Known facts or plausible conjectures

Comparing two types of support for learning with computer simulations

Anna Charlotte Sickers, s1020692

University of Twente

Faculty for Behavioral Science, Psychology

Department of Instructional Technology

Examination Committee

Dr. A. Lazonder | University of Twente

Dr. P. Wilhelm | Enschede, 25 january 2013

Abstract

Aim of this study was to examine the relative effectiveness of offering hypotheses and domain information during inquiry learning. Domain information contained ‘known facts’ about the variables students had to investigate whereas hypotheses offered ideas, either correct or incorrect, students could test.

The study was conducted with 67 undergraduate students who worked on an inquiry task about an unknown topic. Students were randomly assigned to a condition in which they were supported by hypotheses (n=26), by domain information (n=23) or received no support at all (n=18). Learning and performance in all three conditions were then evaluated and compared. It was assumed that offering pre-specified hypotheses would lead participants to use a theory-driven method of investigation. It was also assumed that students in the hypothesis condition would gather more knowledge than participants in both other conditions.

Both hypotheses were partly supported by the results. Data indicated that participants receiving hypotheses information used a theory-driven method of investigation. It could however not be proven that offering pre-specified hypotheses leads to higher knowledge gains than offering domain information. However, the difference of both conditions to the control group was significant.

Abstract

Het doel van deze studie was het effect te onderzoeken van het aanbieden van hypothesen in vergelijking met het aanbieden van domein informatie tijdens onderzoekend leren. De domein informatie bevatte feitelijke informatie over de variabelen die de proefpersonen moesten onderzoeken; de hypotheses bevatte dezelfde informatie in de vorm van ideeën, zowel juist als onjuist, die de proefpersonen konden testen.

Het onderzoek is uitgevoerd met 67 Bachelorstudenten die werkten aan een onderzoekend leertaak over een onbekend onderwerp. Deze proefpersonen zijn willekeurig toegewezen aan een conditie waarin ze werden ondersteund door hypotheses (n=26), domein informatie (n=23), of geen enkele ondersteuning kregen (n=18).

Hun leerprocessen en leeruitkomsten zijn onderling vergeleken. Er werd verondersteld dat het aanbieden van vooraf gespecificeerde hypothesen de proefpersonen zou stimuleren om een theory-driven te gebruiken. Bovendien werd verwacht dat studenten in de hypothesen conditie meer zouden leren dan studenten in de beide andere condities.

Beide hypothesen werden deels bevestigd. Uit de resultaten bleek dat de proefpersonen die hypothesen ontvingen inderdaad een theory-driven onderzoeksmethode gebruikten. Het kon echter niet worden bevestigd dat het aanbieden van vooraf gespecificeerd hypothesen tot meer kennis leidt dan domein informatie. Het verschil tussen beide condities en de controlegroep was echter wel significant.

Table of Contents

Introduction

Promoting the cognitive development of students is one important goal of educational instances (Palles, 2010). Different learning approaches build the basis for policies and programs developed to improve the educational system and to guarantee that students can learn as much as possible. Many approaches about how students learn best exist (Kirschner, Sweller & Clark, 2006). Basically, all of these approaches can be traced back to either a minimally guided approach, a guided discovery approach or a direct instructional approach (Kirschner et al, 2006; Hmelo-Silver, Duncan & Chinn, 2007).

Minimally guided approaches are based on the assumption that people 'construct' knowledge through their personal prior experience and individual learning styles (Kirschner et al., 2006). This 'constructed knowledge' is considered to be more valuable than knowledge generated through direct instruction (Schauble, 1996). The approach corresponds to the conception of Piaget (1970) that 'each time one prematurely teaches a child something he could have discovered for himself, that child is kept from inventing it and consequently from understanding it completely' (p.715). Students are admonished to gather knowledge almost completely through their own investigations. In contrast to guided discovery approaches and direct instructional approaches they are not supported by information about the topic at hand.

Direct instructional approaches, by contrast, are based on the assumption that students should get information that 'fully explains the concepts and procedures […] required to learn' so as to keep working memory load low. Students further receive learning strategy support that is compatible with human cognitive architecture (e.g. worked examples, process worksheets) (Kirschner et al., 2006, p.75). It is neither necessary nor desired for learners to discover things on their own. Not before they know all important facts about a topic they are invited to practice.

One approach combining aspects of minimally guided learning and direct instructional learning is the guided discovery approach. It is generally assumed that students generate more knowledge if they have the possibility to discover it on their own. But they also need some basic knowledge to promote productive learning processes. Students are thus provided with some information that gives an overview of the topic at hand. This 'contrived knowledge' helps overcome learning obstacles: it serves as a scaffold that brings together important facts of the topic and provide students with a starting point for their investigation. By drawing conclusions about the contrived knowledge and pulling the facts together, students can form hypotheses or predictions about the effects of and relationships between different parts of the topic at hand. They can then test their hypotheses and investigate the topic by themselves. By using the given information, students are thus able to find out a substantial part of the topic by themselves.

One basic difference between all three approaches is the point in time when students are invited to practice by themselves. Whereas with minimally guided learning, students are ‘thrown in at the deep end’ and start learning by directly investigating a topic without having any precognition, guided discovery learning and direct instruction focus on giving information and support to the student before and during their investigations. The latter two approaches emphasize the importance of providing information and support to students. However, guided discovery learning focuses on providing only the amount of information that is required to find a meaningful beginning for investigation. Students are thus offered the possibility to explore many aspects of a topic on their own. Support is offered if students need it, either on demand or just-in-time. With direct instruction, however, practice occurs at the end of the learning process when students know every aspect of a topic. Support is thus offered before practice.

Merits and drawbacks of these three approaches are broadly discussed in the literature (Furtak, Seidel, Iverson & Briggs, 2012; Hmelo-Silver et al., 2007; Kirschner et al., 2006; Klahr & Nigam, 2004). Evidence supporting these theoretical claims is divergent (Hmelo-Silver et al., 2007; Kirschner et al., 2006). By and large, research indicates that, in terms of knowledge gains, especially students with low prior knowledge do not benefit from minimally guided approaches (Kirschner et al., 2006) whereas guided discovery approaches and direct instructional approaches lead to favorable results. However, for high prior knowledge students the knowledge gains are equal regardless of which approach is chosen.

Some studies indicate that guided discovery approaches seem superior to direct instructional approaches (Hmelo-Silver et al., 2007). A possible explanation for the failure of minimally guided approaches is offered by cognitive load theory. Cognitive load on working memory is high when dealing with unknown topics and hinders the acquisition of new knowledge (Chandler & Sweller, 1991). The support offered in direct instructional approaches and guided discovery approaches minimizes cognitive load and therefore stimulates knowledge creation.

Based on the three approaches, minimally guided learning, guided discovery learning and direct instructional learning, different pedagogies for promotion of successful learning have been developed. Once such pedagogy is scientific inquiry learning, an instructional approach that aims to foster students' understanding of science content and science processes by engaging them in solving 'realistic' problems (De Jong, 2006, Sandoval & Reiser, 2003). Inquiry learning is defined as 'an approach to learning that involves a process of exploring the natural or material world, and that leads to asking questions, making discoveries, and rigorously testing those discoveries in the search for new understanding.' (De Jong, 2006, p.532). Key aspect of inquiry learning is to help learners build a scientific knowledge base which enables them to predict and explain phenomena of the natural world (Van Jolingen, De Jong & Dimitrakopoulou, 2007). In comparisons of inquiry-based instruction and more traditional teaching methods (such as lectures and textbook-based instruction), researchers have found that inquiry methods help students gain a better understanding of fundamental concepts in science (Sandoval & Reiser, 2003; White & Fredericksen, 1998).

There are different theories of scientific inquiry learning which are usually based on theories of scientific reasoning skills (De Jong & Van Joolingen, 1998). Scientific reasoning skills involve hypotheses generation, experimentation or observation, and evidence evaluation. Research indicates that students often have problems performing these skills. They often tend to find it difficult to identify variables to investigate, generate testable hypotheses and draw valid conclusions from data (De Jong & Van Joolingen, 1998). However, these problems apply only to students without prior domain knowledge. Students with prior knowledge about a topic generally experience fewer difficulties (Kirschner et al., 2006).

A reasonable instructional implication is to provide students with support. There are essentially two forms of support: process support and content support. Process support may address the core scientific reasoning skills as well as the strategic decisions involved in controlling the inquiry process and the process of articulating and reflecting on what has been learned (Lazonder, Hagemans & de Jong, 2010). Content support, by contrast, aims to provide students with the prior domain knowledge needed to make the inquiry productive, or offer information that is too difficult for students to induce from their own investigations.

Most research has concentrated on the effects of offering process. Studies on the effects of content support are comparatively rare, although they could reveal possibilities to further improve inquiry learning. A study by Lazonder et al. (2010) investigated the effect of offering students domain specific content knowledge before and during an inquiry task. Students in this study worked on a fictitious inquiry task so as to ensure that they had no prior knowledge about the topic they investigated. Three conditions, receiving domain specific facts before the task, receiving domain specific facts before and during the task and receiving no domain information at all were compared. Results indicated that students who received domain information generated more knowledge than students receiving no domain information. Furthermore, learning results of students who received domain information before and during the task were superior to the results of students who only received the domain information before the task. This indicates that keeping domain information available during the task has a positive effect on knowledge generation.

Despite these favorable results, it is somewhat counterintuitive to offer students domain-specific facts that they could have otherwise found out by themselves. Still, in order to generate hypotheses, students need some information about the topic they are investigating. Without any prior knowledge, deceleration or even stagnation of learning is probable (Furtak et al., 2012; Hmelo-Silver et al., 2006; Kirschner et al., 2007). One possibility to overcome these learning obstacles and promote effective inquiry learning without offering support through domain specific facts might be to offer students testable hypotheses that could be right or wrong. By offering hypotheses, students are not confronted with true facts but have to investigate the information to establish its truth value. The given hypotheses thus serve as a starting point for their investigations.

Generating and testing hypotheses is one general learning mechanism people use. Following the propositions of the hypothesis testing model (HTM), hypothesis testing is a real-time learning process wherein the student continuously builds and justifies hypothesized data trial by trial (Yu, Smith, Klein & Shiffrin, 2007). In general, students need to have some conceptual understanding in order to create hypotheses (Marzano, Pickering & Pollock, 2001). So, by providing students with pre-defined hypotheses they are offered some conceptual knowledge to begin investigating a topic. In this way, students are able to start directly with an unknown task with only low prior knowledge while at the same time they have the freedom to investigate on their own.

To summarize, a review of the research literature showed, that few studies have examined the educational advantages of offering predefined hypotheses during inquiry learning. This nevertheless seems to be an interesting possibility as it brings together the main aspects of minimally guided approaches (allow exploration) and direct instructional approaches (provide support) in a guided discovery approach.

The present study therefore attends to the effect of offering testable hypotheses to students before and during an unknown inquiry task. To establish if there is a difference between offering hypotheses and offering domain specific facts on knowledge acquisition, both are compared. Before presenting the design and methods of the study, a theoretical framework is described from which research hypotheses were derived.

Theoretical framework

Different models that try to explain scientific reasoning have been developed. Rivers and Vockell (1987) described scientific reasoning as a cycle that is composed of planning (experiment designing), executing (experiment completion and data collection), and evaluating (data analysis and hypothesis generation). Friedler, Nachmias and Linn (1990) also developed a cycle of different steps to explain scientific reasoning. These steps comprise: definition of a scientific problem, generation of a testable hypothesis, designing an experiment, collecting and evaluating data, applying results and drawing conclusions out of results.

The present study was based on the implications of Klahr and Dunbar's (1988) model of scientific discovery as dual search (SDDS) (Figure 1). The SDDS model captures two main properties of scientific learning: generating hypotheses and testing them via experiments. While exploring things, a search in two related problem spaces, the hypothesis space and the experimental space, is required. Scientific learning is divided in three steps that form an iterative cycle. The first step is searching for a testable hypothesis. To conduct such a hypothesis, people can use prior knowledge or generalize from previously conducted experiments. The next step is to test the hypothesis by designing and conducting experiments. The evidence that results from these experiments is then evaluated and further steps are considered. Those steps may include acceptation or rejection of the current hypothesis. Another possibility is that the evidence is inconclusive so that the hypothesis has to be tested again.

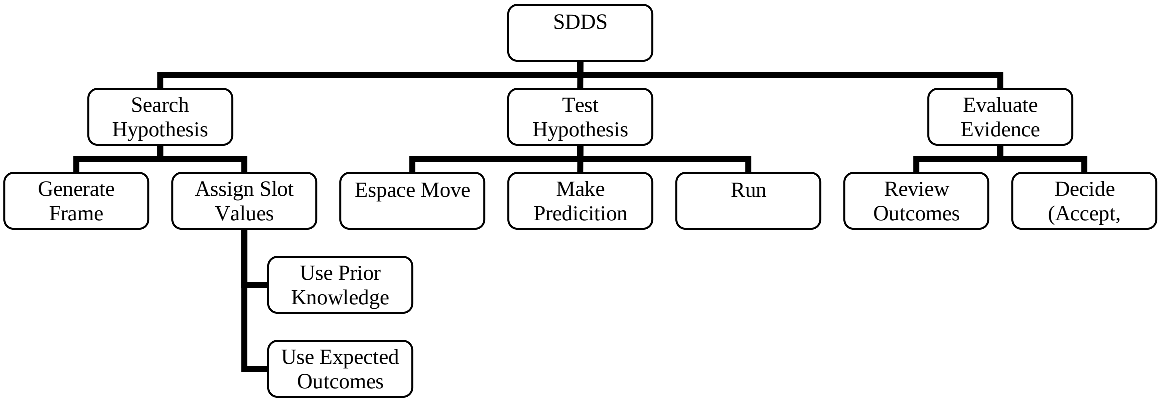

Figure 1. Process hierarchy for the model of scientific discovery as dual search. The figure is an extract of the model presented in Klahr and Dunbar's (1988). Only parts that are important for the given research are displayed.

To develop the SDDS model, Klahr and Dunbar (1988) investigated the connection between students' domain knowledge and their approach to the inquiry task. Students were characterized as experimenters or theorists based on their experimentation behavior. Both experimenters and theorists started searching the hypothesis space but differed in behavior once the initial hypotheses were rejected. Experimenters began to use a data-driven research method whereas theorists kept searching the hypothesis space. It was assumed that prior knowledge enables students to use a theory-driven research approach whereas low prior knowledge causes students to use a data-driven research method. Lazonder, Wilhelm and Hagemans (2008) confirmed this assumption. Other research has shown that theory-driven methods are superior in terms of effectiveness and efficiency. Students using a theory-driven method need less time and fewer experimental trials to complete an inquiry task (Klahr & Carver, 1995).

The previously mentioned study by Lazonder et al. (2010) used the results of Klahr and Dunbar's (1988) study as justification for offering domain knowledge to students. By offering domain knowledge before and during an inquiry task, students were endowed with 'prior knowledge' so that they were able to use a theory-driven method of inquiry. Results confirmed that students who received support before and during the task displayed more hypothesis-driven behavior and acquired more knowledge than students who received support only before the task and students who received no support at all.

Research design - hypotheses

Based on implications drawn from the SDSS model, the present study ought to assess the relative effectiveness of offering hypotheses and domain information. With regard to these implications and the earlier mentioned strengths that pre-specified hypotheses may have, there may be two reasons why offering pre-specified hypotheses can enhance inquiry learning.

The first reason is that offering pre-specified hypotheses helps promote a theory-driven method of inquiry. As mentioned earlier, theory-driven investigations lead to higher knowledge gains. The difficulty of generating testable hypotheses would be removed and thus students are supported in scientific inquiry learning. Although offering hypotheses resembles offering domain information, students have to think actively about the information they receive instead of just accepting it. They have to test if the given hypotheses are right or wrong. By testing hypotheses—instead of just accepting them—students have to think more about the variables and relations in the task domain. It is assumed that more knowledge is generated if people think actively about a topic.

The second reason is that pre-specified hypotheses may serve as example for the creation of students' own hypotheses. Students thus learn how a good, testable hypothesis is formulated and might use this knowledge when building their own hypotheses.

The present study tries to investigate these possible positive effects of offering hypothesis based support to students in order to find ways to enhance knowledge acquisition through inquiry learning.

The assumption underlying the present study was that offering testable hypotheses would improve students' inquiry learning. It was hypothesized that the general performance in an inquiry based task, the knowledge acquired and the hypothesis generation skills of students would be enhanced. To test these assumptions, the study compared the performance of low prior knowledge students offered domain information (*DI-condition*), hypotheses information (*HI-condition*) or no support (*control condition*) during a simulation based inquiry task.

With regard to the aforementioned implications of the SDDS model, two research hypotheses were generated.

The first hypothesis stated that participants in both experimental conditions would use a theory-driven method of investigation *(Hypothesis 1)*. It was assumed that participants in the HI condition and DI condition would base their experiments on specific hypotheses they tried to investigate. Unlike participants in the control group, they received information that they could use to generate testable hypotheses. This information would serve as 'prior knowledge' allowing participants to use a theory-driven method of investigation from the start. Participants in the control condition were expected to start with a data-driven method of investigation which either stays data-driven or develops into a more theory-driven method of investigation.

Several testable predictions were derived from this hypothesis. First, it was assumed that the specificity of hypotheses would increase over time in all conditions. However, participants in both experimental conditions were expected to start off with more specific hypotheses than their control counterparts (*Hypothesis 1a*) because both types of support were assumed to contribute to a theory-driven method of investigation. Offering no support leaves participants with a data-driven method of investigation which is considered to be inferior in terms of effectiveness and efficiency (Klahr & Carver, 1995; Okada & Simon, 1997). Participants in the control group were therefore expected to conduct more experiments and need more time to discover the effects of the variables than participants in both experimental conditions *(Hypothesis 1b)*.

In addition, a difference between both experimental conditions was predicted. Participants in the HI condition would need more experiments, and hence more time, to investigate the topic than participants from the DI condition because they had to rigorously test all information they received *(Hypothesis 1c)*. Participants in the DI condition could concentrate on the parts they had received no information about, and were thus expected to conduct fewer experiments in less time. Concerning exploratory experiments (i.e., experiments not guided by a hypothesis) no difference between the DI condition and the HI condition was expected *(Hypotheses 1d)*.

The second hypothesis stated that participants in the HI condition would acquire more knowledge than participants in both other conditions *(Hypothesis 2)*. This hypothesis was based on two assumptions. First, participants in the HI condition needed to rigorously test all information they received because they were told that the pre-specified hypotheses might not necessarily be true. They would therefore investigate the topic more thoroughly than participants from the other two conditions. It was thus assumed that participants in the HI condition would detect the underlying interaction effect more often than participants in both other conditions *(Hypothesis 2a)*. Secondly, participants in the control group had no support to overcome learning obstacles so they were assumed to generally find out fewer effects than their counterparts in both experimental conditions *(Hypothesis 2b)*.

Method

Participants

The participants of the study were 67 undergraduate students, with a mean age of 20.68 (SD=4.46). Some students participated in the experiment for course credits, others volunteered. At the beginning of the experiment all participants were randomly assigned to the DI condition (n=23), the HI condition (n=26) or the control condition[1] (m=18).

Materials

Inquiry task and learning environment The experiment was designed around an inquiry-based learning environment that involved the sale figures of a shoe store. Participants were asked to find out how each of five factors influenced the store's weekly sales. The baseline sales figure was set at 75 pairs of shoes per week and could be enhanced to 89 pairs of shoes per week through manipulating each of the five factors.

The task was designed to ensure that participants had no prior-knowledge about the topic the topic of inquiry as to guarantee the comparability of all subjects. Although the chosen context had realistic elements, the influences of the factors on a shoe store's weekly sales were arbitrary, thus erasing the possibility to rely on prior-knowledge.

The five factors were type of background music, location of the stockroom, type of floor covering, displayed shoe and the building's roofline style. Every factor contained different changeable values that could be manipulated.

The factor type of background music contained the values jazz, blues and country music. Changing the background music to blues or country enhanced weekly sales by respectively two and four pairs of shoes. Location of the stockroom includes the values first floor and basement. Placing the stockroom in the basement caused an increase of three pairs of shoes. Weekly sales could be increased by two pairs of shoes if floor covering was changed from carpet to wooden floor and by one pair of shoes if floor covering was changed to vinyl floor (interaction effect). The fourth factor, displayed shoe enhances weekly sales by three pairs if the right shoe is presented along with wooden floor covering and enhances weekly sales by two pairs if the right shoe is presented alongside vinyl floor covering. Roofline style is the last factor, contained of the

[1] Data of the control condition was taken from previous study by Lazonder et al. (2010). The procedure and constraints of both studies were the same.

values spout gable, curved gable and stepped gable. Changing the style to curved or stepped gable would enhance weekly sales by one and two pairs of shoes.

Four of these factors were independent and did not influence each other. There was, however an interaction effect between the factors displayed shoe and type of floor covering. This means that the chosen floor covering ‘changes’ the effect of the displayed shoes.

It was tried to create factor values which are comparable so that no assumptions could be made based on their content. This can be illustrated through an example. Participants were asked if weekly sales would be higher if the left or the right shoe is presented in store. Obviously the content of the values, thus left or right shoe is equal. The possibility to draw conclusions from the factor values is thus limited because people have no idea about the differences between left and right shoes that would cause a higher effect on weekly sales. Participants were thus fully dependent on their experimental abilities and the available support to investigate the impact of the presented shoe. However, sometimes it was difficult to create ‘equal’ factor values. As it comes to music there is without a doubt a difference between jazz, country and blues. Considering this, the factor values were created without any reference to the realistic differences between the types of music.

The impact that different factors had on the weekly sales could then be investigated and observed. Once all factor values were set, participants had to predict the outcome and click on the start button to run the experiment. The outcome appeared in a graph window (see Figure 2). By analyzing the outcome, students could reject, accept, or refine their hypotheses as well as generate new ones. The history window below the graph window showed the simulation output along with the selected values, and could be scrolled to review previously conducted experiments. Furthermore, participants had the possibility to take notes or write down their conclusions in a text editor.

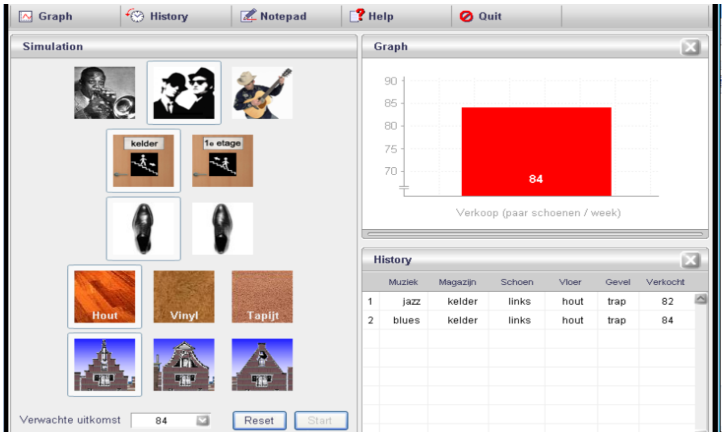

Figure 2. Interface of the learning environment.

In case participants forgot some basic information about the task they could click on a help button. A helpfile viewer showed a general description of the task and its cover story. In both experimental conditions (DI and HI) participants also received some extra information, regarding the magnitude and direction of effect on the weekly sales figure (e.g., Sales increase by 1 if the floor covering is changed from carpet to vinyl.) (See Figure 3). Both groups received the same kind of information; however, participants in the DI condition received facts, whereas participants in the HI-condition received hypotheses that needed testing before their truth value was established.

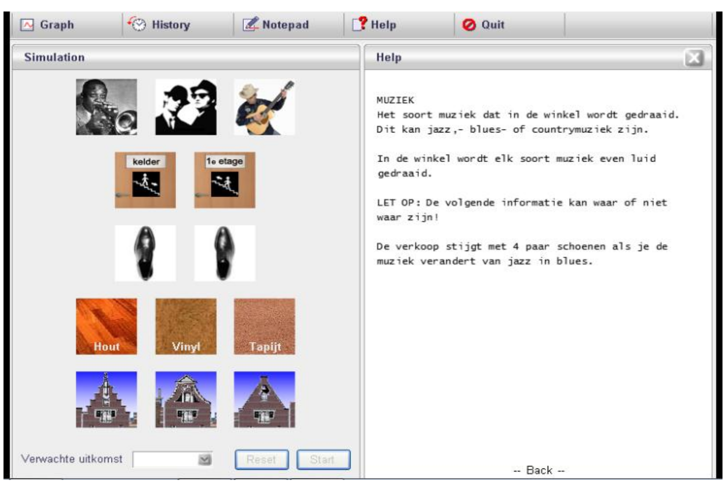

Figure 3. Screenshot of the help file viewer used in the HI condition. The image shows the help file for the factor ‘music’.

Content support - Domain information guide. At the start of the experiment, students in the DI-condition received an information guide introducing the problem statement and the cover story. The guide also included domain information about the factors. Additional information was given about three factor values' magnitude and direction of effect on weekly sales. The information related to the factors background music ('Sales increase by 4 if background music is changed from jazz to country'), displayed shoe ('Sales increase by 2 if the right shoe is displayed and there is a vinyl floor covering instead of a carpet floor covering') and type of floor covering ('Sales increase by 1 if the floor covering is changed from carpet to vinyl'). Because the given information was limited, participants had to find out the largest part by themselves. The content of the domain information was similar to the previously described help files for participants in the DI condition.

Content support - Hypothesis information guide. The guide given to participants in the HI-condition resembled the domain information guide in that it included the cover story and some general information about the variables. The difference with the domain information guide was manifested in the information about the factor values. The participants did not receive facts, but three hypotheses concerning factor values' effects on weekly sales. Participants were told that the given hypotheses could be right or wrong. They did not know which - and how many hypotheses are right or wrong. The two hypotheses concentrating on background music ('Sales increase by 4 if background music is changed from jazz to blues') and floor covering ('Sales increase by 2 if the floor covering is changed from vinyl to carpet') were wrong; the one concentrating on displayed shoe ('Sales increase by 2 if the right shoe is displayed and the floor covering is changed from carpet to vinyl') was correct. The content of this guide was similar to the help files that were available to participants in the HI condition.

Procedure

Demographic characteristics were collected at the point of sign up for the experiment. The experiment was structured so that only one participant at a time could attend. All participants received the same basic instructions. To guarantee that the task and the learning environment were well understood, a little training with a simple inquiry task was given. This task resembled the actual inquiry task in structure so that participants had the possibility to familiarize themselves with all functions of the task. Participants were randomly assigned to one of the two experimental conditions (DI and HI condition) or the control condition. Based on their assignment to conditions, they were given the domain information guide, the hypothesis information guide, or no guide at all. It is important to mention that there was no fixed time limit, so as to ensure that participants had enough time to transfer all information to memory. Participants in the control condition got no guide, thus no additional information.

After this preparation phase, students continued with the inquiry task. The task was introduced through a cover story, presenting the main factors of the task and the underlying problem statement. After this introduction participants were invited to start their investigation of how each of the five factors influenced the store's weekly sales. Each time a participant executed an experiment, the experimenter asked for the underlying hypotheses. Two questions were used, concerning the factor of investigation ('what are you going to investigate?') and the suspected effect on the output variable ('what do you think will be the outcome?'). All answers were written down for later analysis. Participants had 45 minutes maximum to complete the task. It was however possible to quit earlier. All interactions participants had with the digital learning environment were recorded in a log file.

Measures and scoring procedure

Four variables, which are described in detail below, were used to test the two stated hypotheses of the study. These variables were: duration of the experiment, hypotheses stated, experiments conducted, and the final knowledge of the effects of single factors.

The variable 'duration of the experiment' concerned both, the *time* participants needed in the preparation phase and time they needed to complete the task. Time needed in the preparation phase was measured by the experimenter; time needed to finish the task was taken from the log files.

'Hypotheses stated' concentrated on the *quality of the hypotheses* stated by participants through time and across conditions. Participants' hypotheses were arranged chronologically and divided into four quartiles. They were furthermore classified as fully specified, partially specified or unspecified hypotheses. Fully specified hypotheses contained two or more factor values and a prediction of the direction and magnitude of the change in weekly sales (e.g., 'Sales increase by 4 if music is changed from jazz to country'). Partially specified hypotheses comprised the direction of change of two or more factor values (e.g., 'Sales increase if music is changed from jazz to country'). Unspecified hypotheses noted the possible effect of a factor value but neither contained the direction nor magnitude of change (e.g., ‘Changing the music from jazz to blues has an effect on weekly sales’). Scores for each quartile were then calculated and could be compared. Two raters coded the hypotheses of four randomly selected scoring sheets of each condition. Inter-rater agreement (Cohen’s K) reached 1.00.

Via the variable ‘experiments conducted’, the *characteristics of experiments* participants carried out were measured. Experiments were classified as unique or duplicated. Unique experiments were defined as ‘simulation runs testing a new, untried combination of factor values’ (Lazonder et al., 2010). If the same combination of factor values was administrated more than one time, the trial was characterized as duplicated. Furthermore it was assessed whether experiments were based on a hypothesis to calculate the percentage of exploratory experiments compared to the number of experiments that were not guided by a hypothesis to the total number of experiments.

The *overall performance success* was measured through the variables ‘overall score’, ‘interaction recognition’ and ‘average recognition of unknown effects’. The variable ‘overall score’ involved the number of correct conclusions. The log files were used to generate an overall score of success. Results were analyzed through a hierarchical rubric. Every note with reference to a certain factor written down in the log files was classified into one of three categories. According to these categories, participants received one, two or three points. Participants could earn three points for a factor if both the magnitude and direction of an effect were correct for each value of the factor. Two points could be earned if the correct direction of effect was mentioned for each value of a factor. One point was given if participants recognized that certain values have an effect on weekly sales but neither mentioned the direction nor the magnitude of effect. If participants wrote down more than one conclusion for a certain value, the answer was also scored with one point for recognizing that the value has an effect. It was possible to reach an overall score of 15 points. To measure the specificity of investigation, the variable ‘interaction recognition’ indicated the percentage participants per condition that recognized the interaction effect. The difference in knowledge generation between the control group and both experiment groups was measured through the variable ‘average recognition of unknown effects’. This variable was created to detect the average score per variable participants had no information about, thus the amount of information participants found out by themselves. For both experimental groups the average scores of the variables location of the stockroom and roofline style were calculated; for the control group the average score of all variables was taken.

Two raters generated an overall score of six randomly selected sets of log file data for each condition. Inter-rater reliability was satisfactory (Cohen's K= .89).

Results

The descriptive statistics for participants' performance by condition are summarized in Table 1. As the data indicate, there were no significant differences in 'time on task' between the three groups. Time during executing phase was measured for all three conditions. Univariate analysis of variance (ANOVA) showed no significant differences, F (2, 64) = 1.37, p = .26.

Table 1
Means (and SD) for participants' performance by condition

	Condition		
	DI (n = 23)	**HI (n = 26)**	**Control (n = 18)**
Time on task (min.)			
Preparation phase	7.65 (3.19)	7.65 (3.39)	
Execution phase	24.09 (10.02)	27.58 (8.59)	28.76 (10.51)
Number of experiments	24.39 (16.04)	34.08 (17.31)	18.11 (9.67)
Unique experiments (%)	88.05 (9.93)	78.51 (10.76)	93.45 (8.22)
Duplicated experiments (%)	11.95 (9.93)	21.49 (10.76)	6.18 (8.37)
Exploratory experiments (%)	3.65 (5.95)	1.53 (4.21)	41.95 (15.17)
Number of Hypotheses[a]	23.91 (16.15)	33.35 (17.00)	9.78 (5.63)
First Quartile (Q1)	1.43 (0.63)	1.39 (0.38)	0.88 (0.65)
Second Quartile (Q2)	1.77 (0.51)	1.78 (0.35)	1.20 (0.64)
Third Quartile (Q3)	1.85 (0.41)	1.91 (0.40)	1.42 (0.60)
Fourth Quartile (Q4)	1.94 (0.49)	2.11 (0.48)	1.52 (0.79)
Performance success			
Overall score	10.91 (0.86)	11.19 (2.93)	7.67 (2.22)
Interaction recognition (%)[b]	39.13	57.69	0.2
Average points for unknown effects[c]	2.65 (0.65)	2.5 (0.80)	1.53 (0.45)

Note. DI = domain information support; HI = hypothesis information support; control = no support.
[a] Scores represent the mean domain specificity of the participants' hypotheses within each quartile.
[b] Percentage of participants that detected the interaction effect
[c] The average score of effects no information was given about

Additionally, time during the preparation phase was measured for both experimental groups. Again, there was no significant difference detected, F (1, 47) = 0.00, p = 1.00.

The overall number of experiments was displayed through the average number of experiments participants needed in each condition. ANOVA showed that there is a significant difference between the average number of experiments, F (2, 64) = 6.21, $p < .01$. Confidence intervals showed that the number of experiments was significantly higher in the HI condition than in the control condition, $p < .01$. There was no significant difference between the DI condition and the control group $p = .58$. Furthermore there was no significant difference found between both experimental groups $p = .88$.

The overall number of experiments is composed of 'unique experiments' and 'duplicated experiments' which are given as percentages in Table 1. Multivariate analysis of variance (MANOVA) revealed that condition has a significant effect on the mean number of unique and duplicated experiments, Pillai's trace = 0.32, F (4, 128) = 6.12, $p < .01$. Both the DI and the control condition had nearly 90 percent unique experiments. For the HI condition the percentage unique experiments approached 80 percent. ANOVA confirmed that this difference is significant, F (2, 64) = 13.143, $p < .01$. As it comes to duplicated experiments the reverse pattern is observed. ANOVA showed a significant difference, F (2, 64) = 13.58, $p < .01$. Confidence intervals revealed that in the HI condition there were significantly more duplicated experiments than in both other conditions, $p < .01$.

Table 1 gives also the percentage of exploratory experiments. Exploratory experiments were defined as the number of experiments that were conducted without hypothesis. The number of exploratory experiments differed significantly among conditions, F (2, 64) = 127.95, $p < .01$. Approximately 40 percent of all experiments made by respondents of the control group were exploratory. For respondents in the DI and HI condition those percentages were lower than five. Confidence intervals showed that students in the control condition had significantly more exploratory experiments than students from both experimental conditions, $p < .01$. There was no significant difference found between the DI condition and the HI condition, $p = 1.00$.

The mean number of hypotheses, calculated for each condition, can also be found in Table 1. Participants in the control group had the least hypotheses, followed by the DI condition and the HI condition. ANOVA showed that this difference is significant, F (2, 64) = 14.01, $p < .01$. Confidence intervals showed that the difference between the experimental groups and the control group was significant, $p < .01$. There was no significant difference between the DI condition and the HI condition, $p = .08$.

The overall mean domain specificity score for each quartile and per condition can also be found in Table 1; a graphically presentation of the results is given in Figure 4.

To analyze how hypotheses evolved through time and in each condition, a mixed-design ANOVA was carried out. Data indicated that the sphericity assumption was violated, χ^2 (5) = 21.58, $p < .01$. As the Greenhouse-Geisser correction, $\varepsilon = 0.81$, seemed to be higher than 0.75, the correction of Huynh and Feldt (1976) was used for analysis of within subject effects. Results showed a significant within subject effect of time, F (2.62) = 26.12, $p < .01$. Confidence intervals showed significant within-subject effects between the first and second quartile, $p < .01$, and the second and fourth quartile, $p < .01$. There was no significant within-subject difference between the second and third quartile and the third and fourth quartile. Furthermore, analysis showed a between-subject effect of condition, F (2, 64) = 12.96, $p < .01$. Confidence intervals showed a significant between-subject difference between the control group and both experimental conditions, $p < .01$. There was, however, no significant between-subject effect found between HI condition and DI condition, $p = 1.00$. Furthermore, no interaction effect was found between time and condition, F (5.24) = 0.38, $p = .87$.

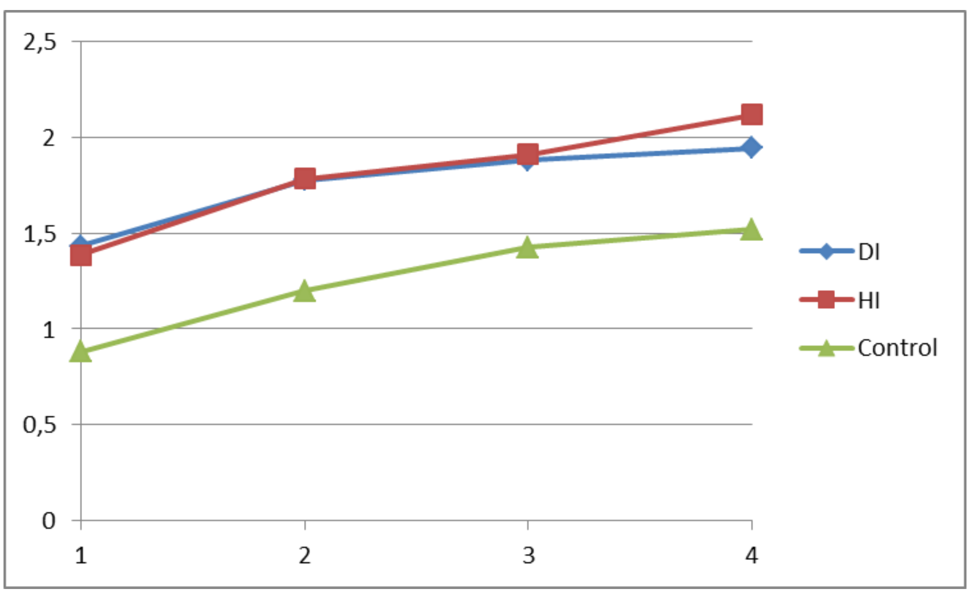

Figure 4. Mean domain specificity of participants' hypotheses through time and by condition. 1 = Domain information condition. 2 = Hypotheses information condition. 3 = Control condition.

Performance success is also displayed in Table 1. It concerned the average overall score participants received, the interaction recognition and the average recognition of unknown effects. ANOVA showed a significant difference in the average overall score, F (2, 64) = 10.22, $p < .01$. Confidence intervals revealed that the HI condition and the DI condition had significantly higher scores than the control group, $p < .01$. There was no significant effect between HI condition and DI condition, $p = 1.00$. Recognition of interaction effect described the percentage of participants in each condition that recognized the interaction effect in the task. Chi-Square test revealed no significant difference between both experimental groups, χ^2 (1, 49) = 1.68, $p = .2$. The difference between DI and control condition, χ^2 (1, 41) = 6.17, $p < .05$, and HI and control condition, χ^2 (1, 44) = 12.49, $p < .01$, was significant. ANOVA showed a significant difference in the average score of unknown effects, F (2, 64) = 14.47, $p < .01$. Confidence intervals revealed that the HI condition and the DI condition had significantly higher scores than the control group, $p < .01$. Again, no significant difference was found between both experimental groups, $p = 1.00$.

Discussion

The present study investigated whether offering pre-specified hypotheses facilitates the knowledge acquisition through inquiry learning in low prior knowledge students. It was expected that either providing hypotheses or domain information would lead participants to use a theory-driven method of investigation, which is considered to be superior to data-driven methods of investigation. It was furthermore expected that participants in the HI condition would gather more knowledge than those in the DI and control condition. Both hypotheses were partly supported by the results.

Evidence for the first hypothesis comes from the comparison of the HI condition with both the DI condition and control condition.

It was assumed that time would have an influence on the specificity of hypotheses. Furthermore the specificity of hypotheses in both experimental conditions was predicted to start at a higher level than the specificity of hypotheses in the control condition (*Hypothesis 1a*). As predicted, time had an influence on the specificity of hypotheses. Data analysis of the quartiles showed that the number of specific hypotheses increased gradually in all conditions. There was no interaction found between the conditions and time. However, even if participants from the control group enhanced the number of specific hypotheses gradually, it was still remarkably lower than in both other conditions. As data of the HI group resembles data of the DI group, it seems likely that both groups used the same method of investigation. Both groups started mostly with formulating partly-specified hypotheses and ended with formulating fully-specified hypotheses. This indicates that participants in both experimental conditions more often used a theory-driven method of investigation, whereas participants in the control condition used more often a data-driven method of investigation which partly enhanced.

It was furthermore assumed that participants in the control group would conduct more experiments and need more time than their counterparts in both experimental conditions (*Hypothesis 1b*). The hypothesis that participants in the control group used more experiments than participants in both other groups was not confirmed. In fact, participants in the control group conducted far less experiments than participants in the DI and HI condition. At first sight this seems to contradict the presumption of a data-driven method of investigation. As Klahr and Carver (1995) and Okada and Simon (1997) found, data-driven methods of investigation are inferior in terms of effectiveness and efficiency. Certainly this is true for investigating all parts of a task. However, during the given inquiry task, it was possible for participants to stop whenever they wanted. It is thus possible that participants in the control group stopped with the task right before they found out all possible information. A look at the data confirms this assumption. Participants in the control group had a significant lower overall score than participants in both other groups which indicates that they gained less knowledge. The same constraints apply to the time participants needed during the execution phase. Again, no significant difference was found between participants in the control group and participants in both experimental groups. However, it can be assumed that the control condition would have needed much more time to reach an overall score as high as both other conditions had.

The third assumption stated that participants in the HI condition would need more experiments and time than participants in the DI condition (*Hypothesis 1c*). Data showed that participants had a higher average number of experiments than both other groups, though this difference was not significant between both experimental groups. However, the average difference between HI and DI condition was almost 10, which may serve as an indication for a general direction of effect.

The assumption that the percentage of exploratory experiments would not differ significantly between both experimental conditions was confirmed (*Hypothesis 1d*). However, the control condition had a significantly higher number of exploratory experiments. This means that comparatively many experiments in the control condition were not based on hypotheses, which may also be an indication of a data-driven approach.

The second hypothesis which stated that participants in the HI condition would acquire more knowledge than their counterparts in both other conditions was partly confirmed. Evidence came from the comparison of the HI condition with both the DI condition and the control condition. Participants in the HI condition had a higher average overall score than participants in the control condition. However no significant difference was found between participants in the HI condition and participants in the DI condition.

This finding is in keeping with the earlier discussed evidence for participants in HI condition and DI condition using more of a theory-driven method of investigation which is assumed to enhance knowledge gains.

The assumption, that participants in the HI condition would detect the interaction effect more often than participants in both other conditions was not confirmed (*Hypothesis 2a*). Although on first view the percentages seemed to differ, there was no significant difference between the percentage interaction recognition between both experimental groups. However, the control group detected the interaction effect significantly less often than the experimental groups, which is in line with the hypothesis that offering support enhances knowledge construction.

The assumption that control participants would generally find out fewer effects than their counterparts in both experimental conditions was confirmed *(Hypothesis 2b)*. This was measured by the average recognition of unknown effects, and substantiates the importance of offering support to students with low prior knowledge. This is in line with the results of earlier studies concerning the effects of inquiry learning support (Hmelo-Silver et al., 2007; Lazonder et al., 2010).

Although the results confirm both hypotheses only partly, they indicate different effects of offering students information they need to investigate instead of giving them facts they directly can accept. First, it was unexpected that no statistically difference was found between ‘overall score’ and the percentage ‘interaction detection’ of both experimental groups. This finding is at odds with the assumption that participants receiving hypotheses information support would investigate a topic deeper and find out more than their counterparts. Another way of looking at this is by inspecting the percentages interaction detection because they offer a view on the depth of investigation. Even if they are not significantly different, they may indicate a general direction of effect. Participants in the HI condition detected the interaction effect in nearly 60 percent of the cases whereas participants in the DI condition detected the effect only in 40 percent of the cases, so it can be assumed that offering hypotheses enhances the specificity of investigation at least gradually.

Furthermore, it was interesting that participants in the HI condition had the highest average number of experiments, which normally is assumed to go along with a data-driven method of investigation (Klahr & Carver, 1995; Okada & Simon, 1997); however, analysis showed that a theory-driven method of investigation was used. The experiments were mostly guided and not exploratory. So why did participants receiving hypotheses conduct more experiments than participants receiving domain information support or no support at all? One possible reason could be that they felt like testing every little bit of the topic twice because they were told that all information they got may be false. It was remarkable that the number of duplicated experiments was especially high for participants in the HI condition, whereas analysis of the hypotheses showed that they evolved to be very specific. It seems thus likely that participants wanted to reassure that their results were coherent. One tentative conclusion is therefore participants in the HI condition investigated the topic more thoroughly than their counterparts in both other conditions. The present findings thus indicate positive effects of offering students hypotheses based support during inquiry learning. It seems clear that offering hypotheses is at least as good as offering domain information in supporting students during inquiry learning. It would be interesting to investigate the effects of offering hypotheses information support to other groups of learners, for instance learners without prior experience in scientific investigation or with high prior knowledge learners. Especially for high prior knowledge students, offering hypotheses information support would offer possibilities to direct their attention to different perspectives of a topic they might otherwise ignore because of their prior knowledge. At the same time, it still offers the freedom for students to investigate each part of a topic themselves.

There are also some limitations to the present study. First, the task requested students to use a science model of investigation which is characterized as investigating cause and effect relations. However, it was noticeable that students used an engineering approach of investigation, thus concentrating mainly on finding the highest effects of variables, thus the ‘best solution’ of the problem. Reason for that may be that the problem statement introduced in the cover story suggested that there is a most desirable outcome, which usually goes along with an engineering investigation approach (Schauble, Klopfer & Raghavan, 1991). One possible consequence is that participants concentrated on values that offered the 'best solution', which sometimes kept them from investigating each value equally accurate and made it difficult to detect the interaction effect. Another limitation relates to the way information about the hypotheses participants had was collected. This happened right from the beginning of their investigations through standardized questioning. Participants thus had no time to familiarize with the task. The first hypotheses were mostly vague and unspecific and may have been trials to understand the program. A possible consequence may be that the low hypothesis specificity measured for the first quartile is inaccurate.

The final limitation, also discussed by Lazonder et al. (2010), refers to the characteristics of the task. To guarantee that participants had no prior knowledge the task was fictional. This makes it somewhat difficult to generalize the findings of the study to inquiry learning in class rooms. However, the task concentrates on hypothesis generation and experimentation which are the main activities during scientific inquiry learning. The important thing in inquiry learning is not the task at hand but the way it is executed. Furthermore, it is often the case that even real scientific tasks make no sense to students (Hart, Mulhall, Berry, Loughran & Gunstone, 2000). The given task resembles real tasks in scientific inquiry learning thus pretty well, even if it is fictious. It can thus be concluded that the results can well be generally applied to classroom settings.

Practical implications can be made for classroom settings where students generate and test hypotheses while learning a new topic. The research focused on scientific inquiry learning in a computer based environment, so the results can especially be generalized to such situations. The actual research suggests that offering students hypotheses information is an effective method to help students performing inquiry tasks. Compared to domain information support, it has the benefit of not offering students knowledge they could construct themselves. As advised, this support should be given right before and during the preparation of an inquiry task (Lazonder et al., 2010).

References

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8,* 293-332.

De Jong, T. (2006). Technological advantages in inquiry learning. *Science, 312*, 532-533.

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

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

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta analysis. *Review of Educational Research, 82,* 300-329.

Hart, C., Mulhall, P., Berry, A., Loughran, J., & Gunstone, R. (2000). What is the purpose of this experiment? Or can students learn something from doing experiments?. *Journal of Research in Science Teaching, 37*, 655-675.

Hmelo-Silver, C. E., Golan Duncan, 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.

Huynh, H., & Feldt, L. S. (1976). Estimation of the box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics, 1*, 69-82.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75-86.

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

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., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15*, 661-667

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research based strategies for increasing student achievement.* Association for Supervision and Curriculum Development.

Okada, T. and Simon, H. A. (1997), Collaborative Discovery in a Scientific Domain. *Cognitive Science*, *21*, 109-146.

Palles, M. A. (2010), Meeting the basic educational needs of children and youth. *Children and Youth Services Review, 32*, 1199-1210.

Piaget, J. (1970). Piaget's theory. In P. Mussen (Ed.), *Carmichael's manual of child psychology* (Vol. 1, pp. 703-772). New York: John Wiley & Sons.

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

Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, *88,* 345-372.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32*, 102-119.

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

Lazonder, A. W., Hagemans, M. G. & De Jong, T. (2010). Offering and discovering domain information in simulation-based inquiry learning. *Learning and Instruction*, *20*, 511-520.

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, 580e592.

Van Jolingen, W. R., De Jong, T., & Dimitrakopoulout, A. (2007). Issues in computer supported inquiry learning in science. *Journal of Computer Assisted Learning, 23*, 111-119.

White, B., & Frederiksen, J. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*, 3–118.

Yu, C., Smith, L. B., Klein, K. A., & Shiffrin, R. M. (2007). Hypothesis testing and associative learning in cross-situational word learning: are they one and the same? *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, August, 2007.