Finding out how they find it out: an empirical analysis of inquiry learners' need for support

Mulder, Yvonne G.; Lazonder, Ard W.; Jong, Ton de

Postprint / Postprint
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:
Mulder, Y. G., Lazonder, A. W., & Jong, T. d. (2009). Finding out how they find it out: an empirical analysis of inquiry learners' need for support. *International Journal of Science Education*, 1-37. https://doi.org/10.1080/09500690903289993

Finding out how they find it out: An empirical analysis of inquiry learners' need for support

Journal:	*International Journal of Science Education*
Manuscript ID:	TSED-2009-0151.R1
Manuscript Type:	Research Paper
Keywords:	model-based learning
Keywords (user):	scientific reasoning, prior knowledge, inquiry learning

Abstract

Inquiry learning environments increasingly incorporate modelling facilities for students to articulate their research hypotheses and (acquired) domain knowledge. This study compared performance success and scientific reasoning of university students with high prior knowledge (n = 11), students from senior high-school (n = 10), and junior high-school (n = 10) with intermediate and low prior knowledge respectively, in order to reveal domain novice's need for support in such environments. Results indicated that the scientific reasoning of both groups of high-school students was comparable to that of the experts. As high-school students achieved significantly lower performance success scores, their expert-like behaviour was rather ineffective; qualitative analyses substantiated this conclusion. Based on these findings, implications for supporting domain novices in inquiry learning environments are advanced.

Finding out how they find it out: An empirical analysis of inquiry learners' need for support

Computer-supported inquiry learning environments essentially enable students to learn science by doing science, offering resources to develop a deep understanding of a domain by engaging in scientific reasoning processes such as hypothesis generation, experimentation, and evidence evaluation. The central aim of this investigative learning mode is twofold: students should develop domain knowledge and proficiency in scientific inquiry (cf. Gobert & Pallant, 2004). Unfortunately the educational advantages of inquiry learning are often challenged by students' poor inquiry skills (e.g., de Jong & van Joolingen, 1998). Researchers and designers therefore often attempt to compensate for students' skill deficiencies by offering support such as proposition tables to help generate hypotheses (Shute, Glaser, & Raghavan, 1989), adaptive advice for extrapolating knowledge from simulations (Leutner, 1993), or regulative scaffolds to assist students in planning, monitoring, and evaluating their inquiry (Davis & Linn, 2000; Manlove, Lazonder, & de Jong, 2006)

Deleted: (de Jong, 2006b)

Although much has been learned from these approaches, the empirical foundations underlying the contents of these support tools often remain hidden to the public eye. The work of Quintana et al. (2004) forms a notable exception. They argued that more insight into the specific problems students face is called for, and accordingly based their scaffolding framework on a descriptive analysis of students' inquiry learning problems. Yet even this well-documented framework lacks a specific frame of reference: if anything, there is an implicit reference to expert behaviour as yardstick of proficiency.

This study therefore sought to gain insight into students' scientific reasoning skill deficiencies by contrasting domain novices' inquiry behaviour and performance to that of a considerably more knowledgeable reference group (hereafter: experts). A group of students with intermediate levels of prior knowledge was included in this comparison to shed more light on the developmental trajectories of students' scientific reasoning and domain knowledge. Before elaborating the design of the study, a brief overview of the literature is given in order to contextualize the design rationale. This overview starts from classic novice-expert literature and results in a descriptive framework of the core scientific reasoning processes.

Theoretical background

Novice-expert differences have been studied extensively in the field of problem solving. This research has identified key characteristics of expert performance, some of which were found to be robust and generalizable across domains. In short, problem solving research has shown that people who have developed expertise in a certain area mainly excel within that area, perceive large meaningful patterns in their domain of expertise, perform fast (even though they spend a great deal of time analysing a problem), and have superior short-term and long-term memory. Experts also represent a problem in their domain at a deeper, more principled level than novices do and have strong self-monitoring skills (Bransford, Brown, & Cocking, 2002; Chi, Glaser, & Farr, 1988).

These general characteristics, although informative, are not specific enough to guide instructional designers and science educators in determining what exactly their support should focus on. A further complicating issue is that novice-expert differences in problem solving do not necessarily generalize to inquiry learning. According to Batra and Davis (1992), most problem solving tasks require participants to find a unique correct solution. In inquiry learning this search for a single optimal outcome (often referred to as an engineering approach) is generally considered less effective in facilitating students' understanding of a domain than a so-called science model of experimentation (Schauble, Klopfer, & Raghavan, 1991). Performing an inquiry task effectively and efficiently might thus require different skills and strategies than proficient problem solving does. As a result, the general instructional implications from problem solving research should be substantiated by, or supplemented with, insights gleaned from novice-expert differences in inquiry learning.

Inquiry learning attempts to mimic authentic scientific inquiry by engaging students in processes of orientation, hypothesis generation, experiment design, and data interpretation to reach conclusions (Shrager & Klahr, 1986; Zimmerman, 2007). While some have argued that the inquiry tasks given to students in schools evoke different cognitive processes than the ones employed in real scientific research (Chinn & Malhotra, 2002), the advancement of computer technology has significantly narrowed this gap. Contemporary electronic learning environments offer a platform for students to examine scientific phenomena through computer simulations. These environments increasingly provide opportunities for students to build computer models of the phenomena they are investigating. As in authentic scientific inquiry, modelling is considered an integral part of the inquiry learning process. Students can use models to express their understanding of a relation between variables (Jackson, Stratford, Krajcik, & Soloway, 1994; White, Shimoda, & Frederiksen, 1999); these propositions can be tested by running the model; evidence evaluation then occurs by weighting model output against prior knowledge or the data from the simulation. These comparisons yield further insight into the phenomenon and assist students in generating new hypotheses.

Deleted: (de Jong, 2006a)

Deleted: (van Joolingen, de Jong, Lazonder, Savelsbergh, & Manlove, 2005; White, Shimoda, & Frederiksen, 1999).

Deleted: .

Deleted: T

The effectiveness and efficiency with which students perform these processes can be expected to differ as function of their level of domain expertise. In the present research, Klahr and Dunbar's (1988) SDDS model was used to describe and explain these differences. This descriptive framework captures the core scientific reasoning processes and is sensitive to students' evolving domain knowledge. SDDS conceives of scientific reasoning as a search in two problem spaces (hence its name: Scientific Discovery as Dual Search): the hypothesis space and the experiment space. The former space comprises the hypotheses a learner can generate during the inquiry process; the latter consists of all possible experiments that can be conducted with the equipment at hand. Search in the hypothesis space is guided by either prior knowledge or experimental results. Search in the experiment space can be guided by the current hypothesis; in case learners do not have a hypothesis they can search the experiment space for exploratory experiments that will help them formulate new hypotheses.

According to the SDDS model, inquiry learning consists of three iterative processes: hypothesizing, experimenting, and evaluating evidence. The way students perform these processes is assumed to depend on their knowledge of the task domain. Students with domain expertise can generate hypotheses from prior knowledge and then test their hypotheses by conducting experiments (i.e., a 'theory-driven' approach). After experimenting, students can evaluate their hypotheses against the cumulative experimental results and prior knowledge. Evaluation has three possible outcomes: the current hypothesis can either be accepted, rejected, or considered further. Depending on this evaluation the student may start a new search for hypotheses, continue investigating the current hypothesis (which generally involves some alteration), or end the inquiry. Students without domain expertise cannot generate initial hypotheses from prior knowledge. They have to search the experiment space for a series of exploratory experiments (i.e., a 'data-driven' approach). Once performed and evaluated, these experiments may help students to formulate an initial hypothesis, which can then be tested through experimentation.

Research has generally confirmed the alleged influence of domain knowledge on scientific reasoning. The original study by Klahr and Dunbar (1988) provides evidence that prior knowledge reduces time on task and the number of experiments conducted. Performance success was independent of prior knowledge: all participants succeeded in discovering how an unknown function of an electronic device worked. Klahr and Dunbar also identified two distinct investigative strategies, a Theorist approach and an Experimenter approach. One of the key differences between the two was that Experimenters conduct more experiments than Theorists and that this extra experimentation is conducted without an explicit hypothesis statement (Klahr & Dunbar, 1988).

However, these results could not be replicated under more controlled circumstances. Wilhelm and Beishuizen (2003) for instance compared learning activities and outcomes across a concrete and abstract inquiry task. These tasks were designed so that participants had no prior knowledge of the abstract task and ample prior knowledge of the concrete task. Participants were found to perform better when their task was embedded in a concrete context. Compared to the students in the concrete condition, students in the abstract condition stated fewer hypotheses, but performed as many experiments (time on task was not assessed). Lazonder, Wilhelm, and Hagemans (2008) replicated these findings in a within-subject comparison. They too found that participants perform better on a concrete task with familiar content. Results also confirmed that participants generate more, and more specific hypotheses on the concrete task. The number of experiments was again comparable on both tasks. Lazonder et al. (2008) also confirmed the existence of two distinct investigative strategies. They argued that as individuals have little domain knowledge they are presumed to start off in a data-driven approach, meaning that they start experimenting without having formulated specific hypotheses, but gradually switch to a more theory-driven mode of experimentation. Individuals who do posses domain knowledge, in contrast, approach the task by generating and testing specific hypotheses, which is the Theorist approach.

Deleted: n

These findings suggest that, although prior knowledge does not reduce the number of experiments per se, it does reduce the number of experiments not guided by a hypothesis. Students with prior knowledge thus engage in more theory-driven experimentation which leads to superior task performance. The latter part of this conclusion was corroborated by Lazonder, Wilhelm, and van Lieburg (in press), who found that the number of hypotheses stated by participants was a strong predictor of performance success. This study further showed that students learning by inquiry benefit little from knowledge of the meaning of variables per se, but it is the knowledge of the relations of the variables that is of pivotal importance.

In line with the previously mentioned studies, the research reported here investigated how prior domain knowledge influences students' scientific reasoning and performance in an inquiry task. In contrast to the previous studies, this study was designed as a novice-expert comparison that aimed to replicate and extend previous findings under more ecologically valid conditions. Toward this end the study utilized a genuine physics task that was situated in a realistic setting, and performed with an inquiry learning environment designed for secondary education –which stands in marked contrast to the fictitious small-scale inquiry tasks used in laboratory studies cited above. Another key difference with prior research is that modelling was treated as integral part of the inquiry process. Toward this end the learning environment housed a modelling tool students could use to articulate their hypotheses and (acquired) domain knowledge.

Research design and hypotheses

This study compared scientific reasoning and performance success of low-level novices, high-level novices and experts on an inquiry task that involved modelling a charging capacitor. Low-level novices had no prior knowledge of the task content, but could induce this knowledge by interacting with a computer simulation so as to build a model of the capacitor. High-level novices were familiar with the physics laws that govern the behaviour of a charging capacitor, whereas the experts' knowledge of capacitors was well beyond the needs to complete the task.

In line with previous findings participants' prior domain knowledge was expected to influence their performance success and scientific reasoning. As participants could infer all knowledge by interacting with the learning environment, the quality of their final models was expected to be comparable and therefore independent of prior domain knowledge. However, it was expected that novices would need more time to create their models than experts.

Scientific reasoning was expected to differ as function of participants' prior domain knowledge. Low-level novices, in absence of prior domain knowledge, were expected to start off in a data-driven mode of inquiry and gradually shift to a more theory-driven approach, resulting in increasingly domain-specific hypotheses. High-level novices possessed some prior domain knowledge, and were therefore expected to approach the beginning of the task more theory driven than low-level novice. Still, high-level novices were expected to show an increase in their hypotheses' domain specificity. Experts on the other hand, were predicted to engage in theory-driven experimentation throughout their inquiry, expressing highly domain-specific hypotheses. As participants engaging in a data-driven approach will conduct more experiments than participants engaging in a theory-driven approach, a negative relationship was expected between prior domain knowledge and the number of conducted experiments.

Relatively many studies have been conducted investigating learners' evidence evaluation. This kind of research generally focuses on developmental differences and reasoning errors people make during evidence evaluation (for an extensive overview see Zimmerman, 2000). However, as the influence of prior domain knowledge on evidence evaluation has remained unexplored, this study does not start from an assumption regarding the process of evaluating evidence, and addressed this scientific reasoning process in an explorative way.

Deleted: (

Deleted:)

Method

Participants

Thirty-one Dutch students participated in this study. They were selected for their levels of prior domain knowledge and classified as either low-level novice, high-level novice, or expert. Low-level novices (*n* = 10) were junior high-school students (aged 14 - 15) who had no prior domain knowledge: as capacitors were not part of their curriculum they were unfamiliar with the relevant formulas. However, they did have modelling experience, as they had recently attended an 8-hour modelling unit in which they built system dynamics models of several phenomena (i.e., influenza, fluid dynamics, and greenhouse gasses). High-level novices (*n* = 10) were senior high-school students (aged 18 - 20) from the science track with some prior domain knowledge (capacitors had been taught in their curriculum and all relevant formulas were addressed), and modelling experience. One year prior to the experiment they had attended the same modelling unit as the low-level novices. Additionally, they had just finished a modelling refreshment course that, among other things, involved modelling a capacitor. Experts (*n* = 11) were university students (aged 20 - 27) who had finished their first year in electrical engineering. They thus had extensive prior domain knowledge (their curriculum involved knowledge about capacitors well beyond the scope of the task), as well as ample modelling experience.

Materials

Participants engaged in an inquiry task in a modified standalone version of the Co-Lab learning environment (van Joolingen, de Jong, Lazonder, Savelsbergh, & Manlove, 2005). The task was to replace parts of the electrical circuit of a speed control camera so it would match new specifications. The cover story told participants that a modification to speed control cameras (adding a transmitter that activates a matrix board) caused too long

Deleted: et al.

recharging times of the capacitor in the electrical circuit. Participants were told that by replacing the resistor in the electrical circuit the recharging times could be influenced. They had to suggest a possible resistance value which would lead to smaller capacitor recharging times.

Deleted: had to solve this problem
Deleted: resistance
Deleted: ,
Deleted: finding
Deleted: ing

In order to tackle the problem, participants first had to investigate how resistance affects the time to charge a capacitor. The behaviour of a charging capacitor could be studied by running experiments with a simulation (see Figure 1). The simulation represented an electrical circuit containing a power source, a resistor, a device that activates a matrix board (which has resistance), and a capacitor. Experiments could be conducted with this electrical circuit to examine the influence of the resistance on the charging of the capacitor. In the simulation the resistor value could be manipulated (five possible values), which changed the current in the circuit. Simulation output of all variables could be inspected through a table and graph.

Deleted: resistance
Deleted: sender
Deleted: s
Deleted: with
Deleted: a
Deleted: resistance
Deleted: ing

Insert Figure 1 about here

Participants could infer knowledge by interacting with the learning environment. Four knowledge components about electrical circuits can be distinguished: Ohms Law, Kirchhoff's law (including its two rules: the junction rule, and the loop rule), and the behaviour of capacitors. Students who are unfamiliar in the domain can generate this knowledge by conducting experiments with the simulation. For instance, from viewing the animation students can grasp the notion that a capacitor is a device where charge is stored (hence the animation was designed including a "peeled off" capacitor, so students could see a potential difference arising across the plates). Furthermore, the knowledge components could be inferred through (systematic) inspection of the results generated from these experiments (in a graph or table). For instance, students can plot the potential difference across the capacitor

Deleted: a

during charging in a graph. From inspection of this graph it can be hypothesized that as the potential difference across the capacitor increases, the charging speed decreases. Therefore, the increase in potential difference across the capacitor should be dependent (among others) on the potential difference across the capacitor itself. Such reasoning concerns knowledge about the behaviour of capacitors and the loop rule.

The model editor (see Figure 1) enabled participants to build and test a model that represents their conceptions of the charging behaviour. (A reference voltage of 0 Volts at the negative battery pole was assumed so that absolute voltages could be used in the model.) The syntax of this system dynamics model makes use of ‘stocks’, ‘auxiliaries’, ‘constants’, ‘flows’ and ‘relations arrows’. A model consists of several components: basic elements (i.e., elements that represent the model ‘input’: constants and stocks), auxiliary elements (i.e., elements that specify the integration of elements) and connecting arrows. An example looks like this: A basic element that changes over time and has an initial value (Charge) is represented in a stock. Connected to a stock are flows, indicating the changes in the stock. These changes are specified from the basic elements that remain constant (i.e., constants) (e.g., capacitance (C), power source (S), resistance (R_1 and R_2)) and auxiliary elements (i.e., auxiliaries) (e.g., potential difference across the capacitor (V_c), potential difference across the resistances (V_r), current (I), resistance total (R)) which are connected by relation arrows.

Deleted: n

Deleted: executable

Deleted: capacitor

Deleted: which

Deleted: in

Deleted: which

Deleted: capacitor capacity

Deleted: voltage in capacitor

Deleted: voltage

Deleted: in circuit,

As explained in van Joolingen et al. (2005), participants could build their initial model early on by selecting pre-specified, qualitative relations from a drop-down menu (not shown in Figure 1). During the later stages, when participants’ knowledge of the capacitor had increased, qualitative relations could gradually be replaced by quantitative ones using scientific formulas. Thus participants could use their models to express propositions about a relation between variables. Hence, students’ modifications to a model were considered hypotheses that could be tested by running the model and analyzing its output through the table and graph. These tools further allowed students to compare model and simulation output in a single window.

The Co-Lab learning environment stored participants' actions in a log file; Camtasia Studio ("Camtasia Studio", 2003) was used to record participants' actions and verbalizations in real time.

Deleted: logfile

Deleted: ,

Procedure

Students participated in the experiment one at a time. As experts had no prior experience with the syntax of the modelling tool, they completed a brief tutorial prior to the assignment. All other instructions and procedures were identical for the three groups of participants.

At the beginning of a session, the experimenter explained the experimental procedures. Participants were then presented with the cover story that introduced them to the inquiry task. Next, the experimenter demonstrated the procedural operation of the simulation, the model editor, and the graph and table tool. During this demonstration, the experimenter handed out a paper instruction manual on the modelling syntax participants could consult at any time during the task. All participants were familiar with this manual: both novices groups used it during their modelling unit and the experts studied the manual during their modelling tutorial prior to the assignment.

Participants were asked to think aloud during the task. Thinking aloud was practiced on a simple task (tying a bowline knot). After this final instruction, participants received the problem statement and started their inquiry. They had 1.5 hours maximum to complete the task.

During task performance the experimenter prompted the participants to think aloud when necessary. Thinking aloud was further encouraged by asking participants to state their hypotheses upon running the simulation and to verbalize their evaluation of evidence upon inspecting experimental results in the table or a graph. Towards this end the experimenter used non-directive probes 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?") that have been shown to have no disruptive influence on participants' inquiry learning processes (Wilhelm & Beishuizen, 2004).

Coding and scoring

Variables under investigation in the study were time on task, performance success, and the three scientific reasoning processes of hypothesising, experimentation, and evidence evaluation. Time on task was assessed from the log-files. Performance success was scored from the participants' final models. Both a model content and a model structure score were calculated. The model content score represented participants' understanding of the four distinct knowledge components about electrical circuits within the task (i.e., Ohms Law: $I = V/R$, resistances connected in parallel: $1/R_t = 1/R_1 + 1/R_2$, the potential difference in the circuit depends on the power source and the potential difference across the capacitor: $\Delta V = V_s - V_c$, and the relationship between the potential difference across the capacitor and the amount of charge that gathers on the capacitor: $C = Q/V_c$). In a correct, fully specified model these components are correctly integrated and meet Equation 1. One point was awarded for each correctly specified component, leading to a four-point maximum score. Two raters scored the models of three randomly selected low-level novices, three randomly selected high-level novices and three randomly selected experts. Inter-rater reliability estimate was 1.0 (Cohen's κ).

Deleted: part

Deleted: voltage in the circuit depends on the voltage source and the capacitor voltage

Deleted: U

Deleted: U

Deleted: U

Deleted: voltage

Deleted: U

Deleted: part

Deleted: e

Deleted: part

Deleted: U

$$(dQ/dt) = (V_s - Q/C) * (1/R_1 + 1/R_2) \qquad (1)$$ [1]

The model structure score was scored in accordance with Manlove et al.'s (2006) model coding rubric. This score represented the number of correctly specified variables and relations in the models. "Correct" was judged from the reference model shown in Figure 1. One point was awarded for each correctly named variable; an additional point was given if that variable was of the correct type. Concerning relations, one point was awarded for each correct link between two variables and one point was awarded for the direction. The maximum model structure score was 38. Two raters coded the models of three randomly selected low-level novices, three randomly selected high-level novices and three randomly selected experts. Inter-rater reliability estimates were .74 (variables) and .92 (relations) (Cohen's κ).

Participants' *simulation hypotheses* concerned statements about variables and relations accompanying simulation runs, and were assessed from the think-aloud protocols. Each hypothesis was classified according to the level of domain specificity using a hierarchical rubric consisting of fully-specified, partially-specified, and unspecified hypotheses (as did Lazonder et al., in press). A fully-specified hypothesis comprised a prediction of the direction and magnitude of the effect ("I think a 10 times larger resistance will extend the capacitors' recharging period by 10"). Partially-specified hypotheses predicted the direction of effect ("I think increasing the resistance will increase the capacitors' recharging period"). Unspecified hypotheses merely denoted the existence of an effect ("I think the resistance influences the capacitors' recharging period"). Statements of ignorance or experimentation plans ("I'll just see what happens") were not considered hypotheses. Two raters coded the simulation hypotheses of three randomly selected low-level novices, three randomly selected high-level

---

[1] Equation 1 can also be written as dQ/dt = (V/R) exp[-t/RC], with R being the total resistance of the parallel resistors. The formula used here was preferred because it is consistent with the system dynamics formalism.

novices, and three randomly selected experts (in total 74 hypotheses). Inter-rater agreement was .77 (Cohen's κ).

In accordance with van Joolingen et al. (2005), model changes were also considered hypotheses. A *model hypothesis* was operationally defined as the changes in a participant's model between subsequent runs. Model hypotheses were coded based on the same hierarchical rubric as simulation hypotheses. Any change to a quantitatively specified relationship between two elements in the model was coded as fully-specified hypothesis. Changes in qualitative relationships were coded as partially-specified hypothesis, and changes to relation arrows not accompanied by a qualitative or quantitative specification was coded as unspecified hypothesis. Two raters coded the models of three randomly selected low-level novices, three randomly selected high-level novices and three randomly selected experts (in total 145 models). Inter-rater agreement was .85 (Cohen's κ).

The number of conducted *experiments* with the simulation and the number of model runs were retrieved from the log files. Every time participants clicked the 'Start' button in the simulation window was considered a simulation experiment. Experiments that were not accompanied by a hypothesis were considered exploratory experiments. Simulation experiments were further classified as unique or duplicated depending on whether the experiment had been previously run with the same resistance value. As the learning environment enabled participants to choose from 5 different resistance values, a maximum of 5 unique experiments could be conducted. Every time participants clicked the 'Start' button in the model editor was considered a model run. If the model had been conceptually altered since the previous run, this run was considered an experiment.

The results of participants' *evidence evaluation* was assessed from the progression of participants' models during their session. This evaluation of evidence process was coded based on participants' subsequent models. Based on cumulative evidence resulting from experimenting (and prior knowledge) participants could decide to (temporarily) accept, reject, or alter their current hypothesis (contrary to Klahr and Dunbar's (1988) study, further consideration of the current hypothesis with different experiments is conceptually not possible when a model is considered an hypothesis). Modifications to the previous version of the model were considered 'alterations', except when these modifications were deletions or additions that were not related to the previous hypothesis. Deletions of elements in prior models were considered 'rejections', as they reject the hypothesis in the prior model specified by this element. Additions of elements in models signalled 'acceptations', as the prior model was (temporarily) accepted as it was, and now a new hypothesis is considered by addition of this new element.

Results

Both groups of novices needed more than 80 minutes to complete the task (low level novices: $M = 81.80$, $SD = 11.39$; high-level novices: $M = 81.30$, $SD = 19.61$); experts took about 20 minutes less time ($M = 63.36$, $SD = 22.12$). Univariate analysis of variance (ANOVA) showed this difference to be statistically significant, $F(2,28) = 3.45$, $p < .05$). Planned contrasts indicated that experts needed less time on task than novices ($t(28) = -18.19$, $p = .01$), whereas the high-level novices and low-level novices needed as much time to complete the task ($t(28) = -.50$, $p = .95$).

Table 1 presents a summary of participants' performance. Performance success was assessed from participants' final models. Multivariate analysis of variance (MANOVA) showed that the quality of the participants' models differed as function of their prior knowledge ($F(4,56) = 9.50$, $p < .01$). Subsequent univariate ANOVA's indicated that prior knowledge influenced both model content ($F(2,28) = 59.105$, $p < 0.01$) and model structure score ($F(2,28)$, $p < .01$). Planned contrasts revealed that experts achieved significantly higher model content ($t(28) = 3.09$, $p < .01$) and model structure scores ($t(28) = 9.05$, $p < .01$) than novices. The comparison among both groups of novices showed that high-level novices had higher model content scores than low-level novices ($t(28) = 1.10$, $p < .01$). However, the model structure score indicated no significant difference between both novice groups ($t(28) = 3.30$, $p = .24$).

Deleted: The model structure scores also differed in favour of the high-level novices, but not to a statistically significant degree

Insert Table 1 about here

From Table 1 it can be seen that participants differed in the number of hypotheses they generated. Although MANOVA with the number of simulation and model hypotheses as dependent variables did not reach significance ($F(4,56)= 2.01$, $p = .11$), the large standard deviations indicate a considerable variation in scores. Therefore, the content of these hypotheses was analysed using the percentages of all stated hypotheses as measure.

Deleted: showed that these differences approached significance

Deleted: .

Deleted: were

Deleted: based on

As few participants (4 low-level novices, 3 high-level novices, and 7 experts) stated hypotheses with both the simulation and the models, data were analyzed with non-parametric Kruskal-Wallis' ranks tests. Results indicated that the groups neither differed in mean model hypothesis' specificity ($\chi^2(2, N = 20) = 5.59$, $p = .06$), nor on their mean simulation hypothesis specificity ($\chi^2(2, N = 20) = .72$, $p = .70$).

Deleted: s differed nearly significantly on their mean model hypothesis' specificity

Deleted: but not on

Deleted: their

Figure 2 depicts the specificity of participants' hypotheses through time (as time on task differed between groups, it was standardized using quartiles). An increase in domain specificity was expected for both novice groups, whereas experts were expected to generate highly domain specific hypotheses throughout the task. Contrary to expectations however, the mean domain specificity of participants' hypotheses remained relatively stable through time. One noticeable finding is that low-level novices had substantially more domain specific simulation hypotheses in the fourth quartile. Yet the domain specificity of their model hypotheses failed to follow this trend.

Insert Figure 2 about here

Participants could experiment either by running the simulation or their models. MANOVA with the number of unique and duplicated simulation experiments as dependent variables produced no significant differences ($F(4,56) = 1.63$, $p = .18$). ANOVA of the number of model experiments was not significant either ($F(2,23) = 1.61$, $p = .22$), and nor was the percentage of these experiments that was exploratory (simulation experiments: $F(2,28) = 0.62$, $p = .55$; model experiments: $F(2,23) = 1.25$, $p = .31$). These results indicate that participants with varying levels of prior knowledge performed as many experiments, and used these experiments as often to test hypotheses.

Participants could perform these experiments during the task as they deemed necessary, resulting in large inter-individual differences in experimenting behaviour over time. Figure 3 depicts the spread of the number of experiments conducted with the simulation and the models over time (as with hypotheses, time was divided in quartiles). As can be seen, in general the number of experiments with the simulation decreased over time, whereas the number of experiments with the models tended to increase. There was also a decline in the number of participants who experimented with the simulation. Even though an initial knowledge base could be acquired by experimenting with the simulation, seven low-level novices chose not to experiment with the simulation in the first quartile. Actually, three low-level novices did not experiment with the simulation at all. Even more participants did not make use of the modelling tool to experiment with, one low-level novice and four high-level novices never executed one of their own models.

Insert Figure 3 about here

For subsequent models, results of participants' evidence evaluating processes were analysed in light of the number of hypotheses. Therefore comparable to hypotheses' data, these data were also converted to percentages and analysed with non-parametric Kruskal-Wallis' ranks test. From Table 1 it can be seen that groups did not differ in percentage of evidence evaluation resulting in accepting (χ^2 (2, N = 20) = 0.10, p = .95) and alteration (χ^2 (2, N = 20) = 2.61, p = .27). However, prior knowledge affected the percentage of evidence evaluation processes resulting in rejection (χ^2 (2, N = 20) = 6.72, p < .05). Low-level novices rejected more model hypotheses than high-level novices and experts.

Qualitative analyses

From these statistical analyses it appears that novices predominantly followed the same approach as experts. Performance success scores suggest that this approach suited experts better than novices. Qualitative analyses of participants' modelling activities were performed to reveal why novices' behaviour was less effective.

When looking at participants' initial models (i.e., the first model they tried to run), it appeared that participants with domain knowledge were only a fraction better at deciding which components to include in their model. Experts' initial models contained nearly all basic elements from the target model (i.e., 1 stock and 4 constants) (M = 4.45, Range = 3-5), indicating that they could oversee the entire problem and correctly identified the relevant pieces of information from the problem statement. Novices included as many elements in their first model (low-level novices: M = 4.33, Range = 2-6; high-level novices: M = 4.00, Range = 3-5). However, low-level novices' initial models contained a few erroneous elements

Deleted: experts

such as ‘loading time’ and ‘switch’ (M = 0.89, Range = 0-2), whereas high-level novices and experts’ models had no such elements. The low-level novices’ final models contained a comparable number of incorrect elements (M = 1.22, Range = 0-4).

Although low-level novices had a pretty good sense of which elements to include in their initial models, they were probably ignorant of the relationships between model elements. The modelling tool in Co-Lab anticipated this by offering participants the possibility to specify relationships qualitatively. Participants could thus specify relationships before they fully grasped the mathematical formula governing the relation between two variables. Surprisingly however, only two low-level novices and one expert made use of this feature. While this may seem a defendable choice for the experts and high-level novices, it may not be a wise decision for the low-level novices. Yet they generally ignored, and sometimes even deliberately rejected qualitative modelling by saying that it produced a less specific model that would not help them to discover the capacitor’s behaviour.

These findings support the idea that low-level novices tried to build their models in an expert manner. But due to their lack of prior knowledge, low-level novices could only base their modelling efforts on insights gained through experimentation, or engage in trial and error activities. Therefore, participants’ think-aloud protocols were analyzed to reveal the reasoning behind subsequent model changes (i.e., model hypotheses). Results indicated that low-level novices hardly reasoned at all. Nine low-level novices utilized the modelling tool to experiment with their models, eight of them also experimented with adjusted models. These eight low-level novices did not motivate 87% of the changes they made to their models at all. The changes to models that were guided by reasoning could be considered ‘data-driven’; this is illustrated in Excerpt 1.

Excerpt 1 (low-level novice)

"They [the resistances] ought to be 4.4 Volts.

Participant inspects model output in the table

Hmmz, 410 kilo Ohm, so with every kilo Ohm there will be approximately 0.1 Volts resisted. Thus this resistance resists 3 Volts and the other 1.1 Volts.

The experts, in contrast, relied heavily on their prior knowledge for their model changes. Eight experts performed more than one model experiment, and 83% of their model changes were motivated from prior knowledge; a typical example is shown in Excerpt 2. Of the remaining model changes, 12% was 'data-driven', often involving statements about previous model runs, 2% was based on logical reasoning, and 3% was not motivated.

Excerpt 2 (expert)

"Now I have the, ehm, source power I've got let's say to the…the source power is influenced by the resistances, from that I've made this current. That is the current behind the parallel resistances. As that is necessary to charge the capacitor. The formula to charge the capacitor is: the value of the capacitor times the current time derivative. So now I'm going, ehm, then you have the current over there…"

Only four high-level novices performed more than one model experiment. In the think-aloud protocols of the four high-level novices who found subsequent experimenting worthwhile, 89% percent of the changes made to the model were motivated. This reasoning was based on prior domain knowledge (28%), data from prior experiments (33%), information found in the assignment (28%; see Excerpt 3), or logical reasoning (11%).

Excerpt 3 (high-level novice)

"With these [the arrows connecting elements in the model] I want to indicate that there is a charge directly towards the capacitor…and that it goes through the sender or the resistance let's say…and then again through the capacitor, like in that circuit [the circuit depicted in the assignment paper]."

Discussion

The aim of this study was to reveal domain novices' need for support by comparing their scientific reasoning and performance success to that of students with higher levels of domain knowledge. The experts' task performance served as standard against which the scientific reasoning and knowledge acquisition of low-level novices and high-level novices were compared. The first comparison in particular elucidates the issues support for students without prior domain knowledge should address. The discussion concludes with implications for the design of such support.

Consistent with problem-solving research, the experts required less time for task completion than both groups of novices. Other findings suggest that these time differences were attributable to the experts' rich knowledge base. That is, experts needed only a few simulation experiments to create comprehensive initial models that generally contained all basic elements from the target model. Their model runs were always intended to test a hypothesis, and nearly all changes to the model were motivated from prior knowledge.

Low-level novices were predicted to perform these scientific reasoning processes in a different way. Contrary to expectations, however, their hypothesizing and experimenting did not differ from that of experts. Although the latter result is consistent with previous laboratory studies (Lazonder et al., 2008, in press; Wilhelm & Beishuizen, 2003), the higher proportion of exploratory experiments found in these studies could not be confirmed. Together these findings suggest that low-level novices based their rather specific hypotheses on mere guesswork. The qualitative analyses bore this out: most low-level novices did not engage in qualitative modelling, and very few of the changes to their models (i.e., model hypotheses) were guided by reasoning. Therefore, many of these hypotheses inevitably were incorrect and should be rejected. This is indeed what appears to have happened since low-level novices rejected a larger proportion of their model hypotheses than experts did.

Performance success scores reflect to what extent participants' scientific reasoning was effective. Based on Klahr and Dunbar (1988), performance success was assumed to be independent of participants' prior knowledge because, contrary to most problem solving tasks, low prior knowledge participants could infer all knowledge by interacting with the learning environment. Results indicate that they did not: the quality of the experts' models was higher compared to that of the high-level novices' models, whereas high-level novices built better models than low-level novices. A closer look at these results shows that the experts achieved an almost perfect model content score; a few minor inaccuracies caused that not every expert produced a fully correct model. Low-level novices, in contrast, had rather low performance success scores. The magnitude of their model content scores indicates that they did not acquire complete understanding of any of the four formulas that governed the behaviour of the charging capacitor. Although the learning environment provided them with all necessary tools to induce this knowledge, low-level novices did not succeed in doing so –which suggests that their scientific reasoning was rather ineffective.

Deleted: proves

From these findings it can be concluded that low-level novices predominantly exhibit expert-like behaviour during an unsupported inquiry task, and that this approach apparently does not suit them that well. This conclusion is consistent with the findings of Lazonder et al. (2008). Their within-subject comparison revealed that students generally adopt a similar approach to inquiry tasks in familiar and unfamiliar domains, but perform better on tasks they possess prior knowledge of. Therefore, it can be concluded that the current results

Deleted: suits experts better than low-level novices complement existing evidence on the influence of prior knowledge on inquiry behaviour. Findings from prior laboratory studies in which prior knowledge was manipulated by differences in task design, can now be generalized to more ecologically valid classroom situations.

This study added an intermediate group (i.e., high-level novices) to the novice-expert comparison. Insight into high-level novices' inquiry behaviour and difficulties is of interest for the design of support because low-level novices will probably encounter the same problems once they have gained some knowledge of the topic they are investigating. As high-level novices' prior knowledge was higher than the low-level novices' and lower than the experts', they were expected to perform better than the low-level novices, though possibly not as good as experts. Contrary to expectations, however, their hypothesizing and experimenting neither differed from that of experts, nor from that of low-level novices. The qualitative analyses suggest that this expert-like behaviour suits the high-level novices as there appeared to be sound reasoning behind the high-level novices' highly specific hypotheses. Consequently, most of their experiments resulted in either acceptation or alteration of the hypotheses, which was comparable to experts' evidence evaluation results.

The high-level novices' performance success scores were higher than low-level novices'. Yet these scores were still fairly low, considering that the high-level novices were familiar with all relevant domain knowledge. It appears that, despite their prior knowledge, performance on this task was difficult for the high-level novices, suggesting that they were unable to effectively apply their knowledge. These findings lead to the conclusion that learners who are somewhat familiar in the domain also need support in order to help them manage their knowledge to effectively perform an inquiry task.

However, there was one slightly a-typical finding. Several high-level novices were found not to perform any model experiment. This could be a result of the task difficulty. If high-level novices had difficulty expressing their knowledge in a model during the task, they probably did have enough domain knowledge to realize that the model was not good enough yet. As such, it would make sense not to run that model as they knew it to be incorrect. Future research might give more insight on this problem and how it can be overcome.

These conclusions lead to implications for support. Bearing in mind that what constitutes effective and efficient inquiry behaviour is dependent on domain knowledge, it can be argued that novices' (having no prior knowledge) unsupported inquiry behaviour was not effective on this task, but could be effective if they were familiar in the domain and would apply and expand this knowledge through iterative cycles of model testing. Conclusions for support for inquiry learning can therefore go into two directions, either providing domain support in order to increase the effectiveness of their students' natural inquiry behaviour, or process support to better attune students' inquiry behaviour to their level of domain knowledge. These two directions correspond with what Quintana et al. (2004) called content support and process support respectively.

Deleted: most likely

Deleted: w

In a literature review, de Jong and van Joolingen (1998) conclude that providing direct access to domain information seems effective as long as the information is presented concurrently with the simulation, so that the information is available at the appropriate moment. Lazonder, Hagemans, and de Jong (in press) found that offering domain support before and during the task is even more effective. Students who received domain information before and during the task not only inferred more knowledge from their investigations, but also exhibited more sophisticated scientific reasoning. This confirms the notion that providing domain knowledge to students is an effective type of support. However, as our low-level novices already exhibited quite sophisticated scientific reasoning, while still being rather unsuccessful on the task, providing domain knowledge appears not to be the most appropriate type of support. Moreover, as Lazonder et al. (in press) also mention, providing domain

Deleted: this more

Deleted: 2009

knowledge is somewhat at odds with the concept of inquiry learning, where learners have to discover domain knowledge themselves.

Therefore, it seems more appropriate to support students' inquiry behaviour by better attuning students' inquiry behaviour to their level of domain knowledge. Directions for such process support can be derived from this study's results. The bottleneck for novice learners was found *not* to be the identification of relevant elements, as it was the inquiry of the nature of the relationship between these elements that caused problems. Novice learners knew quite well which elements to include in the model (even their initial model contained nearly all correct elements and few erroneous elements). However, novice learners attempted to infer the relationships between those elements by means of testing hypotheses that were very specific in nature. Moreover, novices most likely based these hypotheses on guesswork, as there was hardly any underlying reasoning. As such inferring the correct relationships becomes very difficult and it is no surprise that they hardly succeeded in inferring these relationships.

The modelling tool in the learning environment aims to support learners' hypotheses construction in a graphical way (van Joolingen et al., 2005). Learners in this study were given a choice as to how detailed they wanted to specify relationships. They could opt for a self-generated, full-fletched scientific formula (i.e., quantitative relations), or select less detailed pre-specified, qualitative relations from a drop-down menu (i.e., qualitative relations). Qualitatively specified relations are more appropriate at the beginning of the modelling process when learners do not yet have a clear idea about the model they are making (Löhner, van Joolingen, & Savelsbergh, 2003; Sins, Savelsbergh, & van Joolingen, 2005). Therefore it is surprising that in the present study, where participants did not receive any kind of support, only 2 low-level novices made use of the possibility to state qualitative relations.

Deleted: poses

Deleted: which

Deleted: most likely

Deleted: behind them

Deleted: is meant

Deleted: , as propositions about relations between variables can be expressed in these models.

Deleted: the

Deleted: in

Deleted: much

Deleted: the

In view of these findings it might be fruitful to restrain domain novices' natural tendency to engage in quantitative modelling from scratch by first having them create models that are qualitatively specified, and then enabling them to transfer these qualitative relations into quantitative ones. This type of support is in line with the model progression approach described by White and Frederiksen (1990). Model progression was found to lead to higher performance (Rieber & Parmley, 1995; Swaak, van Joolingen, & de Jong, 1998). However, these authors interpret model progression as a type of support where the model at first is not offered in its full complexity, but variables are gradually introduced (or, in terms of White and Frederiksen (1990), a model progression where the degree of elaboration of a model is increased). Our proposed support, as suggested by Gobert and Clement (1999), can be considered a more fine-grained kind of model progression, where the specificity of the models is increased. This kind of model progression resembles what White and Frederiksen (1990) call model progression where the *order* of a model is increased.

To conclude, we propose to support learners on an inquiry learning task with model progression, where the model is progressed in specificity. In line with the coding of the model hypotheses, three increasingly specific stages of modelling can be identified: a stage in which relationships between elements are unspecified, a stage in which relationships between elements are specified qualitatively, and a stage in which these relationships are specified quantitatively. In the first stage of model progression, students investigate a phenomenon (e.g., an electrical circuit containing a capacitor) and have to make a model structure of that phenomenon without having to specify the relationships in the model. In the second stage, students continue to investigate the phenomenon in order to specify the relationships in their model qualitatively. In the third stage, students finalize their investigation of the phenomenon by replacing the qualitatively specified relationships with quantitatively specified relationships.

Deleted: Accordingly, we suggest a type of support where domain novices first create models which are qualitatively specified before they can specify their models quantitatively.

Deleted: as

Deleted: (de Jong et al., 1999; Swaak, van Joolingen, & de Jong, 1998) and is generally applied (e.g., de Jong, 2005; Eysink, Dijkstra, & Kuper, 2001; Gijlers & de Jong, 2005; Swaak & de Jong, 2001; Veenman, 2005; Veermans, van Joolingen, & de Jong, 2006)

Deleted: Although others have suggested that learners should be supported by increasingly specific models (de Jong, 2005; White & Frederiksen, 1990), to the authors knowledge the instructional efficacy of this type of support has never been confirmed.

Deleted: s

Deleted: ;

Deleted: not

Deleted: between elements

Deleted: As such,

Deleted: i

Deleted: have to

Deleted: are requested to

Deleted: about

Deleted: here they do not have

Deleted: subsequent

Deleted: ing

Deleted: and

Deleted: now are required

Deleted: final

Deleted: continue investigating the phenomenon and now are required to

Deleted: e

Deleted: by

One important condition for this form of model progression to be effective is that students should have enough opportunity to build and test hypotheses. The simulation that was used in the present study might not satisfy this requirement: although output of all elements in the simulation interface could be inspected in the table or graph, only one element (the resistor) could be manipulated. Allowing students to change the values of the other elements as well extends the possibilities for students to validate the hypotheses they generate from interacting with the simulation and running their own model. Model progression could then be an effective way to support students' inquiry and modelling process. Validating this assumption in science classrooms is an important topic for future research.

Deleted: Further research is necessary to assess the effectiveness of this support with regard to students' inquiry behaviour and learning outcomes. ¶
¶

References

Batra, D., & Davis, J. G. (1992). Conceptual data modeling in database design: Similarities and differences between expert and novice designers. *International Journal of Man-Machine Studies, 37*, 83-101.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2002). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Camtasia Studio (Version 2.0). (2003). TechSmith.

Chi, M. T. H., Glaser, R., & Farr, M. (1988). *The nature of expertise*. New Jersey: Lawrence Erlbaum Associates.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education, 86*, 175-218.

Davis, E. A., & Linn, M. C. (2000). Scaffolding students' knowledge integration:Prompts for reflection in KIE. *International Journal of Science Education, 22*, 819-837.

de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*, 179-201.

Gobert, J. D., & Clement, J. J. (1999). Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *Journal of Research in Science Teaching, 36*, 39-53.

Gobert, J. D., & Pallant, A. (2004). Fostering students' epistemologies of models via authentic model-based tasks. *Journal of Science Education and Technology, 13*, 7-22.

Jackson, S. L., Stratford, S. J., Krajcik, J., & Soloway, E. (1994). Making dynamic modeling accessible to precollege science students. *Interactive Learning Environments, 4*, 233 - 257.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1-48.

Deleted: de Jong, T. (2005). The guided discovery principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 215-229). Cambridge: Cambridge University Press.¶
de Jong, T. (2006a). Computer simulations: Technological advances in inquiry learning. *Science, 312*, 532-533.¶
de Jong, T. (2006b). Scaffolds for computer simulation based scientific discovery learning. In J. Elen & R. E. Clark (Eds.), *Handling complexity in learning environments* (pp. 107-128). Amsterdam: Elsevier.¶
de Jong, T., Martin, E., Zamarro, J. M., Esquembre, F., Swaak, J., & van Joolingen, W. R. (1999). The integration of computer simulation and learning support: An example from the physics domain of collisions. *Journal of Research in Science Teaching, 36*, 597-615.¶

Deleted: Eysink, T. H. S., Dijkstra, S., & Kuper, J. (2001). Cognitive processes in solving variants of computer-based problems used in logic teaching. *Computers in Human Behavior, 17*, 1-19.¶
Gijlers, H., & de Jong, T. (2005). The relation between prior knowledge and students' collaborative discovery learning processes. *Journal of Research in Science Teaching, 42*, 264-282.¶

Lazonder, A. W., Hagemans, M. G., & de Jong, T. (in press). Offering and discovering domain information in simulation-based inquiry learning. *Learning and Instruction*, doi:10.1016/j.learninstruc.2009.08.001.

Deleted: 2009

Deleted: *Unpublished manuscript submitted for publication. University of Twente*

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. (in press). Unraveling the influence of domain knowledge during simulation-based inquiry learning. *Instructional Science,* 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.

Löhner, S., van Joolingen, W. R., & Savelsbergh, E. R. (2003). The effect of external representation on constructing computer models of complex phenomena. *Instructional Science, 31*, 395-418.

Manlove, S., Lazonder, A. W., & de Jong, T. (2006). Regulative support for collaborative scientific inquiry learning. *Journal of Computer Assisted Learning, 22*, 87-98.

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., Kyza, E., Edelson, D., & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *The Journal of the Learning Sciences, 13*, 337-386.

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

Schauble, L., Klopfer, L., & Raghavan, K. (1991). Students' transitions from an engineering to a science model of experimentation. *Journal of Research in Science Teaching, 28*, 859-882.

Shrager, J., & Klahr, D. (1986). Instructionless learning about a complex device: the paradigm and observations. *International Journal of Man-Machine Studies, 25*, 153-189.

Shute, V. J., Glaser, R., & Raghavan, K. (1989). Inference and discovery in an exploratory laboratory. In P. L. Ackerman, R. J. Sternberg & R. Glaser (Eds.), *Learning and individual differences* (pp. 279-326). New York: Freeman.

Sins, P. H. M., Savelsbergh, E. R., & van Joolingen, W. R. (2005). The difficult process of scientific modeling: An analysis of novices' reasoning during computer-based modeling. *International Journal of Science Education, 14*, 1695-1721.

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

Deleted: Swaak, J., & de Jong, T. (2001). Discovery simulations and the assessment of intuitive knowledge. *Journal of Computer Assisted Learning, 17*, 284-294.¶

van Joolingen, W. R., de Jong, T., Lazonder, A. W., Savelsbergh, E. R., & Manlove, S. (2005). Co-Lab: Research and development of an online learning environment for collaborative scientific discovery learning. *Computers in Human Behavior, 21*, 671-688.

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

Deleted: Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In B. Moschner & C. Artelt (Eds.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* (pp. 75-97). Berlin: Waxmann.¶
Veermans, K., van Joolingen, W., & de Jong, T. (2006). Use of heuristics to facilitate scientific discovery learning in a simulation learning environment in a physics domain. *International Journal of Science Education, 28*, 341-361.¶

White, B. Y., Shimoda, T. A., & Frederiksen, J. R. (1999). Enabling students to construct theories of collaborative inquiry and reflective learning: Computer support for metacognitive development. *International Journal of Artificial Intelligence in Education, 10*, 151-182.

Wilhelm, P., & Beishuizen, 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.

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review, 20*, 99-149.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27*, 172-223.

Figure Captions

Figure 1. *Screen capture of the simulation (left pane) and model editor tool (right pane). Pressing the start button in the simulation started an animation of moving green dots representing current, a flow of charge over time (*see *Equation 1). The charging of the capacitor was visualized by green dots piling up on the top plate of the capacitor. The model editor shows the reference model students had to build from their prior knowledge and/or insights gained through experimenting with the simulation.*

Deleted: *e*

Deleted: *Screencapture of the simulation (left pane) and model editor tool (right pane). Upon pressing the start button in the simulation, little green dots representing current would run from the power source through the capacitor. The charging of the capacitor was visualized by little green dots piling up on the top pole of the capacitor. The displayed model in the model editor tool is the reference model students had to build from their prior knowledge and/or insights gained through experimenting with the simulation.*

Figure 2. *Mean specificity of participants' hypotheses accompanying simulation experiments (left pane) and model experiments (right pane) over time and by group.*

Figure 3. *Mean number of experiments conducted with the simulation (left pane) and with the model (right pane) over time and by group.*

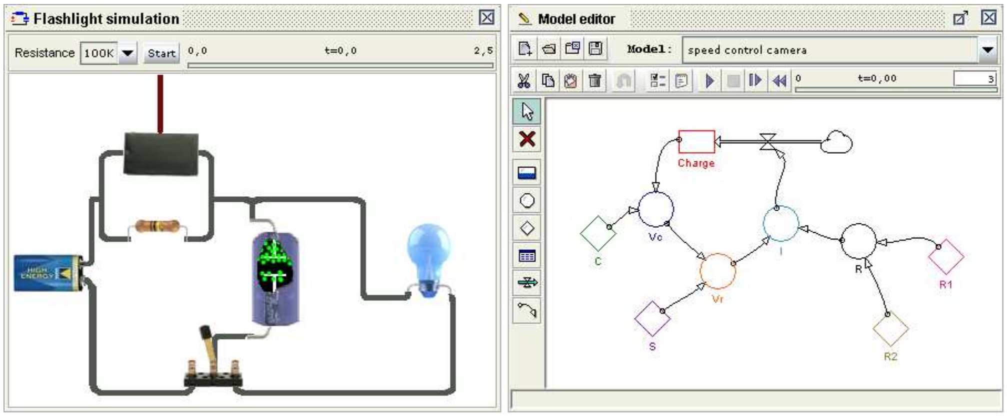

Figure 1. Screen capture of the simulation (left pane) and model editor tool (right pane). Pressing the start button in the simulation started an animation of moving green dots representing current, a flow of charge over time (see Equation 1). The charging of the capacitor was visualized by green dots piling up on the top plate of the capacitor. The model editor shows the reference model students had to build from their prior knowledge and/or insights gained through experimenting with the simulation.
238x98mm (96 x 96 DPI)

	Low-level novices		High-level novices		Experts	
	M	*SD*	*M*	*SD*	*M*	*SD*
Performance success						
Model content[1]	0.00	0.00	1.10	1.20	3.64	0.67
Model structure[2]	13.30	5.74	16.60	6.40	24.00	6.40
Hypothesizing						
Simulation hypotheses	2.10	2.73	3.70	4.35	2.10	1.70
Model hypotheses	6.00	5.42	1.30	2.11	5.91	5.49
Domain specificity simulation hypotheses	1.80	0.57	1.84	0.24	1.89	0.37
Domain specificity model hypotheses	2.10	0.65	2.75	0.50	2.58	0.54
Experimenting						
Unique simulation experiments	1.80	1.87	2.50	1.90	2.45	1.21
Duplicated simulation experiments	2.60	3.69	4.90	5.92	1.91	1.64
Exploratory simulation experiments (%)	58.06	36.52	58.69	31.19	55.52	32.97
Model experiments	7.11	4.60	3.50	3.02	4.91	3.83
Exploratory model experiments (%)	10.89	22.99	7.87	12.22	0.00	0.00
Evaluating evidence						
Accepted hypotheses (%)	32.18	15.60	37.50	47.87	29.00	22.42
Reject hypotheses (%)	20.96	14.95	4.17	8.33	5.18	8.38
Altered hypotheses (%)	46.86	21.18	58.33	50.00	65.82	23.18

[1] Maximum score = 4

[2] Maximum score = 38

Table 1. Summary of participants' performance.

173x178mm (96 x 96 DPI)

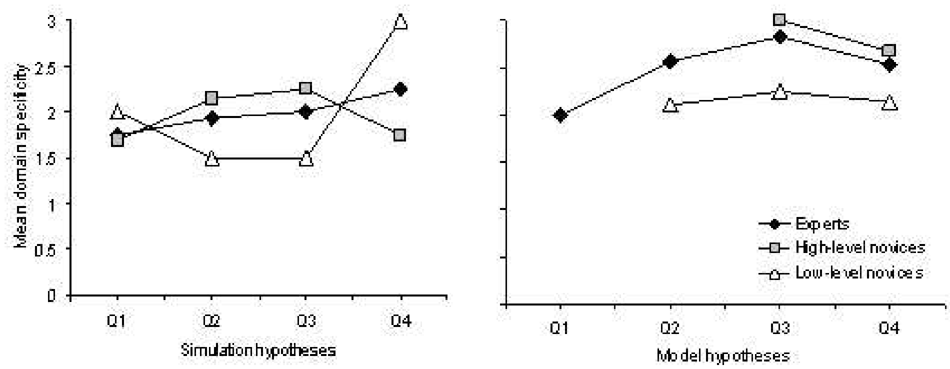

Figure 2. Mean specificity of participants' hypotheses accompanying simulation experiments (left pane) and model experiments (right pane) over time and by group.
146x59mm (96 x 96 DPI)

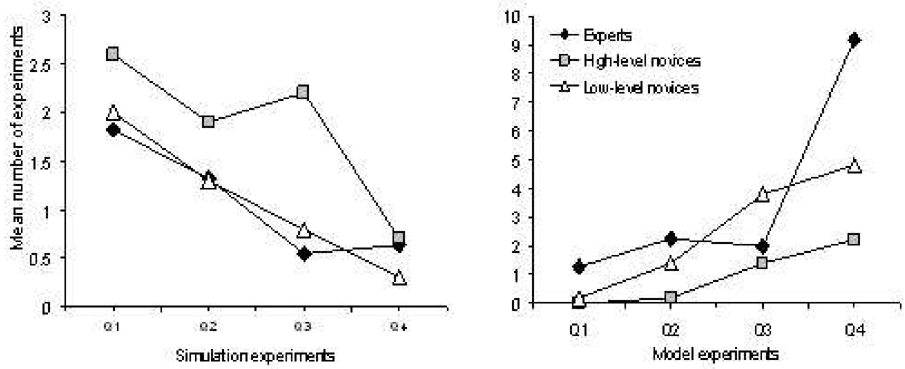

Figure 3. Mean number of experiments conducted with the simulation (left pane) and with the model (right pane) over time and by group.
146x57mm (96 x 96 DPI)