CHAPTER SIX

Research into components of *Physhint*[4]

6.0 Abstract

Many students experience difficulties in solving applied physics problems. Researchers claim that the development of strategic knowledge (analyse, explore, plan, implement, verify) is just as necessary for solving problems as the development of content knowledge. In order to improve these problem-solving skills, it might be profitable to know at what time during problem-solving is the use of instructional support most effective: before, during or after problem-solving.

In an experiment with fifth-year secondary school students, one experimental group (n = 18) received hints during and worked examples after problem-solving, and another experimental group (n = 18) received worked examples only after problem-solving. Both groups used versions of a computer program to solve a variety of problems. The control group (n = 23) used a textbook. There was a pre-test to estimate the measure of prior expertise of the students in solving physics problems. The results of a problem-solving post-test indicated that the version of the program providing hints during and examples after problem-solving was the most effective, followed by the version which only supplied examples afterwards. There was no difference in effect for students with more than average prior knowledge or less prior knowledge.

6.1 Introduction and research questions

In developing computerized instruction for problem-solving, several design issues have to be addressed. The most important design issues relate to the type of knowledge to be developed in problem-solving, the delivery of problem-solving instruction content, its timing, and student characteristics that may influence

---

[4] This chapter is based on Pol, H.J., Harskamp, E.G. & Suhre, C.J.M. (2008). The effect of the timing of instructional support in a computer-supported problem-solving program for students in secondary physics education. *Computers in Human Behavior*, *24*, 1156-1178.

learning to solve problems (De Jong, 1986; Mestre, 2002). This paper addresses the issue of timing of instruction.

6.1.1 Type of problems and knowledge involved

Learning to solve physics problems is an important goal of secondary education. Students, however, have great difficulty in solving problems in which physics knowledge about facts and rules of physics (declarative and procedural knowledge) has to be applied. Although they often possess this knowledge because it has recently been taught, they still cannot apply it to fresh problems dissimilar to the ones they have studied. For instance, when the formula 'velocity = distance / time' is taught, many students find it hard to apply this knowledge to a situation in which the runners start from opposite directions with different average speeds and the question is where they will meet. Students can use several strategies to solve such a problem (Bransford et al., 2000). Problems that ask for the application of knowledge are difficult for students because they may not recognize the knowledge to be used or they may not be able to combine different pieces of their knowledge into a solution plan. De Jong (1986) calls the ability to match a new problem situation with existing knowledge 'strategic knowledge'. He defines strategic knowledge as the ability to use solution methods in relatively new situations for the students. Strategic knowledge tells students how to analyse a problem situation, find relevant content knowledge, make a plan and solve a problem. Other researchers (e.g. Mestre, 2002) also stress the importance of strategic knowledge in learning to apply physics knowledge to new problems.

6.1.2 The timing of instruction

An important question in designing instruction to develop problem-solving abilities is: 'At what time is support most effective for the development of strategic knowledge in solving problems?' Supporting students in solving applied problems can be accomplished by giving instructions or examples before the problem-solving process begins, during the problem-solving process or after the student's final answer.

Supporting students in advance can be done with worked-out examples in which students are clearly instructed about the way to solve types of problems *before* starting to work out these problems for themselves (Owen & Sweller, 1985; Renkl, 2002). In using worked-out examples, a student learns to solve one type of problem at a time, step-by-step with less reliance on help, thus learning to solve problems independently (the decrease of scaffolded support is called ‘fading’, see Renkl et al., 2002). Van Gog (2006) showed a significant effect of the use of worked-out examples in changing from process-oriented to product-oriented worked-out examples during the learning process. In this way, students first have the chance to develop knowledge about the solution procedures for solving a set of problems, followed by the opportunity to investigate possible solutions, and are no longer hampered by a lack of knowledge of the correct procedure. The success of worked-out examples can be explained by the fact that knowing how to solve a certain type of problem in a general sense gives the student more mental capacity to solve a particular problem. Worked-out examples did show good results in cases where the problems used for the instruction were similar to the worked-out examples. However, in cases where problems were very different from the instructed problems, the effect of worked-out examples disappeared, or even became negative (Kalyuga et al., 2001; Renkl et al., 2002). When students need to learn how to solve different types of problems, worked-out examples in advance seem to be ineffective.

The second option for timing support is *during* the problem-solving process (just-in-time). Examples can be found in different tutoring systems, which give students help according to their actions. If a student does not succeed in finishing a problem, or gives a wrong answer, the program provides a hint which should help the student to continue with the problem-solving process (Albecate & VanLehn, 2000; Koedinger et al., 1997). Tutoring systems are based on the principle that students learn by developing their own initiative and combining the different kinds of knowledge they already have when searching for a solution to a new problem. In helping students when they are struggling, they guide them through the solution process and thus develop their strategic knowledge. Some researchers claim efficiency for tutoring programs when tested using problems which are similar (that is, the method of solving them is equivalent) to problems the students have already practised (see also Corbett & Anderson, 2001). In using these kinds of tutoring systems for the development of problem-solving abilities, one has to address two issues. Firstly, many tutoring programs are based on one superior procedure to solve problems. However, in solving applied problems, not all students use the same procedure. In other words, when using a tutoring program, a student is forced to follow one procedure when they might be better suited to another. The other drawback of tutoring during the problem-solving process is that students might become dependent on the tutoring program. In solving applied problems, students need to learn to find their own way to an answer (Fox, 1993; Schoenfeld, 1992).

The third option of timing support is giving help *afterwards*, also called delayed feedback. Feedback afterwards has proven to be effective (Anderson et al., 1995; Bloom, 1984; Mathan & Koedinger, 2005). The main disadvantage of delayed feedback is related to the amount of time that elapses between solving the problem and receiving feedback. Feedback can be effective when provided directly after solving a problem. However, when the delay becomes greater, it is not effective because students may have forgotten part or all of their solution method (Mathan & Koedinger, 2005).

6.1.3 Program or student control of instruction

In computer environments with program control the program decides when instruction is presented and also decides the content of the instruction. The procedure for solving problems is clear from the start. Examples can be found in tutoring programs for problem-solving such as devised by Albecate & VanLehn (2000) or Koedinger et al. (1997). The supply of instruction is predictable and unambiguous. Fine-tuning of the instructions to overcome imperfections in non-standard but otherwise adequate solution methods is often not provided. An important drawback is that program control leads to programs that are effective only in cases where the solving of similar problems is being taught (see Owen & Sweller, 1985; Renkl, 2002).

The drawback of being unable to provide instruction that fits in with adequate alternative problem-solving approaches does not apply to student-controlled computer environments. In these environments students can be given control over whether and when to consult the information offered by the program and which instruction method to follow. Instruction can be designed in a way that leaves room for the students to choose from different solution methods (Mathan & Koedinger, 2005; Mestre, 2002; Reif, 1995; Teong, 2003). One reason for preferring student control in a computer program is that it gives students sufficient room to develop strategic knowledge that fits their way of learning (Reif, 1995). Students need to acquire a flexible problem-solving strategy with which they are able to tackle different types of problems.

Examples of the effectiveness of student control on problem-solving abilities can be found in the research of Mathan and Koedinger (2005) and Mestre (2002). Mestre found positive effects on problem-solving when asking students to pose problems based on concepts and contexts of a certain domain. In working out these tasks, students were given clues during the process by combining different parts of the tasks posed. Mathan and Koedinger (2005) created a model of the 'intelligent novice'. When using this model, students first worked out problems without help. If they wanted to move on before having solved a problem correctly, they would be advised to accept help in finishing the first problem correctly. Users of this model learned faster and performed better on a conceptual understanding test and on a transfer test than a control group.

Schoenfeld (1992) is an important proponent of the approach to problem-solving where students take the initiative. He investigated expert and novice problem-solving behaviour. On the basis of his research he distinguished between five 'episodes' in the process of problem-solving:

- survey the problem (read, analyse)
- activate student's prior knowledge (explore)
- make a plan (plan)
- carry out the plan (implement)
- check the answer (verify)

Experts and novices differ in their approach to solving problems; novices almost immediately start to work out a poorly defined plan whereas experts take time to analyse the problem and gather information before making and implementing a plan. Schoenfeld (1992) argued that novices need to learn to work through the different episodes more effectively. He showed how to teach students the use of the episodes through questions and hints.

An important question is whether all students, regardless of their prior knowledge, can best be instructed in the way Schoenfeld suggested.

6.1.4 Student characteristics: gender and prior knowledge

Hyde, Fennema and Lamon (1990) and Orenstein (1995) showed that the disadvantage in science for female students has its onset at the approximate age of sixteen. This finding has been replicated and the disadvantage is still present in our secondary schools (Robinson & Gillibrand, 2004). The disadvantage may be ascribed to a decrease in confidence and academic risk-taking by girls as they get older. Female students often have a lower self concept of their skills in science than males, and females prefer support and interaction with others while males like to work independently (Sanders & Cotton-Nelson, 2004). That may be the reason why females may find it difficult to voice their ideas and opinions when they have to co-operate with males on a common task. Male students perceive themselves more as successful problem solvers while many female students perceive science as a masculine affair. In addition, teachers often think females have more difficulty in solving problems independently as females students prefer support and interaction with others compared to male students (Gallagher & Kaufman, 2005). It is assumed that mixed gender class interactions play an important role in the development of a gender gap. However, it is not clear if the negative effects of the classroom will also occur when females study physics problem-solving individually. In this situation, the interactions with male students do not come into play, though the general classroom atmosphere may still be in favour of males.

There is substantial evidence that learner expertise influences the effect of instruction in problem-solving. In a program-controlled environment, learning to solve problems with worked examples may be profitable to some extent for learners with little experience (novices), at least when new problems are similar to the problems students were trained to solve. This is in line with the findings we have discussed above. However, for students with more experience in problem-solving (and with more strategic knowledge), worked examples may be counterproductive. For instance, Kalyuaga et al. (2001) found that studying worked examples in advance hindered experienced students in solving problems. They call this 'the expertise reversal effect': if students know how to solve a certain type of problem, extra information about how to solve the problem will induce extra cognitive load in working memory and interfere with the students' solution processes. Reduced instructional guidance will be more beneficial when learners have more experience in solving problems in a certain domain (Renkl et al., 2002). If students have experience in a domain of knowledge it may be more profitable to let the students solve a problem by themselves and provide them with a worked example as feedback afterwards. This is also recognized by proponents of instruction by worked-out examples. In an explorative study, Reisslein, Atkinson, Seeling and Reisslein, (2006) compared three instruction conditions in learner-controlled environments: a) first example then problem, b) first problem then example, and c) fading of example during problem-solving. The students in each condition had free choice whether to study the examples or use the fading instructions. The results showed no differences in learning outcome or time on task between the conditions. As expected, students who started with more expertise in solving problems of the domain had higher scores on a problem-solving post-test than their counterparts who started with less expertise. The group with more expertise profited more from the 'first problem then example' condition than those with less expertise. In this way the worked example may function as feedback that asks the student to reflect on the episodes in solving the problem and strengthen strategic knowledge.

Moreno (2006) concludes from a review of recent research on worked examples that learning from worked examples is highly dependent on the quality of students' cognitive processing of the examples. Learner-controlled environments in particular are suited to learning to solve diverse problems and it is uncertain what role worked examples may play and how their effectiveness can be enhanced.

Our conclusion, in line with Schoenfeld's suggestion, is that in learning to solve problems strategic knowledge of how to structure a problem-solving process and how to regulate one's activities is very important. Environments that support students' free choice of just-in-time instruction during problem-solving and deliver worked examples afterwards may be more effective than environments with program-controlled instruction and worked examples. It needs to be investigated whether students with different prior expertise in solving problems in a domain will profit in the same way from just-in-time instruction and worked examples afterwards.

6.1.5 Sequencing problem-solving tasks and just-in-time instruction

One possible drawback of computer programs with student control is that students with lower prior expertise may not fully use the instruction available in the computer program. To overcome this drawback it is imperative that a computer contains ample problems of a given domain presented in a sequence of increasing level of difficulty. In fitting the level of the first tasks to the prior knowledge of the students, one should be conscious of not making the tasks, including help, too difficult. Joshua & Dupin (1991) showed that the teacher very often regards problems as being easy to solve, while the students feel unable to solve the same problems. On the other hand one should not make the level of the tasks too easy. Tasks are most effective for learning when complexity gradually increases. The hints during problem-solving need to be adjusted to task complexity. Problem-solving can only be learned in situations where students indirectly have all the information needed at their disposal, but are still challenged (see, for example, Van Heuvelen, 1991 a, b).

Anzai & Yokoyama (1984) and Maccini et al. (1999) indicated that students are semantically sensitive to hints. Students are able to use and comprehend one hint, while at the same time not understanding and thus ignoring another one. In giving a heuristic hint, one needs to clarify how to use this heuristic hint in solving the problem. For example, when telling the student to make a drawing one needs to give first-hand help in making the drawing for the problem concerned. The effect of the work of Van Gog (2006), in which the worked-out examples were supported with process-oriented information, should be seen in this light (see also Van Gog et al., 2006). In practice this means that hints should contain directions that not only provide a piece of content information but also information about how to use the content information.

6.1.6 Research questions

The research described above indicates that attempts have been made to combine various ways of supporting problem-solving by students. We expect student-controlled environments to be most effective for learning how to solve diverse problems. From this starting point two issues are important: at what time should support be provided and what kind of support should it be? It would be interesting to discover whether support during and after problem-solving is more effective than support only after problem-solving.

We expect that the timing of when instruction is provided will be a particularly relevant factor in improving students' problem-solving skills. The research described in this paper aims to discover whether hints during problem-solving in combination with worked examples afterwards are more effective than only worked examples after solving a problem. We assume that when students receive timely support with advice on how to apply content knowledge in different episodes of problem-solving (see Schoenfeld's episodes, above), the students will acquire relevant strategic knowledge. The combination of timely support and worked examples afterwards may further improve strategic knowledge, leading to better problem-solving. It needs to be investigated whether students with less prior expertise profit just as much from the combination of support as students with more prior expertise.

To test our expectations, two versions of a computer program (Pol et al., 2005) were developed: a version with hints during and worked examples after processing a problem, and a version with worked examples only after processing a problem. A third condition had students working on the same problems from a textbook with a manual that provided correct answers for the students to check. We intended to answer the following questions:

1. Does a computer program with hints during and after problem-solving improve students' problem-solving abilities more than a computer program with hints afterwards only?

Hypothesis 1: Students using the full version of the computer program will improve their problem-solving abilities more than students who have access to the version with worked examples only afterwards (see Mestre, 2002; Teong, 2003 for possible effects of strategic knowledge training).

2. Do either of the two versions of the computer programs improve students' problem-solving abilities more than the use of the traditional textbook with model answers?

Hypothesis 2: Students using either version of the computer program will improve their problem-solving abilities more than students using the textbook with model answers only (see Mathan & Koedinger, 2005, Van Heuvelen, 1991 a, b for possible effects of supportive instruction).

We did not postulate hypotheses about the effect of 'gender' or 'prior problem-solving expertise' on learning to solve problems. As stated before, very little is known about the possible disadvantage of females in learning to solve physics problems with help of a computer program. There is also little known about the different effects of a student-controlled learning environment on students with little prior expertise in problem-solving in a domain or more prior expertise. In learning to solve diverse problems, a well-designed student-controlled learning environment may be just as beneficial for students with little expertise as for students with more expertise. However, we know that in program-controlled environments, experienced learners are hampered in their learning if worked examples are offered before or during problem-solving (Moreno, 2006).

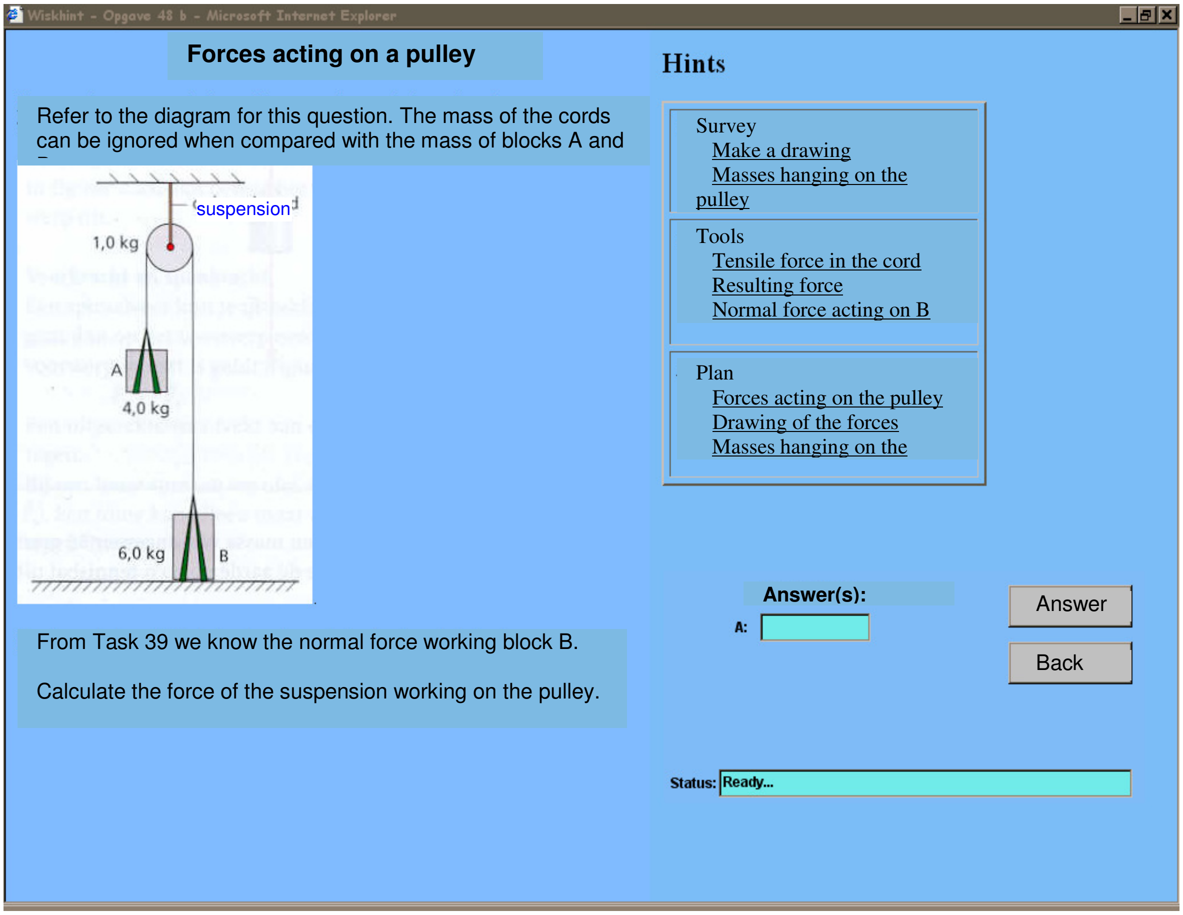
In our research we will take the possible effects of gender and difference in prior expertise into account when studying the effect of a new problem-solving program.

We also discuss the three types of knowledge involved in problem-solving: declarative knowledge, procedural knowledge and strategic knowledge (De Jong, 1986). It is possible that by learning to solve problems, students also improve their content knowledge. This may be the case if students have little content knowledge or their knowledge is not well organized. In our study students will first be instructed in the topic of forces and thereafter they will use a program to practise problem-solving. As students are already taught declarative and procedural knowledge, we do not expect them to increase this knowledge when using a problem-solving program. We only expect them to improve their strategic knowledge. In order to check this assumption we will include a content knowledge pre-test and post-test in our study.

6.2 The computer program

Pol et al. (2005) developed a computer program inspired by the five episodes of problem-solving Schoenfeld used to structure students' problem-solving. To allow both strong and weak problem solvers to profit from the computer program, the program entails 57 problem tasks, selected based on an increasing level of difficulty. Whether this set of problems is enough to ensure that both weak and already fairly strong problem solvers both profit from the use of the program will be evaluated by inspecting the relationship between problem-solving capacity prior to and after use of the program.

In solving a problem, students can choose additional instruction by clicking on a toolbar with short descriptions of different kinds of hints available. The idea is that when students are solving problems in a productive way, they are in fact running through Schoenfeld's episodes. Within each of the episodes, students may encounter blockages in the problem-solving process. By offering hints for each episode, using different solution methods the student can continue the solving process and finish. In Figure 6.1 we give an illustration of the computer screen as seen by the students. The problem can be found on the left of the screen. On the right a menu with hints is shown.



Refer to the diagram for this question. The mass of the cords can be ignored when compared with the mass of blocks A and

From Task 39 we know the normal force working block B.

Calculate the force of the suspension working on the pulley.

Figure 6.1: Task 40 from the computer program Physhint (originally from Middelink et al., 1998).

The hints are: Survey (Schoenfeld's episodes: 'read' and 'analyse'), Tools ('explore') and Plan ('plan'). After answering, students get the chance to check their solution and to reflect on it ('verify'). Students are given three opportunities to check an answer, during which time they can continuously consult the hints. The hints to be consulted for Task 40 from Figure 6.1 can be read in Figure 6.2.

Survey

Make a drawing

Hint: Copy the figure and draw the forces acting on the pulley.

Masses hanging on the pulley

Hint: Which masses are hanging on the cord? In choosing the method of calculation, take into account the normal force working on block B.

Tools

Tensile force in the cord

The tensile force of a cord works on the pulley at the place where the cord leaves the pulley.

Resulting force

Independently of the method of solution, when using forces you also need to apply Newton's first law, which states that if a resulting force acting on an object is 0 N, the object is not experiencing acceleration. This is the case in this situation.

Normal force acting on B

From Task 39 it is known that the normal force acting on B is 20 N.

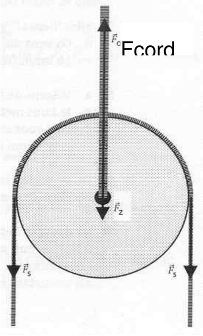
Plan

Forces acting on the pulley

Four forces are working on the pulley. The tensile forces of the cord are working at each place where the cord leaves the pulley. In addition, we have gravity working on the pulley, and in the opposite direction the force of suspension is also working on the pulley. Use Newton's first law to find the missing force.

Drawing the forces

Look at the drawing. The tensile force of the cord works on the pulley where the cord leaves the pulley. Do not forget the effect of gravity on the pulley.



Masses hanging on the cord

Instead of looking at the forces working on the pulley, you can also look at the masses hanging from the pulley. When you choose this way of working out the task, take into account the normal force acting on B and the mass of the pulley.

Figure 6.2: Hints for Task 40 using the episodes Survey, Tools and Plan.

Hints that accompany the first episode discuss the problem situation, mostly using informal methods such as a scheme, table or simple numerical calculation. The intention of these hints is not only to give help, but also to show the usefulness of these informal methods and to stimulate their use. On the other hand, in the computer program, hints will not only be common descriptions of a certain action, but will almost always be linked to the content knowledge needed for the task. Several researchers have emphasized the need to link strategic knowledge to declarative and procedural knowledge (Maccini et al., 1999; Wood & Wood, 1999).

In 'Tools' the student needs to choose which content knowledge is needed to solve the problem; the use of certain definitions is discussed under this menu. Declarative hints are especially important because in this phase of the problem-solving process students also need to be helped further so that they are directed away from an incorrect physical representation of the problem to a correct one. Not only are heuristic hints needed for this process, there also can be a simple lack of content knowledge.

'Plan' hints are focused on the different ways of solving the problem and on helping the student through the halfway stage of the solution process. If the help offered under the survey and tools episodes is not sufficient, a hint in this episode is called for. Students can very often choose from various solution methods, in this case there are three.

Using the computer program, students can give an answer up to three times, with the computer comparing their answer with the correct answer. After their last attempt, students receive a menu with descriptions of worked examples. They can consult one or more of these models. The different worked examples can cover informal solution methods (table, numerical calculation, etc.) or formal solution methods (formula, algebraic equation, etc.). The function of the worked examples is to support reflection on the solution process. Figure 6.3 gives possible formal solutions for Task 40.

Model

Forces acting on the pulley
$F_{S,TOT}$ = Total upward force due to the suspension
$F_{S,L}\downarrow$ is tensile force on the left
$F_{S,R}\downarrow$ is tensile force on the right; $F_{ZK}\downarrow$ is gravity of the pulley
The pulley is in equilibrium => Σ F = 0; $F_{S,TOT} = F_{S.L} + F_{S,R} + F_{Z,K}$
$F_{S,TOT}$ = 39 + 39 + 1.0 x 9,8 = 88 N
Drawing the forces
$F_{S,TOT} = F_{S.L} + F_{S,R} + F_{Z,K}$ (see drawing)
$F_{S,TOT}$ = 39 + 39 + 1.0 x 9.8 = 88 N
Masses hanging from the pulley
Hanging from the suspension point is a total mass
1.0 + 4.0 + 6.0 = 11 kg This produces a gravitational force of
$F_{Z,TOT}$ = 11 x 9.8 = 107.8 N On the ground it is 20 N ('resting')
(see the normal force of 20 N). The remaining force is
107.8 – 20 = 88 N working on the suspension.

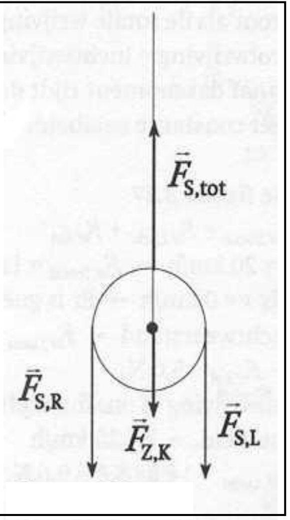


Figure 6.3: Model answers for Task 40.

6.2.1 Embedding the use of the computer program in the lessons about forces

The students participated in the experiment over 14 lessons of 45 minutes in which they were taught as a class, did practical work, were given demonstrations and worked independently. Figure 6.4 gives a schedule of the lessons taught. Teaching periods as well as practical work and demonstrations were the same for both experimental groups (using the computer program) and the control group (textbook), and were given to the group as a whole.

The tasks were done independently, with students from the control group staying in the classroom while students from the experimental group went to the computer room. The schedule in Figure 6.4 shows that instruction and individual practice in the three research groups were alternated during the whole sequence. Students from the control group undertook up to 57 tasks from the textbook. In the textbook version, difficult tasks were made easier by the addition of a total of 33 sub-questions, thus giving the students some direction in finishing the task. The tasks which could be undertaken by the students in the experimental groups were the same as the tasks of the control group; however, the sub-questions were left out of the sequence, making a total of 57 tasks available for the students of the experimental groups. All students of the experimental groups and the control group worked out the tasks at school as well as at home.

To check the answers and verify the solutions, a solution manual was available to the students in the control group.

Lesson	Content of the lesson and homework	Subject content
1	Instruction (whole lesson) Tasks: 1 – 2	Explanation of the principle of superposition - Vector versus scalar - Composing forces, goniometry and construction
2	Instruction (15 min.) Independent working Tasks: 3 – 14	Explanation of - Resolving forces in rectangular components - Resolving forces in other components
3	Independent working Tasks: 15 – 16	
4	Instruction and demo (15 min.) Independent working Tasks: 17 – 19	Explanation of - Newton's first law Demonstration: Cart on an air track
5	Instruction (5 min.) Independent working Tasks: 20 – 24	Explanation of - Newton's second law
6	Demonstration and Practical work (whole lesson)	Demonstration: Computer measurement table: Acceleration of a car by a small weight. Practical work: Carrying out measurements at the table in groups.
7	Independent working Tasks: 25 – 30	
8	Instruction (10 min.) Independent working Tasks: 31 – 38	Explanation of - Centre of gravity of an object and torque
9	Independent working Tasks: 39 – 45	
10	Working out tasks as a class (whole lesson)	4 Tasks about torques were presented. Students were given time to work out the tasks. Check and reflection within the group afterwards.
11	Instruction (10 min.) Independent working Tasks: 45 – 50	Explanation: - Check and reflection on problem 4 from lesson 10
12	Practical work (whole lesson)	Practical work: working in groups of 3 – 4 students on torque. Different situations to be calculated and measured.
13	Independent working Tasks: 51 – 57	
14	Instruction (whole lesson)	Explanation of - Fixed pulley versus loose pulley - Calculation of forces at a turning point

Figure 6.4: Schedule of the lessons.

6.3 Methodology

An experimental pre-test/post-test randomized group design was used to answer the research questions. The 15 to 16-year old participants (first year upper-level) in the experiment were taken from five fifth-year classes from four secondary schools in the Netherlands with average physics examination results. In each class, students were randomly assigned to one of the three groups:

- The 'during and after' group (DA) had support from the complete computer program: students were able to use hints, to use the control option of the computer program and, after answering, to compare their solution to the problem with the worked examples given by the computer program.
- The 'after' group (A) was given the same computer program, but students were restricted to a view of the tasks, the use of the control option of the computer program and the program's worked examples. The difference with the first group is an absence of hints for the survey, tools and plan episodes during the problem-solving process.
- The control group (C) received the tasks on paper, accompanied by 33 sub-questions, and students were able to check the answers and solutions with the help of the textbook and the solutions manual. This group's support consisted only of access to the model answers provided by the manual. Group (C) represents the usual practice in physics education classes.

6.3.1 Sample survey

The experiment started with a total of 78 students. A total of 19 students did not complete the experiment. The primary reason for this was students dropping out of school. In the Dutch secondary school system students must complete four terms which are each tested. After three terms, some students realize that they have no chance of passing, so they drop out and often start again the following academic year. Another reason for the experiment being incomplete in groups DA and A was computer malfunction at home. The total dropout rate was not significantly different for the three groups (Pearson's Chi-square = 0.16, p = 0.92). The total number of students participating in the experiment was: 18 in the DA group (10 boys, 8 girls), 18 in the A group (11 boys, 7 girls) and 23 in the control group (12 girls, 11 boys).

6.3.2 Procedure

The procedure followed consisted of a pre-test of problem-solving ability, and a domain-test on understanding the subject of forces, followed by the treatment and finally a post-test for problem-solving ability, plus a domain-test on understanding the subject of forces. A schedule for the procedure can be found in Figure 6.5.

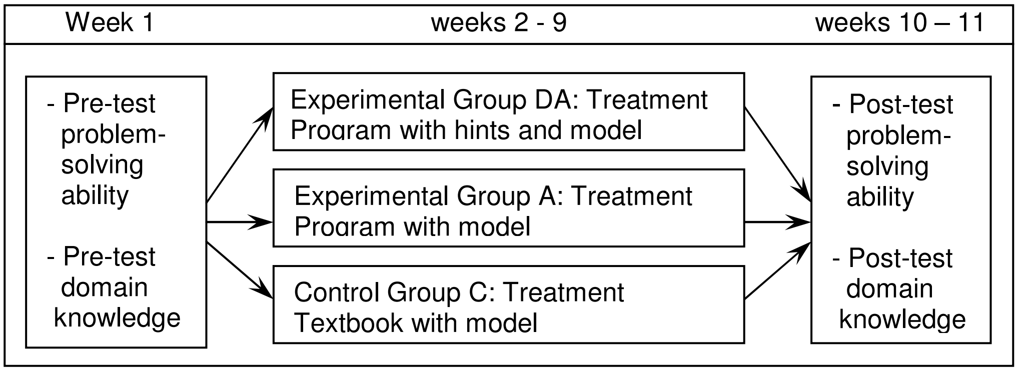


Figure 6.5: Schedule for the procedure of the experiment

All students took a pre-test, in addition to a post-test on applied problems. Their scores were used to analyse the differences between the three conditions vis-à-vis their ability to solve problems. Two content-knowledge tests were also taken before and at the end of the experiment.

6.3.3 Problem-solving tests

The problem-solving pre-test consisted of six applied problems on topics which had been taught during the previous two years. The problems were set in situations not previously encountered by the students. The subjects of the problem-solving pre-test were distance, velocity and acceleration. Examples of tasks assigned in the pre-test can be found in Figure 6.6.

Test item 4: Trains from Amsterdam to Paris

Two trains maintain the connection between Amsterdam and Paris. For both trains, the normal departure time is 7.30 a.m. from Paris and Amsterdam. The train from Paris is composed of four carriages, the train from Amsterdam of three. Due to electrical problems on the Dutch railways, the train from Amsterdam only departed at 8.30 a.m. The train from Paris departed at the normal time. The distance in a straight line between both cities is 475 km, but by rail it is 523 km. The Paris train travels at an average speed of 150 km/h, but because of a sleepy engine driver, the train from Amsterdam only travels at an average speed of 120 km/h.

Question: At what time do the trains meet?

Figure 6.6: Problem from the problem-solving pre-test.

When solving the problems in the pre-test, the students were explicitly asked to write down how they *analysed* the problem, came up with a solution *plan*, and how they *checked* their solution. The different episodes were graded, with a maximum of 2 points for analysis, 6 points for (working out) the plan and 2 points for the analysis respectively, resulting in a maximum of 60 points for the problem-solving pre-test.

The problem-solving post-test consisted of five applied tasks on the subjects of forces and torque. The tasks were set in situations not previously encountered by the students. An example of the tasks in the post-test can be found in Figure 6.7. As in the pre-test, the students were able to demonstrate their ability to fulfil the episodes: 'analyse', 'plan' and 'verify the solution'.

Test item 2: Go flat out

On a long straight road a car is moving at a constant speed. At a certain moment, the driver speeds up. The force of the motor as a function of time can be found in the first figure. Mass of the car, including driver, is 1.0 x 103 kg. The velocity of the car as a function of time can be found in the second figure.

Question: Calculate the total resistance experienced by the car at time t = 30 s.

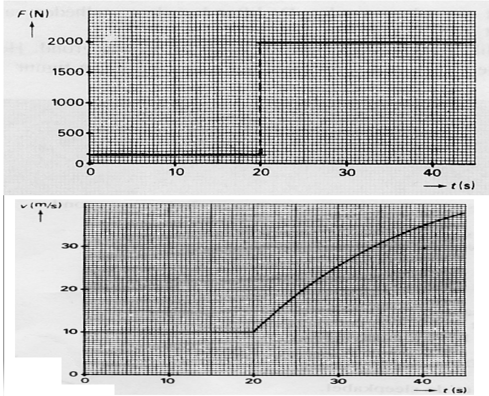


Figure 6.7: Item 2 from the problem-solving post-test.

Again, grading was done with a maximum of 2 points, 6 points and 2 points respectively, making a possible total of 50 points for the problem-solving post-test. An example of a scoring model can be found in Figure 6.8 for the problem in Figure 6.7. In this model, every possible item in the solving process has been scored with one or more points. Working out the tasks correctly gives a score of 10 points for every question.

A random check by a second observer was conducted, by which 20 tests were scored independently according to the scoring model. Comparison of the results of the different tasks gave correlations of between 0.60 and 0.95 for the different items, with an average of 0.82. The overall reliability was sufficient to use the scoring model for scoring both tests.

Correction model of Test item 2
Survey: From text: $m_{car} = 1.0x10^{3}$ kg; t = 30 s and question: $F_{friction}$?
- nothing: - 0 points; - 2 items, no question: - 1 point; question: - 1 point;
- complete: - 2 points

Plan: a = 37.5/30 = 1.25 m/s^2; $F_{res} = 1.25x1.0x10^{3} = 1.25x10^{3}$ N (2p); F_{motor} = 2000 N;
$F_{friction} = F_{motor} - F_{res} = 2000 – 1.25x10^{3} = 7.5x10^{2}$ N (2 p)
- wrong calculations of a (m/s^2): - 1 point; - correct a: - 2 points;
- calculation of F_{res}: - 2 points; - calculation of difference: - 2 points;
- correct remark about friction: - 2 points; - calc. F_{res} no a, state $F_{friction}$: - 2 points

Check and conclusion:
- check (for example back calculation): - 1 point; - conclusion: - 1 point

Figure 6.8: Scoring model for item 2 of the problem-solving post-test.

6.3.4 Knowledge-based tests

Both pre-test and post-test knowledge-base tests consisted of 20 items, the subjects of which were as follows:

- resultant and composite forces (vectors)
- Newton's first and second laws
- torque

The level of the tasks from the pre-test was taken from chapters used in lower classes. For examples of items of the knowledge-base pre-test see Figure 6.9.

Test item 1
Paul is cycling to school. His velocity does not change. Name three forces which are working on the bicycle.

Test item 11
Jet is hanging her case, with a mass of 75 g, on scales.
How many Newtons are measured by the scales?

Figure 6.9: Two problems from the knowledge-base pre-test.

The knowledge-base post-test had a comparable level to the tasks in the program. See Figure 6.10 for examples of problems from the post-test. For both tests, students could earn up to 3 points for every question, making a possible total of 60 points for the knowledge-base pre-test as well as the knowledge-base post-test.

Test item 2
Imagine two forces of 5 N and 3 N respectively. We want to compute the resulting force, not knowing the angle between both forces. What can you say about the resulting force?

Test item 21
An elevator moves down with a deceleration of 2.0 m/s^2. A man with a mass of 65 kg is standing on scales in the elevator. Compute the reading of the scales.

Figure 6.10: Two problems from the knowledge-base post-test.

6.3.5 Reliability and correlations of the tests

All tests showed a sufficient level of reliability ($\alpha = 0.70$ or higher). The correlation between the knowledge pre-test and post-test was 0.31 ($p < 0.05$), the correlation between the problem-solving pre-test and post-test was 0.40 ($p < 0.01$).

6.3.6 Treatment

The subject content that was supplied to all three groups was based on lessons about 'Forces', using the same chapter in the same textbook. Entering the upper level in Dutch secondary schools, all students who have chosen the science profile are offered the same subject matter, usually starting with mechanics. The topic 'Forces' is the second chapter to be covered. In the experiment, all students were offered the same tasks to assist them in processing the information from the chapter. These tasks were all taken from the textbook *Systematic Physics* (Middelink et al., 1998). In total, the students in the experimental group were offered 57 tasks. Students from the control group were also offered stepping-up tasks, which were left out for the experimental groups, thus making 80 tasks available for the students in the control group. Regarding the use of the program, the students from the experimental groups were given brief instructions on how to access the program. The students from the experimental groups were not given any special instructions about problem-solving.

During the project, data on how the tasks were worked out by the students was collected. Data about the use of the computer program by both experimental groups was also collected by checking a log file stored on the computer. All tasks, hints and solutions were located in different files in the computer program, which could be accessed by the students. All the files accessed and all the attempts to fill in an answer were stored in this log file. For the students in the control group, at the end of the project the teacher checked the number of tasks worked out by each student.

6.3.7 Analyses

Descriptive techniques were used to describe the data on the use of hints, and the results of the tests (content analysis on usage patterns of hints, frequency of distribution for the use of hints, averages and standard deviations of the test results of both research groups, etc.).

The effect of the different versions of the computer program on the problem-solving and content-knowledge post-test was tested using analysis of covariance. The independent variables were the three conditions, and the dependent variables were the students' results on the problem-solving post-tests and a domain-knowledge post-test. The co-variables were the two pre-test scores on solving applied problems and the domain knowledge.

6.4 Results

6.4.1 General use of the chapter by the different groups

Students in the different experimental conditions were given different kinds of help. One experimental group was given help during and after problem-solving (DA group). A second experimental group was given help after problem-solving (A group). For the students in the control group, no digital support was available. Students in this group only had access to their textbook and a model answer manual. The students in both experimental groups were given 57 tasks in the field of Forces. The students in the control group were given the same tasks on paper. For the students in the experimental groups, a log file was kept containing all the student's actions within *Physhint*. Data from the control group was gathered on forms filled in by the students and checked by the teacher. Below, a general impression of the use of the program and the material is given. As shown in Table 6.1, the average relative number of tasks worked through is comparable for all groups. An ANCOVA with the group as the fixed factor and the relative number of tasks carried out as the dependent variable showed no significant differences between the conditions (F = 0.15, p = 0.86).

Table 6.1: Overview of use performance students (in using Physhint)

Group Average in each group	During+After group (n=18) (SD)	After group (n=18) (SD)	Control group (n=23) (SD)
Tasks carried out (% of total tasks)	79.2 (22.5)	82.4 (18.9)	82.5 (20.5)
Correct with use of hints (% of tasks done)	37.5 (19.3)	-	-
Correct, no use of hints (% of tasks done)	42.0 (18.1)	81.9 (10.2)	68.7 (11.1)
Not correct, with use of hints (% of tasks done)	13.6 (6.3)	-	-
Not correct, no use of hints (% of tasks done)	6.9 (8.3)	18.1 (10.2)	31.3 (11.1)

To find out whether the relative number of tasks worked out correctly was related to the availability of the program, a second ANCOVA was carried out with the group as the fixed factor. The difference in relative number of tasks answered correctly seemed to be significant (F = 8.50, p < 0.001). This means that students working out tasks with the help of the program on average finished a higher number of tasks correctly than students from the control group.

6.4.2 Use of the program by the experimental groups

As can be seen in Table 6.2, the methods of help used by the students in the experimental group covers the whole range. Of interest is the percentage of 'use of model' for both groups, which is not significantly different (an ANCOVA between both groups, with both pre-tests as co-variates and the relative use of the model as the dependent factor gave F = 0.066; p = 0.799), despite students from the During/After group being able to discover many aspects of the tasks to be solved from the hints available.

The issue of whether the use of different kinds of help assisted students in answering more tasks correctly was investigated by means of partial correlations. In these partial correlations the percentage of correctly answered tasks was corrected for the scores on both pre-tests. Table 6.3 mainly shows the tendency for students to answer more tasks correctly when they use the different kinds of help. Only 'use of survey' contributed significantly to the percentage of correctly answered tasks. There is, however, little difference in the relationship between the different types of help and the percentage of correct answers, as all partial correlations lie around the level of significance.

Table 6.2: Implementation of the computer program and different episodes

	During/After group (n=18) (SD)	After group (n=18) (SD)
Number of tasks carried out	45.2 (12.8)	46.9 (10.8)
Correct (%)	78.2 (10.5)	81.9 (10.2)
Hints used during process (%)	51.1 (22.5)	-
Survey used (%)	35.8 (21.4)	-
Tools used (%)	36.2 (23.1)	-
Plan used (%)	38.1 (23.4)	-
Model used (%)	41.9 (29.7)	43.3 (27.6)
Total computer time (sec.)[1]	8364 (3312)	7986 (2975)
Average computer time per task (sec. per task)	183 (41)	168 (40)

[1] Total time-registered use during lessons. Due to a software error, registration of student activity did not start after logon, but after the first user action. As a result the total computer time registered is expected to be slightly higher for the D/A group.

Table 6.3: Partial correlations between percentage of tasks answered correctly and the use of the different kinds of help, corrected for both knowledge-base and problem-solving pre-test

	Survey, tools and/or plan used (%), n =18	Use of survey (%), n = 18	Use of tools (%), n =18	Use of plan (%), n = 18	Use of model (%), n = 18 (group DA)	Use of model (%), n = 18 (group A)
% correctly answered	0.35 (ns)	0.43 ($p < 0.05$)	0.40 (ns)	0.33 (ns)	0.35 (ns)	-0.05 (ns)

6.4.3 Students' test scores and differences between the groups

Before the hypotheses about the effectiveness of the two versions of the program are discussed below, we will first look at the average scores of the three research groups and those of male and female students on the pre- and post-tests. These scores, displayed in Table 6.4, indicate to what extent there may be a possible bias in group assignment and to which extent male and female students differ. In Table 6.4 we see that all groups scored about half of the 60 points available for the problem-solving pre-test, and about two-thirds of the 60 points available for the knowledge-base test. The average scores hardly differ between the three groups. The average scores for male students are slightly higher than the averages for female students. A multivariate analysis with both tests as dependent variables and group and gender as fixed factors revealed that there are no significant differences between groups nor between female and male students. We can therefore be confident that the three groups did not differ from each other before the start of the experiment.

Table 6.4: Test scores on pre-tests and post-tests

	Group			Gender	
	DA group	A group	Control group	Female	Male
Pre-test problem-solving (0–60)	32.8 (10.1)	32.3 (9.7)	32.4 (10.6)	30.9(9.8)	33.1(10.3)
Pre-test knowledge (0–60)	40.4 (5.9)	39.6 (9.1)	40.3 (7.8)	38.4(6.8)	41.6(8.0)
Post-test problem-solving (0–50)	30.7 (6.1)	26.1 (10.5)	21.9 (8.0)	25.2(9.8)	26.4(8.4)
Post-test knowledge (0–60)	30.6 (7.5)	32.4 (7.2)	31.7 (8.7)	27.5(7.0)	35.0(6.9)

As the students, under all conditions, had the possibility to check model answers and learn how to apply content knowledge to solve the tasks, we expected roughly the same amount of domain-knowledge learning in all three groups after the experiment. This seems to be the case. Table 6.4 shows the differences between groups in their average knowledge post-test scores to be small. The higher score of the A group can be attributed to the fact that the A group had relatively more boys than either of the other groups. As can be seen in Table 6.4, the average score for male students is much higher than for female students. Table 6.4 also shows that the three groups seem to differ in their average scores on the problem-solving post-test. Female and male students do not appear to differ on the problem-solving post-test.

Now we turn to the central questions of this paper. Do the research groups differ from each other in average student post-test scores and do the differences between conditions apply to both students with lower prior knowledge and higher prior knowledge in the same way? This second question is important because we expect both types of students to profit from the computer support. To answer these questions we first carried out a multivariate analysis of covariance. The aim of the multivariate analysis of covariance was to probe for the presence of group and gender effects on both post-tests. In this analysis the knowledge-base post-test and the problem-solving post-test were treated as the dependent variables, the experimental condition and gender as fixed factors and both pre-tests as co-variates. We simultaneously tested for the presence of interactions between the fixed factors and both co-variates. We needed this test for interactions to resolve the question whether group and gender differences in post-test scores actually vary between students with low and high pre-test scores. The multivariate analysis showed the interaction effects to be insignificant. This indicates the absence of significant differences in the relationships between pre-tests and post-tests between the three conditions as well as between female and male students. Both the gender effect (Wilks L=0.76; p=.001) and the group condition effect (Wilks L=0.74; p=.003) turned out to be significant. These effects apply to both students with lower prior knowledge and higher prior knowledge as indicated by their scores on the pre-tests.

Further univariate ANCOVA analyses of the gender effect revealed that the gender effect had only a significant effect on the knowledge-base post-test scores, but not on the problem-solving test scores. Male students significantly *outperformed* female students on the knowledge post-test. The experimental condition effect only applied to the problem-solving test. Further univariate analyses by means of ANCOVA with the experimental condition as the fixed factor, the problem-solving post-test as the independent factor and both pre-tests as co-variates proved to be significant (F = 6.19; p = 0.004), as expected.

To evaluate the two specific hypotheses of this paper we used a planned comparison test procedure and evaluated the corresponding t-tests at 5% level of significance. The results of the planned comparisons are displayed in Table 6.5.

Table 6.5: Estimates of the mean problem-solving post-test scores of the three research groups, with both pre-tests used as ‘covariates’

Group differences	Contrast estimate	Significance
During/After vs After	4.283	< 0.05
During/After vs Control	8.369	0.001
After vs Control	4.085	< 0.05

The statistical results of the planned comparisons confirm the expectations formulated in hypotheses 1 and 2. The mean of the DA group on the problem-solving post-test differs significantly from the means of the A group and the control group. The DA group shows nearly half a standard deviation gain compared to the A group. The difference for the experimental DA group compared with the A group is 30.7 – 26.1 / 10.5 = 0.44 standard deviation. This is a medium effect size. The measured gain for group A compared with the control group is 26.1 – 21.9 / 8.0 = 0.53 standard deviation. This is also a medium effect. These effect sizes can be interpreted in terms of the percent of no overlap of the treated group's scores with those of the untreated group or group with alternative treatment (Cohen, 1988). An effect size of .44 indicates a no overlap of 28% in the two group distributions. An effect size of .55 indicates a no overlap of 35% in the two group distributions.

6.5 Conclusion and Discussion

The first research question of this article was 'Do students using a computer program with hints during and worked examples after the solving of applied physics problems improve their problem-solving abilities as compared with students using the same computer program, but only receiving worked examples afterwards?' The idea is that by supporting students with hints listed according to the episodes of Schoenfeld (1992), students could choose to click on a hint if they became stuck during the problem-solving process. The hints could help the student to continue the problem-solving process and make it more efficient.

An experimental research design with two experimental groups and a control group was used to evaluate the effect of the computer program. In the experimental conditions, students were supported with worked examples afterwards (After group) or with hints during and worked examples after the problem-solving process (During/After group). The only difference between the two experimental conditions was the support by hints during the solving process.

Students of both experimental groups processed about the same number of tasks, and spent on average the same time solving tasks with the help of the computer program. The students of the During/After group used one or more hints with 51% of the tasks.

The analysis of the post-tests showed a significant difference between the two experimental conditions and the experimental and control condition on the post-test problem-solving, but not on the knowledge-base test. This outcome confirms our assumption stated in the Introduction of this paper that the program does not support growth in content knowledge more than the usual training of problem-solving with a textbook (the control condition). Although the hints provide content knowledge, the knowledge does not go beyond the subject matter taught before. The overall meaning of the hints is the organization of prior knowledge into problem-solving episodes (Pol et al., 2005). That is why extra strategic knowledge is gained but not extra content knowledge.

The analysis indicates there is a gender effect on the knowledge-base test. There is no interaction effect of gender x conditions on the knowledge-base post-test or the problem-solving post-test. In all three conditions females acquire less content knowledge than males. However, in the three conditions there is no gender difference in strategic knowledge. The gender difference seems to indicate that female students learn less well when they are in a class setting than in an individual learning setting. In a physics class the teacher usually expects males to take the initiative and solve problems and females often feel unsure about their knowledge and are more apt to ask questions and for clarification. This may cause the gender gap in physics in the higher grades of secondary education (Gallagher & Kaufman, 2005). Further study is needed to analyse which group processes and self-perception mechanisms hinder the development of females' knowledge in physics.

In line with our first hypothesis, we can conclude that students supported by hints during problem-solving and worked examples afterwards improve their problem-solving abilities more than students who only have worked examples available after problem-solving. In the During/After group there is a relationship between the number of hints used and the tasks answered correctly. Such a relationship does not occur in the After group between the use of worked examples and number of program tasks answered correctly. From this evidence we may conclude that providing model answers after the problem-solving process is especially effective when they are part of a total program of help available during and after the problem-solving process. Here one might expect an interaction effect as found by Reisslein et al. (2006). They found that students with more prior knowledge than the other students gained more from the worked examples after answering a problem. In our research no interaction effects of prior knowledge of strategic or content knowledge was found at all. Students with more prior expertise than average profited in the same measure from the program as students with less expertise.

Is there a relationship between the use of hints and the problem-solving abilities of students in the DA group? At first sight, one would expect to see a decrement in the use of hints as students become more acquainted with the problems in the program. However, there is no such fading effect. The effect of fading in the case of worked-out examples (Renkl et al., 2002) is based on the idea of learning to work out one type of task and gradually learning to use less help in solving such a task. In our program, heterogeneous tasks are offered and in order to carry them out different concepts of force and problem-solving methods are needed. We therefore do not expect a decrement in the use of hints during the use of the program. What can be expected is an increment in the systematic use of the hints according to the episodes of Schoenfeld, as described by Schoenfeld. Wood & Wood (1999) found that students who became better able to solve test problems also showed an increase in the systematic use of problem-solving episodes through practice. A further analysis of our data from the DA group could possibly demonstrate a significant difference between students showing an increment in the systematic use of hints compared with students not showing a growth in the systematic use of hints. If there is such a difference, then a relationship with problem-solving test scores could be explored.

The second research question was ‘Do students using one of the two versions of the computer program improve their problem-solving abilities more than students using the textbook with sub-questions and model answers only?’ This question concerns the influence of the level of tasks in the experimental groups versus the level of tasks in the control group on problem-solving abilities. From the literature we know that in using tasks to develop problem-solving abilities, the instructor should be wary of making the level of the tasks too easy (Van Heuvelen, 1991 a, b). To stimulate the use of hints in the program, the students were offered the same problems found in the textbook, but without the sub-questions designed to help the student to solve the problems, thus making the problems in the computer program less easy to solve. Looking at the number of tasks worked out correctly, there is a difference between the experimental DA and A groups versus the control group in favour of the experimental groups. Finally, the students in the DA and A groups outperformed the students in the control group on the problem-solving post-test, thus confirming our second hypothesis that offering tasks in a more complex way can positively affect the development of problem-solving abilities.

However, this second conclusion has to be put into perspective. The students in the experimental group were using the computer and students from the control group the textbook with model answers. Literature supports the idea that a difference in media (computer versus textbook) may cause a difference in test scores (Clark, 1994; Woodrow, 1998). In using the computer, students might work differently from those working with the textbook. For example, in using the computer program, students filled in their answers, received feedback and were required to try again if their answers were not correct. The students from the control group checked their solutions by comparing their answers with the answers at the back of their textbook. The feedback system provided by the computer could influence problem-solving abilities and cause a difference between the experimental groups and control group, thus demanding further research using the version of the program in which both hints and model answers are left out.

The first practical implication of these results is that it is possible to train problem-solving abilities without explicit guidance by the teacher. In the case of this experiment, students in the experimental group were supported digitally, with hints which helped them to work through problems with success. In this way students experienced a more efficient form of learning than that available through working independently using a textbook and the teacher as a possible background source of help.

The second point which can be made is that problem-solving can be supported by giving students appropriate tasks that provide them with problems requiring a degree of effort to solve. The experiment showed that the use of hints increases the capacity to solve the program tasks correctly and helps to improve problem-solving abilities to a greater extent than only providing worked-out examples after the solving process.