The Integration of Computer Simulation and Learning Support: An Example from the Physics Domain of Collisions

Ton de Jong,[1] Ernesto Martin,[2] Jose-Miguel Zamarro,[2]
Francisco Esquembre,[3] Janine Swaak,[1] Wouter R. van Joolingen[1]

[1]*Faculty of Educational Science and Technology, University of Twente, Enschede, The Netherlands*

[2]*Department of Physics, and* [3]*Department of Mathematics, Universidad de Murcia, Murcia, Spain*

Received 12 May 1997; revised 6 March 1998; accepted 16 June 1998

Abstract: Discovery learning is generally seen as a promising but demanding mode of learning that, in most cases, can only be successful if students are guided in the discovery process. The present article discusses a study on discovery learning with a computer simulation environment in the physics domain of collisions. In the learning environment, which is called Collision, students learned about collisions where two particles move in the same direction and interact via a conservative force in such a way that the total mechanical energy is conserved. In the experiment we conducted with Collision, we evaluated the effects of adding two different ways to guide students: model progression, in which the model is presented in separate parts; and assignments, small exercises that the student can choose to do. The effect of providing assignments and model progression was evaluated by comparing the learning behavior and learning results over three experimental conditions in which different versions of the simulation environment were presented: pure simulation, simulation plus assignments, and simulation plus model progression and assignments. Students' use of the environment was logged, their subjectively experienced workload was measured on-line, and their learning was assessed using a number of assessment procedures. Providing assignments with the simulation improved students' performance on one aspect of a so-called intuitive knowledge test. Providing the students with model progression did not have an effect. A subjective workload measure indicated that expanding the simulation with assignments and model progression did not raise the workload experienced by the students. © 1999 John Wiley & Sons, Inc. J Res Sci Teach 36: 597–615, 1999

Discovery learning with computer simulations is generally seen as a promising area for learning and instruction. There are a number of studies in the literature that compare the effects of simulation-based learning to some kind of expository teaching. These studies cover a variety of domains such as biology (Rivers & Vockell, 1987), economics (Grimes & Willey, 1990; Shute & Glaser, 1990), decision support theory (De Jong, De Hoog, & De Vries, 1993), Newtonian

Correspondence to: T. de Jong
Contract grant sponsor: European Commission
Contract grant numbers: D2007, ET 1020

 CCC 0022-4308/99/050597-19

mechanics (Rieber, 1990; Rieber, Boyce, & Assad, 1990; Rieber & Parmley, 1995; White, 1993), and electrical circuits (Carlsen & Andre, 1992; Chambers et al., 1994). As an overall picture, these studies show that learning with computer simulations is as effective as expository instruction. These and other studies, however, also suggest that effects of simulation-based learning are stronger when the simulation is embedded in an environment that intends to support specific aspects of discovery learning.

Embedding simulations into supportive environments seems necessary because learners lack proficiency in discovery learning processes. Discovery learning in science domains is characterized by a number of specific learning processes such stating hypotheses and designing experiments (e.g., Friedler, Nachmias, & Linn, 1990). Furthermore, a paramount characteristic of discovery learning is that the learner is in control of the learning process. This learner centeredness of discovery learning makes regulative processes (for example, planning and monitoring) play a crucial role in the learning process (e.g., Njoo & De Jong, 1993). A relatively large number of studies point to regulation of the learning process as a key problem for learners engaged in discovery learning (e.g., Charney, Reder, & Kusbit, 1990; Glaser, Schauble, Raghavan, & Zeitz, 1992; Lavoie & Good, 1988; Simmons & Lunetta, 1993; Shute & Glaser, 1990; Schauble, Glaser, Raghavan, & Reiner, 1991; Teodoro, 1992).

Supporting learners in their regulative processes can be done in several ways. One way to help learners in the planning process is to give them assignments. Assignments tell the learner what to do, and in this way support the planning process. Assignments are small exercises that point the learner to specific elements of the simulation model. Assignments can also take the form of questions (Showalter, 1970; Tabak, Smith, Sandoval, & Reiser, 1996; Zietsman & Hewson, 1986) or games (White, 1984; 1993). De Jong et al. (1994) described different types of assignments that can be used in combination with simulations. For example, investigation assignments prompt learners to find the relationship between two or more variables, while specification assignments ask learners to predict the value of a certain variable. Explicitation assignments ask learners to explain a certain phenomenon in the simulation environment. White (1984) reported that learners working with games gained a higher performance than those who worked with the pure simulation on a test of qualitative problems. Zietsman and Hewson (1986) used questions with a simulation on velocity to confront students with their misconceptions. They were successful in changing students' models of velocity and position.

A second way to support regulative processes is to restrict the simulation environment, so that learners do not have to cope with the simulation in its full complexity from the start. This way of organizing a simulation environment is called model progression. White and Frederiksen introduced it in their QUEST system, a computer simulation environment in the domain of electricity (White & Frederiksen, 1989, 1990). In model progression, the simulation model is offered in separate steps in which learners gain control over an increasing number of variables. A number of studies on model progression show that the introduction of model progression can lead to higher performance by learners (Alessi, 1995; Rieber & Parmley, 1995). However, it has been found that model progression may not help learners and that it is better to present them with the simulation in its full complexity all at once (Quinn & Alessi, 1994).

In the present study, we performed an evaluation of a discovery environment built around a simulation in the physics domain of collisions. In the discovery environment, we concentrated on the effect of providing support for the planning process by means of presenting students with assignments and model progression. In one experimental condition, students worked with a simulation that was enhanced with assignments. Students were free to select these assignments themselves. In a second experimental condition, we introduced model progression in addition to assignments. For this condition, the simulation was presented in discrete steps and the assignments were organized according to the steps in the model progression. As a control condition, we used an unsupported simulation.

Learning was evaluated by using a number of assessment procedures. Traditional methods of assessment ignore the fact that in scientific discovery learning the learning goal is more in the area of intuitive, qualitative knowledge than formalized knowledge (Berry & Broadbent, 1984; Laurillard, 1992; Lindström, Marton, Ottosson, & Laurillard, 1993; White, 1993). An analysis of experimental studies on simulation-based discovery learning (De Jong & Van Joolingen, 1998) showed that the benefits of discovery learning are more manifest when learning performance is assessed with tests intended to measure qualitative or intuitive knowledge. For this reason, we included in our experiment a test aimed at measuring this intuitive quality of knowledge following the guidelines by Swaak and De Jong (1996). We expected that the students would gain intuitive knowledge in the simulation environment, but not traditional, definitional knowledge. We also expected the differential effects of the experimental conditions to be revealed on the intuitive knowledge test. Finally, for a more detailed analysis of the results, we registered students' use of the environment and the workload students experienced while working with the environment.

Method

Domain

When two or more bodies interact during a certain amount of time, changes in their movements can be observed corresponding to exchanges of momentum, angular momentum, and energy. In a broad sense, this process is called a *collision*. Processes of this kind are very common. An obvious example is the collision between billiard balls. In physics, phenomena such as the shooting of a bullet and the scattering of an electron by the nucleus of an atom are seen as collision processes. It is, however, not only the frequency with which these processes are found in nature that makes the study of collisions so important. Collisions are among the most basic phenomena in which a fundamental quantity is conserved, and in physics conservation theorems play a fundamental role. If we can distinguish a before state (State 1 at Time t1) and an after state (State 2 at Time t2) for a given process, and we know that a given quantity, say P, is conserved, no matter how complex the process between those states could be, the value of P at time t1 can be used to predict the value of P at time t2 without a detailed knowledge of what is happening between t1 and t2. Within the present study, we are concerned with collisions that occur in a short time interval. Therefore, we neglect any momentum transfer coming into the system from the external world. The possible transfer of angular momentum, resulting from friction between the different objects when they are in contact, is also neglected. The system consists of two particles moving in the same direction (1D head-on collision) and interacting in such a way that the total mechanical energy is conserved (elastic collision). The conservation of momentum is the basic principle applied to these situations. It takes into account that during the collision, Newton's third law holds, so that the forces acting between the two interacting bodies are equal in magnitude and have an opposite direction.

The Simulation Environment: Collision

In the present study, subjects learned with a computer-based learning environment in which the central part was a simulation on collisions. In addition to the simulation, the learning environment contained two instructional support measures: model progression and assignments. In the environment, explanations provided students with the necessary background knowledge necessary to understand the phenomena they were observing. Since the experiment was conducted in Spain, the simulation environment was in Spanish. Examples given in this article are translated.

Model Progression. Model progression divides the domain into a number of levels which can be studied one at a time to make sure that the student does not receive too much new information at the same time. In the present simulation environment, five progression levels were implemented. The first model progression level introduced the subjects to the interpretation of space versus time diagrams, which is basic information for the remaining levels. The characteristics of the uniform motion were stressed because it is this type of motion that is found in collisions. At the second level, the 1-D collision between two particles was represented in terms of the position versus time for both particles. This level contained an introduction of momentum conservation through the analysis of kinematical data. The third level also gave information concerning the position versus time of the center of mass (CM) of each particle. The main objective of this level was to introduce the CM concept as an alternative means of describing the dynamics of a multiparticle system. At the fourth level, information on position and of momentum versus time was included in two separate graphs. The momentum exchange and conservation of the total momentum were stressed at this level as well. The fifth and final level presented information on position and kinetic energy versus time in two separate graphs. At this level, the energy exchange and conservation of the total (mechanical) energy for these elastic collisions were stressed. It was shown that kinetic energy is not conserved during the collision, but that the initial kinetic energy is equal to the final kinetic energy.

As an example, Figure 1 presents the interface of the simulation part of model progression Level 3. Students could change input variables in a numerical way. Output was presented in a graph, as numerical values, and in an animation of two colliding objects. *LO* corresponds to a fixed length. If the distance between the particles is greater than *LO*, the particles do not interact. If this distance is less than *LO*, a repulsive force starts acting between the particles.

Assignments. At each model progression level, subjects could choose to complete one or more assignments, small exercises that help to perform a sensible action and that may point to specific aspects of the simulation model. Collision offered different kinds of assignments, each type designed to support a different kind of learning process (De Jong et al., 1994). Investigation assignments asked subjects to investigate the relation between two or more variables. For instance, the student could be asked to investigate the relationship between the total momentum before the collision and the same quantity after the collision (to find that it is constant). Such an assignment stimulated the student to investigate relationships that may otherwise be overlooked. Explicitation assignments asked the student for an explanation of a phenomenon presented in the simulation. The assignment offered one or more phenomena to the student by manipulating the simulation state, and asked the student to provide an explanation of the phenomena. For instance, the student could be shown a number of different values for the initial speed, see that the graph is always a straight line, and explain this by the observation that constant speeds yield a straight line in a graph. Specification assignments presented the student with a situation and asked the student to predict the value of a specific variable. For instance, given the velocities for the two particles and the fact that both masses were equal, the student had to predict the velocity values after the collision.

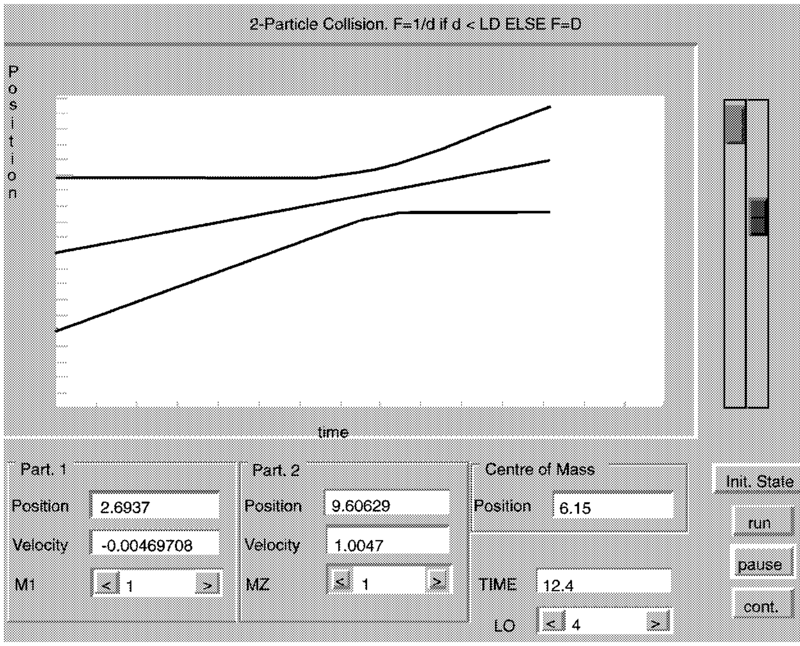

Figure 1. Interface of model progression Level 3.

In Collision, students could, at their own initiative, select from a list of available assignments. When a student selected an assignment, the simulation was put in a state that was relevant for the assignment. Upon selection, an assignment activated itself by presenting the text and setting the simulation in an appropriate way. For investigation and explicitation assignments, the student could choose an answer from a multiple choice list, and feedback was presented for each (right or wrong) answer. The student completed the specification assignments by typing in values. The simulation then was played allowing the student to check the answer.

Explanations. In Collision, we used explanations to provide students with necessary background information. We introduced explanations because we did not want students to get stuck in the simulation owing to a mistake or lack of basic understanding. Explanations therefore mainly provided straightforward definitions of central variables in the simulation. Students could select an explanation from a list that was permanently available. For example, if a student selected the explanation that was indicated with the name *acceleration,* a simple graph was shown with the text: “The acceleration (green curve) is the time derivative of the velocity (in red).”

Design

In the study, we compared three versions of basically the same environment. The first condition (Condition I) was an environment with model progression and assignments. This environment contained all five model progression levels as described above, and at each level there was a number of associated assignments. Subjects were allowed to progress through the model progression levels at their own pace and could go back to a level that was already visited. However, they were required to initially stay at a level for at least 10 min. Assignments at each level were available for the subjects without restrictions. In the second condition (Condition II), subjects saw only the last two model progression levels (Levels 4 and 5). These two levels provided different views on the domain; in one, there was an emphasis on momentum, and in the other, the emphasis was on kinetic energy. Subjects could toggle between the two levels. Assignments from the other three model progression levels, as far as they were applicable to Levels 4 and 5, were also available, to keep the number of assignments between Conditions I and II in the same range. In the third condition (Condition III), assignments were omitted and subjects saw only model progression Levels 4 and 5 as in Condition II. In each condition and at each model progression level, subjects had access to explanations. For Condition I, some explanations were the same at different model progression levels. A total of 20 different explanations were available in Condition I. The basic design of the study and an overview of the collision application for each experimental condition are displayed in Table 1.

Subjects

Subjects were first-year university students from the Computer Science and Biology Departments. Owing to practical circumstances, it was impossible to distribute students from these different backgrounds evenly over experimental conditions. Consequently, subjects in Condition I were from the Biology ($n = 8$) and Computer Science ($n = 7$) Departments, while subjects in the other two groups were from the Biology Department. Biology and computer science students have comparable backgrounds in high school but the students from computer science passed through a more strict selection process to be allowed to study computer science. As a control, we made separate analyses of the data comparing biology and computer science students.

Tests

To assess students' performance in the environments, two types of test were used: a definitional knowledge test and an intuitive knowledge test. Both tests were presented as pre- and posttests. For the definitional knowledge test, the same test was used at pre- and posttest, while parallel versions of the intuitive knowledge test were used. The intuitive knowledge test that we used is a relatively new test format. To gain insight into the place of this test in the nomological network, we used a test that aimed to assess the structure of knowledge. This test was used as a posttest only, since at that point we expected students to have more intuitive knowledge.

Table 1
Overview Collision for each experimental condition

Condition	Subjects	MP level	Assignments	Explanations
I	15	1	10	4
		2	4	8
		3	3	8
		4	3	8
		5	4	8
II	16	4 + 5	15	14
III	15	4 + 5	0	14

Definitional Knowledge. The tests for definitional knowledge concerned knowledge of individual elements from the domain. Multiple choice items (presenting three answer alternatives) were used to assess the definitional knowledge about the facts and concepts of the domain. An example of a definitional knowledge test item is: "We have a system with two particles with kinetic energies of KE1 and KE2, respectively. The total kinetic energy is: (a) KE1·KE2; (b) (KE1+KE2)/2; (c) KE1+KE2." The definitional knowledge test consisted of 36 multiple choice items, each with three alternatives.

Intuitive Knowledge. To measure intuitive knowledge, we created the speed what-if test. In this test, each test item contains three parts: conditions, actions, and predictions. The conditions and predictions are states in which the system can be. They are displayed as a simulation output, and if necessary for students' understanding, are accompanied with a minimum amount of text. The action, or the change of a variable within the system, is presented in text. The speed what-if task requires the student to decide as accurately and quickly as possible which of the predicted states follows from a given condition as a result of the action that is displayed. The items of the task are kept as simple as the domain permits. The design of this type of test is discussed in Swaak and De Jong (1996). Figure 2 presents two examples of items tapping intuitive knowledge. To determine the level of intuitive knowledge, both correctness and answer time were used. The intuitive knowledge test consisted of 37 items, each with three alternatives.

Structural Knowledge. A concept-mapping task was used as a structural knowledge test. The concept-mapping task required students to sort the concepts (variables) of the domain into clusters of concepts in a way that represents their view of the domain. The instructions for this test emphasized that there was no correct or incorrect solution. We used a computerized task in which

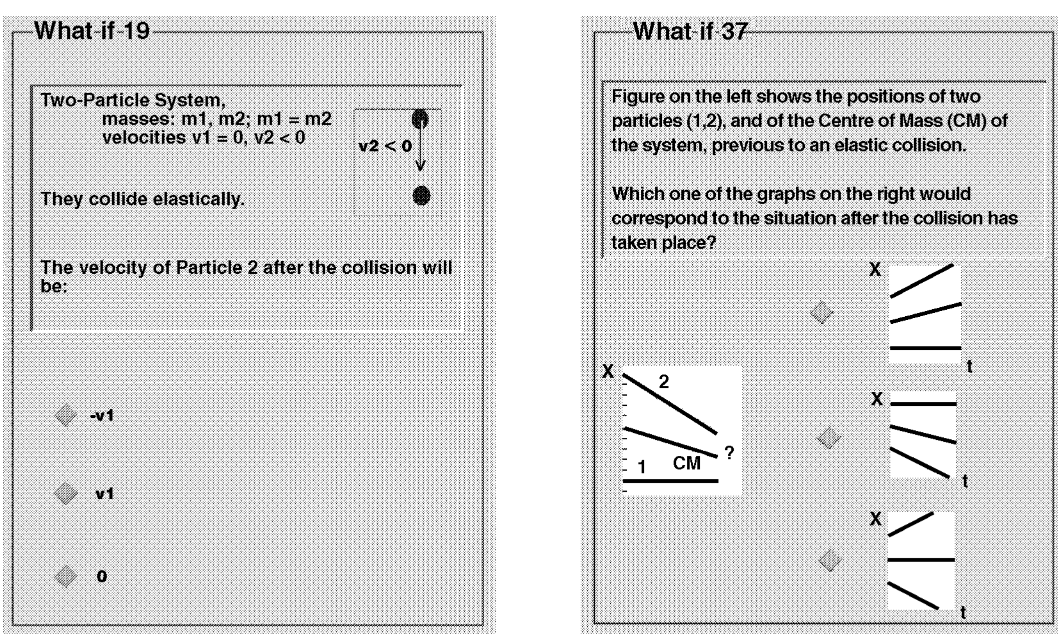

Figure 2. Two examples of intuitive knowledge test items.

concepts of the domain were presented on the computer screen. The student could point with the mouse to a concept. When the student clicked the mouse, the concept was selected and could be dragged to the part of the screen where the student wished to create the relevant cluster. The student could then select another concept and add it to the already existing cluster or create a new cluster. The interface allowed for an easy rearrangement of clusters until the subject was satisfied. Subjects could also use color to indicate the differences between clusters. A total of 22 concepts were used. These concepts were: kinematics, reference frame, position (r), time (t), velocity (v), acceleration (a), system of particles, Newton's third law, center of mass (CM), collisions, particle dynamics, Newton's first law, mass (m), force (F), linear momentum (P), Newton's second law, $dP(\mathrm{CM})/dt = \mathrm{sum}(F_{\mathrm{external}})$, work, energy, kinetic energy, potential energy, and conservation laws.

Measuring Interaction Behavior

We registered all actions students made while interacting with the simulation. This provided us with data on the use of the simulation and the supportive measures that were present. We introduced a second type of interaction measurement to assess subjectively experienced workload. At regular moments, a small electronic questionnaire appeared and subjects had to complete it before they continued working with the environment. By pulling sliders, subjects could indicate their perceived difficulty of the subject matter, their perceived difficulty of the interaction with the environment itself (this concerns the human–machine interaction aspects), and their view on the helpfulness of the instructional measures (the functionality of the tools) that they had used. The questionnaire was set to pop up every 20 min, but display was always postponed until an event occurred that marked the end of a coherent subject's action, such as closing an explanation or completing an assignment. This was done to prevent this measurement from interfering with the discovery behavior. The general assumption behind a subjective technique is that learners are able to introspect on their cognitive processes. Several studies indicate that learners seem to have no difficulty in assigning numerical values to their experienced cognitive load (e.g., Hill et al., 1992; Paas, van Merriënboer, & Adam, 1994). Moreover, Paas et al. (1994) compared the reliability and sensitivity of a rating scale and a cardiovascular measure, and concluded that in instructional research the subjective technique met these requirements better.

Procedure

Each experimental session had a duration of approximately 3 h 30 min. It consisted of the following parts, in chronological order:

1. Introduction. Subjects received a general introduction in the form of a written document explaining the procedure, sessions, and general idea behind discovery learning. The different support measures as applicable to the experimental condition were then explained. A short overview of the pre- and posttest, together with a sample item from each test, were displayed. Finally, students were told that everything they did in the environment was logged and the meaning of the subjective workload questionnaire was explained. Students could read this written introduction and ask the experiment leader questions. They could consult this written introduction throughout the course of the experiment. The introduction was approximately 10 min long.
2. Pretests. After the general introduction, the definitional and intuitive pretest were administered. This took approximately 45 min.
3. Practice. After completing the pretests, students received a practice simulation environment. This environment was built around a simple harmonic oscillator mass–spring sys-

tem, in which both the mass and the spring constant could be controlled, like the initial conditions of the collision system (position and velocity of the mass). In this way, motions with different amplitudes, phases, and frequencies could be obtained and investigated. In this training application, a sample of every kind of assignment used in Collision was implemented. The appearance of the practice environment was also adapted to the appearance of the Collision application condition that subjects would receive (e.g., no assignments were present if the Collision application also would not contain assignments). The experimental leader made a demonstration using the display of a screen on an overhead projector. Subjects could ask questions and, upon request, receive personal assistance. Practicing the simulation environment took approximately 10 min.

4. Interaction with Collision. After the introduction, subjects learned with the Collision environment on their own. The experimental leader was present and could assist with questions concerning the operating of the environment, but not on the subject matter content. A minimum time of 45 min and a maximum time of 90 min were set for the interaction. Twenty-five minutes before the end of the 90-min period, the experimental leader announced that 25 min remained and repeated how many model progression levels were present. The decision to move quickly to the final level (if subjects were not yet there) or stay at a lower level was in the hands of the subjects.
5. Posttests. After the interaction with the simulation environment, posttests were presented. The sequence of presentation was definitional test, intuitive knowledge test (parallel version), and structural knowledge test. All tests were presented electronically (as were the pretests). A total of 60 min was allotted for the posttests.

Results

In this section, we report the results on the different knowledge tests and give an account of the interaction behavior and the subjective workload measure.

Definitional Knowledge Test

The definitional knowledge test was given in the same form at pre- and posttest. It consisted of 36 multiple choice items each with three alternative answers. Table 2 gives the average number of correct items for the definitional pre- and posttests, for the three experimental conditions averaged across students.

The data showed that the overall difference between the definitional pre- and posttest was significant in favor of the posttest, $F(1, 43) = 37.56$, $p < .05$. This means that students gained definitional knowledge in the experiment. The gain, however, was not large—an average of two

Table 2
Average number of correctly answered items on definitional pretest and definitional posttest

	Condition							
	I (sim, asn, mp)		II (sim, asn)		III (Only sim)		Overall	
Test	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Definitional pretest	24	5	21	4	23	4	23	4
Definitional posttest	27	5	24	4	25	4	25	4

Note. $n = 36$ items. asn = assignments; mp = model progression; sim = simulation.

items. The biology students in Condition I began with an average score of 21 [standard deviation (SD) = 3]. Students in this condition had an average score of 24 (SD = 4) at the posttest. Their computer science counterparts started with a higher average score of 27 (SD = 3) and also gained an average of three items at posttest (score was 30, SD = 3). Comparing scores across conditions for the complete sample, we did not find differences on definitional posttest scores between conditions, using an analysis of covariance (ANCOVA) to correct for pretest scores. This implies that all learning environments were equally effective for definitional knowledge.

Intuitive Knowledge Test

For the intuitive test, items were scored on both the correctness of the answer and the time used for giving the answer. For response times, we removed outliers using a procedure similar to the one used in Swaak, Van Joolingen, and De Jong (1998). With this procedure, no more than 1.6 % of the data were excluded from further analyses. The number of items over which analyses were conducted differed between students because the removal of outliers was performed based on individual data. For Condition I, an average of 35.6 items remained, while the average remaining number of items for Conditions II and III was 35.5. Table 3 presents the results on the intuitive knowledge test.

A repeated measurement analyses showed significant overall differences between pre- and posttest on number of correct items, $F(1, 43) = 36.34$, $p < .05$, and also on test completion times, $F(1, 43) = 125.09$, $p < .05$. This indicates that the students improved on both correctness score and completion time.

An ANCOVA on the what-if posttest correctness scores, using the pretest as a covariate, did not reveal differences between the experimental conditions, $F(2, 42) = 1.38$, $p > .10$. As was the case for the definitional test, the computer science students in Condition I started with better scores than did the biology students. The differences between the scores, however, were rather small [average scores 21 (SD = 4) and 20 (SD = 5)]. However, the gain in scores was much higher for the computer science students [from 21 to 30 (SD = 3)] compared to the biology students [from 20 to 22 (SD = 6)]. This may point to a differential learning effect based on characteristics of the students.

An ANCOVA on the what-if posttest completion times, using the pretest as a covariate, revealed differences between the experimental conditions, $F(2, 42) = 3.66$, $p < .05$. In compar-

Table 3
Means and standard deviations for intuitive pre- and posttests (time in seconds to answer an item)

	Condition							
	I (sim, asn, mp)		II (sim, asn)		III (Only sim)		Overall	
Test factors	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
What-if pretest correctness	21	5	17	3	19	3	19	4
What-if posttest correctness	26	6	21	5	22	4	23	5
What-if pretest time	29	6	36	9	37	6	34	8
What-if posttest time	19	4	22	6	27	8	23	7

Note. asn = assignments; mp = model progression; sim = simulation.

ing the two groups of students in Condition I, we saw different starting level for the completion time averages of 26 s ($SD = 5$) and 31 s ($SD = 5$) for computer science and biology students, respectively. Both the computer science and biology students in Condition 1 gain an average of 10 s in their answering time. When we removed the computer science students from Condition I, an ANCOVA on the what-if completion times did not show overall differences between the experimental conditions at an α level of .05, but did at an α level of .10, $F(2, 35) = 2.80$, $p = .07$. Removing these seven students, however, lowered the power of the test. An ANCOVA on the what-if completion times comparing just Conditions II and III showed a significant difference, $F(1, 28) = 4.52$, $p < .05$, in favor of Condition II. An ANCOVA on the what-if completion times comparing Conditions I and III only showed a significant difference at an α level of .10, $F(1, 27) = 3.57$, $p = .07$.

Following our definition of intuitive knowledge (we defined *intuitive* as being correct and quick on items of the what-if type), we expected a low correlation between answer time and correctness score. As expected we found no tradeoff between correctness and speed. [The correlation between answer time and correctness had a value of $r = -.17$, $p > .10$ for the what-if pretest items and a value of $r = -.23$, $p > .10$ for the what-if posttest items (within students, across items)]. Computed within items and across students, the values were $r = .02$, $p > .10$ and $r = -.27$, $p > .10$ for the pretest and posttest, respectively. When we looked at the item response times for correct and incorrect items separately, the data showed small differences in answer time between correct and incorrect items at both the pretest [average of 33 s ($SD = 8.3$) for correct items and 35 s ($SD = 8.3$) for incorrect items] and posttest [average of 22 s ($SD = 9.0$) for the correct items and 24 ($SD = 6.9$) for the incorrect items]. Because of these small differences, separate analyses were not made.

Structural Knowledge Test

To assess students' performance on the structural knowledge test, we used concept maps created by two experts (physics teachers and coauthors of this article) as criteria. Subjects' concept maps were transformed into similarity matrices. Similarity of concepts was given a one if two concepts appeared in the same group and a zero if two concepts belonged to different groups. The same procedure was followed for the two expert concept maps. For each subject, the proximity between the subject's matrix and the two expert matrices were calculated using a formula introduced by De Jong and Ferguson-Hessler (1986). The proximity measure equals one when the expert's grouping is identical to the subject's grouping, and may reach a small negative value with very dissimilar groupings. The proximity between the two expert groupings was high and had a value of .84. Table 4 gives the average proximity measures with both experts

Table 4
Average proximity measures of subjects' sorting and two experts

	Condition							
	I (sim, asn, mp)		II (sim, asn)		III (Only sim)		Overall	
Expert	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	.44	.15	.42	.22	.40	.22	.42	.20
2	.45	.15	.41	.23	.41	.22	.42	.20

Note. asn = assignments; mp = model progression; sim = simulation.

across the three experimental conditions. The table shows that there were no average differences between the two experts as criteria.

Univariate analyses of variance revealed no difference between conditions for using either Expert 1 or Expert 2 as a criterion, $F(2, 43) < 1$ for both scores.

Relations between the Different Tests

Table 5 displays the correlations between the three knowledge tests across the three conditions. For the what-if speed test, the results for correctness and time are given separately. The proximity scores for the concept-mapping test are taken separately for the two experts (here indicated as Concept Map I for Expert 1, and Concept Map II for Expert 2).

The pattern that emerges from the analysis of the three measures is two clear clusters. The first one consists of the definitional test and what-if correctness. The second is the what-if time aspect. The test for structural knowledge, as measured by concept maps, correlated moderately with the first cluster but not with the second one, and could be regarded as a third cluster. These data show that the what-if speed test indeed measured an additional aspect of knowledge which came out especially at the time aspect of this test. The structural knowledge test had a partial overlap with the what-if correct aspect, but had no relationship with the what-if speed aspect.

Interaction Behavior

We registered all actions students made while interacting with the simulation. This provided data on the use of the simulation and the supportive measures that were present. Because of technical issues resulting in a loss of log files, it was not possible to collect the interaction data from all students. In the subsequent analyses, the complete interaction data of 34 subjects were used. As a result 15, 12, and 7 subjects remained over Conditions I, II, and III, respectively.

Time on Task. In the experimental procedure, a minimum of 45 min and a maximum of 90 min were set for interacting with the simulation. In Condition I (simulation, model progression, and assignments) subjects spent an average of 79 min ($SD = 21$), in Condition II (simulation and assignments) 82 min ($SD = 11$), and in Condition III (only simulation) 75 min ($SD = 18$). An analysis of variance (ANOVA) revealed differences between the experimental conditions, $F(2, 31) = 6.37$, $p < .01$. The Scheffé confidence intervals for locating pairwise differences showed that only Condition II significantly differed from Condition III (t value $= 2.39$, $p < .05$). To determine whether the differences in time could be responsible for the effects of experimental condition on the intuitive test results, we performed an ANCOVA with interaction

Table 5
Correlations between different aspects of knowledge over all three conditions on posttest

	What-if correct	What-if speed	Concept map I	Concept map II
Definitional	.80 ($p < .01$)	−.11 ($p < .10$)	.35 ($p < .05$)	.35 ($p < .05$)
What-if correct		.02 ($p > .10$)	.37 ($p < .05$)	.37 ($p < .05$)
What-if speed			.00 ($p > .10$)	.03 ($p > .10$)
Concept map I				.97 ($p < .001$)

Note. Levels of significance are in parentheses.

time and time on the intuitive pretest as covariates. This yielded similar results as those described in the section on the intuitive knowledge test.

Number of Runs. Students were active in the simulation. In Condition I (simulation, model progression, and assignments), students used an average of 49 runs ($SD = 23$). In Condition II (simulation and assignments), subjects used 47 runs ($SD = 41$) on average. For Condition III students (only simulation), an average of 57 runs were employed ($SD = 29$). We might expect a higher number of runs in Condition III, since this condition contained no instructional measures. The figures show that the number of runs was slightly higher in Condition III, but it was not significantly different from the other conditions.

Number of Assignments and Explanations Used. In general, subjects made extensive use of assignments and explanations. However, differences between subjects' use of explanations were considerable, with a few subjects using almost no explanations. In Condition I, there was almost a maximal use of assignments for all students, while students in Condition II used fewer assignments. Owing to large individual differences, it was not meaningful to examine the relationships between use of instructional measures (assignments and explanations) and scores on the knowledge tests.

Subjective Workload

Subjects' subjective workload was measured by means of the pop-up electronic questionnaire. Subjects' evaluation of three aspects of the environment were gathered: subject matter difficulty (Is the subject matter seen as easy or difficult?), system usage (Is working with the system easy or difficult?), and usability of tools (Do tools make the learning task easier or more difficult?). In Condition III, *tools* referred only to the explanations that were present in all three experimental conditions. For each aspect, subjects' scores could range from 0 to 100, where 100 indicated that the subject matter was extremely difficult, the environment was extremely difficult to work with, or the tools made the task much more difficult. In Condition I the pop-up questionnaire appeared an average of 6.33 times, with a range of 3–9. The average for Condition II was 3.42 times and the range was 0–5. In Condition III, the pop-up questionnaire appearance had an average of 2.6 times, with a range of 0–3. Across all three conditions, scores on difficulty correlated .22 with system use ($p > .10$) and .53 ($p < .05$) with tools. System use and tools correlated .47 ($p < .05$). The correlations, though significant on two of three occasions, were moderate, thus indicating that the three measures assessed different aspects of subjective workload.

A multivariate ANOVA (MANOVA) (using the three ratings as dependent variables) with the data of 32 subjects revealed differences between the experimental conditions, $F(6, 54) = 2.91$, $p < .05$. Subsequent univariate tests showed that the ratings on system appreciation, $F(2, 29) = 5.62$, $p < .01$, and on helpfulness of the tools, $F(2, 29) = 6.58$, $p < .01$, were responsible for the differences. The ratings on perceived difficulty did not differ between the experimental conditions. Using Scheffé confidence intervals for locating pairwise differences, we found that in Condition I (simulation, model progression, and assignments), system use was seen as less troublesome than in Condition II (simulation and assignments). Also, the tools were appreciated more in Condition I than in Condition II. For helpfulness of the tools, a significant difference, $t = 2.97$, $p < .01$, was found only between Conditions I and III, with the tools being more appreciated in Condition I. Again, making a distinction between the two groups of students in Condition I, we found that most of the differences are due to differences between the scores of the Computer Science and Biology students. A MANOVA (using the three ratings as dependent variables) with just the subjects of Condition I revealed that the scores for the biology and computer science students were different, $F(3, 11) = 7.19$, $p < .05$. Subsequent univariate tests showed that the ratings on system appreciation, $F(1,13) = 13.02$, $p < .01$, and on helpfulness of the tools, $F(1,13) = 14.21$, $p < .01$, were responsible for the differences. The ratings on perceived difficulty did not differ between biology and computer science students. The general conclusion from this is that computer science students and biology students had similar perceptions regarding domain difficulty, but computer science students appreciated the environment and the tools more than the biology students did. This seems plausible if we assume that the computer science students have a greater computer affinity. A MANOVA with the data without the computer science students (a total of 25 subjects remain in this analysis) revealed no differences between the experimental conditions, $F(6, 42) < 1$. This shows that for the biology students, adding tools to the environment did not raise their workload.

We also computed correlations between the workload measure and the posttest scores. They were computed across the experimental conditions and we controlled for the pretest scores. This control was introduced because preknowledge might influence the perceived difficulty. The partial correlations are displayed in Table 6.

From Table 6, we can see that only correlations for the what-if completion times were significant at the .05 level. This indicates that subjects who estimated the subject matter as easier (low score) had lower response times than subjects who estimated the subject matter as difficult. Subjects who thought the tools were very helpful (low score) also had relatively low completion times.

Discussion

Our main objective was to determine whether integrating instructional measures, in the form of model progression and assignments, into a simulation environment would lead students to higher performance compared to a simulation environment that did not include these instructional measures. In discussing the results of this study, we will also mention data we gathered in two other studies, conducted around the same time, in other domains of physics: transmission lines (De Jong, Härtel, Swaak, & Van Joolingen, 1996) and harmonic oscillations (Swaak et al., 1998). Although these studies were different in domain, type of students, and setup, they also examined the effect of model progression and assignments.

In De Jong et al. (1996) and Swaak et al. (1998), we found that adding assignments to a simulation had clear effects. In the current study, this effect was present but much less pronounced. On only one aspect of intuitive knowledge—time—there was a clearly significant difference on a comparison of two of the conditions. Also, in the current study the learning effects,

Table 6
Partial correlations between knowledge scores (posttests) and measures of workload

	Perceived difficulty	System appreciation	Helpfulness of tools
Definitional posttest scores	−.26 ($p > .10$)	.27 ($p > .10$)	−.17 ($p > .10$)
What-if posttest correctness scores	−.12 ($p > .10$)	.17 ($p > .10$)	−.30 ($p > .10$)
What-if posttest completion times	.46 ($p < .05$)	.32 ($p < .10$)	.42 ($p < .05$)

though significant, were not very large. We can think of several reasons for this. First, the study time was not very long. Second, there could have been a larger number of assignments and the quality of the assignments could have been improved. Based on our experience, we have now developed a new version of Collisions, which, apart from a better interface, e.g., more natural animations, now has more assignments. These assignments now emphasize reading and understanding graphs, including matching graphs and animations. In addition, the assignments contain real-life investigations, such as asking the students to make one particle a beach ball and the other a bowling ball, and run the simulation comparing the momentums of the two balls.

In all three studies, assignments were very popular with students; together with their influence on the acquisition of knowledge, this makes them good candidates to include in simulation learning environments. Some caution, however, should be exercised, since assignments also take away students' responsibility for planning the discovery behavior (see also De Jong et al., 1998-b). An interesting study would be to examine, in a longitudinal way, the effects when assignments gradually disappear from simulation learning environments used by students over a longer period in time.

The overall effect of model progression is less clear. In the study by Swaak et al. (1998), we compared a condition consisting of a pure simulation to a condition where only model progression was added. In this study, the effect of introducing model progression was close to significance on response times for an intuitive knowledge test. In the current study, providing model progression did not demonstrate an effect. As we already mentioned in the introduction, the literature is inconclusive on the effects of model progression. We can think of two main reasons for these inconclusive results. First, the complexity of the domain involved may play a role, with model progression only of use for complex domains. Our collision domain was not a very complex domain. Second, the initial knowledge of the students may have had an influence. In this study, the students' prior knowledge, as measured at the pretest, was substantial. This may have affected the effectiveness of model progression. In Collision, the first level of model progression introduced students to basic concepts and graph characteristics. As the results on the pretest were quite high, it might be that the students did not really need the first level. Levels 2 and 3 introduced elastic collisions and conservation of momentum, respectively. Here, the transition between levels was rather small and both levels were based on the same underlying model. Thus, the way we introduced model progression might have been too easy for our subjects. Instead of introducing more complexity, model progression Levels 4 and 5 (both based on the same model) gave two different perspectives, a momentum and an energy perspective. The rationale behind Levels 4 and 5 was different from that behind the first three levels. Therefore, model progression levels are presented not only to support the regulative (planning) processes but also to provide multiple perspectives and thus help students build integrated and flexible knowledge structures (for other examples, see Bennett, 1992; De Jong et al., 1998; Moyse, 1991; Sime, 1998; White & Frederiksen, 1989, 1990). In the current study, however, this was not reflected in the structural knowledge test. We designed the new Collision environment so that each model progression level except the introductory one contained all perspectives. In other words, concepts such as position, velocity, kinetic energy, and momentum of the particles were treated both within each level and across levels of model progression. As a consequence, both a complete and coherent understanding per level of model progression could be supported. More specifically, at the level of elastic collision against a fixed wall, the concepts of kinetic energy and momentum before and after the collision were introduced to the learners. At the level of complete elastic collisions, it was shown to learners that the elastic collision against a fixed wall is a special instance of elastic collisions, i.e., with one mass (the wall) far greater than the other. Then, at complete inelastic collisions, the differences and similarities between elastic and inelastic collisions were emphasized. Understanding of connections between the levels of model progression is achieved by using parallel assignments at these three levels.

For complex (learning) environments, workload might be an important factor. Some authors speculate that adding tools to a learning environment increases the complexity and thus increases workload. This may hinder the positive effects that are expected from adding cognitive tools as we did here (e.g., Gruber, Graf, Mandl, Renkl, & Stark, 1995). In this study, we measured subjective workload by means of an electronic questionnaire that appeared on the screen. Three different aspects of subjective workload were measured: subject matter difficulty, use of the environment, and usability of the instructional measures. The electronic questionnaire appeared to assess the different aspects of subjective workload as evidenced by the correlations between the three different measures. The environments from the three experimental conditions did not differ with respect to subject matter difficulty, but did differ with respect to helpfulness of instructional measures. Students in Condition I (that included both assignments and model progression) rated the instructional measures to be more helpful than did students in Condition II (that did not include assignments) or Condition III (the simulation as such with only explanations available). It was also found that Condition II students appreciated the instructional measures more than Condition III students did. In any case, there was no indication that a higher subjective workload resulted from adding instructional measures to the environment. Instead, evidence for the opposite effect was present. This is in line with the results found in Swaak et al. (1998) when using the same questionnaire.

We questioned whether it made sense to introduce new ways of measuring knowledge in addition to traditional, definitional knowledge tests. In the current study, we saw a pattern emerge similar to what was seen in the two related studies (De Jong et al., 1996; Swaak et al., 1998). First, progress was higher on the test of intuitive knowledge as compared to definitional knowledge. Although we expected no increase on definitional knowledge, there was a small but significant increase that may be attributed to the presence of explanations in all conditions. Second, differential effects between conditions came out only at the test of intuitive knowledge. In the present study, effects came out at the time aspect of the test for intuitive knowledge. This made a coherent pattern with other measures. In the correlational analysis, we found that the time scores formed a separate cluster from what-if correct scores and definitional test scores, and performance on the structural knowledge test. We also found that the data we gathered on the three aspects of subjective workload (subject matter difficulty, system usage, and usability of tools) correlated sensibly with what-if time scores. Low completion times were associated with rating the subject matter difficulty as easier, greater appreciation of the tools, and (close to significance) greater appreciation of the environment. Based on the results of this and previous studies, we conclude that to estimate the effects of simulation-based discovery environments fairly, it is necessary to introduce measures that differ from the traditional definitional tests.

In conclusion, we feel that this study has contributed to the understanding of how to support learners in their discovery learning process. As other studies have shown, assignments help learners to gain intuitive knowledge, while the role of model progression is less clear. Finally, this study gives some indication that learner characteristics may play a role in discovery learning.

The current study was performed within the SMISLE Project, partially sponsored by the European Commission in the DELTA program as Project D2007. The SMISLE Project is currently continued in the SERVIVE Project also partially sponsored by the European Commission (Project ET 1020). The authors thank the reviewers from this journal for their extensive commentaries. They suggested, among other things, additional analyses that helped the authors to improve the study.

References

Alessi, S.M. (1995, April). Dynamic vs. static fidelity in a procedural simulation. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Bennett, K. (1992). The use of on-line guidance, representation aiding, and discovery learning to improve the effectiveness of simulation training. In J.W. Regian & V.J. Shute (Eds.), Cognitive approaches to automated instruction (pp. 217–242). Hillsdale, NJ: Erlbaum.

Berry, D.C., & Broadbent, D.E. (1984). On the relationship between task performance and associated verbalizable knowledge. Quarterly Journal of Experimental Psychology, 36A, 209–231.

Carlsen, D.D., & Andre, T. (1992). Use of a microcomputer simulation and conceptual change text to overcome students preconceptions about electric circuits. Journal of Computer-Based Instruction, 19, 105–109.

Chambers, S.K., Haselhuhn, C., Andre, T., Mayberry, C., Wellington, S., Krafka, A., Volmer, J., & Berger, J. (1994, April). The acquisition of a scientific understanding of electricity: Hands-on versus computer simulation experience; conceptual change versus didactic text. Paper presented at the AERA 1994, New Orleans, LA.

Charney, D., Reder, L., & Kusbit, G.W. (1990). Goal setting and procedure selection in acquiring computer skills: A comparison of tutorials, problem solving, and learner exploration. Cognition and Instruction, 7, 323–342.

de Jong, T., Ainsworth, S., Dobson, M., van der Hulst, A., Levonen, J., Reimann, P., Sime, J.-A., van Someren, M., Spada, H., & Swaak, J. (1998-a). Acquiring knowledge in science and mathematics: the use of multiple representations in technology based learning environments. In M. Van Someren, P. Reimann, H. Boshuizen, & T. de Jong (Eds.), The role of multiple representations in learning and problem solving (pp. 9–41). Oxford: Elsevier Science.

de Jong, T., de Hoog, R., & de Vries, F. (1993). Coping with complex environments: The effects of overviews and a transparent interface on learning with a computer simulation. International Journal of Man–Machine Studies, 39, 621–639.

de Jong, T., & Ferguson-Hessler, M.G.M. (1986). Cognitive structures of good and poor novice problem solvers in physics. Journal of Educational Psychology, 78, 279–288.

de Jong, T., Härtel, H., Swaak. J., & van Joolingen, W. (1996). Support for simulation-based learning: The effects of assignments in learning about transmission lines. In A. Díaz de Ilarazza Sánchez & I. Fernández de Castro (Eds.), Computer aided learning and instruction in science and engineering (pp. 9–27). Berlin: Springer-Verlag.

de Jong, T., van Joolingen, W., Scott, D., de Hoog, R., Lapied, L., & Valent, R. (1994). SMISLE: System for multimedia integrated simulation learning Environments. In T. de Jong & L. Sarti (Eds.), Design and production of multimedia and simulation based learning material (pp. 133–167). Dordrecht: Kluwer Academic.

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

de Jong, T., van Joolingen, W.R., Swaak, J., Veermans, K., Limbach, R., King, S., & Gureghian, D. (1998-b). Self-directed learning in simulation-based discovery environments. Journal of Computer Assisted Learning, 14, 235–246.

Friedler, Y., Nachmias, R., & Linn, M.C. (1990). Learning scientific reasoning skills in microcomputer-based laboratories. Journal of Research in Science Teaching, 27, 173–191.

Glaser, R., Schauble, L., Raghavan, K., & Zeitz, C. (1992). Scientific reasoning across different domains. In E. de Corte, M. Linn, H. Mandl, & L. Verschaffel (Eds.), Computer-based learning environments and problem solving (pp. 345–373). Berlin: Springer-Verlag.

Grimes, P.W., & Willey, T.E. (1990). The effectiveness of microcomputer simulations in the principles of economics course. Computers & Education, 14, 81–86.

Gruber, H., Graf, M., Mandl, H., Renkl, A., & Stark, R. (1995, August). Fostering applicable knowledge by multiple perspectives and guided problem solving. Paper presented at the conference of the European Association for Research on Learning and Instruction, Nijmegen, The Netherlands.

Hill, S.G., Iavecchia, H.P., Byers, J.C., Bittner, A.C., Zakland, A.L., & Christ, R.E. (1992). Comparison of four subjective workload rating scales. Human Factors, 34, 429–439.

Laurillard, D. (1992). Learning through collaborative computer simulations. British Journal of Educational Technology, 23, 164–171.

Lavoie, D.R., & Good, R. (1988). The nature and use of predictions skills in a biological computer simulation. Journal of Research in Science Teaching, 25, 335–360.

Lindström, B., Marton, F., Ottosson, T., & Laurillard, D. (1993). Computer simulations as a tool for developing intuitive and conceptual understanding in mechanics. Computers in Human Behavior, 9, 263–281.

Moyse, R. (1991). Multiple viewpoints imply knowledge negotiation. Interactive Learning International, 7, 21–37.

Njoo, M., & de Jong, T. (1993). Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. Journal of Research in Science Teaching, 30, 821–844.

Paas, F.G.W.C., van Merriënboer, J.J.G., & Adam, J.J. (1994). Measurement of cognitive load in instructional research. Perceptual and Motor Skills, 79, 419–430.

Quinn, J., & Alessi, S. (1994). The effects of simulation complexity and hypothesis generation strategy on learning. Journal of Research on Computing in Education, 27, 75–91.

Rieber, L.P. (1990). Using computer animated graphics in science instruction with children. Journal of Educational Psychology, 82, 135–140.

Rieber, L.P., Boyce, M., & Assad, C. (1990). The effects of computer animation 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, 14, 359–374.

Rivers, R.H., & Vockell, E. (1987). Computer simulations to stimulate scientific problem solving. Journal of Research in Science Teaching, 24, 403–415.

Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. Journal of the Learning Sciences, 1, 201–239.

Showalter, V.M. (1970). Conducting science investigations using computer simulated experiments. Science Teacher, 37, 46–50.

Shute, V.J., & Glaser, R. (1990). A large-scale evaluation of an intelligent discovery world: Smithtown. Interactive Learning Environments, 1, 51–77.

Sime, J.-A. (1998). Model switching in a learning environment based on multiple models. Journal of Interactive Learning Environments, 5, 109–125.

Simmons, P.E., & Lunetta, V.N. (1993). Problem-solving behaviors during a genetics computer simulation: Beyond the expert/novice dichotomy. Journal of Research in Science Teaching, 30, 153–173.

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

Swaak, J., van Joolingen, W.R., & de Jong, T. (1998). Support for simulation-based learning; the effects of model progression and assignments on definitional and intuitive knowledge. Learning and Instruction, 5, 235–253.

Tabak, I., Smith, B.K., Sandoval, W.A., & Reiser, B.J. (1996). Combining general and domain-specific strategic support for biological inquiry. In C. Frasson, G. Gauthier & A. Lesgold (Eds.), Intelligent tutoring systems (pp. 288–297). Berlin: Springer-Verlag.

Teodoro, V.D. (1992). Direct manipulation of physical concepts in a computerized exploratory laboratory. In E. de Corte, M. Linn, H. Mandl, & L. Verschaffel (Eds.), Computer-based learning environments and problem solving (NATO ASI series F: Computer and Systems Series) (pp. 445–465). Berlin: Springer-Verlag.

White, B.Y. (1984). Designing computer games to help physics students understand Newton's laws of motion. Cognition and Instruction, 1, 69–108.

White, B.Y. (1993). ThinkerTools: Causal models, conceptual change, and science education. Cognition and Instruction, 10, 1–100.

White, B.Y., & Frederiksen, J.R. (1989). Causal models as intelligent learning environments for science and engineering education. Applied Artificial Intelligence, 3, 83–106.

White, B.Y., & Frederiksen, J.R. (1990). Causal model progressions as a foundation for intelligent learning environments. Artificial Intelligence, 42, 99–157.

Zietsman, A.I., & Hewson, P.W. (1986). Effect of instruction using microcomputers simulations and conceptual change strategies on science learning. Journal of Research in Science Teaching, 23, 27–39.