sijtsma, K., & Hamers, J. H. M. (in press-a). Comparison of four IRT models when analyzing two tests for inductive reasoning. *Applied Psychological Measurements*.
- De Koning, E., Sijtsma, K., & Hamers, J. H. M. (in press-b). Construction and validation of a test for inductive reasoning. *European Journal of Psychological Assessment*.
- Demery, J. (2000). The relationship between teacher's perceptions of school climate, racial composition, socioeconomic status, and student achievement in reading and mathematics.

- Dissertation Abstracts International Section A: Humanities and Social Sciences*, **61**(3A), 957.
- Elshout-Mohr, M., Van Hout-Wolters, B., & Broekkamp, H. (1999). Mapping situations in classroom and research: Eight types of instructional-learning episodes. *Learning and Instruction*, **9**, 57–75.
- Ennis, R. (1990). The extent to which critical thinking is subject-specific: Further clarification. *Educational Researcher*, **19**(4), 13–16.
- Goswami, U. (1998). *Cognition in children*. Hove, UK: Psychology Press.
- Hager, W., & Hasselhorn, M. (1998). The effectiveness of the cognitive training for children from differential perspective: A meta-evaluation. *Learning and Instruction*, **8**(5), 411–438.
- Halford, G. S. (1993). *Children's understanding: The development of mental models*. 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.
- Hessels, M. G. P., & Hamers, J. H. M. (1993). A learning potential test for ethnic minorities. In Hamers, J. H. M., Sijtsma, K., & A. J. J. M. Ruijsseenaars. *Learning potential assessment: Theoretical, methodological and practical issues* (pp. 285–313). Amsterdam: Swets & Zeitlinger.
- 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 Univ. Press.
- Janssen, J., Van der Schoot, F., Hemker, B., & Verhelst, N. (1999). *Balans van het rekenwiskundeonderwijs aan het einde van de basisschool 3* [Review of math education results at the end of primary education 3]. Arnhem: CITO.
- Klauer, K. J. (1989). *Denktraining für Kinder I* [Cognitive training for children I]. Göttingen: Hogrefe.
- Leseman, P. P. M., & De Jong, P. F. (2001). How important is home literacy for acquiring literacy in school? In L. Verhoeven & C. Snow (Eds.), *Literacy and motivation: Reading engagement in individuals and groups* (pp. 71–93). Mahwah, NJ: Erlbaum.
- Ma, X., & Klinger, D. A. (2000). Hierarchical linear modelling of student and school effects on academic achievement. *Canadian Journal of Education*, **25**(1), 41–55.
- McShane, J. (1991). *Cognitive development: An information processing approach*. Oxford, UK: Basil Blackwell.
- Mercer, N. (1996). The quality of talk in children's collaborative activity in the classroom. *Learning and Instruction*, **6**(4), 359–375.
- Nisbett, R. E. (1993). *Rules for reasoning*. Hillsdale, NJ: Erlbaum.
- Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, **1**, 117–175.
- Piaget, J. (1970). Piaget's theory. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (Vol. 1). New York: Wiley.
- Raven, J. C. (1958). *Standard Progressive Matrices*. London: Lewis.
- Resing, C. M., & Roth van der Werf, A. (1997). *Training inductive reasoning skills in learning disabled children*. Paper Presented at the EARLI Conference, Athens, Greece.
- Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist*, **24**, 113–142.

- Scardamalia, M., & Bereiter, C. (1986). Research on written composition. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 778–803). New York: Macmillan.
- 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. 47–103). Hillsdale, NJ: Erlbaum
- Sternberg, R. J. (1998). When will the milk spoil?: Everyday induction in human intelligence. *Intelligence*, **25**(3), 185–203.
- Sternberg, R. J., & Gardner, M. K. (1983). Unities in inductive reasoning. *Journal of Experimental Psychology: General*, **112**(1), 80–116.
- Stipek, D. J. (2001). Pathways to constructive lives: The importance of early school success. In A. C. Bohart & D. J. Stipek (Eds.), *Constructive and destructive behavior: Implications for family, school, and society* (pp. 291–315). Washington, DC: American Psychological Association.
- Verhoeven, L. (1992). *Toetsen voor Begrijpend Lezen: Handleiding [Tests for reading comprehension: Manual]*. Arnhem: CITO.
- Verhoeven, L. (1995). *Drie-Minuten-Toets: Handleiding [Three-Minute-Test: Manual]*. Arnhem: CITO.
- Verhoeven, L. (1996). *Woordenschattoets I: Handleiding [Vocabulary Test I: Manual]*. Arnhem: CITO.
- Vermunt, J. D., & Verloop, N. (1999). Congruence and friction between learning and teaching. *Learning and Instruction*, **9**, 257–280.
- Wegerif, R., Mercer, N., & Dawes, L. (1999). From social interaction to individual reasoning: An empirical investigation of a possible socio-cultural model of cognitive development. *Learning and Instruction*, **9**, 493–516.
- 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, UK: Cambridge Univ. Press.
- Wijnstra, J. M. (1997). *Specialistisch Bulletin: Verwijzing van leerlingen uit etnische en culturele minderheidsgroepen naar het speciaal onderwijs [Referral of pupils from ethnic and cultural minorities to special education]* (Vol. 42). Arnhem: CITO.
- Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension: An experimental investigation*. Cambridge, UK: Cambridge Univ. Press.

Received January 10, 2002