Original article

The effects of discovery learning and expository instruction on the acquisition of definitional and intuitive knowledge

Janine Swaak*, **Ton de Jong**† & **Wouter R. van Joolingen**‡
*Telematica Instituut, AN Enschede, The Netherlands
†Faculty of Behavioral Sciences, University of Twente, Enschede, The Netherlands
‡Graduate School of Teaching and Learning, University of Amsterdam, Amsterdam, The Netherlands

Abstract

Types of learning with a strong emphasis on the responsibility of the learner (such as discovery learning) are gaining popularity over traditional forms of (expository) instruction. Discovery learning distinguishes itself by the central role of learning processes such as hypothesis generation (induction), experiment design, and data interpretation. Expository instruction pays more attention to directly 'exposing' definitions and equations to learners. In the current study, students worked with either a simulation (discovery learning) or a hypertext learning environment (expository instruction) with the same domain content. Each of the environments contained a large number of assignments. The study followed a pre-test, post-test design. To measure the knowledge acquired a definitional knowledge test, an intuitive knowledge test (where both correctness and speed of answering are aspects that are measured) and a test in which relations needed to be explained were administered. It was predicted that the hypertext group would outperform the simulation group on the definitional knowledge test and it was expected that the simulation group would perform better on the intuitive knowledge test. Results showed that both the interaction with the simulation and with the hypertext resulted in substantial learning gains. It was found that the hypertext group performed better on the definitional knowledge test. On the intuitive knowledge test the hypertext scored better than the simulation group on the correctness of the items but not on the time needed to answer items. On the explanation test there was no difference between the two groups. An analysis of interaction processes as recorded in the logfiles indicates that the differences between both environments in their actual usage were less distinctive than expected. In the simulation group many students followed the assignments given and did not engage in self-guided discovery. Since the assignments were rather directive, this resulted in 'discovery behaviour' that focused on generating outcomes; outcomes that were also, and more directly presented in the hypertext environment. For research and practice, this implies that simulations are to be considered only when clear benefits of discovery are expected, and only with complex domains, sufficient learning time and freedom for students in the assignments to engage in discovery.

Keywords assessment, comparative study, discovery learning, simulations

Accepted: 27 May 2004
Correspondence: Janine Swaak, Telematica Instituut, PO Box 589, 7500 AN Enschede, The Netherlands.
Email: janine.swaak@telin.nl

Introduction

The current study compares discovery learning with learning from expository instruction. Discovery learning and expository instruction are based on dissimilar models of the learner. Expository teaching creates a fairly passive role for learners who are expected to receive information and reproduce them at some point (see e.g. Jones *et al*. 1990; Cunningham 1991; Jonassen 1991). In contrast, discovery learning assumes that learners takes an active role and construct their own knowledge base. Besides emphasizing a passive or active role of the learner, discovery learning can be distinguished from expository instruction in the type of learning processes. For expository learning (e.g. learning from text) learners (especially the poor ones) rely, to a certain extent, on 'superficial processes' such as reading and memorizing (Ferguson-Hessler & de Jong 1990). For discovery to be meaningful, the processes that make up the empirical cycle (see e.g. de Groot 1969) should take place. By using processes like collecting and classifying information, stating hypotheses, making predictions, and interpreting outputs of experiments, learners infer knowledge from the information given. This is essential, as a coherent knowledge base is not directly available in 'discovery situations' and knowledge is to be inferred (see e.g. Shute *et al*. 1989).

A few studies investigated the differences between discovery and expository strategies in teaching and learning. Shute (1990) made a comparison between inductive and deductive simulation-based learning environments (on basic principles of electricity). The difference between the environments was the nature of the feedback to the students. In the deductive environment, the feedback to assignments included the rule to solve an assignment, whereas in the inductive environment, not the complete rule, but only the variables of the rule were given in the feedback. In the study, students worked for an extended amount of time (i.e. 45 h) with one of the two learning environments. Four post-test measures were applied, assessing, respectively, declarative knowledge of the different components and devices of electrical circuits, qualitative understanding of laws involved, application of laws, and learners' ability to generalise knowledge and skills. Main effects of the deductive-inductive distinction were not found on any of the four tests.

Rieber and Parmley (1995) compared a tutorial on the principles of the simulated physics domain with practice in either a structured or unstructured simulation. The tutorial was highly structured and 'designed to exclude as much formal mathematics as possible and instead concentrated on concept formation and application' (p. 363). The practice in the structured simulation consisted of four assignments in which participants were given more and more control over a simulated free-floating object, and in which concepts of motion were step-by-step introduced. In the unstructured simulation, learners were given a number of trials in flying a free-floating object to a space station. Participants had full control over the floating object and no concepts of motion were introduced. Rieber and Parmley found that learners who worked with the structured simulation without tutorial performed not significantly different (mean 74.5% correct) on a multiple choice post-test consisting of 'rule-using' items, from learners who received only the tutorial (mean 85.1% correct). Learners who worked with the unstructured simulation without the tutorial performed significantly worse (mean 67.6%).

In the study by Shute, the difference between the two learning environments tested was only found in the feedback to students; in both environments students could manipulate the simulation and thus used a discovery approach. The study by Rieber and Parmley (1995) indicates that differences between an expository approach (of the tutorial) and a discovery approach (of the simulation) are complex. First, they confirmed that pure simulations are not very fruitful learning environments (see also de Jong & van Joolingen 1998) and, second, they found that adding a simulation to a tutorial obviously does not increase knowledge acquisitions. Both studies emphasised that the way of measuring learning results is pivotal (see also Swaak & de Jong 1996; 2001). A more direct comparison of a discovery and an expository environment using different ways of measuring knowledge could help to gain a better insight into the advantages and disadvantages of both approaches.

Zacharia and Anderson (2003) compare expository instruction with simulation as a preparation for practice. They find that the simulation contributes to *conceptual change* in physics concepts, as measured on a standard paper and pencil test. They do not specify the precise nature of the simulations, nor the kind of instructional support that was offered, so it is not clear to what aspect of the simulation these differences can be attributed.

In the present study, and analogue to the Rieber and Parmley study, a discovery learning environment was implemented in the form of a simulation-based learning environment and an expository environment was realised as a hypertext-based learning environment. The content in both environments was the same. Learners in the simulation condition were expected to acquire knowledge with an intuitive quality, as measured by an intuitive knowledge test. In contrast, learners in the hypertext condition should score higher on the mere recognition of factual knowledge as measured by an explicit knowledge test. Apart from investigating the test results, the interaction processes, as recorded in the logfiles, were examined.

Method

Participants

One hundred and twelve participants participated in the study. The participants came from high school (pre-scientific education), were taught physics, and had sufficient computer experience. Their age was 16–17 years. The participants were randomly assigned to one of the two experimental conditions, one participant ended up in the wrong room, and as a result the simulation condition contained 57 participants and the hypertext condition contained 55 participants. Participants took part in the study on a voluntary basis.

Learning environments

The learning environments of the current study covered the physics topic of head-on elastic and inelastic collisions. The simulation environment contained simulation models of moving and colliding particles. Participants could control a number of input variables (mass and initial velocity of the particles) and watch the behaviour of the particles as it was expressed in graphs (displaying velocity, position, and kinetic energy of the particle(s)), in numerical output, and in an animation of the system.

Figure 1 displays an exemplary interface in the environment. The hypertext environment contained static graphical displays and static pictures of moving and colliding particles. In this condition it was not possible for the participants to manipulate the input variables. The simulation and hypertext environments contained the same information and the same support measures. These were model progression, assignments, feedback to the assignments, and explanations with equations.

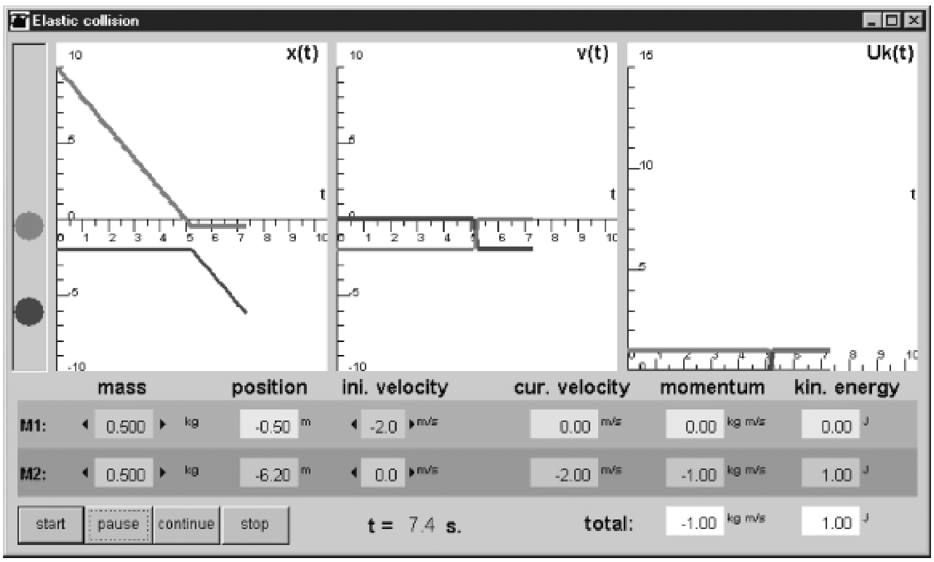

Fig. 1 An exemplary simulation interface of the simulation environment, taken from the third level of model progression, perfect elastic collisions (translated from Dutch). The graphs of the hypertext environment looked the same, only no interaction was possible.

Model progression

In the simulation, four levels of model progression were present: non-accelerated moving particles (Level 1), perfect elastic collisions with a fixed wall (Level 2), perfect elastic collisions of two particles in one dimension (Level 3), and perfect inelastic collisions of two particles in one dimension (Level 4). The number and kind of input variables that could be controlled and the output variables that could be observed varied across the levels. At the simplest level, the learner could control only two input variables, and two output variables were displayed. At Levels 3 and 4, the learner could control four input variables, and at these levels, 10 output variables were displayed. The hypertext contained the same levels of model progression and precisely the same number and kind of variables were included in the levels. However, within the hypertext environment no input variables could be controlled and only examples of values for output variables were displayed in the static graphical and numerical outputs. For every level of model progression in the hypertext environment, one page with a triplet of static graphs, including the numerical outputs, was included (similar to the simulation, see Fig. 1).

Assignments

At each level of model progression, a number of assignments guided the learner in the exploration of the domain covered at that specific level. These assignments had a multiple-choice format and prompted the learner to start an inquiry on the relationship between two given variables. The contents of the assignments were identical for the hypertext and simulation (i.e. they concerned the same variables of the domain). The phrasing, however, was different. In the simulation environment, learners were advised to perform experiments with the simulation. More precisely, it was suggested to manipulate the variables included in the assignments and carefully look at the output displayed. Assignments in the simulation environment emphasised reading and understanding graphs, including matching graphs and animations. Also, some of the assignments contained 'real-life' investigations, for example, asking the students to make one particle a beach ball and the other a bowling ball and to compare the simulated momentum of the two balls. In contrast, the assignments in the hypertext environment encouraged learners to have a close look at the equations included in the environment and/or to study the static graphs. In both the simulation and hypertext environment a number of explanations contained equations or background information on reading the graphs. At the first model progression level, seven assignments were available for the learner. Eight assignments were present at Level 2, 14 at Level 3, and the fourth level included 11 assignments.

Feedback

Feedback explanations appeared as feedback on each response alternative of all assignments. As with the assignments, the contents of the feedback explanations were identical for the hypertext and simulation environment, the nature of the feedback differed for the two environments. Examples from both environments are displayed below.

An example of feedback in the simulation:

> This is not the good response. Please go back to the simulation. Give the two balls similar masses. Give one of the balls no initial velocity, this is $0\,\mathrm{m\,s^{-1}}$, and the other an initial velocity >0 m/s. Is the end velocity of the two balls higher than the initial velocity of the one ball?'

An example of feedback in the hypertext:

> 'This is not the good response. Please go back to the page with the graphs. Also, pay attention to the values underneath the graphs. What was the initial velocity of the balls? What is the end velocity? Is the end velocity of the two balls higher than the initial velocity of the one ball?

'Non-obligatory' and directive

In the learning environments, at any point in time, the learner could choose (and was free to do so) to examine the set of assignments or equations, manipulate the simulation (simulation), or study static graphs (hypertext). No constraints were included for learners to go to a new level, and once at a more complex level, there were no restrictions for the participants to go back to the simpler models. In other words, the instructional measures were used in a 'non-obligatory' fashion in both the simulation and the hypertext environment. However, in both the simulation and hypertext, the way the assignments were phrased and the inclusion of feedback to their responses can be characterised as 'directive' (i.e. learners were steered in certain directions).

Assessments

To assess the participants' knowledge three tests were used, a definitional knowledge test, an intuitive knowledge test – called the WHAT-IF test – and a test in which learners' ability to explain – called the WHAT-IF-WHY test – was tested. Furthermore, all of the actions participants made while interacting with the environments were registered.

Definitional knowledge test

The definitional knowledge test consisted of multiple choice items with three response alternatives and aimed to measure recognition of conceptual knowledge of facts, definitions and equations. As an example, one question asked the students for the formula of the total momentum after an elastic collision. The definitional test was given both as pre- and post-test. When answering, test learners were allowed to go back to previous responded items. The definitional knowledge test consisted of 20 items.

WHAT-IF test

To measure intuitive knowledge about the relationships between the variables of the domain, a so-called 'WHAT-IF tests' were created. Two parallel versions of the WHAT-IF test were developed. Each version consisted of 24 three-choice questions (nine out of the 24 items were identical in both versions). The versions were created in such a way that a one-to-one mapping existed between the WHAT-IF pre-test and the WHAT-IF post-test items. Parallel items covered the same content and had similar difficulty but differed on the direction or size of the induced change in the items. Participants were instructed to respond as accurately and quickly as possible. In determining the level of intuitive knowledge both correctness and response time were used. Two examples of WHAT-IF items are depicted in Fig. 2.

WHAT-IF-WHY test

Like the WHAT-IF test items, the so called WHAT-IF-WHY test' items required the learners, in the first part of the response, to decide which of the three predicted situations followed from a given condition, given the

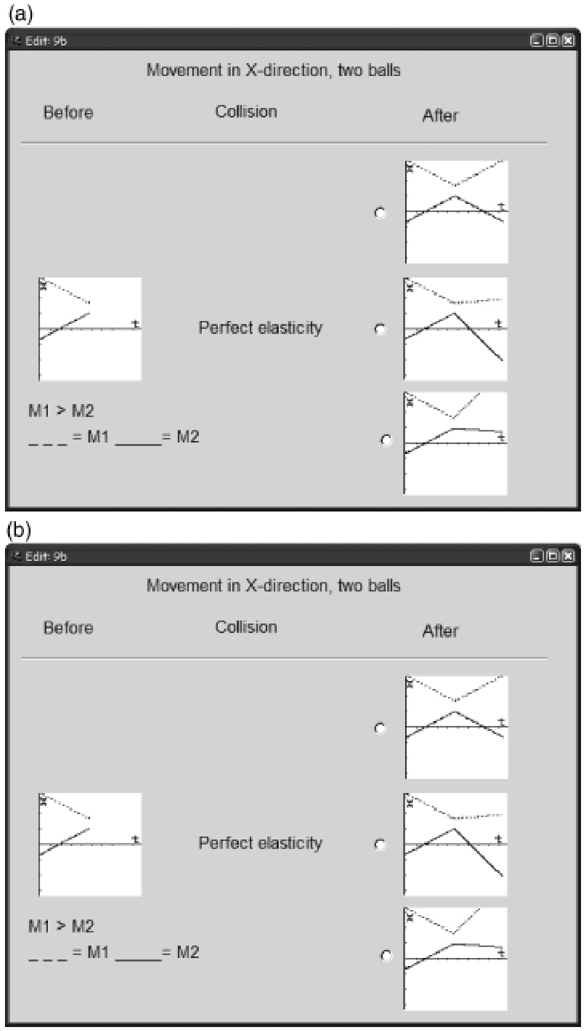

Fig. 2 Two exemplary WHAT-IF items used in the current study (translated from Dutch).

action that was displayed. For this purpose, a sub-test of 13 items from the WHAT-IF post-test was used. In the second part of the response, the learner was asked to explain in his or her own words why the action resulted in the predicted situation. The WHAT-IF-WHY test was presented as post-test only.

Process measures

All of the actions participants made while interacting with the environments were registered. This provided us with data on the use of the simulation environment and hypertext environment, including the assignments, feedback to responses to assignments, and equations that were present. These data were used to make a comparison between the simulation and hypertext condition, and also to relate specific interaction patterns within an experimental condition and with post-test outcomes.

Procedure

Each experimental session lasted approximately 3 h. It consisted of a 30 min pre-test, a 10-min introduction to the simulation or hypertext environment, 90 min working with the experimental environment, and 45 min of post-test. First, a short questionnaire was administered in which the students had to report their view on their own knowledge (results are not included in this article), followed by the students' completed definitional knowledge test, the WHAT-IF test, and finally the WHAT-IF-WHY sub-test.

Results

This section first reports the results on the different knowledge tests and then an account is given of the interaction measures.

The definitional knowledge test

The definitional knowledge test was given in the same form as pre- and post-test. It consisted of multiple choice items, each with three response alternatives. Due to technical issues, the pre-test data for two participants (in the simulation condition) were lost.

The upper part of Table 1 gives the number of correct items for the definitional pre- and post-tests for the two experimental conditions averaged across participants. To examine whether the students improved overall on the definitional test, a repeated measurement analysis was performed. The repeated measurement analysis on the definitional test-scores showed a significant within-participant effect for the number of correct items ($F_{1,108} = 224.99$, $P < 0.001$). The effect size d for the increase in correctness scores yielded a value of 1.40 SD's. Moreover, a variance analysis on the pre–post test differences score revealed that the hypertext group gained more than the simulation group ($F_{1,108} = 5.67$, $P < 0.05$).

The WHAT-IF test

For the WHAT-IF test, items are scored on both the correctness of the response and on the time it took to give the response.

The average number of correct items and the average time to respond to WHAT-IF items is given in the middle section of Table 1.

To examine whether the students improved overall on the WHAT-IF test, repeated measurement analyses were performed. A repeated measurement analysis on theWHAT-IF test scores showed a significant within-participant effect for the number of correct items ($F_{1,110} = 166.74$, $P < 0.001$). The effect size d for the increase in correctness scores yielded a value of 1.31 SD's. Moreover, an effect of experimental condition on the pre–post test differences scores was found ($F_{1,110} = 7.92$, $P < 0.01$) indicating that participants in the hypertext condition improved more than participants in the simulation condition.

Table 1. Means and standard deviations for knowledge measures

Knowledge tests	Condition		
	Simulation	Hypertext	Overall
Definitional pre-test (of 20)	11.8 (SD = 2.9)	12.0 (SD = 2.5)	11.9 (SD = 2.7)
Definitional post-test (of 20)	15.3 (SD = 2.9)	16.6 (SD = 2.1)	15.9 (SD = 2.6)
WHAT-IF pre-test correctness (of 24)	15.0 (SD = 3.6)	14.4 (SD = 3.5)	14.7 (SD = 3.5)
WHAT-IF post-test correctness (of 24)	18.3 (SD = 4.0)	19.6 (SD = 3.0)	19.0 (SD = 3.5)
WHAT-IF pre-test time*	17.9 (SD = 7.1)	18.3 (SD = 7.2)	18.1 (SD = 7.1)
WHAT-IF post-test time*	14.4 (SD = 5.0)	15.9 (SD = 6.4)	15.1 (SD = 5.8)
WHAT-IF-WHY prediction (of 13)	10.8 (SD = 1.9)	11.7 (SD = 1.5)	11.3 (SD = 1.7)
WHAT-IF-WHY explanation (of 13)	8.0 (SD = 3.0)	8.9 (SD = 2.9)	8.5 (SD = 2.3)

*Time measured in seconds.

With respect to the latencies collected, repeated measurements showed a significant within participant effect on average item response time ($F_{1,110} = 31.84$, $P < 0.001$). The effect size *d* for the increase in correctness scores yielded a value of 0.53 SD's. No effect of experimental condition on test scores was found ($F_{1,110} = 1.23$, $P > 0.10$).

The WHAT-IF-WHY test

To assess the participants' performance on the WHAT-IF-WHY test, the correctness of the prediction given and the correctness and completeness of the written explanation were scored. This resulted in two different measures, the prediction correctness and the explanation correctness. The prediction could either be correct (score '1') or incorrect (score '0'). The explanation could be correct and complete (score '1'), correct but partly complete (score '0.5'), or incorrect (score '0'). The explanations were scored by two independent raters. The inter-rater reliability yielded a κ of 0.60 (i.e. amount of agreement corrected for amount of agreement expected by chance), which can be considered moderate to substantial (Landis & Koch 1977). The average number of correct predictions and the average correctness of the explanations of the WHAT-IF-WHY test are given in the lower part of Table 1. A MANOVA including both the prediction and explanation scores showed differences between the experimental conditions ($F_{2,109} = 3.41$, $P < 0.05$). A subsequent ANOVA on the prediction scores showed a difference between the simulation and the hypertext condition ($F_{1,110} = 6.77$, $P < 0.05$) in favour of the hypertext condition. An ANOVA on the explanation scores yielded no significant difference – at an α-level of 0.05 – between the conditions ($F_{1,110} = 3.49$, $P = 0.065$).

Process measures

All actions participants made while interacting with the environments were recorded. This yielded data on the use of the simulation environment and hypertext environment. These data included the number of assignments taken and the feedback on assignment answers viewed and the number of equations viewed. Moreover, for the simulation condition also the number of simulation runs was counted and its counterpart in the hypertext environment, the number of graph pages, viewed. For the simulation condition, the number of runs with the simulation was recorded, and for the hypertext condition, the number of times pages with the graphs were consulted was registered. Table 2 gives a summary of the interaction measures within and across experimental conditions. In the following, the separate measures and the comparisons between conditions are more closely examined.

Most participants made extensive use of assignments. The averages and standard deviations are given in Table 2. An ANOVA on the average number of assignments indicated significant differences between the two experimental conditions ($F_{1,110} = 9.10$, $P < 0.05$) with the hypertext condition doing more assignments.

For every selected response alternative, students got feedback on its correctness and hints on how to proceed. An ANOVA on the average number of total feedback indicated significant differences between the two experimental conditions ($F_{1,110} = 9.74$, $P < 0.05$), again in favour of the hypertext condition.

Table 2. Means and standard deviations for interaction measures within the simulation and within the hypertext condition

Interaction measures	Condition	
	Simulation	Hypertext
Assignments (out of 41)	30.0 (SD = 9.8)	34.8 (SD = 6.9)
Feedback to responses	51.4 (SD = 24.5)	64.9 (SD = 21.0)
Equations (out of 13)	5.9 (SD = 3.3)	–
Equation pages (out of 4)	Not present in simulation	3.9 (SD = 0.40)
Equations, total frequency of use (i.e. including repeated use)	9.3 (SD = 7.2)	21.1 (SD = 10.5)
Graph pages (4) total frequency of use (i.e. including repeated use)	Not present in simulation	22.1 (SD = 10.1)
Runs	104.5 (SD = 52.5)	Not possible in hypertext

In the simulation condition, participants could consult thirteen equations. The minimum number of equations consulted was zero and the maximum number was twelve. In the hypertext condition, at each level of model progression one page with equations was included, together containing the same 13 equations. All students in the hypertext condition looked at least once at each of the four pages of equations. We counted both the number of different equations viewed as the total number of times equations were looked up.

In the hypertext environment, no simulation window was available for the learners. Instead, learners could look at pages with static graphs, one page at each level of model progression. All participants looked at least once at every graph page at each level. In Table 2, the total frequency of graph use across the levels is given.

Participants were rather active with the simulation in the environment. The individual differences, however, were substantial. The minimum number of runs was 14 and the maximum was 317. The average and standard deviation are given in Table 2.

Conclusion and discussion

The study presented in this article compared discovery learning with learning from expository instruction. A first assumption of this work was that discovery learning with simulations would result in intuitive knowledge (see also Swaak & de Jong 1996; de Jong & van Joolingen 1998; Swaak *et al*. 1998; de Jong *et al*. 1999). Furthermore, it was expected that expository instruction with hypertexts would result in explicit, definitional, knowledge. Hence, intuitive knowledge was compared with the mere reproduction of facts and definitions as measured by a definitional knowledge test. As simulations are expected to lead directly (i.e. without a declarative stage) to intuitive knowledge that is hard to verbalise, it was predicted that the learners in the simulation condition would have more problems explaining the relations between variables than the learners in the hypertext condition.

In both experimental conditions, medium to large learning gains (from pre- to post-test) were found on both the definitional and intuitive knowledge tests. For the latter test, both an increase in correctness and a decrease in time were observed. As expected, learning from hypertexts was beneficial for the acquisition of definitional knowledge; the participants of the hypertext condition outperformed the participants of the simulation condition on the definitional knowledge test. Learning from hypertext also appeared to be more advantageous as compared to learning from simulation for acquiring intuitive knowledge as reflected in the higher correctness score of the WHAT-IF test. However, in line with the ideas on intuitive knowledge, learning from simulations seemed to result in quicker (not significantly) response times to WHAT-IF items as compared to learning from hypertexts. On theWHAT-IF-WHY-test, high correctness scores on the prediction part of the items (on the whole an average of 11.3 out of 13) were found (with again an advantage for the hypertext condition). The data also showed high correctness scores on the explanation part of the items (an average of 8.5 out of 13).

In this study, learning from expository instruction was operationalised as learning from a hypertext. Discovery learning was operationalised by having learners interact with a simulation. Three main characteristics attributed to a simulation are *perceptual richness* (expressed as dynamic graphs), *low transparency* (the domain content is not presented overtly), and *opportunities for active experiences* (i.e. doing experiments) (see Swaak & de Jong 1996). These three features had to be distinctive of the simulation environment and should not have been applicable to the hypertext environment. When, however, the interaction with the simulation environment is more closely examined, it can be argued that the environment is 'verbally', as well as perceptually rich, because of the many assignments and feedback explanations used by the learners. Moreover, whereas the hypertext environment did not contain dynamic graphs, static graphs were available and frequently consulted by the learners. In other words, learners made the hypertext environment perceptually rich and the simulation environment verbally rich. And as the equations and feedback explanations (displaying domain information directly) were very frequently used in the simulation environment, the minimal transparency of the environment can be questioned, levelling out another distinctive feature of simulations. Furthermore, it can be commented that, taking into account only the frequencies of the learners' interactions, learning from the hypertext environment was at least as 'active' as learning from the simulation environment.

In both the simulation and hypertext environments, the way the assignments were phrased (see also Method Section) and the inclusion of feedback to their responses can be characterised as 'directive'. The assignments of other studies (Swaak & de Jong 1996; Swaak *et al*. 1998; de Jong *et al*. 1999) were posed in a non-directive way. It may be that the power of the assignments' directive character in the present study has overruled the distinctive features of the simulation and the hypertext environment. An implication is that learners in the discovery environment hardly made use of the opportunity to design experiments themselves, testing their own spontaneous hypotheses. Instead they followed the assignments and used the experiments merely to produce graphs, graphs that were also available in the hypertext environment. It appears therefore that the learners have made the interaction with the simulation environment and the hypertext environment very similar.

This leaves us with the intriguing question of why the hypertext condition students outperformed the simulation condition students at the definitional test and the correctness of the intuitive test. One explanation could be that students work more efficiently with the hypertext as compared to the simulation environment. The way our participants worked meant that students in the simulation condition had to generate graphs following the assignments, whereas for the students in the hypertext environment these graphs were ready-made. A related explanation is that an interaction between learning time and environment – simulation *vs.* hypertext – exists in terms of learning efficiency. In other words, had the interaction time been longer, then the simulation condition students could have outperformed the hypertext students, assuming they would engage in genuine discovery. Yet another explanation of the results could be that the surplus value of dynamic graphs (i.e., the simulation) was small. Similar questions were investigated by Rieber *et al*. (1990) and Rieber (1991) who also compared static graphics with dynamic graphics, and investigated correctness scores and latencies in their post-test measures. Rieber *et al*. (1990) did not find main effects types of presentation format on correctness, but found significant effects on the response time of the post-test items in favour of the dynamic graphs. However, Rieber (1991), who combined either static graphics or animated graphics with practice including a structured simulation (i.e. with increasing level of control) and questions with feedback, found main effects of presentation format on post-test correctness scores but not on post-test latencies. In the current study, a trend comparable with results of Rieber *et al*. (1990) was found, though no significant advantage on response time was present. This lack of surplus value of dynamic graphs may be due to the fact that the domain complexity was not high enough, especially when considering the moderate to high entry knowledge level of the students of this study. Finally, a possible clarification from an instructional point of view entails that the directive assignments and feedback turned both the hypertext and simulation environments in instances of direct instruction, which may be the more natural mode for hypertext.

These results of this study again emphasise that the design of simulation-based discovery learning environments that are integrated with support is a delicate process. In a recent study, Reid *et al*. (2003) found that supplying students with what they call 'interpretative support' helps to foster knowledge of different kinds. Interpretative support activates students' prior knowledge and intends to enhance problem representations. In the study by Reid *et al*. (2003), support that presented the students with general ideas (and not specific directions as in this study) on the discovery process ('experimental support') only had an effect on top of the interpretative support. These results emphasise again that when introducing support for learners it is important to estimate if this will not take away the discovery character and that it leaves sufficient time and sufficient freedom for students in the assignments to engage in a discovery mode.

References

Cunningham D.J. (1991) Assessing constructions and constructing assessments: a dialogue. *Educational Technology* **3**, 13–17.

de Groot (1969) *Methodology, Foundations of Inference and Research in the Behavioural Sciences*. Mouton, The Hague.

de Jong T., Martin E., Zamarro J-M., Esquembre F., Swaak J. & van Joolingen W.R. (1999) The integration of computer simulation and learning support; an example from the physics domain of collisions. *Journal of Research in Science Teaching* **36**, 597–615.

de Jong T. & van Joolingen W.R. (1998) Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research* **68**, 179–202.

Ferguson-Hessler M.G.M. & de Jong T. (1990) Studying physics text; differences in study processes between good and poor performers. *Cognition and Instruction* **7**, 41–54.

Jonassen D.H. (1991) Objectivism versus constructivism: do we need a new philosophical paradigm? *Educational Technology Research & Development* **39**, 5–14.

Jones M.K., Li Z. & Merrill M.D. (1990) Domain-knowledge representation for instructional analysis. *Educational Technology* **30**, 7–32.

Landis J.R. & Koch G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174.

Reid D.J., Zhang J. & Chen Q. (2003) Supporting scientific discovery learning in a simulation learning environment. *Journal of Computer Assisted Learning* **19**, 9–20.

Rieber L.P. (1991) Animation, incidental learning, and continuing motivation. *Journal of Educational Psychology* **83**, 318–328.

Rieber L.P., Boyce M.J. & Assad C. (1990) The effects of computer animations on adult learning and retrieval tasks. *Journal of Computer Based Instruction* **17**, 46–52.

Rieber L.P. & Parmley M.W. (1995) To teach or not to teach? Comparing the use of computer-based simulations in deductive versus inductive approaches to learning with adults in science. *Journal of Educational Computing Research* **13**, 359–374.

Shute V.J. (1990) *A comparison of inductive and deductive learning environments: which is better for whom and why?* Paper presented at the American Educational Research Association (AERA) Annual Meeting, Boston, USA.

Shute V.J., Glaser R. & Raghavan K. (1989) Inference and discovery in an exploratory laboratory. In *Learning and Individual Differences: Advances in Theory and Research* (eds P.L. Ackerman, R.J. Sternberg & R. Glaser), pp. 279–326. W.H. Freeman & Company, New York.

Swaak J. & de Jong T. (1996) Measuring intuitive knowledge in science: the development of the what-if test. *Studies of Educational Evaluation* **22**, 341–362.

Swaak J. & de Jong T. (2001) Discovery simulations and the assessment of intuitive knowledge. *Journal of Computer Assisted Learning* **17**, 284–295.

Swaak J., van Joolingen W.R. & de Jong T. (1998) Supporting simulation-based learning; the effects of model progression and assignments on definitional and intuitive knowledge. *Learning & Instruction* **8**, 235–252.

Zacharia Z. & Anderson O.R. (2003) The effects of an interactive computer-based simulation prior to performing a laboratory inquiry-based experiment on students' conceptual understanding of physics. *American Journal of Physics* **71**, 618–629.