



ELSEVIER

Learning and Instruction 20 (2010) 511–520

Learning and  
Instruction

www.elsevier.com/locate/learninstruc

# Offering and discovering domain information in simulation-based inquiry learning

Ard W. Lazonder\*, Mieke G. Hagemans, Ton de Jong

*Department of Instructional Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands*

Received 9 March 2009; revised 8 June 2009; accepted 2 August 2009

## Abstract

The present study investigated how presenting domain information influences scientific reasoning and knowledge acquisition in low prior knowledge students. Fifty-five college freshmen received an inquiry task in an unfamiliar domain and were randomly assigned to a condition in which domain information was available before and during the task, before the task, or not at all. Students in the first two conditions exhibited more hypothesis-driven behavior and acquired more knowledge than students without access to domain information. The comparison among the two conditions with domain information yielded similar results in favor of the before and during condition. Together these findings confirm the predicted superiority of the before-plus-during format.

© 2009 Elsevier Ltd. All rights reserved.

*Keywords:* Scientific reasoning; Inquiry learning; Computer simulations; Prior knowledge

## 1. Introduction

The ability to reason scientifically appears as a goal in contemporary science curricula of many European countries (Abd-El-Khalick et al., 2004; Trends in International Mathematics and Science Study [TIMSS], 2007). Scientific reasoning essentially involves the skills implicated in hypothesis generation, experimentation or observation, and evidence evaluation (Zimmerman, 2007). Gaining proficiency in performing these skills is one of the central aims of inquiry learning, a pedagogy in which learners infer knowledge about a domain by formulating hypotheses and designing and executing experiments to validate these hypotheses. Schools increasingly implement this mode of learning by engaging students in scientific inquiry tasks that involve the use of computer simulations. By interacting with these programs, students practice scientific reasoning in the context of

performing their own investigations, while at the same time develop knowledge of the content modeled by the simulation.

The concurrent development of domain knowledge and scientific reasoning skills requires that students learning by inquiry possess at least a rudimentary understanding of the task or domain they are investigating (De Jong, 2006). The reason underlying this condition is that inaccurate or incomplete domain knowledge can impede the learning process (Hmelo, Nagarajan, & Roger, 2000; Lavoie & Good, 1988; Schauble, Glaser, Raghavan, & Reiner, 1991). In all of these studies, learners with high levels of prior domain knowledge displayed sophisticated scientific reasoning in the inquiry task. Their less knowledgeable counterparts, however, acted less proficiently, and apparently could not seize the opportunity to bring their scientific reasoning to a higher level. Their mediocre scientific reasoning in turn caused low prior knowledge students to infer less domain knowledge from their inquiries than high prior knowledge students did.

The mutual influence of domain knowledge and scientific reasoning points to two directions for supporting low prior knowledge students during inquiry learning, both of which are integral to the concept of scaffolding (Hmelo-Silver, Golan, &

\* Corresponding author. Tel.: +31 53 489 3082; fax: +31 53 489 2849.  
E-mail address: a.w.lazonder@utwente.nl (A.W. Lazonder).

Chinn, 2007). One is to promote students' scientific reasoning by offering process support. This type of assistance aims to help learners in accomplishing scientific inquiry tasks and enables them to learn from their efforts (Reiser, 2004). Following Quintana et al. (2004), process support can address the basic operations of scientific inquiry (e.g., experimentation hints), the strategic decisions involved in controlling the inquiry process (e.g., planning guidance, cues for monitoring), and the process of articulating and reflecting on what has been learned (e.g., reflection prompts). The second option is to provide content support, for instance by giving learners access to conceptual organizers, expert guidance, annotated examples, or by directly offering domain information that discloses the meaning of variables and relations in the inquiry task. The latter type of support has received little attention in educational research, and is central to the work presented in this article.

There is, however, something paradoxical to offering information that learners could otherwise find out by themselves. A seemingly successful solution is to give students just enough information to overcome learning obstacles, thus enabling them to investigate and learn the largest possible part of a domain as efficiently and effectively as possible. Shute (1993), for instance, found that offering definitions and explanations of key task concepts promotes the acquisition and transfer of domain knowledge. Rieber, Tzeng, and Tribble (2004) showed that these effects generalize to students with higher levels of prior knowledge. Hulshof and De Jong (2006) combined information on core concepts in their simulation with experimentation hints. Students who had access to this support achieved higher learning gains than students who did not receive this information. Similar benefits were reported by Reid and colleagues who offered domain information in conjunction with experimental and reflective support (Reid, Zhang, & Chen, 2003; Zhang, Chen, Sun, & Reid, 2004).

The cited research nevertheless leaves it somewhat obscure whether the presence of domain information promotes both knowledge acquisition and scientific reasoning. One reason is that most studies did not assess the influence of domain information in isolation. Other studies have shown that domain information has little added value when offered in combination with other types of support such as experimentation hints, assignments, and reflection prompts (Fund, 2007), or is virtually neglected by students in favor of more process-oriented guidance (Manlove, Lazonder, & De Jong, 2007). A second reason is that the cited studies offer little insight into students' scientific reasoning. In absence of this data it is neither clear what accounts for the observed knowledge gains, nor whether the presence and use of domain information promotes scientific reasoning.

The aim of the present study is a more controlled assessment of how inquiry learners benefit from being given access to domain information. The study took Klahr and Dunbar's (1988) model of Scientific Discovery as Dual Search (SDDS) as starting point to predict when and why offering domain information is effective. These predictions were then tested empirically to establish the influence of domain information

on low prior knowledge students' scientific reasoning and knowledge acquisition.

### 1.1. Theoretical and empirical foundations

Klahr and Dunbar's (1988) SDDS model is a descriptive framework of the reasoning skills involved in scientific inquiry. The model characterizes a student's inquiry activities as a search in two related problem spaces: the hypothesis space and the experiment space. The hypothesis space contains a student's knowledge and conceptions of the relations between the variables in the domain; the experiment space contains all possible experiments that can be conducted with the equipment at hand. Scientific inquiry then proceeds in iterative cycles by searching the hypothesis space for a testable hypothesis, searching the experiment space for experiments to test the hypothesis, and evaluating evidence to verify or refine the hypothesis. Alternatively, if students are unable to retrieve a testable hypothesis from the hypothesis space, they can search the experiment space for exploratory experiments that will help them formulate new hypotheses.

Klahr and Dunbar (1988) used the SDDS model to examine the dynamic interplay between students' domain knowledge and their approach to the inquiry task. Participants in this study were characterized as Theorists or Experimenters based on their experimentation behaviors. Both groups initially used hypotheses to guide the selection of experiments, but their approaches diverged once the initial hypotheses were abandoned. Theorists kept searching the hypothesis space for new hypotheses; Experimenters applied a data-driven approach and explored the experiment space to see if they could induce regularities from experimental outcomes. From these findings it was inferred that Theorists had more prior knowledge which might have influenced the strength of their initial hypotheses and facilitated their search for alternative hypotheses. This supposition was generally confirmed by Lazonder, Wilhelm, and Hagemans (2008) who examined students' performance on two inquiry tasks. Within-subject comparisons of the hypotheses students generated on both tasks showed that most participants adopted a theory-driven approach on the task they had considerable prior knowledge of. Surprisingly, however, approximately one third of these participants acted as Experimenters and changed their initial hypothesis-testing strategy into a data-driven mode of experimentation halfway through their inquiry. A reverse pattern was observed on the task that was situated in an unfamiliar domain. Here most participants started off in a data-driven mode of inquiry and gradually shifted to a theory-driven strategy.

Although both strategies are common in authentic scientific inquiry, a theory-driven mode of inquiry is superior in terms of efficiency and effectiveness. Klahr and Dunbar (1988) found that Theorists acted more efficiently than Experimenters in that they needed less time and fewer experimental trials to successfully complete an inquiry task. Concerning effectiveness, Lazonder, Wilhelm, and Van Lieburg (in press) showed that hypothesis testing was the strongest predictor of knowledge gains in simulation-based inquiry learning, whereas data-

driven experimentation was found to have a strong but negative effect on this measure.

These findings point to implications for the presentation of domain information in scientific inquiry learning. Students with little prior domain knowledge should receive domain information before they start their investigations. This would fill their otherwise empty hypothesis space and thus enable them to engage in a theory-driven mode of experimentation. Domain information should be kept available during the learning process to prevent students from falling back on a data-driven mode of inquiry once they examined their initial hypotheses. Students with high prior knowledge already have a well-equipped hypothesis space that enables them to generate and test specific hypotheses. For these students domain information should merely be offered during the inquiry task to ensure sustained hypothesis-testing behavior. As one never knows when students cannot formulate a new hypothesis, domain information should be available on demand throughout the entire inquiry process.

Empirical evidence on the effectiveness of these presentation formats is scant. And the studies that do exist address only some of the underlying theoretical assumptions. Leutner (1993), for instance, reports two experiments in which domain information was presented before or during the students' inquiries. Although both information-presentation formats showed higher knowledge gains on a written posttest than the control group (who did not receive any domain information), they were neither compared among themselves, nor with a combined before-plus-during condition. More importantly, what accounted for the observed knowledge gains remains unclear because the students' inquiry activities were not included in the analyses. It could be that the domain information enhanced students' scientific reasoning which, in turn, increased knowledge acquisition. But it is just as conceivable that students simply recalled the information they received before or during their inquiry when completing the posttest.

Kester, Kirschner, Van Merriënboer, and Baumer (2001) proposed an information-presentation model that distinguished between declarative information on the conceptual organization of the domain in general, and procedural information about the rules, concepts and facts in a given task within that domain. Participants in their study voiced a need for both types of information before as well as during task performance. Subsequent studies showed that concurrently presenting declarative and procedural information before the task is as effective as making this information jointly available during task performance (Kester, Kirschner, & Van Merriënboer, 2004a, 2004b, 2006). Results for alternative presentation formats (e.g., declarative information before, and procedural information during task performance) were mixed and inconclusive across studies. As with Leutner (1993), the before-plus-during format was not included in these studies, and the impact of consulting declarative and procedural information on students' reasoning was not assessed.

The research described above confirms some of the alleged benefits of presenting domain information before and during scientific inquiry. Questions remain however as to whether and

why the combined before-plus-during presentation format is effective for low prior knowledge students. A possible complicating factor is that students often neglect available support during task performance (Aleven, Stahl, Schworm, Fischer, & Wallace, 2003; Clarebout & Elen, 2006). If so, the before-plus-during format essentially becomes a "before" format, which was deemed less effective for domain novices. The present study took these issues into account by investigating how the presence and timing of domain information influences the scientific reasoning skills "hypothesis generation" and "experimentation" of low prior knowledge students as well as the domain knowledge they acquire from engaging in these processes.

## 1.2. Research design – hypotheses

The basic premise underlying the present research was that domain information should always be available during the inquiry task, and needs to be offered before task performance in case students have insufficient prior knowledge. To test both assumptions, the study examined scientific reasoning and acquired knowledge of low prior knowledge students engaged in a simulation-based inquiry task. Three versions of this task were compared that differed only with regard to the presence and timing of domain information. In the Domain Information Before and During ( $DI_{B+D}$ ) condition domain information was available before as well as during the task. Students in the Domain Information Before ( $DI_B$ ) condition only had access to this information before task performance, and students in the control condition had access to domain information neither before, nor during the task.

For the sake of consistency, it could have been possible to include a fourth condition in which domain information was presented during the task. There are, however, two reasons why this option was rejected. One is a complete lack of theoretical evidence on the benefits of this presentation format for low prior knowledge students. Offering domain information *before* an inquiry was deemed essential to bootstrap low prior knowledge students' scientific reasoning. Without this information, their performance is bound to be suboptimal. A related methodological concern is that low prior knowledge students who face this situation are likely to start their inquiry by reading the domain information, which essentially makes their treatment comparable to that of the  $DI_{B+D}$  condition.

The SDDS model was used to predict cross-condition differences in scientific reasoning and knowledge acquisition. The presence of domain information was expected to provide students with a reasonably well-equipped hypothesis space that enables them to generate and test specific hypotheses. The thus-acquired knowledge is incorporated in the hypothesis space and can be used to generate and test subsequent hypotheses. Students in the  $DI_{B+D}$  and  $DI_B$  condition were therefore predicted to generate more, and more specific hypotheses than students in the control condition. As testing hypotheses is a more efficient and effective mode of inquiry (Klahr & Dunbar, 1988; Lazonder et al., 2008), students in both DI conditions were further assumed to perform fewer

exploratory experiments and acquire more domain knowledge than their control counterparts (Hypothesis 1).

Concerning the timing of domain information, a comparison of the  $DI_{B+D}$  and  $DI_B$  condition was performed to validate the alleged advantages of the before-plus-during presentation format for low prior knowledge students. As both groups had studied the domain information before the task, they were expected to start off in a theory-driven mode of experimentation. However, due to the absence of domain information during the task, students in the  $DI_B$  condition may reach a point in their inquiry where they can no longer generate new hypotheses (cf. Lazonder et al., 2008). They would then have to apply a data-driven approach by searching the experiment space to induce regularities in the domain from exploratory experiments. In contrast,  $DI_{B+D}$  students could consult the domain information in case their investigations failed to produce new hypotheses. Therefore,  $DI_{B+D}$  students were predicted to (a) generate more and more specific hypotheses, (b) conduct fewer exploratory experiments, and (c) develop more domain knowledge than students from the  $DI_B$  condition (Hypothesis 2).

## 2. Method

### 2.1. Participants

Participants were 55 undergraduate students in social sciences who volunteered to participate in the experiment for course credits. There were 18 males and 37 females with a mean age of 20.58 years ( $SD = 1.71$ ). Participants were randomly assigned to either the  $DI_{B+D}$  ( $n = 19$ ), the  $DI_B$  ( $n = 18$ ) or the control condition ( $n = 18$ ). The male-female ratio in these conditions was, respectively, 7:12, 4:14, and 7:11.

### 2.2. Materials

#### 2.2.1. Inquiry task and learning environment

Participants in all three conditions engaged in an inquiry task that invited them to find out how each of five factors influenced a shoe store's weekly sales. These factors were type of background music (jazz, blues, or country), location of the stockroom (basement or first floor), displayed shoe (left or right), type of floor covering (wood, vinyl, or carpet), and the store building's roofline style (stepped, curved, or spout gable). The baseline sales figure was set at 75 pairs of shoes per week (jazz music, stockroom on first floor, left shoe on display, carpet, spout gable). Sales could increase to 89 pairs of shoes depending on the factors' values. Changing the background music to blues or country increased weekly sales by 2 and 4 pairs of shoes respectively. Three extra pairs of shoes were sold each week if the stockroom was situated in the basement. Replacing the carpet floor by a vinyl or wooden floor caused an increase of, respectively, 1 and 2 pairs of sold shoes. A similar increase occurred if the spout gable was replaced by a curved and stepped gable. The effect of displaying the right shoe depended on the type of floor covering.

It improved sales by 3 pairs of shoes if there was a wooden floor, caused an increase of 2 pairs with a vinyl floor, and had no effect in case of a carpet floor.

The inquiry task was designed to ensure that participants would have minimal prior knowledge of the relationships between dependent and independent variables. The task therefore involved factors that, although common in many shoe stores, in reality have an indeterminate impact on weekly sales. Factor values were selected for comparability so that the difference among a factor's values would not evoke expectations about its influence on weekly sales. To illustrate, jazz, blues, and country music were included in the task because of their broad and almost universal appeal. If, for argument's sake, jazz were replaced by a musical genre that does not appeal to mainstream tastes (e.g., heavy metal), participants could infer from everyday knowledge that this factor value is less likely to have a positive impact on weekly sales than blues or country music.

The inquiry task was performed within a learning environment in which participants could experiment through a computer simulation (see Fig. 1). Participants could investigate the impact of a single factor by manipulating its values and observing the effect on the shoe store's weekly sales. Factor values could be set by clicking the corresponding icon. Once all factors were set, participants had to predict the outcome by selecting a value from the pull-down menu. They could then click the Start button to run the experiment; results appeared in the Graph window. The History window showed simulation output along with the selected values, and could be scrolled to review previously conducted experiments. Clicking the Notepad button on the menu bar opened up a simple text editor that allowed participants to take notes and write down their final solutions.

Likewise, clicking the Help button opened up a Helpfile viewer on top of the Graph and History window. Helpfiles in the  $DI_B$  and control condition merely contained the description of the inquiry task and its cover story; helpfiles in the  $DI_{B+D}$  condition, in addition, also conveyed information about all factors within the task. Each factor was covered in a separate helpfile that could be accessed via the helpfile viewer's homepage (see Fig. 2). Helpfiles had a fixed format consisting of (a) the meaning of the factor and its values, (b) a few specifics about the factor (e.g., "All three types of music are played equally loud") to prevent participants from making false assumptions and irrelevant predictions, and (c) the factor's general direction of effect on the weekly sales figure (e.g., "A stockroom in the basement leads to a greater increase in weekly sales than a stockroom on the first floor"). The magnitude of the effect was left undisclosed and had to be inferred by experimenting with the simulation.

#### 2.2.2. Domain information guide

Before being given access to the learning environment, participants in the  $DI_{B+D}$  and  $DI_B$  condition received a self-study guide that introduced them to the variables in the inquiry task. This guide was the paper analog of the domain helpfiles within the  $DI_{B+D}$  version of the learning environment, and

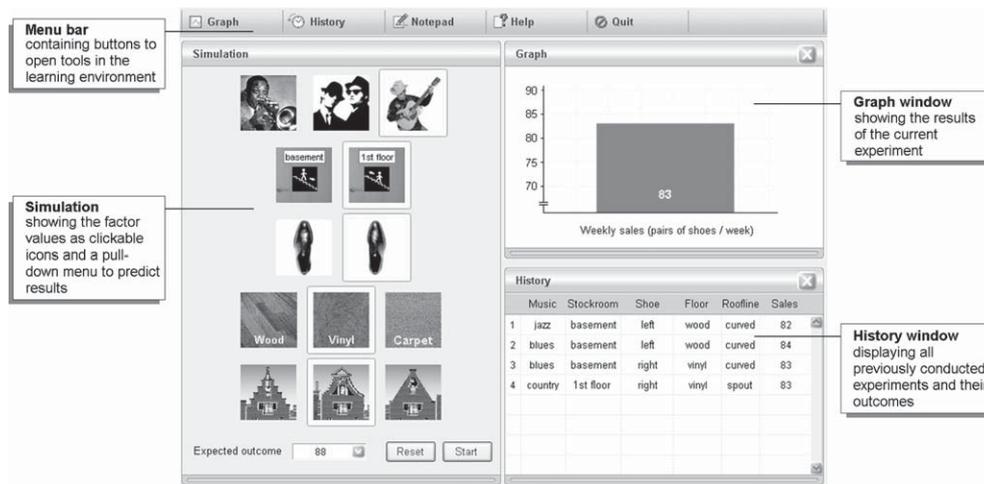


Fig. 1. Annotated interface of the learning environment.

contained the exact same information on the inquiry task and the factors that could be manipulated in the simulation.

### 2.2.3. Comprehension exercise

To ensure that the contents of the domain information guide were well read and well understood, participants in the  $DI_{B+D}$  and  $DI_B$  conditions engaged in a reading comprehension exercise consisting of 11 short-answer questions about the variables in the inquiry task. Six items involved either the meaning of a variable (e.g., “What does ‘weekly sales’ mean?”) or its specifics (e.g., “Are all three types of music being played equally loud?”). The remaining five items addressed the influence of one of the independent variables on weekly sales (e.g., “Which type of floor covering leads to a greater increase in weekly sales than vinyl does?”).

Participants could consult the domain information guide while answering these questions, which enabled them to reconsider pieces of information they were unsure about or could not recall. Although this opportunity increased the likelihood that participants in both DI conditions would have a well-equipped hypothesis space, it ruled out the possibility to use the comprehension exercise as an assessment of participants’ familiarity with the task content (i.e., comprehension exercise scores represent the upper bounds of participants’ prior domain knowledge).

Prior knowledge was not assessed in the control condition for three reasons. One is that all relations between the variables in the simulation were arbitrary and therefore unknown to participants. Second, gauging control participants’ knowledge of these relationships might evoke expectations about the

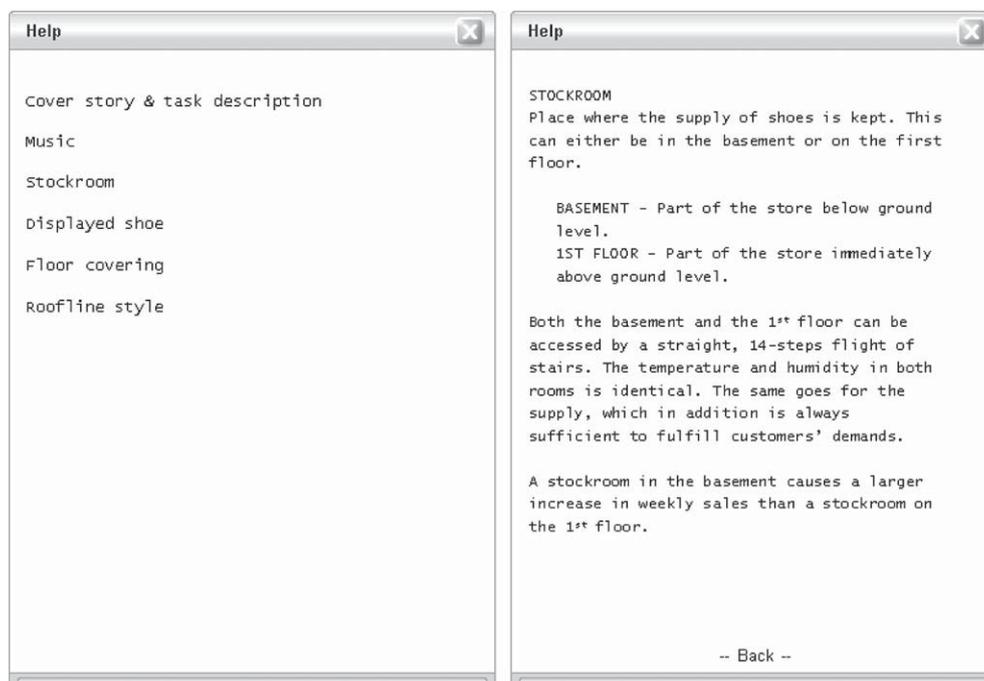


Fig. 2. Screen shots of the Helpfile viewer used in the  $DI_{B+D}$  condition. The left image displays the homepage from which all helpfiles could be accessed. The right image shows the helpfile for the factor ‘stockroom’; details on the types of information within this page appear in the text.

factors' influence on weekly sales, which could subvert the required between-condition differences at the start of the experiment. Finally, although it is probably safe to assume that most control participants were familiar with the factors' meaning, this knowledge need not be assessed because it provides an insufficient basis for hypothesis generation (Lazonder et al., in press).

### 2.3. Procedure

Students participated in the experiment one at a time, receiving the same instructions and following the same experimental procedures. At the beginning of a session, the experimenter explained the experimental procedures and demonstrated the operation of the tools within the learning environment by means of a simple inquiry task (determining the costs of a skiing holiday). (Participants' demographic characteristics were collected when they signed up for the experiment.)

Participants in the  $DI_{B+D}$  and  $DI_B$  condition then entered the preparation phase in which they were asked to study the domain information guide and complete the comprehension exercise. Neither of these activities was time constrained, and participants received the comprehension exercise once they indicated they were done reading. To ensure that participants in both  $DI$  conditions paid similar attention to the contents of the domain information guide,  $DI_{B+D}$  participants were not informed that this information would also be available in their version of the learning environment until after they had handed in their answers to the comprehension exercise. Participants in the control condition skipped the preparation phase and went straight to the execution part of the session.

Upon entering the execution phase, participants logged into the environment and started reading the cover story that introduced the factors in the simulation and the task's problem statement. They then started experimenting with the simulation to investigate how each of the five factors influenced the shoe store's weekly sales. All participants could consult an index card to see how each factor was visually represented in the simulation interface. They used the Notepad facility to take notes during the activity and write down their final solutions. Participants were allowed 45 min maximum to complete the inquiry task. Their interactions with the environment were recorded in a logfile.

During the execution phase the experimenter asked participants about their hypotheses. A hypothesis was defined as a statement of the factor under investigation and the presence, direction, or magnitude of its effect on the dependent variable. While questioning has been criticized for prompting participants to an underlying goal structure for systematic experimentation (Klahr & Carver, 1995), research has shown that non-directive probes have no influence on participants' inquiry learning processes (Wilhelm & Beishuizen, 2004). Hence two non-directive questions were used to elicit the factor under investigation ("What are you going to investigate?"), and its alleged effect on the output variable ("What do you think will be the outcome?"). These questions were

asked every time participants could be testing a hypothesis. That is, questioning occurred whenever a participant had set-up a new experiment and selected an expected outcome value from the pull-down menu. The experimenter wrote down the participant's responses on a scoring sheet. The reliability of this registration method was assessed by having two raters simultaneously record the responses in five experimental sessions. Analysis of their scoring sheets demonstrated 90% inter-rater agreement.

### 2.4. Measures and scoring procedure

Participants' answers to the 11 items of the comprehension exercise were checked against the contents of the domain information guide and scored as true or false. Data on the use of the domain helpfiles was gathered from the logfiles.

The main variables of the study were time on task, the experiments participants conducted, the hypotheses they proposed, and their final knowledge of relations between the variables in the simulation (i.e., performance success). With *time*, a distinction was made between the preparation phase and the execution phase. During the preparation phase the experimenter used a stopwatch to record the time to read the domain information guide and complete the comprehension exercise. Time during the execution phase concerned the duration of the learning environment session and was assessed from the logfiles.

Participants' *experiments* were also recorded from the logfiles, and classified as either unique or duplicated. Simulation runs testing a new, untried combination of factor values were coded as a unique experiment; if this combination reappeared in subsequent trials it was considered a duplicated experiment. The maximum number of unique experiments was 108 since there were this many distinct combinations of factor values in the inquiry task; the minimum number of unique experiments to discover the simulation's underlying model was 11. Each experiment was further classified according to the presence or absence of a hypothesis (the definition of a hypothesis is given below). This categorization was used to calculate the percentage of exploratory experiments as the ratio of the number of experiments that were not guided by a hypothesis to the total number of experiments.

Participants' *hypotheses* were scored from their responses to the probing questions. A hierarchical rubric was used to classify the hypotheses according to their level of domain specificity. A distinction was made between fully-specified, partially-specified, and unspecified hypotheses (cf. Van Joolingen & De Jong, 1991). A fully-specified hypothesis comprised two or more factor values and a prediction of the direction *and* magnitude of the effect ("I think blues music improves weekly sales by 3 pairs of shoes compared to jazz music"). Partially-specified hypotheses predicted the direction of effect of two or more factor values ("I think blues music improves weekly sales compared to jazz music"). Unspecified hypotheses merely denoted the existence of an effect ("I think blues music influences weekly sales"). Statements of experimentation plans ("I am going to investigate blues music") or

ignorance (“I have no idea”) were not considered hypotheses. Two raters coded the hypotheses of 4 randomly selected scoring sheets of each condition. Together these 12 sheets contained 195 hypotheses; inter-rater agreement (Cohen’s  $\kappa$ ) reached 1.00.

*Performance success* was assessed from the participants’ saved Notepad files. A hierarchical rubric was used to transform this data into an overall performance success score. Up to 3 points could be earned for each factor mentioned, leading to a maximum score of 15 points. Three points were awarded for a factor if both the magnitude and direction of the effect were correct for each value of that factor. Two points were given if the direction of the effect was correct but its magnitude was (partly) incorrect or incomplete. One point was awarded if the answer expressed that a factor affected weekly sales, but neither the magnitude nor the direction of this effect was specified correctly. In all cases, “correct” was judged from the simulation’s underlying model. Two raters used this rubric to score a randomly selected set of 6 Notepad files for each condition. Inter-rater reliability was satisfactory (Cohen’s  $\kappa = .79$ ).

### 3. Results

Table 1 summarizes the descriptive statistics for participants’ performance by condition. Data for time on task

Table 1  
Means (and *SD*) for participants’ performance by condition.

	Condition		
	DI <sub>B+D</sub> ( <i>n</i> = 19)	DI <sub>B</sub> ( <i>n</i> = 18)	Control ( <i>n</i> = 18)
Time on task (min.)			
Preparation phase <sup>a</sup>	10.44 (1.94)	10.98 (1.97)	–
Execution phase	31.27 (8.46)	33.41 (6.35)	28.76 (10.51)
Comprehension exercise			
Overall score <sup>a</sup>	10.74 (0.45)	10.56 (0.51)	–
Number of experiments			
Unique experiments	12.95 (2.04)	14.94 (3.47)	16.39 (7.24)
Duplicated experiments	0.68 (1.11)	1.22 (2.53)	1.67 (3.33)
Exploratory experiments (%)	1.92 (3.98)	4.40 (5.49)	41.95 (15.17)
Hypotheses <sup>b</sup>			
First quartile (Q1)	1.90 (0.21)	1.77 (0.33)	0.88 (0.65)
Second quartile (Q2)	2.14 (0.25)	1.92 (0.37)	1.20 (0.64)
Third quartile (Q3)	2.31 (0.29)	2.02 (0.35)	1.42 (0.60)
Fourth quartile (Q4)	2.45 (0.41)	2.14 (0.49)	1.52 (0.79)
Performance success			
Overall score	13.21 (1.93)	10.44 (3.31)	7.67 (2.22)
Magnitude of effect <sup>c</sup>	4.16 (0.90)	2.67 (1.41)	0.78 (1.26)

DI<sub>B+D</sub> = domain information before and during the task; DI<sub>B</sub> = domain information before the task; Control = control condition (no domain information).

<sup>a</sup> Students in the control condition skipped the preparation phase (including the comprehension exercise).

<sup>b</sup> Scores represent the mean domain specificity of the participants’ hypotheses within each quartile.

<sup>c</sup> Scores represent the mean number of factors of which the magnitude of effect was successfully discovered.

indicated that participants in the DI<sub>B+D</sub> and DI<sub>B</sub> condition needed about as much time during the preparation phase to study the domain information guide and complete the comprehension exercise. Univariate analysis of variance (ANOVA) confirmed that this difference in time was not significant,  $F(1, 35) = 0.71$ ,  $p = .41$ . Performance on the comprehension exercise further showed that the domain information was well read and well understood. That is, 24 participants achieved a perfect score; the remaining participants answered all but one item correctly. An ANOVA produced no cross-condition difference in scores,  $F(1, 35) = 1.31$ ,  $p = .26$ .

Participants in the DI<sub>B+D</sub> condition could consult the contents of the domain information guide at will during their inquiry via helpfiles embedded within the learning environment. Helpfile viewing was observed in 17 of the 19 DI<sub>B+D</sub> participants, with frequency scores ranging from 1 to 9 and an average duration of 1.15 min ( $SD = 0.80$ ). One-sample *t*-test showed that the mean number of helpfile views ( $M = 4.32$ ,  $SD = 2.94$ ) differed significantly from zero,  $t(18) = 6.39$ ,  $p < .01$ , Cohen’s  $d = 1.47$ , which confirmed the difference in treatment across the two DI conditions. This disparity notwithstanding, participants in all three conditions needed approximately 30 min to complete the inquiry task during the execution phase. A univariate ANOVA<sup>1</sup> showed that the relatively minor deviations in time reported in Table 1 were not statistically significant,  $F(2, 33.33) = 1.34$ ,  $p = .28$ .

Table 1 also gives an account of the experiments participants performed with the simulation. Multivariate analysis of variance (MANOVA) yielded no significant effect of instructional condition on the mean number of unique and duplicated experiments, Pillai’s trace = 0.09,  $F(4, 102) = 1.19$ ,  $p = .32$ . There was, however, a significant univariate difference in the percentage of exploratory experiments,  $F(2, 30.13) = 57.59$ ,  $p < .01$ , partial  $\eta^2 = .80$ . Planned contrasts revealed that participants from both DI conditions together performed relatively fewer exploratory experiments than participants from the control condition,  $t(18.68) = 10.60$ ,  $p < .01$ , Cohen’s  $d = 7.79$ . The difference in the percentage of exploratory experiments between the DI<sub>B+D</sub> and DI<sub>B</sub> condition was not significant,  $t(30.94) = 1.57$ ,  $p = .13$ . Within the DI<sub>B+D</sub> condition, the percentage of exploratory experiments was associated with the frequency and time of helpfile viewing ( $r = .37$ ,  $p = .06$ , and  $r = .41$ ,  $p = .04$ , respectively). While these correlations seem to suggest that participants who could not generate new hypotheses turned to the helpfiles after having conducted some exploratory experiments, the percentage of exploratory experiments was too low to validate this pattern in scores from the logfiles.

The hypotheses participants generated were analyzed through time and across conditions. Toward this end every participant’s hypotheses were sequenced chronologically.

<sup>1</sup> In all univariate ANOVAs, the Welch *F*-ratio is reported in case the homogeneity of variance assumption was violated; this is reflected by the decimal value for the residual degrees of freedom.

Hypotheses were then divided into quartiles according to their order of appearance, and the mean domain specificity scores of the hypotheses within each quartile were computed (see Table 1). These scores characterize domain specificity as the mean number of constituent elements the hypotheses in a given quartile contained (i.e., presence, direction, and magnitude of an effect; see Section 2.4).

A mixed-design ANOVA was performed to analyze how the domain specificity of participants' hypotheses evolved through time in each condition. As the sphericity assumption appeared to be violated,  $\chi^2(5) = 11.12$ ,  $p < .05$ , the Greenhouse-Geisser coefficient was used for the within-subject analysis. Results showed a within-subject effect of time,  $F(2.66, 138.24) = 17.61$ ,  $p < .01$ , partial  $\eta^2 = .25$ , a between-subject effect of condition,  $F(2, 52) = 4.41$ ,  $p < .01$ , partial  $\eta^2 = .61$ , and no significant interaction,  $F(5.32, 138.24) = 0.53$ ,  $p = .79$ . As visualized in Fig. 3, the hypotheses of participants in all conditions gradually became more domain specific. Planned contrasts bore this out in that significant within-subject differences were found between the first and second quartile,  $F(1, 52) = 10.43$ ,  $p < .01$ , Cohen's  $d = 0.44$ , and the second and third quartile,  $F(1, 52) = 7.60$ ,  $p < .01$ , Cohen's  $d = 0.38$ . The increase in domain specificity between the third and fourth quartile did not reach significance,  $F(1, 52) = 2.98$ ,  $p = .09$ . Fig. 3 also shows that participants from both DI conditions consistently generated more specific hypotheses than participants from the control condition. Planned contrasts showed that these differences were significant for all four quartiles (see Table 2). Differences in domain specificity between the  $DI_{B+D}$  and  $DI_B$  condition were significant for all quartiles except the first.

An overall performance success score reflected the extent to which participants' knowledge of the relations between the variables in the inquiry task matched the simulation's underlying model. A univariate ANOVA showed that the overall means displayed in Table 1 differed across conditions,  $F(2,$

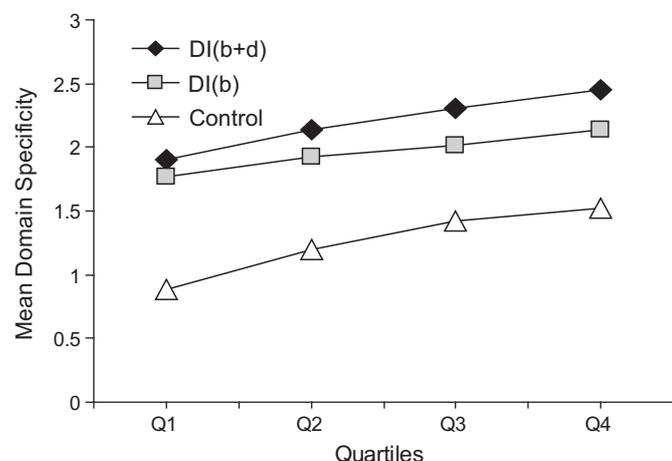


Fig. 3. Mean domain specificity of participants' hypotheses through time and by condition.  $DI_{B+D}$  = domain information before and during the task;  $DI_B$  = domain information before the task; Control = control condition (no domain information).

Table 2

Summary of planned contrasts for the effects of condition on the domain specificity of participants' hypotheses.

	$(DI_{B+D} + DI_B)$ vs. Control		$DI_{B+D}$ vs. $DI_B$	
	<i>t</i>	Cohen's <i>d</i>	<i>t</i>	Cohen's <i>d</i>
First quartile (Q1)	5.97**	2.23	1.36	—
Second quartile (Q2)	5.21**	1.87	2.08*	.72
Third quartile (Q3)	4.90**	1.71	2.67*	.93
Fourth quartile (Q4)	3.87**	1.34	2.08*	.71

$DI_{B+D}$  = domain information before and during the task;  $DI_B$  = domain information before the task; Control = control condition (no domain information).

\*\* $p < .01$  \* $p < .05$ .

52) = 21.88,  $p < .01$ , partial  $\eta^2 = .46$ . Planned contrasts revealed that participants who had access to domain information before and/or during their inquiry (i.e.,  $DI_{B+D}$  and  $DI_B$  together) outperformed participants from the control condition,  $t(52) = 5.68$ ,  $p < .01$ , Cohen's  $d = 1.54$ . Participants from the  $DI_{B+D}$  condition achieved higher performance success scores than  $DI_B$  participants,  $t(52) = 3.30$ ,  $p < .01$ , Cohen's  $d = 1.06$ . Performance success was negatively correlated with exploratory experiments,  $r = -.59$ ,  $p < .01$ , indicating that participants who exhibited more hypothesis-testing behavior also acquired more domain-specific knowledge.

As participants in both DI conditions were informed about the presence and direction of effects, acquired knowledge of the magnitude of effects (which was left undisclosed to all participants) was also analyzed separately. A univariate ANOVA showed a significant effect for condition,  $F(2, 52) = 36.40$ ,  $p < .01$ , partial  $\eta^2 = .58$ . Planned contrasts indicated that on average participants from both DI conditions discovered the magnitude of the factors' effect more often than their control counterparts,  $t(52) = 7.60$ ,  $p < .01$ , Cohen's  $d = 2.00$ .  $DI_{B+D}$  participants, in turn, discovered the magnitude of effects more often than participants in the  $DI_B$  condition,  $t(52) = 3.76$ ,  $p < .01$ , Cohen's  $d = 1.30$ . Again a substantial negative correlation was found with exploratory experiments,  $r = -.65$ ,  $p < .01$ .

#### 4. Discussion

The present study investigated how the presence and timing of domain information influences scientific reasoning and knowledge acquisition in low prior knowledge students. Presenting domain information per se was predicted to enhance hypothesis-driven inquiry and knowledge acquisition. Presenting domain information at the right time (i.e., before as well as during the inquiry) was further expected to yield similar advantages over a situation where domain information is merely available before the inquiry. Both hypotheses were supported by the results.

Evidence for Hypothesis 1 comes from the comparison of both DI conditions with the control condition. As predicted, participants, who had access to domain information before and/or during the task, performed fewer exploratory experiments and generated more specific hypotheses than

participants in the control condition did. Participants in both DI conditions also acquired more knowledge – even about parts of the task *not* revealed by the domain information – which confirms findings from prior research in which domain information was offered in conjunction with other types of support (Hulshof & De Jong, 2006; Reid et al., 2003; Shute, 1993; Zhang et al., 2004). The results also extend those from prior research by demonstrating that differences in acquired knowledge are not simply due to the fact that DI-participants copy domain information to their answer forms (and the experimenter witnessed no such behavior during any of the sessions), but actually use this information to generate and test more, and more specific hypotheses which allows them to induce more knowledge from their interactions with the simulation.

Hypothesis 2 was confirmed by the comparison between the two DI conditions. Consistent with expectations, participants in the  $DI_{B+D}$  condition performed fewer exploratory experiments overall, and generated more specific hypotheses after the first few experiments had been conducted than their  $DI_B$  counterparts. This suggests that keeping domain information available during the task can help students to maintain a hypothesis-driven mode of inquiry. This was substantiated by the complete absence of substantial decreases (i.e., half a point or more) in domain specificity of participants' hypotheses in the  $DI_{B+D}$  condition, whereas such drops in domain specificity occurred 4 and 8 times in the  $DI_B$  and control conditions respectively. The domain-specificity scores of the hypotheses in the third and fourth quartile further indicate that participants in the  $DI_{B+D}$  condition had specific predictions about the magnitude of effects, which was less often the case in the  $DI_B$  condition. Magnitudes of effect were also discovered more often in the  $DI_{B+D}$  condition, and overall performance success was higher in this condition as well.

These results, although consistent with expectations, are particularly noteworthy because students have often been found to disregard available supports (Clarebout & Elen, 2006). Participants in the  $DI_{B+D}$  condition, in contrast, did consult the domain helpfiles even though they had studied this information thoroughly before the task. Why did the participants in the  $DI_{B+D}$  condition decide to revisit this information? A possible explanation could be that studying the domain information guide made them aware of the applicability of the domain information to their task (Aleven et al., 2003). While the domain information was still fresh in their minds, participants from the  $DI_{B+D}$  condition apparently found it useful to check their understanding – and the time and frequency scores indicate that this is how they used the domain helpfiles. These quick checks ostensibly helped to generate more, and more specific hypotheses, which substantiates the added value of keeping domain information available during the task. Still, helpfile usage may have an even more pronounced impact if the lag between preparation and execution phase is extended. This increase in time causes learners to recall less of the information studied beforehand, which presumably increases the time and frequency of helpfile viewing. Future research should investigate this assumption.

The present findings thus confirm the alleged benefits of the before-plus-during presentation format for low prior knowledge students. The next logical step would be to validate the proposed ideal format for high prior knowledge students. As these students already have a well-equipped hypothesis space, merely offering domain information during the task was deemed sufficient to ensure sustained hypothesis testing and enhance performance (cf. Rieber et al., 2004). The present findings also point to an extension of the SDDS model. Klahr and Dunbar (1988) assumed that hypotheses can be generated either from prior knowledge or by generalizing from the results of prior experiments. The present study suggests that domain information is a third source of hypotheses. Its effectiveness to bootstrap hypothesis generation seems to hinge on the incompleteness of the offered information for it stimulates students to supplement the given facts with insights gained from their own, hypothesis-driven experiments.

However, several methodological limitations merit attention. First, data on participants' hypotheses were collected by concurrent probing. As this method is insensitive to unconscious cognitive processes, participants could have had more, or more specific expectations than they actually reported. Although this shortcoming does not rule out cross-condition comparisons, obtained data do not allow for definitive conclusions on the hypotheses' level of domain specificity per se. A related concern is whether the probing questions affected helpfile viewing. Probes could have stimulated learners to attend to information that would otherwise remain unattended, and therefore have the potential to affect learning. While the chances for this to happen are few (see Wilhelm & Beishuizen, 2004), they cannot be ruled out completely. Direct comparisons with other studies on helpfile usage should therefore be made with some caution. Third, the inquiry task is somewhat in the realm of fiction. Although this allows for maximum control of participants' prior knowledge, it also challenges the generalizability of this study's findings to inquiry learning in science classrooms.

There is, however, reason to be optimistic about the generalizability of findings. While the inquiry task may be a little contrived, it does call for hypothesis formation and experimentation, thus reflecting the prevailing research activities conducted in science classrooms (Wilhelm & Beishuizen, 2003). The way students go about performing these activities is relatively independent of the task domain. Veenman, Wilhelm, and Beishuizen (2004) showed that students' use of the control-of-variables strategy is consistent across tasks and domains (cf. Klahr & Li, 2005). Lazonder et al. (2008) demonstrated that students in general, and Theorists in particular, follow the same approach to inquiry tasks they are knowledgeable and incognizant of. Similar conclusions were reached by Mulder, Lazonder, and De Jong (submitted for publication), who used a genuine school science task (charging a capacitor in an electrical circuit) to portray differences in scientific reasoning among domain novices and experts. Together these studies suggest that the present findings generalize to more ecologically valid settings. Testing this assumption is an interesting avenue for future research that will contribute to the ongoing debate about

whether laboratory studies can generate usable knowledge for educational practice.

Practical implications, although tentative, pertain to classroom situations where students generate and test hypotheses to learn about a new topic or domain. This pedagogical approach prevails in science classes and is generally performed with the aid of computer simulations or in a school lab. The present study suggests that students should first receive a short introduction into the subject matter, and be able to consult the information used therein during their inquiry. The latter could be achieved by embedding domain information in the learning environment, offering handouts, or allowing learners to bring their own notes to the task. The importance of a topical introduction is generally acknowledged (e.g., De Jong, 2006); the importance of sustained access to domain information is not. Yet consulting domain information during the inquiry has a distinct added value, and might become even more important if the time gap between the introduction and inquiry task performance increases – which is already the case when students prepare themselves for class at home.

## References

- Abd-El-Khalick, F., BouJaoude, S., Duschl, R., Lederman, N., Mamiok-Naaman, R., Hofstein, A., et al. (2004). Inquiry in science education: international perspectives. *Science Education*, 88, 397–419.
- Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, 73, 277–320.
- Clarebout, G., & Elen, J. (2006). Tool use in computer-based learning environments: towards a research framework. *Computers in Human Behavior*, 22, 389–411.
- De Jong, T. (2006). Scaffolds for scientific discovery learning. In J. Elen, & R. E. Clark (Eds.), *Handling complexity in learning environments: Theory and research* (pp. 107–128). London: Elsevier.
- Fund, Z. (2007). The effects of scaffolded computerized science problem-solving on achievement outcomes: a comparative study of support programs. *Journal of Computer Assisted Learning*, 23, 410–424.
- Hmelo, C. E., Nagarajan, A., & Roger, S. (2000). Effects of high and low prior knowledge on construction of a joint problem space. *Journal of Experimental Education*, 69, 36–56.
- Hmelo-Silver, C. E., Golan, D. R., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: a response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42, 99–107.
- Hulshof, C. D., & De Jong, T. (2006). Using just-in-time information to support discovery learning about geometrical optics in a computer-based simulation. *Interactive Learning Environments*, 14, 79–94.
- Kester, L., Kirschner, P. A., & Van Merriënboer, J. J. G. (2004a). Information presentation and troubleshooting in electrical circuits. *International Journal of Science Education*, 26, 239–256.
- Kester, L., Kirschner, P. A., & Van Merriënboer, J. J. G. (2004b). Timing of information presentation in learning statistics. *Instructional Science*, 32, 233–252.
- Kester, L., Kirschner, P. A., & Van Merriënboer, J. J. G. (2006). Just-in-time information presentation: improving learning a troubleshooting skill. *Contemporary Educational Psychology*, 31, 167–185.
- Kester, L., Kirschner, P. A., Van Merriënboer, J. J. G., & Baumer, A. (2001). Just-in-time information presentation and the acquisition of complex cognitive skills. *Computers in Human Behavior*, 17, 373–391.
- Klahr, D., & Carver, S. M. (1995). Commentary: scientific thinking about scientific thinking. *Monographs of the Society for Research in Child Development*, 60, (4, Serial No. 245).
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12, 1–48.
- Klahr, D., & Li, J. (2005). Cognitive research and elementary science instruction: from the laboratory, to the classroom, and back. *Journal of Science Education and Technology*, 14, 217–238.
- Lavoie, D. R., & Good, R. (1988). The nature and use of prediction skills in a biological computer simulation. *Journal of Research in Science Teaching*, 25, 335–360.
- Lazonder, A. W., Wilhelm, P., & Hagemans, M. G. (2008). The influence of domain knowledge on strategy use during simulation-based inquiry learning. *Learning and Instruction*, 18, 580–592.
- Lazonder, A. W., Wilhelm, P., & Van Lieburg, E. Unraveling the influence of domain knowledge during simulation-based inquiry learning. *Instructional Science*, (in press). doi: 10.1007/s11251.008.9055.8.
- Leutner, D. (1993). Guided discovery learning with computer-based simulation games: effects of adaptive and non-adaptive instructional support. *Learning and Instruction*, 3, 113–132.
- Manlove, S., Lazonder, A. W., & De Jong, T. (2007). Software scaffolds to promote regulation during scientific inquiry learning. *Metacognition and Learning*, 2, 141–155.
- Mulder, Y., Lazonder, A. W., & De Jong, T. Finding out how they find it out: an empirical analysis of inquiry learners' need for support, Manuscript submitted for publication.
- Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., et al. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences*, 13, 337–386.
- Reid, D. J., Zhang, J., & Chen, Q. (2003). Supporting scientific discovery learning in a simulation environment. *Journal of Computer Assisted Learning*, 19, 9–20.
- Reiser, B. J. (2004). Scaffolding complex learning: the mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences*, 13, 273–304.
- Rieber, L. P., Tzeng, S., & Tribble, K. (2004). Discovery learning, representation, and explanation within a computer-based simulation: finding the right mix. *Learning and Instruction*, 14, 307–323.
- Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *The Journal of the Learning Sciences*, 1, 201–238.
- Shute, V. J. (1993). A comparison of learning environments: all that glitters. In S. P. Lajoie, & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 47–75). Hillsdale, NJ: Erlbaum.
- Trends in International Mathematics and Science Study. (2007). *TIMSS 2007 assessment frameworks*. Retrieved 9.06.08, from [http://timss.bc.edu/TIMSS2007/PDF/T07\\_AF\\_chapter2.pdf](http://timss.bc.edu/TIMSS2007/PDF/T07_AF_chapter2.pdf).
- Van Joolingen, W. R., & De Jong, T. (1991). Supporting hypothesis generation by learners exploring an interactive computer-simulation. *Instructional Science*, 20, 389–404.
- Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction*, 14, 89–109.
- Wilhelm, P., & Beishuizen, J. J. (2003). Content effects in self-directed inductive learning. *Learning and Instruction*, 13, 381–402.
- Wilhelm, P., & Beishuizen, J. J. (2004). Asking questions during self-directed inductive learning: effects on learning outcome and learning processes. *Interactive Learning Environments*, 12, 251–264.
- Zhang, J., Chen, Q., Sun, Y., & Reid, D. J. (2004). Triple scheme of learning support design for scientific discovery learning based on computer simulation: experimental research. *Journal of Computer Assisted Learning*, 20, 269–282.
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 27, 172–223.