Computer based instructional support during physics problem solving
Pol, Hendrik

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2009

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
CHAPTER THREE

Design of the main study of Physhint

3.1 Research questions

The goal of this thesis is to evaluate a digital support system which encourages students who possess declarative and procedural knowledge in the domain of vectors and forces to learn how to solve new problems and which aims to improve their strategic knowledge. Answers are sought to the following questions:

1. Will the use of a computer program that supports students during and after the solving of physics problems in school practice improve students’ strategic knowledge?

2. In what way does such a computer program influence the development of strategic knowledge during students’ problem-solving practice?

In Chapter 1 and the preliminary research described in Chapter 2, the characteristics of a computer program were described that can support the development of strategic knowledge during and after problem-solving. The following chapters will describe two experiments involving the analysis of student use of such a program and the measurement of the program effects by means of a pre-post-test design with one (the first experiment) or two experimental groups (the second experiment) and a control group. These experiments will provide answers to both questions. This chapter discusses the design of the two main experiments on the effect of computer-supported problem-solving in physics education.

3.2 The Physhint program

The Physhint program was based on a computer-supported problem-solving program for mathematics called Wiskhint, developed by Harskamp and Suhre (2006). The program was successful in supporting student self-regulated learning of mathematics tasks. Like Wiskhint, the program Physhint guides students in learning a strategic way to approach tasks. The tasks, the hints provided to solve the tasks and the model answers in Physhint are based on the results of the preliminary study as described in Chapter 2. The results of this pilot study were
used to make a final selection of problems for the program and to modify and supply the hints that can be found in the *Physhint* program.

The program is a practical application of the theoretical outline given in Section 2.6. Extensive descriptions of the program, as well as samples of the program screen and menu are given in Sections 4.2, 5.2, 6.2 and 7.2.

### 3.3 Implementation of the computer program in lessons about forces

In order to follow common school practice, the problems in the *Physhint* program are taken from the chapter about forces and torque from a frequently used textbook in the Netherlands (Middelink et al., 1998).

In order to study the effectiveness of the new program, students participated in 14 lessons of 45 minutes in which they received classroom instruction, did practical work and worked independently. Teaching periods, practical work and demonstrations were the same for both experimental groups (using the computer program) and the control group (using the 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 groups went to the computer room. The implementation of *Physhint* thus incorporated blended learning: students received instruction from a teacher, which was the same for the different experimental and control groups, but there was also an individual instruction component, in which the experimental groups worked with *Physhint* and the control group with the textbook with model answers.

The instruction was about three topics within the domain of forces: vectors, Newton’s laws and torque. The following items were included:

- **Vectors**: this is a block of three lessons that starts with an introductory lesson. The principle of superposition is explained, and the difference between a scalar and a vector is discussed. Examples of calculations and constructions are worked through in class. Instruction is given on how to compose forces with the help of goniometry and construction. The resolution of forces into rectangular and other components is discussed and example problems are worked through. After these group activities, students have 60 minutes to work on problems independently.
• Newton’s laws: this is a block of four lessons giving an introduction to the concepts of Newton’s first and second laws; some example exercises are worked out in class and discussed. Demonstrations are given of a cart on an air track and of a computer measurement of an accelerating car. Practical group work is done by the students on how to measure and calculate accelerations. After these group activities, students have 160 minutes to work on problems independently.

• Torque: the last block of seven lessons starts with an introduction to the concepts of centre of gravity and torque. Conservation of torque is introduced and discussed by presenting four problems about torques in class. Students work out the problems in the lesson. Checking and reflection is done in the class under the guidance of the teacher. Practical work for this subject consists of practically working out problems on the subject of torque in groups of three to four students. This part ends with the discussion of the differences between a fixed pulley and a loose pulley and the calculation of forces at a turning point in a system. After these group activities, students have 140 minutes to work on problems independently.

A description of the pre-test of the suitability of the set of problems used in the lesson scheme was given in Chapter 2. The final version of the lesson plans used in the experiments can be found in the following chapters.

The lessons described here indicate that instruction by the teacher and guided practice always precedes individual practice on problem-solving in the series of lessons. Only individual practice is different for the experimental and control groups. In summary, the students from the control group and the experimental group worked on the same problem-solving tasks. The students in the control group took the tasks from the textbook and the students in the experimental group from Physhint. To check the answers and verify the solutions, students in the control group could consult a solution manual. For the students in the experimental group the program would fulfill this function.

3.4 Methodological choices

Two experiments were undertaken in which student use of the program was analysed, and the effects of the program were measured with a pre-post-test
randomized design, with one (the first experiment) or two experimental groups (the second experiment) and a control group.

3.4.1 Type and number of students

In both experiments, the program was implemented in secondary schools in the Netherlands: in the first experiment a school was selected in the north of the Netherlands with average results in the national examinations. The second experiment included four secondary schools with average national examination results. These were a ‘gymnasium’ and three regular secondary schools. In both experiments, students were attending pre-university education and were 15 to 16 years of age. Only mixed-gender groups were considered. Socioeconomic status (SES) was not explicitly measured, as no significant differences were expected between the schools and no relevant differences within classes.

Power analysis shows that in comparing groups with intermediate effect sizes (Cohen’s $d = 0.50$) and significance level $p < .05$, the number of students per group should be 25 or more (Faul, Erdfelder, Lang & Buchner, 2007). If the expected effect size is larger (e.g. differences between the means of the groups are larger) then fewer students per group are needed to find a significant effect. But group sizes are limited for practical reasons. The first experiment in 2003 was done in one school in order to pilot test the program and gain experience in its implementation in a school setting. The aim of the first experiment was to find indications of the effect of Physhint on student problem-solving and to gain ideas for improving the experimental setting so as to implement the program in the best possible manner in the final study. In the second experiment, students came from a number of schools. Convenience sampling was used for the selection of schools, teachers and students. Table 3.1 gives an overview of the spread of students over schools, teachers and gender.

In the first experiment, students ($n = 37$) came from two different groups at the same school and received physics lessons from two different teachers. Both teachers had received academic teacher training in physics. One teacher had seven years’ experience and the other fifteen. Students were 16 years of age and were from the pre-university stream. In the second experiment, students ($n = 59$) came from six different groups, with five different teachers and from four different schools.
Table 3.1: Background of students participating in the experiments

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Type of school</th>
<th>Situated in NL</th>
<th>Girls (N)</th>
<th>Boys (N)</th>
<th>Group A</th>
<th>Group C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Experiment 1</td>
<td>Pre-university</td>
<td></td>
<td>18</td>
<td>19</td>
<td>11</td>
<td>26</td>
</tr>
<tr>
<td>School 1 (Teacher 1)</td>
<td>Pre-university</td>
<td>North</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>School 1 (Teacher 2)</td>
<td>Pre-university</td>
<td>North</td>
<td>9</td>
<td>12</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>Overall experiment 2</td>
<td>Pre-university</td>
<td></td>
<td>27</td>
<td>32</td>
<td>18</td>
<td>23</td>
</tr>
<tr>
<td>School 1 (Teacher 1)</td>
<td>Pre-university</td>
<td>North</td>
<td>9</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>School 1 (Teacher 2)</td>
<td>Pre-university</td>
<td>North</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>School 2</td>
<td>Pre-university</td>
<td>East</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>School 3</td>
<td>Pre-university</td>
<td>West</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>School 4</td>
<td>Pre-university</td>
<td>North</td>
<td>8</td>
<td>10</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

Three schools incorporated both general and pre-university education, but one school was a gymnasium, which only included pre-university education (School 4 in Table 3.1). Students were 16 years old and all came from the pre-university level. The schools are situated in the north, east and west of the Netherlands. All teachers except one were experienced teachers who had received academic teacher training in physics. One teacher was still taking her university teacher degree during the experiment, but was supervised by an experienced teacher.

In order to reduce the effect between groups, students within groups were assigned to different experimental conditions. In this way, the influence of group composition or the style of a teacher in a group was spread among the research conditions.

3.4.2 Research conditions

All students in the classes involved received the same instructional materials and explanations from their teacher. Students within a class were only placed in research conditions during the part of the lesson where they worked on the problems independently. Differences among the different classes were reduced in four ways:

- A concise booklet gave the teacher instructions on how to present the subject knowledge, how to train the students in applying this knowledge in introductory exercises, when to plan laboratory work, and when to let students solve novel problems in order to learn to apply their domain
knowledge. Before the experiment started, the researcher met the teachers and computer network experts at the schools in order to eliminate in advance possible problems with testing, instruction and/or implementation of the Physhint program.

- Teachers handed in their personal lesson plans and the researcher checked whether the desired schedule of the experiment would be followed.
- The researchers regularly visited the lessons. Consultation by phone was also carried out.
- The computer logged the clicking behaviour of the students, giving the researcher insight into the use of the program by the students from the different schools.

Because the program was implemented as an experiment in school, some specific issues needed to be addressed and conditions fulfilled:

- In the experiment, there would be a comparison of students who were offered different instruction. Although precautions were taken to standardize the instruction of the different teachers as much as possible, one teacher’s instruction method could still be different from another’s. In order to minimize this effect, the different groups of students were taught in the same groups and by the same teachers. Each teacher thus taught students from the different experimental and control groups.
- All students, not just volunteers, were included in the experiment. Thus the instruction matches regular education.
- Experimental conditions would be compared with regular education. The only difference between the experimental condition and regular school practice is that the former has the experimental instruction as mentioned above.
- Different textbooks are used in the Netherlands. In order to compare students from different schools, the textbooks used by the different students are the same.

3.4.3 Design of the experiments

The aim of the first experiment is to find out if the computer program with hints during problem-solving and model answers afterwards is more effective in teaching students to solve problems than the traditional way of practising
problem-solving using problems from the textbook and model answers (worked out solutions to a problem) from a solution manual. We also wish to find out whether the program proves effective in increasing strategic knowledge and whether this correlates with how students use the hints and model answers in the computer program.

There are two research conditions in this first experiment:

- An experimental group (n = 11) with hints ‘during and after’ (DA group). In this condition, students solve problems with help from the computer program. Help consists of hints during problem-solving and model answers after problem-solving.
- A control group (n = 26) in which students receive the problems on paper and can check the answers and solutions with the help of the solutions manual (C group). This group represents usual practice in physics classes in the Netherlands.

The second experiment tests the surplus value of hints during problem-solving. If an extra effect can be found, it would be interesting to discover whether the way the students use hints in the program might be related to successfully finishing problems, and how the use of model answers can improve the solution process while using the program. There were three research conditions in this second experiment:

- The ‘during and after’ group (DA group) (n = 18) received support from the complete computer program. Help consists of hints during and model answers after problem-solving. Students are able to use hints, to check whether their answers are correct and, after answering, to compare their solution with the worked examples given by the computer program.
- The ‘after’ group (A group) (n = 18) used the same computer program, but students only have the option of checking their answers and consulting model answers (this option is only available in the second experiment). Students thus do not have hints available during problem-solving.
- The control group (C group) (n = 23) received the problems on paper and students were able to check the answers and solutions with the help of the textbook and the solutions manual.
3.4.4 Data collection

In both experiments, we collected information about student subject knowledge and problem-solving skills by giving two pre-tests to the students followed by the treatment, and finally two post-tests measuring the same categories. A schedule of the procedures in the experiments is found in Figure 3.1.

During the program, the use of episodes and hints by the students in both experimental groups was registered by means of log files. In addition, the completed answers were logged by the program. The program also recorded all steps made by the students over time. For the students in the control group, the teacher checked the number of problems worked out and kept a record of the problems solved correctly. The students kept a record of the time they spent on the problems and handed this to the teacher.

3.5 Performance tests

Declarative and procedural knowledge as well as strategic knowledge was pre- and post-tested. Chapter 1 showed that routine problems can be solved by using declarative and procedural knowledge, with ‘routine’ meaning that the same kind of problem has already been solved and students are thus able to solve them. This is often the case with problems presented in textbooks (see Chapter 2). Strategic knowledge is necessary for solving more complicated and unknown problems. To
solve complicated problems, students not only need strategic knowledge, but also declarative and procedural knowledge in order to analyse a problem and create a plan to solve it. Having students work out complicated problems, tests the combination of applying declarative and procedural knowledge, as well as strategic knowledge. In this thesis, measuring strategic knowledge means the application of strategic knowledge during problem-solving. This was measured by evaluating students’ written reports on the use of episodes in the problem-solving, pre and post-test (Veenman, 2005).

In order to test the students’ specific strategic knowledge, their declarative and procedural knowledge of the topic under consideration were also tested. Four tests were used in the experiments. The first two – the knowledge pre and post-test – mainly tested declarative and procedural knowledge of the topic, while the last two – the problem-solving pre and post-test – tested strategic knowledge. For full versions of the tests see Appendices A to D.

Figure 3.2 presents examples of problems which test the declarative and procedural knowledge of the students.

<table>
<thead>
<tr>
<th>Problem 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>In the picture, you see a plane with two forces working on it. The small force has a magnitude of 3 Newton, the large one 5 Newton. How big is the total of both forces?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem 16</th>
</tr>
</thead>
<tbody>
<tr>
<td>A spaceship is flying through space, without propulsion. The spaceship does not feel any influence from planets or stars. What is happening to the velocity of the spaceship? Explain why.</td>
</tr>
</tbody>
</table>

Figure 3.2: Items 6 and 16 in the knowledge pre-test

The description of the declarative and procedural knowledge needed to solve these problems can be found in Figure 2.3, which presents a concept map of vectors and forces. To solve Problem 6, students need to use their knowledge about adding two vectors, and the mathematics knowledge involved (Items 1, 2 and/or 3 in the concept map). To solve Problem 16, students need to use Newton’s first law of forces (Item 4 in the concept map). The subject-specific knowledge of vectors and forces was tested in the knowledge test.

The strategic knowledge pre and post-test presented students with a problem requiring the above declarative and procedural knowledge, but also
strategic knowledge. Students were expected to be familiar with the declarative and procedural knowledge needed to solve the problem, but the problem was dissimilar to the problems practised before and needed more solution steps to solve. Students had to apply strategic knowledge about tackling problems (what steps to take and how to check progress) in order to solve the problem. An example of a problem in the strategic knowledge test requiring this declarative and procedural knowledge is given in Figure 3.3.

![Figure 3.3: Problem 1 in the problem-solving post-test (Pol, Harskamp & Suhre, 2005)](image)

While solving this problem, students need strategic knowledge in order to have an overview of the different steps needed to solve the problem and metacognitive knowledge when setting up and working out a solution plan. The students also need subject-specific declarative and procedural knowledge of vectors and forces, as described earlier in this section. This test thus incorporated all knowledge needed to solve problems. From now on, this test will be called the problem-solving pre and post-test.

In drawing conclusions about the strategic knowledge of students, we need to look at the scores on the problem-solving tests. However, because these scores are influenced by the students’ declarative and procedural knowledge, the scores on these aspects of knowledge are taken into account when analysing the scores on the problem-solving tests.

### 3.5.1 Knowledge tests

Both the knowledge pre-test and post-test consist of 20 items, the subjects of which are resultant and composite forces (vectors), Newton’s first and second laws, and torque. The pre and post-test were at the expected level of the students and therefore differ in the cognitive level of the tasks. The task level for the pre-test was
taken from chapters used in earlier classes. Examples of items used in the knowledge pre-test can be found in Figure 6.9. Scoring the knowledge pre-test was done by giving a point to every correct aspect needed to solve the problem. The knowledge post-test has a comparable level to the tasks in the program. Examples of items in the knowledge post-test can be found in Figures 4.4 and 6.10. As with the pre-test, students received a point for certain steps in the solution process in the post-test. For both tests, students could earn up to three points for every question, making a possible total of 60 points for the knowledge pre-test and post-test. In the first experiment, the internal consistency of the knowledge post-test as shown by Cronbach’s alpha is 0.56. The low level of internal consistency was taken into account, as can be seen in Chapter 4. In the second experiment, both the knowledge pre-test and post-test show a sufficient level of reliability (α = 0.70 or higher). In the second experiment, the correlation between the knowledge pre-test and post-test was 0.31 (p < 0.05). A full version of the knowledge pre-test is given in Appendix A while a full version of the knowledge post-test is given in Appendix B.

3.5.2 Problem-solving tests
Strategic knowledge was tested by the problem-solving pre and post-test. 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 these tests can be found in Figures 4.2, 5.4 and 6.6. When solving the problems in the problem-solving pre-test, the students were explicitly asked to write down how they analysed the problem, came up with a solution plan and checked their solution. An example of a scoring model for problems in the problem-solving test is given in Figure 6.8. As we see in this model, the different episodes are graded, with a maximum of 2 points for analysis, 6 points for the plan (or working out the plan) and 2 points for the analysis, 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. Examples of problems in the problem-solving post-test can be found in Figures 4.3, 5.5 and 6.7. The scoring model is given in Figure 6.8. As in the
problem-solving pre-test, the students were able to demonstrate their ability to fulfil the episodes ‘analyse’, ‘plan’ and ‘verify the solution’. Again, grading was done with a maximum of 2, 6 and 2 points for the respective parts of the test, making a possible total of 50 points for this post-test. 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 to score both tests (α = 0.70 or higher). In both experiments, the correlation between the problem-solving pre-test and post-test was significant (p < 0.01). A full version of the problem-solving pre-test is given in Appendix C while a full version of the problem-solving post-test is given in Appendix D.

3.6 Analyses

This study was based on the results of two experiments that analysed the effect of computer-supported training. It was expected that students who used the program to do better on a problem-solving test afterwards. However, it was also expected that students would use hints in a more systematic way while working with the program, which would account for at least part of the effect. The results of the two experiments on problem-solving will be discussed in Chapters 4 and 6. The relationship between different aspects of the problem-solving process, use of hints in the program and the scores on the problem-solving post-test will then be discussed in Chapters 5 and 7.

Although the students had lessons in groups, these groups were divided up and individuals assigned to a research condition. The treatment in the research conditions was at the level of individual (students working individually on Physhint or their textbook). Therefore, analysis of the effect of the research conditions was done on an individual basis. Descriptive techniques were used to describe the data on the use of hints and the results of the tests for the different groups in the experiments. To test the differences in problem-solving post-test scores between the experimental and control groups, analysis of covariance was applied with pre-test scores as the covariate (Cohen, 1988; Van Peet, Van den Wittenboer & Hox, 2000).

In order to find possible developments in the systematic use of program hints, student use of the program in the first part of the project was compared with
their use in the last part of the project. This comparison made it possible to show if there was a change in the way hints were used in the experimental groups. Furthermore, possible relationships within the experimental groups, between the use of hints, the number of correctly solved problems in *Physhint*, and scores on the post-test, were studied using correlational analysis (partial correlations), taking into account students’ pre-test scores. More explicit descriptions of the analyses used are given in the following chapters.