Learning to Solve Mathematical Application Problems: A Design Experiment With Fifth Graders

Lieven Verschaffel, Erik De Corte, Sabien Lasure, Griet Van Vaerenbergh, Hedwig Bogaerts, and Elie Ratinckx

Center for Instructional Psychology and Technology
University of Leuven, Belgium

Recent research has shown that many upper elementary school children do not master the skill of solving mathematical application problems. In this design experiment, a learning environment for teaching and learning how to model and solve mathematical application problems was developed and tested in 4 classes of 5th graders. Pupils were taught a series of heuristics embedded in an overall metacognitive strategy for solving mathematical application problems. Meanwhile, pupils of 7 control classes followed regular mathematics classes. The implementation and effectiveness of the experimental learning environment were tested in a study with a pretest–posttest–retention test design with an experimental and a control group. The results indicate that the intervention had a positive effect on different aspects of pupils’ mathematical modeling and problem-solving abilities.

In recent reform documents, such as the Curriculum and Evaluation Standards for School Mathematics (National Council of Teachers of Mathematics, 1989) in the United States, Mathematics Counts (Cockcroft, 1982) in the United Kingdom, A National Statement on Mathematics for Australian Schools (Australian Education Council, 1990), the Dutch Proeve van een Nationaal Programma voor het Reken/Wiskundeonderwijs op de Basisschool (Treffers, De Moor, & Feijs, 1989), and the Eindtermen en ontwikkelingsdoelen basisonderwijs (Ministerie van de Vlaamse Gemeenschap, 1997) in Flanders, there is a strong emphasis on the ac-
Acquisition of mathematical problem-solving and reasoning skills and attitudes, and the ability to apply these skills in real-life situations as one of the major objectives of mathematics education at the elementary school level.

There are, however, reasons to be skeptical about the attainability of these ambitious general goals for the majority of pupils at the end of elementary school. Indeed, a large number of research findings show that, after several years of current mathematical education, many upper elementary school children do not, or at least insufficiently, master the different aptitudes required to approach mathematical application problems in an efficient and successful way (De Corte, Greer, & Verschaffel, 1996; Lester, Garofalo, & Kroll, 1989; Schoenfeld, 1992; Verschaffel, in press).

First, there are deficiencies in (upper) elementary school children that 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 a problem to be solved (e.g., mathematical symbols, concepts, formulas, algorithms, specific problem-solving schemes). As an illustration of how insufficiencies in the domain-specific knowledge base can negatively interfere with pupils' problem-solving ability, we refer to the “multiplication makes bigger, division makes smaller” misconception, which has been observed in many upper elementary school children (but also in older students and adults) in a diversity of countries and typically leads to inversed-operation errors on particular types of multiplication and division word problems involving decimal numbers smaller than 1 (De Corte, Verschaffel, & Van Coillie, 1988; Greer, 1992).

Besides insufficiencies in the domain-specific knowledge base, many upper elementary school children suffer from shortcomings in the heuristic, metacognitive, and affective aspects of mathematical competence. When confronted with unfamiliar complex problem situations, many pupils do not spontaneously apply valuable heuristic strategies like making a sketch or drawing of the problem situation, decomposing the problem into parts, or guessing and checking (De Bock, Verschaffel, & Janssens, 1998; De Corte & Somers, 1982; Lester et al., 1989; Schoenfeld, 1992; Van Essen, 1991). With respect to metacognition, several studies revealed that in a vast majority of pupils' solution attempts, self-regulatory activities such as analyzing the problem, monitoring the solution process, and evaluating its outcome are completely absent. A typical approach used by many pupils can be summarized as follows: The student glimpses the problem, quickly deciding what calculations to perform with the numbers given in the problem statement, and then proceeds with these calculations without considering any alternatives even if no progress is made at all (De Corte et al., 1996; Greer, 1992; Lester et al., 1989; Schoenfeld, 1985, 1992; Verschaffel, in press).

Finally, with respect to the affective aspect, research has revealed that many pupils—even the better ones—hold inadequate domain-related beliefs about and atti-
tudes toward mathematics and mathematics learning, teaching, and problem solving. These beliefs exert a strong negative 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 solution attempts, and on the way they evaluate their success or failure to solve problems (De Corte et al., 1996; Lester et al., 1989; McLeod, 1992; Schoenfeld, 1988, 1992; Vermeer, 1997; Verschaffel, in press). Typical examples of such counterproductive beliefs and attitudes are: (a) mathematics problems have only one right answer, (b) there is only one correct way to solve any mathematical problem, (c) ordinary students cannot solve unfamiliar mathematics problems by themselves, (d) being able to solve a word problem is a mere question of luck, and (e) there is a gap between the mathematics learned in school and the mathematics required in the real world.

These insufficiencies in pupils’ abilities to solve context-based mathematical application problems are related to the following characteristics of the current practice and culture of teaching and learning world problem solving:

1. The nature of the problems used in the lessons in word-problem solving. In the current practice of school mathematics, pupils are mostly confronted with so-called standard problems, in which the relation between the problem context and the required calculation(s) is straightforward. Nonstandard problems inviting or even forcing pupils to reflect on the complex and problematic relation between mathematics and reality and on the difficulties that can be raised in using mathematics to solve real-life situations are very rare (Burkhardt, 1994; Greer, 1997; Greer & Verschaffel, 1997; Lave, 1992; Lester et al., 1989; Nesher, 1980; Reusser, 1988; Treffers et al., 1989; Verschaffel & De Corte, 1997b).

2. The way in which the teacher and pupils are dealing with these problems in the word-problem-solving lessons. In the current practice of teaching and learning mathematical applications, pupils mainly solve word problems individually by means of fixed, standard problem-solving schemes and procedures that are explained and demonstrated by the teacher. Extensive and systematic use of powerful and interactive instructional techniques, aimed at the development of valuable heuristic and metacognitive skills (e.g., modeling, scaffolding, coaching, articulation, reflection, and exploration; see Collins, Brown, & Newman, 1989) and appropriate beliefs and affects, is lacking (Burkhardt, 1994; De Corte et al., 1996; Greer, 1997; Schoenfeld, 1992).

3. The classroom culture. Researchers have documented how several subtle, invisible aspects of the daily mathematics classroom rituals and practices can also contribute to unwanted mathematical behavior and learning outcomes, such as the use of stereotyped and mindless strategies for coping with word problems or the construction of erroneous conceptions and beliefs about what mathematics and mathematics learning and teaching is all about (Cobb, Yackel, & Wood, 1992; Da-
According to these researchers, pupils' learning of mathematics occurs to a large extent through being immersed and socialized in a particular classroom culture, and what pupils learn as a result of their daily involvement with the rituals and cultural practices does not always correspond to the explicitly intended learning goals.

In recent years, several scholars began to design and evaluate instructional environments aimed at the development of mathematical problem-solving skills (Becker & Shimada, 1997; Cognition and Technology Group at Vanderbilt, 1997; De Corte & Somers, 1982; English, 1998; Gravemeijer, 1994; Lampert, 1986, 1990; Lee, 1982; Lesh & Lamon, 1992; Lester et al., 1989; Nelissen, 1987; Resnick & Nelson-LeGall, 1987; Schoenfeld, 1985, 1992; Siemon, 1992; Silver, Smith, Lane, Salmon-Cox, & Stein, 1990; Van Lieshout & Jaspers, 1990; Verschaffel & De Corte, 1997a; Yackel & Cobb, 1996). Some of these studies—especially the older ones—departed from a purely cognitive-rationalist perspective (Greeno, Collins, & Resnick, 1996), whereby a detailed cognitive model of the concepts and cognitive abilities underlying competent problem solving in mathematics is specified and used to guide the development of teacher and student activities that will lead to the acquisition and use of these conceptual and cognitive structures. In more recent design experiments, this cognitive-rationalist perspective was replaced by or enriched with a situative orientation, a social-constructivist orientation, or both. The situative perspective focuses on the way knowledge is distributed in the world among individuals; the tools, artifacts, and books that they use; and the communities and practices in which they participate. In accordance with this view on knowing as practices of communities and the abilities of individuals to participate in those practices, learning is conceived of as the strengthening of those practices and participatory abilities, like in the "cognitive apprenticeship" approach (Collins et al., 1989). Studies taking the social-constructivist perspective bring to the fore the processes involved in the negotiation of the "taken-as-shared" understandings that are necessary for establishing intersubjectivity among the pupils and the teacher with regard to culturally held ideas about mathematics and mathematics learning and teaching, as exemplified in the design experiments of Yackel and Cobb (1996) and Gravemeijer (1994). We briefly discuss three intervention studies that were most influential for this design experiment. Whereas the first one can be considered as representative for the cognitive-rationalist paradigm, the second and the third studies also embody major features of the situative and the social-constructivist view of learning and teaching.

In a teaching experiment with seventh graders, Lester et al. (1989) studied the effects on mathematical problem solving of an instructional program in which pupils (a) exercised the use of some valuable heuristics (strategy training), (b) learned to articulate and reflect on their problem-solving strategies (awareness
training), and (c) were taught to monitor their problem-solving activities (self-regulation training). The instructional program consisted of a set of appropriate problems and tasks and a series of lesson plans describing the different roles and activities for the teacher. The instruction involved a combination of small-group problem solving, whole-class discussions, and individual assignments. The program was administered by one of the investigators in a regular level and in an advanced level seventh-grade class for 15 hr spread over 12 weeks. A collective test involving a set of five nonroutine application problems was administered before and after the intervention. Both the regular class and the advanced class realized a considerable gain in the total score from pretest to posttest, albeit the progress was not as large as expected. The results of clinical interviews revealed no substantial differences between pupils' regulatory activities before and after instruction.

Starting from the results and conclusions of a series of studies documenting the absence of real-world knowledge and sense making during pupils' modeling and interpreting of school arithmetic word problems (e.g., Verschaffel, De Corte, & Lasure, 1994; for an overview, see Greer & Verschaffel, 1997), Verschaffel and De Corte (1997a) set up an exploratory teaching experiment to test the possibility of developing upper elementary school children's disposition toward more realistic mathematical modeling. Whereas the regular mathematics curriculum was taught in the two control classes, the pupils of the experimental class followed the instructional program consisting of five teaching-learning units of 2½ hr each. Each teaching–learning unit focused on one prototypical, problematic topic of realistic modeling, such as interpreting the outcome of a division problem involving a remainder or learning to discriminate among cases where solutions based on direct proportional reasoning are appropriate or inappropriate. A comparison of the results of the experimental class and the two control classes on pretest, posttest, and retention test—all involving a set of 10 new word problems with problematic modeling assumptions—revealed that the experimental program had a considerable positive effect on pupils' disposition toward realistic modeling and interpretation of arithmetic word problems. Nevertheless, the overall percentages of realistic responses of the experimental class on both the posttest and the retention test were still relatively low.

Van Haneghan et al. (1992) used video technology to anchor rich, authentic, and complex problem situations in realistic contexts providing ample opportunities for problem posing, modeling, self-regulation, and interpretation. Participants in this controlled teaching experiment were a fifth-grade class of above-average students. Both the experimental group and the control group were given three teaching sessions. Whereas the experimental group worked out the authentic problems under guidance of one of the researchers, the control students were instructed in solving traditional word problems. Before and after instruction, two tests were administered: a traditional word-problem-solving test and a test in which the prob-
lems were presented on videotape. No differences were found between both groups on the traditional word-problem-solving test, but on the videotape-based test the experimental group significantly outperformed the control group.

The positive, although limited, learning effects obtained in these three teaching experiments provide promising initial support for the possibility of enhancing pupils’ mathematical reasoning and problem-solving skills by means of appropriate instruction. However, these studies suffer also from a number of shortcomings. First, each study focused on particular aspects of the instructional environment and in doing so neglected other relevant components. In line with its cognitive-rationalist roots, the program of Lester et al. (1989) aimed at the acquisition of a number of heuristics embedded in an overall metacognitive strategy. The focus of Verschaffel and De Corte’s (1997a) program was on changing pupils’ counterproductive approach toward superficial and nonrealistic mathematical modeling, and Van Haneghan et al. (1992) were mainly interested in exploring the usefulness of videotapes for designing complex and realistic problem situations that afford pupils the opportunity to develop and apply a deep understanding of mathematical concepts and skills. Second, in each of the aforementioned studies, the instructional program was implemented by one of the researchers instead of the regular classroom teacher. This raises the question of the ecological validity of the results. Third, the evaluation of the effect of the experimental programs was mainly based on pupils’ results on a written test administered before and after the intervention. Although more process-oriented data were collected as well, these additional data were not very convincing because they were obtained and analyzed in a less systematic way. Finally, the aforementioned studies did not systematically address the question of whether the instructional interventions not only led to positive effects on the problem-solving skills and processes of pupils of high and medium ability but also on those of low ability.

In this article, we report on a design experiment in which a learning environment for solving mathematical application problems for upper elementary school pupils was developed and afterwards implemented and tested in 4 fifth-grade classes. The theoretical underpinnings of the study were derived from traditional cognitive-rationalist analyses of mathematical problem solving (Lester et al., 1989; Polya, 1957; Schoenfeld, 1985), the cognitive apprenticeship approach to mathematical problem solving (Collins et al., 1989; De Corte et al., 1996; Schoenfeld, 1992), and the socioconstructivist perspective, wherein the development of an individual’s mathematical reasoning and sense-making processes is seen as “unseparatedly interwoven with their participation in the interactive constitution of ‘taken-as-shared’ (socio-)mathematical meanings and norms” (Yackel & Cobb, 1996, p. 460). As such, the theoretical basis for this study was in line with the synthetic view on knowing, learning, and teaching recently advocated by Greeno and Middle School Mathematics Through Applications Project Group (1998). At a more specific level, we relied heavily on the results and experiences of
the three aforementioned teaching experiments (Lester et al., 1989; Van Haneghan et al., 1992; Verschaffel & De Corte, 1997a). An attempt was made to avoid the weaknesses and to combine the strong elements from each of these three studies, with a view to increase the instructional power of the learning environment. Moreover, in line with recent insights concerning the interplay between research on learning and instruction and educational practice, the design experiment was developed and implemented in partnership with the participating practitioners (De Corte, 1998).

**DESCRIPTION OF THE LEARNING ENVIRONMENT**

**Aim of the Learning Environment**

Generally speaking, the major goal of the learning environment was to transform pupils into more active, more strategic, and more motivated solvers of mathematical application problems. Although we did not exclude that the learning environment may have yielded transfer effects beyond the domain of mathematical application problems (e.g., purely symbolically presented puzzles with numbers, logical reasoning problems that do not involve numbers), such a transfer was not intentionally and systematically pursued in the program.

More specifically, the first aim of the learning environment was pupils’ acquisition of an overall metacognitive strategy for solving mathematical application problems. This strategy consisted of five stages and involved a set of eight heuristic strategies that were especially valuable in the first two stages of this strategy (see Table 1). Acquiring this metacognitive strategy involves (a) becoming aware of the different phases involved in a competent problem-solving process (awareness training), (b) developing an ability to monitor and evaluate one’s actions during the different phases of this problem-solving process (self-regulation training), and (c) gaining mastery of the eight heuristic strategies that can be successfully applied during the first two stages of the solution process (heuristic strategy training). The names and delineations of the distinct stages of the competent solution model as well as the selection of the eight heuristics and their matching with a particular stage of the problem-solving model, were based on an extensive review of the available scientific literature on this topic (e.g., De Corte et al., 1996; Lester et al., 1989; Polya, 1957; Schoenfeld, 1992; Verschaffel, in press). However, because this literature is inconclusive with respect to each of these issues, some decisions had to be made on rather intuitive or pragmatic grounds. For instance, the list of heuristic strategies that we obtained at the end of our review of the problem-solving literature was much longer than the set of eight heuristics that was finally addressed in the learning environment. Because this review did not yield an unequivocal, research-based answer to the question of which heuristics were the most effective for upper elementary school children, a meeting of expert mathe-
TABLE 1
The Competent Problem-Solving Model Underlying the Learning Environment

<table>
<thead>
<tr>
<th>Step 1: Build a mental representation of the problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heuristics:</td>
</tr>
<tr>
<td>Draw a picture</td>
</tr>
<tr>
<td>Make a list, a scheme or a table</td>
</tr>
<tr>
<td>Distinguish relevant from irrelevant data</td>
</tr>
<tr>
<td>Use your real-world knowledge</td>
</tr>
<tr>
<td>Step 2: Decide how to solve the problem</td>
</tr>
<tr>
<td>Heuristics:</td>
</tr>
<tr>
<td>Make a flowchart</td>
</tr>
<tr>
<td>Guess and check</td>
</tr>
<tr>
<td>Look for a pattern</td>
</tr>
<tr>
<td>Simplify the numbers</td>
</tr>
<tr>
<td>Step 3: Execute the necessary calculations</td>
</tr>
<tr>
<td>Step 4: Interpret the outcome and formulate an answer</td>
</tr>
<tr>
<td>Step 5: Evaluate the solution</td>
</tr>
</tbody>
</table>

Mathematics education practitioners was organized in the early stage of the project to determine, among other things, the appropriateness of the proposed heuristics selection.

With the strong relation between pupils’ use of (meta)cognitive strategies when solving mathematical application problems and their domain-related beliefs and attitudes (see Lester et al., 1989; Schoenfeld, 1992) taken into account, a second aim of the learning environment was to make pupils explicitly aware of their erroneous beliefs about and their negative attitudes toward mathematical problem solving. An attempt could then be made to replace these beliefs by more appropriate and positive ones.

Major Pillars of the Learning Environment

Starting from the aforementioned goals and from our critical analysis of the current literature, we designed, in collaboration with the teachers of the experimental classes, a learning environment based on three pillars.

The first pillar of the learning environment was the use of a varied set of carefully designed realistic, complex, and open problems. The term *realistic* means that the problems referred to contexts taken from the experiential worlds of fifth-grade pupils and that the questions raised were also meaningful for children of that age (Gravemeijer, 1994; Lesh & Lamon, 1992). *Complex* refers to the requirement that the problems must have invited and even “forced” pupils to apply the intended heuristic and metacognitive strategies. Another feature of the problems was their *openness*, implying that they could be defined, modeled, solved, and answered appropriately in different ways, encouraging reflection and discus-
sion among pupils (Becker & Shimada, 1997). Finally, there was variation in the presentation of the problems. Some problems were presented in a purely textual format, others in the form of a story told by the teacher, a newspaper article, a brochure, a comic strip, a table, or a combination of two or more of these presentational formats. Actually, some problems were borrowed or adapted from existent materials (e.g., Lester et al., 1989; Van Haneghan et al., 1992; Verschaffel & De Corte, 1997a), but others were created by the research team. A provisional problem set was made by the research team and evaluated, adapted, and complemented during the meetings with the expert practitioners.

The second pillar of the learning environment was the use of a varied set of powerful instructional techniques. Most of the lessons followed an instructional model similar to the one applied in Verschaffel and De Corte’s (1997a) exploratory teaching experiment discussed previously. This model consists of the following sequence of classroom activities: (a) a short, whole-class introduction; (b) two group assignments solved in fixed heterogeneous groups of three or four pupils, each of which is followed by a whole-class discussion; and (c) an individual assignment, which is also followed by a final whole-class discussion. During all of these activities, the teacher’s role is to encourage and scaffold pupils to engage in and to reflect on the kinds of cognitive and metacognitive activities involved in the model of competent mathematical problem solving. These encouragements and scaffolds are gradually withdrawn as pupils become more competent in and aware of their problem-solving activity and thus take greater responsibility for their own learning and problem solving (cf. the “cognitive apprenticeship” model of Collins et al., 1989, and the social-constructivist approach developed by Yackel & Cobb, 1996).

The third pillar of the learning environment was the creation of a classroom culture aimed at the establishment of new social and sociomathematical norms about teaching and learning mathematical problem solving. Some typical characteristics of this new classroom culture follow:

1. Stimulating pupils both during the small-group activities and the whole-class discussions to articulate and to reflect on their personal beliefs, misconceptions, problem-solving strategies, and feelings with respect to mathematical application problems.
2. Debating on new norms about what counts as a good mathematical problem (e.g., “Many problems can be interpreted and solved in many different appropriate ways”), a good response (e.g., “Sometimes it is better to respond with a rough estimation than with a precise numerical answer”), or a good solution procedure (e.g., “In case of uncertainty even an expert problem solver may count on his fingers”).
3. (Re)defining the role of the teacher and the pupils in the mathematics classroom (e.g., “Don’t expect the teacher to decide autonomously which of the generated solutions is the optimal one; this decision is taken by the whole class as a com-
munity of practice after an evaluation of the pros and cons of all distinct candidate solutions").

The elaboration of this third pillar was based on the work of socioconstructivists such as Cobb et al. (1992; see also Yackel & Cobb, 1996), Gravemeijer (1994), and Lampert (1986, 1990). As these studies have shown, the teacher plays a central role in establishing these norms, not only implicitly by creating a general classroom atmosphere in which children are stimulated to develop productive beliefs about the aforementioned aspects of mathematics and mathematics learning and teaching but also explicitly by negotiating with his or her pupils about certain norms (Yackel & Cobb, 1996). In line with the socioconstructivist perspective, these norms are not predetermined criteria introduced into the classroom from the outside. Instead, these normative understandings are coconstructed and continually modified by the pupils and the teacher through their ongoing activities and interactions (Yackel & Cobb, 1996).

Organization of the Learning Environment

The learning environment consisted of a series of 20 lessons designed by the research team in consultation with the teachers of the experimental classes (for the complete set of the 20 lessons designed for the experimental classes [in Dutch], see Verschaffel, De Corte, Lasure, & Van Vaerenbergh, in press). Three major parts can be distinguished in the series of lessons:

1. Introduction to the content and the organization of the learning environment and reflection on the difference between an application problem that can be considered as a routine task and a real problem (Lesson 1).

2. Systematic acquisition of the five-step problem-solving model (Lessons 2–16). In Lessons 2 through 6, Stage 1 and the four embedded heuristics are taught. Stage 2, together with its four heuristics, is addressed in Lessons 7 through 11. Each heuristic from Stage 1 or Stage 2 is the focus of one lesson, and the last lesson of both stages (i.e., Lesson 6 and Lesson 11) is an "integration lesson" in which pupils learn how to choose among the distinct heuristics learned so far and how to integrate them into a single problem-solving process. In Lesson 12, Stage 3 of the model is addressed. Stage 4 of the problem-solving model is the focus of Lessons 13 and 14, and the final stage is addressed in Lessons 15 and 16. Examples of problems used in these lessons are given in Figures 1 through 4.

3. Learning to use the competent problem-solving model in a spontaneous and flexible way in four "project lessons" (Lessons 17–20). Each lesson involves a more complex application problem, the exploration, solution, and discussion of which takes a whole lesson. An example of such a more complex problem is given in Figure 5.
Wim would like to make a swing at a branch of a big old tree. The branch has a height of 5 meter. Wim has already made a suitable wooden seat for his swing. Now Wim is going to buy some rope. How many meters of rope will Wim have to buy?

FIGURE 1 Example of a word problem used in the lesson about the heuristic “Use your real-world knowledge” (Step 1).

At the end of the last whole-class discussion of each lesson, pupils received a sheet for their personal notebook, summarizing the essentials about the what, the how, and the why of each stage; the heuristic of the competent problem-solving model; or both (for an example of a sheet, see Figure 6). Although these sheets were constructed in advance by the research team (in consultation with the expert practitioners), teachers and pupils were allowed and even encouraged to complete the given information with their personal notes.

Components of Teacher Support

Because the lessons were not taught by the researchers but by the regular classroom teachers, these teachers were prepared for and supported in implementing the learning environment. The model of teacher development adopted reflected our views about pupils' learning by emphasizing the creation of a social context wherein
Every piece of a "dominoes game" is divided in two sides, each of which counts either 0, 1, 2, 3 or 4 dots. Some examples are given.

- What's the total amount of pieces in this game if there is one piece of every possible combination?
- What would be the total amount if each side of a piece can have a maximum of 6 instead of 4 dots?

FIGURE 2 Example of a word problem used in the lesson about the heuristic "Look for a pattern" (Step 2).

teachers and researchers learn from one another, rather than a model wherein the researchers transmit knowledge to the teachers (De Corte, 1998). Moreover, because the mathematical teaching—learning process is too complex to be prespecified and that teaching as problem solving is mediated by teachers' thinking and decision making, the focus of teacher development and support was not on making them perform in a specific way but on preparing and equipping them to make informed decisions (see also Carpenter & Fennema, 1992; Yackel & Cobb, 1996). This involved the following components:

- Provision of a general teacher guide containing an extensive description of the background, aims, overall content, and structure of the learning environment. A list of 10 general guidelines for the teachers, comprising actions they should take before, during, and after the individual or group assignments to increase the power of the learning environment, was also provided. In the teacher guide, each of these 10 guidelines (see Table 2) was followed by an explanation of its purpose as well as by several worked-out examples of what it meant to act accordingly.
- Provision of a specific teacher guide for each lesson, containing the overall lesson plan with the different mathematics application problems to be addressed in the lesson. Also included were the preferable sequence of instructional activities, the specific suggestions for appropriate teacher interventions before, during, and after the group or individual assignments, and examples of anticipated correct and incorrect solutions and solution methods.
100 children are being transported by minibus to a summer camp at the seaside. Each minibus can hold a maximum of 8 children. How many minibusses are needed?

The children assemble in the gymnasium where the sporting material is stored in big cases. This cases have to be carried to the sports grounds. To carry one case, 8 children are needed. How many cases can be carried by this 100 children from the gymnasium to the sports grounds at one time, if they all cooperate?

When the children have sported the whole day, they all are very hungry so they gather in the dining-room. The cook has prepared 100 litres of hotchpotch in 8 equal big kettles, all filled to the rim. How many litres of hotchpotch are contained in each kettle?

After dinner the children have to form rows of 8 for the physical training. How many children are left after having made a maximum number of rows?

FIGURE 3 Example of a worksheet used in the lesson about “Interpret the outcome and formulate an answer” (Step 4).
Below, three incorrect solutions for this problem are given. Find the mistake in each one of them.

**Solution 1:**
\[ 8 + 8 + 4 = 20 \]

No, this man can not store the 200 boxes with bulbs in this case, because the case can only hold 20 boxes.

**Solution 2:**
\[ 8 \times 8 \times 4 = \]
\[ 8 \times 8 = 64 \]
\[ 64 \]
\[ \times 4 \]
\[ 116 \]

No, this man can not store this 200 boxes with bulbs in this case, because the case can only hold 116 boxes.

**Solution 3:**
\[ 8 \times 8 \times 4 = \]
\[ 8 \times 8 = 64 \]
\[ \times 4 \]
\[ 256 \]

No, because 256 is a lot more than 200.

FIGURE 4 Example of a worksheet used in the lesson about “Evaluate the solution” (Step 5).
Pete and Annie are building a miniature town with cardboard. The space between the church and the town hall seems the perfect location for a big parking lot. The available space has the format of a square with a side of 50 cm and is surrounded by walls except for its street side. Pete has already made a cardboard square of the appropriate size. What will be the maximum capacity of their parking lot?

1. Fill in the maximum capacity of the parking lot on the banner.
2. Draw on the cardboard square how you can best divide the parking lot in parking spaces.
3. Explain how you came to your plan for the parking lot.

FIGURE 5 Example of a problem used in one of the project lessons.

- Provision of all the necessary concrete material for the pupils (e.g., worksheets, hands-on materials related to the application problems).
- Presence of one member of the research team in each lesson. This researcher did not intervene during the lesson itself, but before and after the lesson he or she had a 5- to 10-min preparatory or evaluative talk with the teacher.
- Organization of regular 2-hr meetings attended by all members of the research team and by all teachers and principals of the experimental school. Two
STEP 2: I PLAN MY SOLUTION PROCESS

WORK WITH EASY NUMBERS

WHAT?

When I don't have any idea of what to do to solve the problem, I try to simplify the problem by replacing the large or difficult numbers by small or easy ones.

HOW?

- I replace the large or difficult numbers by small or easy ones
- I try to solve the problem with the small or easy numbers
- I reflect upon the way in which I have solved the problem with the small or easy numbers, and I try to apply the same reasoning on the original problem

WHY?

- Because it's much easier to get a good understanding of the problem situation when it involves small or easy numbers
- Because it's much simpler to make a plan of the necessary solution steps when the numbers are small and easy
- Because the execution and control of mathematical operations with small and easy numbers is simpler too

FIGURE 6 Example of a sheet from the notebook for the pupils.
TABLE 2

General Guidelines for the Teachers Before, During, and After the Group and Individual Assignments

Before
1. Relate the new aspect (e.g., heuristic, problem-solving step) to what has already been learned before
2. Provide a good orientation to the new task

During
3. Observe the group work and provide appropriate hints when needed
4. Stimulate articulation and reflection
5. Stimulate the active thinking and cooperation of all group members (especially the weaker ones)

After
6. Demonstrate the existence of different appropriate solutions and solution methods for the same problem
7. Avoid imposing solutions and solution methods onto pupils
8. Pay attention to the intended heuristics and metacognitive skills of the competent problem-solving model and use this as a basis for the discussion
9. Stimulate as many pupils as possible to engage in and contribute to the whole-class discussion
10. Address (positive as well as negative) aspects of the group dynamics

meetings took place before the actual start of the experimental program, three during the course of the program, and one afterwards. In the first two meetings the teachers and principals were invited to comment on first drafts of the general and the specific teacher guides (which were given to them before the meeting), and possible extensions and improvements were proposed and discussed. The goal of these meetings was to give the teachers “co-ownership” of the general and specific teacher guides. In the next three meetings, teachers and researchers exchanged positive experiences as well as difficulties with the implementation of certain aspects of the learning environment and searched for appropriate solutions for these difficulties. During the last meeting the major outcomes of the implementation and evaluation study were presented to the teachers and the principals of the experimental schools, who also shared their own experiences and impressions.

RESEARCH METHOD

Participants

Four experimental fifth-grade classes (27, 19, 21, and 19 pupils, respectively) and seven comparable control classes (29, 22, 19, 21, 20, 17, and 18 pupils, respectively) participated in the study. All these classes belonged to different, sex-mixed elementary schools in the neighborhood of Leuven, Belgium.
Design of the Study

A pretest–posttest–retention test design with an experimental group and a control group was used as the overall research design (see Table 3). The first group received the 20 lessons of the experimental learning environment during school hours and, more specifically, during hours regularly allocated for mathematics. Each lesson lasted 50 to 60 min. These lessons were spread over a period of about 4 months. During these months there were no other lessons in solving mathematical application problems in the experimental classes.

During the same period the pupils from the control classes continued to follow the regular mathematics curriculum, which also involved a considerable number of lessons in arithmetic word-problem solving (for a description of the traditional instructional approach of word-problem solving followed in these control classes, see De Corte & Verschaffel, 1989). We were not able to pursue a systematic and detailed analysis of what happened in each of the control classes. However, informal discussions with the teachers of these classes and an analysis of the textbooks used provided us with a good overall view on what happened there during the time of the intervention in the experimental classes. This revealed that little or no attention was paid to the intentional and systematic teaching of heuristics and metacognitive skills. In line with what we found in previous, more systematic studies of current instructional practice (De Corte & Verschaffel, 1989), these lessons typically involved (a) demonstration of a particular solution method for well-defined types of word problems that are explicitly listed in the curriculum (e.g., problems about gross-net-tare weight; problems about time, speed, and distance; problems about unequal partition), followed by (b) individual paper-and-pencil training in the use of the demonstrated method on a series of very
similar problems and (c) presentation of the correct answer and of the learned appropriate solution method for these problems by the teacher or by one of the pupils. Furthermore, whereas the total number of word problems administered was more or less equal in the control and the experimental classes, the number of lessons mainly devoted to word-problem solving during that period in the control classes was about 15 lessons (instead of 20). Because it was practically impossible to control in a very precise way for the total number and the nature of the word-problem lessons in the control classes, we decided to work with seven instead of only four control classes, assuming that this would result in a control group that was representative of the way teaching and learning of word-problem solving is done in a typical nontreatment Flemish fifth-grade class. In the next section, we briefly describe the different instruments used to evaluate the implementation and the effects of the experimental learning environment.

Instruments

Before and after the intervention, three evaluation instruments were collectively administered in the experimental as well as in the control classes.

A self-made written word-problem test (WPT) consisting of 10 difficult nonroutine tasks was constructed in three parallel versions that were administered before the intervention (Test Moment 1), after the intervention (Test Moment 2), and as a retention test (Test Moment 3). Although the first concern in constructing the test items was their unfamiliar and complex nature, we also deliberately searched for problems that, according to a rational analysis and a trial with older, more experienced problem solvers, lent themselves to the efficient application of the heuristics taught in the experimental program (without showing a superficial similarity with any of the training items.) Table 4 presents some example items from the pretest version of this WPT.

Each answer on an item of these tests was scored either as correct answer, wrong answer, technical error, or no answer. However, because the learning environment was especially aimed at the improvement of pupils’ modeling and problem-solving skills (rather than at computational precision), purely technical errors were also considered as correct. This scoring system had a very high reliability. Three independent raters scored all answers of the pupils of two randomly chosen classes on two different test moments, and their interrater reliability reached a kappa coefficient of .99. Cronbach alphas for the three parallel versions of this test were .56, .75, and .79, respectively. The relatively low internal consistency of the pretest (i.e., .56) can be explained by the small variation in the scores on this test, which were in general extremely low (see further in the article for more information). In addition to all of the answers being scored in one of the four aforementioned response categories, pupils’ response sheets were also scrutinized for traces
TABLE 4
Examples of Items From the Word-Problem Pretest

1. Martha is reading a book. Suddenly she finds out that some pages are missing because page 135 is immediately followed by page 173. How many pages are missing?

2. Lies has two doll's houses. The square floor of the small doll's house has a side of 40 cm and consists of 16 tiles. The square floor of the large doll's house has a side which is exactly twice the side of the small doll's house. How many tiles are needed for the floor of the large doll's house if the same tiles are used?

3. Catherine builds little houses with matches. To construct 2 houses, she needs 9 matches. To build a row of 5 houses, she needs 21 matches. How many matches will she need to build a row of 10 houses?

of the application of heuristics. More specifically, for every pupil and for every problem from the pretest, the posttest, and the retention test versions of the WPT, we indicated whether the pupil's notes showed clear evidence of the application of one or more of the eight heuristics from the competent problem-solving model mentioned in Table 1. The scoring of the use of heuristics also had a high interrater reliability. Three independent raters analyzed all pretest, posttest, and retention test items of one randomly chosen experimental class on the use of heuristics and reached a kappa coefficient of .96.

The second instrument was a self-made questionnaire aimed at assessing pupils' beliefs and attitudes about the teaching and learning of mathematical word-problem solving (Beliefs and Attitudes Questionnaire [BAQ]). This questionnaire was derived from an initial questionnaire consisting of 40 Likert-like items dealing with various aspects of pupils' beliefs about and attitudes toward the teaching and learning of mathematical problem solving. A factor analysis of the responses of a large group of 10- to 11-year-olds (including the pupils from the experimental and control classes at Test Moment 1) to the 40 items of this initial questionnaire revealed two relatively reliable factors (explaining 22% of the overall variance). These factors were finally used as the two subscales of the BAQ: a first subscale, containing 7 items dealing with pupils' pleasure and persistence in solving word problems (e.g., "I like to solve word problems," "Difficult problems are my favorites"; Cronbach's $\alpha = .70$), and a second subscale with 14 items expressing a problem- and process-oriented view on word problem solving (e.g., "There is always only one solution to a word problem," "Listening to explanations of alternative solution paths by other pupils is a waste of time"; Cronbach's $\alpha = .73$). With respect to each item of both subscales, pupils had to respond by indicating whether they strongly agreed, agreed, were uncertain, disagreed, or strongly disagreed with the statement. Each response was awarded 1 to 5 points, with a
higher score always reflecting a professed belief or attitude that was more in line with the program's philosophy. This resulted in a maximum factor score of 35 (7 × 5) for the first subscale and of 70 (14 × 5) for the second one.

Third, to assess pupils' general mathematical knowledge and skills, an existing standard achievement test was administered. The test consists of 58 multiple-choice items belonging to eight subscales (i.e., Logical Operations, Sets, Relations, Numbers, Arithmetic Operations, Word Problems, Measuring, and Geometry); each item was scored 1 or 0, resulting in a total score between 1 and 58. Reliability and validity assessments on large groups of pupils from the upper grades of the elementary school revealed that this test had a high internal consistency (i.e., .87 for fifth-graders) and validity in terms of a correlation with the results for mathematics at the end of the school year (i.e., .76 for fifth graders; Stinissen, Mermans, Tistaert, & Vander Steene, 1985).

Fourth, to get a better insight into the qualitative changes in pupils' problem-solving processes as a consequence of the intervention, three pairs of children were selected from each experimental class (one pair of high ability, one pair of medium ability, and one pair of low ability). Dyads of equal ability were asked before and after the intervention to solve five new nonroutine application problems. Unfortunately, it was not possible to organize similar interviews in the control classes. The problem-solving processes of these dyads were videotaped and afterwards analyzed by means of a self-made scoring scheme consisting of three aspects: (a) the final result of the problem-solving process (correct answer, wrong answer, technical error, or no answer), (b) the use of the eight heuristics taught in the program, and (c) the frequency of occurrence of four valuable metacognitive activities (i.e., orientation, planning, monitoring, and evaluation). This scoring system for the four metacognitive activities was largely inspired by similar systems developed and applied by Schoenfeld (1992), Artzt and Armour-Thomas (1992), and Goos and Galbraith (1996). The scoring of the videotapes was done by two "blind judges," who jointly decided when and how to score a particular part of a dyad's problem-solving protocol.

Fifth, to assess the implementation of the learning environment by the teachers, a sample of four representative lessons was videotaped in each experimental class and analyzed by means of a self-made observation scale to determine an implementation profile for each experimental teacher. This implementation profile consisted of 10 categories of teacher activities that we considered essential for the successful implementation of the learning environment and that corresponded more or less to the 10 general guidelines mentioned in Table 2 (e.g., "Provide a good orientation to the task," "Observe the group work and provide appropriate hints when needed"). Based on inspection of the videotapes, every lesson was given a score of 0 (no or inadequate implementation), ½, 1, 1½, or 2 (very good implementation) for each of these 10 categories of the observation scale by two independent judges (maximum score = 20). The interscorer reliability, expressed in a
kappa coefficient, was .85, which is reasonably high for an observation scale of this type.

Finally, the four teachers of the experimental classes were interviewed shortly after the end of the intervention but prior to receiving feedback about the learning outcomes of their pupils. During this interview, which took about 2 hr, the teachers were asked about their impression of the different aspects of the learning environment, about the difficulties they had experienced in implementing it in their classroom, and about possible suggestions for improving it.

**HYPOTHESES AND RESEARCH QUESTIONS**

The first and most important hypothesis of the study was that the learning environment would have a significant positive influence on pupils' skill in the solution of nonroutine mathematical application problems. Therefore, we expected the experimental group to score significantly higher on the written WPT than the control group during the second and the third test moment but not on the first test moment.

Moreover, we expected that the learning environment would have a positive effect on pupils' beliefs and attitudes toward mathematical problem solving. This hypothesis was tested by comparing the answers of the pupils from the experimental classes and the control classes on the two subscales of the BAQ, which was administered to all these pupils before and after the intervention. The prediction was that the experimental group would reach a significantly higher score than the control group on both subscales after the intervention, whereas the scores of both groups would not differ on the pretest.

There was no prediction formulated for the results of the standard achievement test. One could argue that the extra attention in the learning environment at problem solving might negatively influence pupils' mastery of the other parts of the mathematics curriculum of these fifth graders. However, one could also anticipate a positive transfer effect from what the pupils had learned in the experimental learning environment to these other parts of their mathematics curriculum.

We also hypothesized that the increase in the problem-solving capacity of the pupils of the experimental classes (as indicated by the WPT scores) would be paralleled by an increase in the quality of their problem-solving processes. More specifically, it was predicted that, during the posttest (and the retention test), the pupils from the experimental classes would apply the heuristic and the metacognitive strategies more often than before the intervention and more often than the pupils from the control group after the intervention.

Finally, we analyzed if the anticipated overall effects of the intervention on the tests (WPT and standard achievement test) and on the questionnaire (BAQ) were found in each of the four experimental classes. Likewise, we analyzed whether gains from the learning environment were restricted to participants who were of high and medium ability, as is not uncommon in research on teaching (mathemati-
SOLVING APPLICATION PROBLEMS

Problem solving (Resnick, 1987; Wittrock, 1986) or were also found among low-ability pupils.

ANALYSIS AND RESULTS

The impact of the learning environment on pupils' results on the three collective instruments (i.e., WPT, BAQ, and standard achievement test) was analyzed by means of univariate analyses of variance (ANOVAs) with a hierarchical factorial design with three repeated measures (factor test: Pretest, Posttest, and Retention Test) for the WPT and with two repeated measures (factor test: Pretest and Posttest) for the BAQ and the standard achievement test. The factors Group (experimental vs. control group) and Classes were nested in Group (four experimental and seven control classes). Main and interaction effects were further analyzed with a posteriori Tukey HSD tests. Moreover, to get a better idea of the statistical power of the obtained effects, we also calculated effect sizes (Cohen, 1988). The most important outcomes of these analyses as well as those of the additional analyses related to the other research questions and hypotheses, are presented in this section.

First, as expected, the ANOVA on the results of the three WPTs revealed a significant interaction effect between Group and Test, $F(2, 396) = 30.60, p < .001$. Figure 7 presents the mean scores of the pupils from the experimental and control groups on

![Figure 7](image)

**Figure 7** Mean scores of the experimental and the control group on the three versions of the word-problem test (pretest, posttest, and retention test).
these three tests (which were 1.6, 3.6, and 4.3, respectively, in the experimental group and 1.6, 2.1, and 2.7, respectively, in the control group on the pretest, posttest, and retention test). Although no significant difference was found between the experimental and control groups on the WPT during the pretest, the former group significantly outperformed the latter during the posttest, and this difference in favor of the experimental group continued to exist on the retention test (a posteriori Tukey tests, \( p < .05 \)). This effect was a medium effect size of .31 (Cohen, 1988).

Second, the ANOVAs of the results on the two BAQ subscales also revealed significant interaction effects between the factors Group and Test for the Pleasure and Persistence subscale, \( F(1, 205) = 11.67, p < .001 \), and for the Problem- and Process-Oriented View on the Word-Problem Solving subscale, \( F(1, 205) = 10.73, p < .01 \). The mean scores of the pupils from the experimental classes on both subscales evolved more in the expected direction (from 23.6 on the pretest to 26.4 on the posttest for the Pleasure subscale and from 54.2 on the pretest to 58.6 on the posttest for the View subscale) than the mean scores of the pupils of the control classes (from 23.2 on the pretest to 23.3 on the posttest for the Pleasure subscale and from 54.7 on the pretest to 55.8 on the posttest for the View subscale). Additional Tukey tests showed that, as expected, the experimental group scored significantly higher \(( p < .05 \) than the control group on both scales after the intervention, whereas the scores of both groups did not differ significantly before the intervention. However, these effects, although significant, were very small. In both cases an effect size of only .04 was found (Cohen, 1988).

Third, a significant interaction effect between Group and Test was also found for the total score on the standard achievement test, \( F(1, 209) = 8.12, p < .01 \). The mean score of the experimental group increased from 28.2 on the pretest to 36.1 on the posttest, whereas the control group had a mean score of 28.5 on the pretest and of 34.4 on the posttest. According to Cohen (1988), this is a medium effect (effect size = .38). Additional Tukey tests showed that, although there was no significant difference between the pretest results on the standard achievement test of the experimental and the control groups, the results on the posttest revealed a significant difference \(( p < .05 \) in favor of the former group. This suggests that the greater attention on mathematical problem solving in the experimental classes (at the expense of the other subject-matter topics in mathematics) had no negative side effect and even a small positive (transfer) effect on pupils’ mathematical knowledge and skills as a whole.

Fourth, to find out if the significant improvement in the experimental pupils’ scores on the WPT was accompanied by an increased use of the heuristics taught in the program, a univariate ANOVA was run with Group, Test Moment, and Class (nested in Group) as independent variables and the number of word problems in the WPT, for which at least one of the eight taught heuristics was visibly used, as the dependent variable. The ANOVA revealed a significant interaction between Group and Test, \( F(2, 414) = 98.87, p < .001 \). As expected, there was no significant
difference between the experimental and the control groups on the pretest, but there was a significant difference in the number of solution processes involving the application of at least one of the heuristics during the posttest and the retention test (Tukey, \( p < .05 \)). Figure 8 gives a graphic representation of this interaction effect, which is, according to Cohen (1988), very large (effect size = .76). A further comparative analysis between the pupils from the experimental and the control groups on the three test moments for each heuristic separately revealed that the greatest improvements were found for the following heuristics: "Make a drawing," "Look for a pattern," "Distinguish relevant from irrelevant information," and "Guess-and-check." However, this finding does not warrant a definitive conclusion about the efficacy of the program for children's acquisition of the different heuristics. Indeed, we cannot exclude that these heuristics differed also in their applicability to the items of the WPT as well as in the ease with which they could manifest themselves in a written protocol (Verschaffel et al., 1998).

Fifth, a significant Class (in Group) × Test interaction effect was found for the different dependent variables, namely the scores on the WPT, \( F(18, 396) = 4.58, p < .0001 \); the scores on the Pleasure, \( F(9, 205) = 2.40, p < .05 \), and the View subscales of the BAQ, \( F(9, 205) = 2.18, p < .05 \); and the scores on the standard achievement test, \( F(9, 208) = 2.84, p < .01 \). This means that the aforementioned general effects of the learning environment were not found to the same extent in all

![Graph showing mean frequency of the use of heuristics](image_url)
four classes. More specifically, one of the experimental classes, namely Class 3, realized considerably less improvement on all measures than the three other classes (for more details, see Verschaffel et al., 1998).

Sixth, we analyzed whether the aforementioned general effects of the learning environment were found to the same extent for the different ability levels (i.e., high, medium, and low). To make this analysis possible, every pupil from the experimental and the control groups was put into a high-ability, medium-ability, or low-ability group based on his or her pretest score on the standard achievement test. Pupils with a percentile score on the standard achievement test of more than 66 were put in the high-ability group, those with a percentile score between 33 and 65 in the medium-ability group, and those with a percentile score below 33 in the low-ability group. Figure 9 gives the mean scores on the WPT for the three ability levels in the experimental and the control groups on the three test moments. An ANOVA with the factors Group, Test Moment, and Ability Level as independent variables and the number of correct answers on the WPT as the dependent variable revealed no significant triple interaction, $F(4, 370) = 1.37, ns$. This result suggests that all three ability levels contributed significantly to the aforementioned Group × Test interaction effect (presented in Figure 7). In other words, the learning environment resulted not only in a significant increase in the problem-solving skills of the high-ability and medium-ability pupils but also of those of low ability. Also, on
the other dependent variables, namely pupils' scores on the two BAQ subscales, the standard achievement test, and the frequencies of their use of the heuristics on the WPT, no significant interactions between Group, Test Moment, and Ability Level were found, $F(2, 192) = 1.17, \text{ns}$; $F(2, 189) = .72, \text{ns}$; and $F(4, 382) = 1.04, \text{ns}$, respectively. Taken as a whole, these results indicate that all three ability levels contributed in a significant way to the increased overall learning results of the experimental group reported previously.

The results of the qualitative analysis of the videotapes of the problem-solving processes of the three dyads from each experimental class revealed that on the posttest, these dyads made nearly twice as much spontaneous use of heuristics as during the pretest, from 42 manifestations of application of any of the eight heuristics to 81 such manifestations during the posttest. The greatest improvements were observed for "Make a table" and "Use your real-world knowledge" (for more details, see Verschaffel et al., 1998). The finding that the data about the use of the different heuristics in the written test did not correspond with those in the interviews with the dyads supports the warning given earlier. That is, the findings about the increase in the use of a particular heuristic may be related to the nature of the test items as well as to the kind of data-gathering technique; moreover, these data are based on only a small number of dyads that is probably not representative of the entire group of children in the experimental classes. As explained before, we also scrutinized the dyads' problem-solving protocols for signs of four typical metacognitive activities (i.e., orientation, planning, monitoring, and evaluation) using a self-developed scoring scheme. The occurrence of these metacognitive activities increased also from pretest to posttest: from 26 to 35 for orientation, from 18 to 26 for planning, from 89 to 116 for monitoring, and from 3 to 11 for evaluation (for more details, see Verschaffel et al., 1998).

The videotapes of a series of experimental lessons were analyzed by means of the so-called implementation profile consisting of 10 different categories scored on a scale from 0 (insufficient or wrong implementation) to 2 (very good implementation). This led to an average implementation score of 12.4 (on a maximum of 20) for all four experimental teachers together, indicating that, as a whole, the learning environment was implemented in a satisfactory way by the teachers. There were clear differences in the teachers' mean scores on the distinct components of the implementation profile, $F(9, 108) = 8.67, p < .001$. Tukey tests revealed that the mean score on Component 4 ("Stimulate reflection during the group work"), Component 5 ("Stimulate the task involvement of all group members"), and Component 10 ("Pay attention to the group dynamics during the whole-class discussion") were significantly lower than on the other components (for more details, see Verschaffel et al., 1998). Moreover, an ANOVA on these four teachers' implementation scores revealed significant differences in the extent to which they had implemented the learning environment, $F(3, 12) = 10.227, p < .01$. The average individual total scores of the four teachers were 10.1, 13.8, 10.9,
and 14.6. Additional Tukey tests showed that the differences between the total scores of Teachers 2 and 4 on the one hand and Teachers 1 and 3 on the other were significant ($p < .05$). The relation between the teachers’ individual implementation profiles and their pupils’ learning outcomes was somewhat puzzling. The relatively low score of Teacher 3 on the implementation profile (10.9) provides a good explanation for the relatively small improvements of the pupils from Class 3 on all dependent measures (compared to those of the pupils of the three other classes). However, the pupils of the other teacher with a low implementation profile (Teacher 1, who had a comparable mean implementation score of 10.1) made the largest gain on the WPT.

Finally, the results of the interviews with the four experimental teachers can be summarized as follows. First, they considered the five-step competent problem-solving model that lies at the basis of the learning environment as appropriate and attainable for fifth graders. Moreover, they appreciated the content and the organization of the lessons, they were very satisfied with the support and help they received before and during the lessons, and they were enthusiastic about their active involvement and participation in the research project. However, the teachers also indicated some problems. First, they thought that three of the eight heuristics from the problem-solving model ("Make a flowchart," "Look for a pattern," and "Simplify the numbers") were very difficult for fifth graders, at least in the form in which they had been worked out in this learning environment. Second, certain application problems used in the learning environment needed adjustments, especially to make them more authentic and attractive to children of that age. Concerning the group work, some teachers expressed serious doubts about the efficacy of grouping children in heterogeneous groups for the purpose of the program. Furthermore, lack of time resulting in premature endings of group activities, classroom discussions, and individual assignments was unanimously reported as the major problem in the implementation of the learning environment. Finally, two teachers stressed that they would have liked to participate in an even more active and more partnership-based way in the design of the learning environment and in the realization of the whole research enterprise. However, they also admitted that such a form of cooperation, which could be typified in Wagner’s (1997) terms as a colearning agreement between researchers and practitioners, would have been very difficult to effectuate, taking into account their actual workload as full-time elementary school teachers.

**CONCLUSIONS AND DISCUSSION**

In this article, we presented a design experiment aimed at the development, implementation, and evaluation of a powerful environment for learning mathematical modeling and problem solving in the upper elementary school. A set of carefully
designed application problems, a collection of highly interactive teaching methods, and the introduction of new sociomathematical classroom norms were combined in an attempt to create a substantially modified learning environment. This environment focused on fostering a mindful approach toward mathematical modeling and problem solving. The environment was implemented in four classes of the fifth grade, and its effects were evaluated in a teaching experiment with a pretest–posttest–retention test design involving an experimental group and a comparable control group. Whereas the theoretical background of the study goes back to the cognitive-rationalist perspective on teaching and learning mathematical problem solving, this perspective was substantially enriched with insights from more recent theoretical orientations, namely situated cognition and socioconstructivism.

According to the results of the written WPT pretest and posttest, the learning environment had a significant positive effect on the development of pupils' problem-solving skills. The results on the retention test revealed that this positive effect did not disappear after the end of the experimental lessons. The learning environment also had a positive impact on pupils' enjoyment of and persistence in solving mathematical application problems and on their beliefs and attitudes toward the learning and teaching of mathematical problem solving (BAQ). The results on the standard achievement test showed that the extra attention on problem-solving strategies, beliefs, and attitudes in the experimental classes had no negative influence on the learning results for the other, more traditional aspects of the mathematics curriculum; to the contrary, there was even a small but significant positive transfer effect. The analysis of pupils' written notes on their response sheets of the WPT showed that the improved learning outcomes of the experimental group were paralleled by a very substantial increase in their spontaneous use of the heuristics being taught to them. The comparison of the results for the different ability levels showed that not only the high- and medium-ability pupils but also low-ability pupils benefitted significantly from the learning environment in all the aforementioned respects. Overall, these positive results are in line with the conclusions of several other investigations (Lester et al., 1989; Van Haneghan et al., 1992; Verschaffel & De Corte, 1997a), which demonstrated the practicability and the efficacy of learning environments aimed at the development of valuable (meta)cognitive strategies embedded in the mathematics curriculum. However, an important difference between these other investigations and ours is the greater ecological validity of this study, especially because the teaching was not done by the researchers but by the regular classroom teachers (Verschaffel, in press). Another difference relates to the greater variation in the instrumentation for data gathering and data analysis.

In our view, the positive results can be attributed to the combination of the following three factors: (a) the learning environment was not aimed at one specific aspect of mathematical competence (e.g., particular heuristics or metacognitive skills) but at the simultaneous and integrated acquisition of different categories of aptitudes; (b) systematic attention was paid to the different aspects of that environ-
ment, especially the nature of the problems, the instructional techniques, and the classroom culture; and (c) the teachers were involved in the elaboration of the environment and were given substantial help and support during its implementation. Due to the kind of quasi-experimental design of this study, it is impossible to draw conclusions about the relative contribution of these three different possible causal factors (and of their constituent elements) to the obtained results. Therefore, further research is needed in which different versions of this learning environment are systematically compared to identify the aspects that contribute especially to its success and make it even more powerful.

Besides this major methodological concern, there are some other problematic aspects of the research methodology. First, the learning environment was realized and evaluated in only four experimental classes. Although this number of classes is already considerably more than in many other design experiments (e.g., Lester et al., 1989; Van Haneghan et al., 1992; Verschaffel & De Corte, 1997a), a larger number of experimental classes would not only have resulted in more reliable and generalizable conclusions about the effectiveness of the learning environment but also would have allowed a more systematic study of the relation between the teachers' implementation of this environment and their pupils' learning outcomes. Furthermore, critical questions can be raised about some of the evaluation instruments. Because of the aims and scope of this study, it was not possible to borrow existent instruments with well-documented and widely acknowledged psychometric qualities. Therefore, we had to develop these instruments ourselves, but because of lack of time, reliability and validity could not always be sufficiently documented. This latter criticism holds especially for the classification scheme used for the analysis of the metacognitive activities of the dyads from the experimental classes as well as for the observation scheme underlying the experimental teachers' implementation profiles.

A final concern has to do with the absence of an in-depth analysis of what went on in the control classes. Although the available evidence suggests that the teaching and learning of mathematical problem solving in the control classes was representative for the current typical approach in Flemish schools, it was practically impossible in the framework of this project to collect systematic data about the actual instruction in all of these control classes.

Besides these methodological shortcomings, some partial findings also warn against overly optimistic conclusions. In this respect, we are reminded that—except for the drastic increase in the use of heuristics during the WPT—none of the positive outcomes of the learning environment on the experimental classes was large in terms of Cohen's (1988) effect size measure. Moreover, on all measures, the results of one of the four experimental classes (Class 3) were considerably lower than in the other three classes. Finally, the learning environment did not lead to a decrease in the initial differences between pupils of high and low ability on the dependent measures.
Taking all this into account, we end this article with a discussion of some problematic aspects of the learning environment as it was developed and implemented that may help to explain why the obtained effects were not greater. At the same time, some suggestions for adjustments of the learning environment and for further research are given.

A first question relates to the model of competent problem solving and to the eight embedded heuristics. It seems clear that some components or aspects of this model need refinement or adjustment to make them more understandable and accessible for pupils of that age. In this respect, we refer to Siemon’s (1992) comment on models of competent problem solving like the one used in this study, namely, that the language referring to the different stages (e.g., understand, orient, plan, execute, verify, check) as well as the suggested linear order of these stages, may be too unfamiliar and too alienating to most upper elementary school children and may need to be transformed into more user friendly terms in a way that reflects better the cyclical nature of the problem-solving process. In this regard, we also point to the finding from this study that the pupils seemed to have greater difficulty with acquiring and applying some heuristics (i.e., “Look for a pattern” and “Make a flowchart”) than with others (e.g., “Draw a picture” and “Guess and check”; see Verschaffel et al., 1998).

Second, although we recognized the necessity of the establishment of a new classroom culture that fits within the aims and scope of the learning environment, and although we observed some significant effects of the teachers’ attempts to realize such a new classroom culture in enhancing their pupils’ attitudes and beliefs, this pillar was still not addressed in a sufficiently systematic and effective way. In this respect, the literature did not provide us with many helpful suggestions for the elaboration of classroom activities and action plans for the teacher that would guarantee a beneficial effect on pupils’ attitudes and beliefs. A rare recent exception is a study by English (1998) containing several worked-out classroom activities and action plans (e.g., the use of “debating and discussion cards”) aimed at stimulating pupils to articulate, reflect on, and change their beliefs and attitudes about mathematical problems and mathematical problem solving.

However, the two other pillars of the learning environment (i.e., the tasks and the instructional techniques) need improvement, too. With respect to the tasks, we need to search for a better balance between the requirement that the problems can elicit the intended (meta)cognitive activities and the realistic or authentic nature of these problems. With regard to the instructional techniques, a major challenge is how to organize and support the small-group activities so that all pupils—including the low-ability ones—interact and cooperate in a task-oriented high-quality manner. Another challenge is to try to integrate the promising outcomes of recent studies on the development and the implementation of problem-posing programs (e.g., English, 1998; Silver, 1994) into learning environments focusing on mathematical problem solving.
Shortcomings in the teachers' implementation of the learning environment may also partly explain why no stronger effects were obtained. One point this study has made very clear is that the effective realization of such a learning environment puts extremely high demands on the teacher (see also Carpenter & Fennema, 1992; Schoenfeld, 1992; Yackel & Cobb, 1996). Although the teachers were prepared and supported as much as possible (within the constraints of the research project), the main focus of the study was on the design and the evaluation of the learning environment, not on developing and testing a scientifically-based system of teacher development and support. Both from a theoretical and a practical point of view, attention to this issue in future research is extremely important (Carpenter & Fennema, 1992; De Corte et al., 1996; Gravemeijer, 1994; Yackel & Cobb, 1996).

Finally, we remind the reader that the intervention consisted of only 20 lessons given rather separately from the rest of the mathematics curriculum. Presumably, the results would have been better if we could have increased the available instructional time and if we could have integrated the learning environment better within the regular mathematics lessons. In this respect, we stress that our decision to design, implement, and evaluate a restricted series of lesson plans and materials was based on practical considerations and involved no plea whatsoever for the implementation of such a special program in the mathematics curriculum. Ultimately, the development of mathematical problem-solving skills, beliefs, and attitudes should not emanate from a specific part of the curriculum but should permeate the entire curriculum as recommended by the different reform documents mentioned at the beginning of this article.

REFERENCES


