Comparing the effectiveness of two support approaches in simulation-based inquiry learning: Can pre-defined hypotheses compare with domain information?

S. Zierbock

(s0157562)

University of Twente, Enschede

Faculty of Social Sciences

Department Instruction Technology

First supervisor: Dr. A. W. Lazonder

Second supervisor: Dr. P. Wilhelm

Summary

The present study examined whether providing pre-defined hypotheses supports learners equally well as providing additional domain information in a simulation-based inquiry learning task. The effects on both knowledge gain and strategy use were examined in a posttest-only control group design. 58 students from the University of Twente were randomly assigned to either one of the support conditions or the control group, Results showed no significant differences between neither of the support conditions and the control condition indicating that the support measures did not benefit learners. A comparison between both support approaches found no significant differences either, indicating no superiority of one support measure compared with the other. However, the heterogeneity of the sample and the associated large standard deviations within the conditionsmay have hindered the detection of benefitial effects of both support approaches, and differences in effectiveness between them.

Samenvatting

Dit onderzoek heeft onderzocht of het aanbieden van hypotheses net zo effectief is als het aanbieden van domein informatie tijdens onderzoekend leren met een computer simulatie. Hierbij werd de invloed van beide supporten op het verwerven van domein kennis en op het gebruik van strategiëen onderzocht binnen een posttest-only control group design. 58 studenten van de Universiteit Twente werden willekeurig verdeelt op of één van de support condities, of op de controle groep. De resultaten lieten geen significante verschillen zien tussen geen van de support condities en de controle conditie. Dit zou erop kunnen wijzen dat de supporten de lerenden niet ondersteunden tijdens de taak. Ook een vergelijking tussen beide support condities liet geen significante verschillen zien. Dit impliceert dat er geen verschillen in effectiviteit waren tussen de twee supporten. Ondertussen is aan te nemen dat de heterogeniteit van de steekproef en de daarmee verbondene hoge standard deviaties in de groepende opsporing van de effectiviteit van de supporten, en verschillen tussen deze, belemmerde.

1. Introduction

Engaging in scientific inquiry is an important part of contemporary science education. (National Research Council, 1996). By actively engaging in scientific inquiry, students not only acquire scientific knowledge and develop scientific reasoning skills, but they also learn how to think critically about scientific findings (National Research Council, 1996). When engaging in simulation-based inquiry learning, which is widely used in the science classrooms nowadays, students act like scientists: they orient themselves to the task domain, generate hypotheses, conduct experiments, draw conclusions and make evaluations of their own learning process (de Jong, 2006; Swaak, de Jong, & van Joolingen, 2004; van Joolingen, 1999). The use of simulations as compared to hands-on inquiry in the laboratorium or in a natural environment has the advantage that a large variety of scientific phenomena can easily be brought to the classroom (van Joolingen, de Jong, & Dimitrakopoulou, 2007).

Past research into the effectiveness of unguided inquiry learning has, however, shown that it often does not yield the expected learning outcomes (Alfieri, et al., 2011; Mayer, 2004). Especially, learners with a low level of prior knowledge experience difficulties when completing an inquiry task (Hmelo, Nagarajan, & Day, 2000; Kalyuga, 2007; Kalyuga, Chandler, Tuovinen, & Sweller, 2001; Schauble, Glaser, Raghavan, & Reiner, 1991). One of the reasons seems to be that low prior-knowledge learners' strategy use is impaired (Hmelo, Nagarajan, & Day, 2000; Lazonder, Wilhelm, & Hagemans, 2008; Schauble et al., 1991). They often are less proficient in experimentation processes such as managing data, reasoning goal orientedly, generating evidence, interpreting evidence and evaluating their own progress (Hmelo, et al., 2000; Schauble et al., 1991). As there appears to be a reciprocal relationship between strategy use and domain knowledge (Klahr & Dunbar, 1988) one way to support low prior knowledge students in an inquiry task is by compensating for missing domain knowledge. A recent study (Lazonder, Hagemans, & de Jong, 2010) investigated the effectiveness of offering low prior-knowledge learners domain information before and during an inquiry task. This approach turned out to be successful: learners who received the domain information gained more knowledge, and showed improved strategy use. Yet, revealing a large part of the information learners are supposed to find out on their own contradicts the constructivist nature of scientific inquiry learning, where learners are supposed to acquire new knowledge as independently as possible and through their own experimentation. This calls for a support measure that promotes knowledge acquisition and strategy use but avoids giving away too much information about the domain.

The aim of the present study was to examine whether the provision of pre-defined hypotheses with an unknown truth value is as effective as the provision of domain information in stimulating strategy use and knowledge gain. The SDDS model (Klahr & Dunbar, 1988), which is a general model of scientific reasoning, and the Extended SDDS model (van Joolingen & de Jong, 1997) have been used to predict the differential effects both approaches may have on strategy use and knowledge gains.

2. Theoretical framework

The Scientific Discovery as Dual Search (SDDS) model (Klahr & Dunbar, 1988) gives a general description of scientific reasoning. It is a hierarchical model in which the three main processes are: generating a hypothesis, testing it, and evaluating the evidence to either accept or reject the hypothesis, or to continue examining it. To both generate a hypothesis and conduct an experiment, two problem spaces need to be searched: the *hypothesis space* and the *experiment space*. The hypothesis space consists of the hypotheses generated during the reasoning process, and the experiment space consists of all possible experiments that can be conducted with the equipement or simulation at hand. Hypotheses can be generated in two different ways: either on the basis of prior knowledge or by conducting exploratory experiments and deriving a hypothesis from the experimental outcomes. According to Klahr and Dunbar, especially at the beginning of a scientific reasoning task, the learner will generate hypotheses from prior knowledge. Later in the task, after some experimentation has taken place, learners will also be able to generate their hypotheses from the results of experiments.

To test their model, Klahr and Dunbar (1988) conducted two experiments in which students had to discover a single rule that governed the function of a ‘repeat’ button on a robot tank. When analyzing their students’ strategies, Klahr and Dunbar found that the students could be divided into two groups. One group, which Klahr and Dunbar called Theorists, applied a more theory-driven strategy. Their experimentation was usually guided by a hypothesis which they initially derived from memory. During the course of their investigation, new hypotheses were induced from the data obtained through experimentation. In contrast, Experimenters applied a more data-driven approach. They too started by generating hypotheses from their prior knowledge. However, after the initial hypotheses had been rejected, they continued experimenting without a hypothesis in mind so as to generate new experimental outcomes from which new hypotheses could be formulated. The theory-driven strategy proved to be faster and more efficient in that Theorists needed fewer experiments to find out how the ‘repeat’ button worked. Although Klahr and Dunbar did not explicitly assess their students' prior knowledge, they assumed that most of the students who applied a hypothesis-driven strategy had prior knowledge in programming.

To further describe the internal processes of low and high prior knowledge learners, the extended SDDS model (van Joolingen & de Jong, 1997; see also: Lazonder, Wilhelm, & Hagemans, 2008) is described here. Van Joolingen and de Jong (1997) extended Klahr and Dunbar's (1988) SDDS model to more complex domains by giving a more detailed description of the hypothesis space. Van Joolingen and de Jong divided the hypothesis space into several subspaces. The *universal hypothesis space* contains all possible hypotheses about a domain; the *learner hypothesis space* contains all hypotheses a learner knows of; the *effective learner search space* contains all hypotheses a learner decides on testing. The domain is described by the *space of true hypotheses*, containing all true hypotheses about a domain of which the learner in general has to find a subset, the *target conceptual model*. The more a learner learns about the domain, the more the *effective learner search space* approaches the *target conceptual model* (see Figure 1). A learner's domain knowledge is depicted as a configuration of hypothesis space.

Little prior knowledge about the domain constrains the initial generation of hypotheses because the number of hypotheses that can be derived from memory is limited. New hypotheses can initially be generated from the outcomes of hypothesis-driven experimentation, but have to be induced from the outcomes of exploratory experiments as

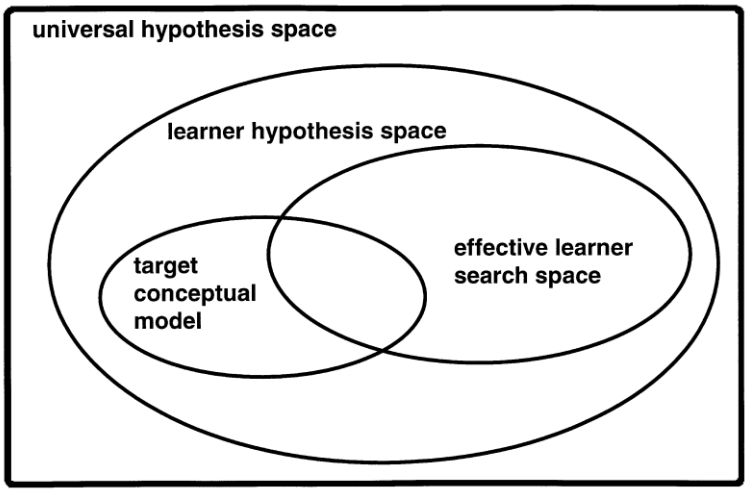

Figure 1. The different regions in the hypothesis space representing the knowledge of the learner (from van Joolingen & de Jong, 1997)

soon as the *effective learner search space* does no longer contain hypotheses the learner wishes to test (van Joolingen & de Jong, 1997). Data-driven exploration is, however, regarded as inefficient (Klahr & Dunbar, 1988; Shute & Glaser, 1990). To a learner with much domain knowledge, generating hypotheses, and thus finding the 'target conceptual model', should be easier because the learner possesses more knowledge from which he or she can infer hypotheses about the domain (Lazonder, Wilhelm, & Hagemans, 2008).

The influence of prior knowledge on the generation of hypotheses during simulation-based inquiry learning has been examined by Lazonder et al. (2008) and by Lazonder, Wilhelm, and Lieburg (2009). In both studies, learners had to discover the influence of four independent variables in either a familiar or an unfamiliar domain, thus creating both low- and high prior knowledge learners. It was found that learners who could use their prior knowledge stated more and more specific hypotheses, conducted fewer exploratory experiments and were more successful learners. Furthermore, when learners could not use their prior knowledge, they started off in a data-driven mode of experimentation; they conducted several exploratory experiments to acquire knowledge about the domain. Subsequently, they switched to a more theory-driven mode, where experimentation was more often guided by a hypothesis. These finding imply that learners successfully integrated acquired domain information into their *effective learner search space*and use it to switch to a more theory-driven mode of experimentation.

How can low prior-knowledge learners be supported in an inquiry-based learning environment? The effectiveness of offering additional domain information has been examined in several studies (for an overview see: de Jong & van Joolingen, 1998). Lazonder, Hagemans and de Jong (2010) have examined the effect of providing domain knowledge *before* and *during* an inquiry task. Their learners had to discover the magnitudes of the effects that multiple variables had on the sales of a simulated shoe store. The underlying rules were chosen arbitrarily to create a situation where learners' prior domain knowledge was low. Three conditions were investigated. One of the groups received domain information *before* the task; one group received it *before* and *during* the task; and the control group received no domain information at all. Results showed that the learners who received the domain information *before* and *during* the task did not only outperform the control group, but also the group who received domain information *before* the task only. The former learners stated more and more specific hypotheses, conducted fewer exploratory experiments and gained more knowledge about the domain. These findings imply that supporting learners with domain information facilitates theory-driven inquiry and thus benefits knowledge gain in an inquiry task.

How can these findings be interpreted in terms of the SDDS model? Offering domain information at the beginning of an inquiry task fills up the empty hypothesis space of low prior-knowledge learners, thus facilitating hypotheses-driven inquiry. Hypothesis-driven inquiry has been found to be more efficient than data-driven inquiry (Klahr & Dunbar, 1988), where exploratory experiments have to be conducted to generate new hypotheses. Furthermore, keeping the domain information available during the task aids in keeping the learner hypothesis space filled, thus preventing the learner from falling back into a data-driven mode of experimentation when hypotheses can no longer be generated from prior knowledge and exploratory experiments have to be conducted instead.

The central aim of this study was to investigate whether the provision of hypotheses can have the same beneficial effects on hypothesis generation and knowledge gain as the provision of domain information, but without giving away too much of the learning content, thus preserving the constructivist nature of scientific inquiry learning. In past research, offering pre-defined hypotheses has been found to have a positive effect on learning outcomes in simulation-based inquiry tasks (Gijlers & de Jong, 2009; Njoo & de Jong, 1993a; Njoo & de Jong, 1993b). Gijlers and de Jong (2009) found that by being provided with pre-defined hypotheses learners saved crucial time and explored hypotheses they would not have thought of themselves, thus exploring a wider range of the domain. Njoo and de Jong (1993a) found that students who received a structured overview of pre-defined hypotheses together with information about the 6 exploratory processes, namely identifying and relating variables and parameters, hypothesis generation, designing an experiment, making a prediction, data interpretation, and stating a conclusion, performed significantly better on a posttest. By comparing the provision of domain information with the provision of pre-defined hypotheses, the present study aimed to find an answer to the question whether the provision of hypotheses can keep up with the provision of domain information in supporting both knowledge gain and theory-driven inquiry.

The SDDS model (Klahr & Dunbar, 1988) and the extended SDDS model (Van Joolingen & de Jong, 1997) were used to make predictions about the differential effectiveness of both support approaches in supporting theory-driven inquiry and increasing knowledge gains. Three conditions were created, two experimental groups and one control group. One of the experimental groups received pre-defined hypotheses with an unknown truth value before and during the task (H), the other experimental group received domain information before and during the task (DI). The control group did not receive any support.

In terms of the Extended SDDS model, both types of support aim at filling the *learner hypothesis space*, either by providing domain information from which hypotheses can be generated or by providing pre-defined hypotheses. This way it is expected that both support options facilitate theory-driven inquiry, and hence increase knowledge gains. It was therefore predicted that learners in both experimental conditions would gain more domain knowledge, generate more and more specific hypotheses, and conduct fewer exploratory experiments than learners in the control condition (Hypothesis 1).

The main difference between the two support conditions is that the *learner domain space* of learners in the DI condition will initially have a greater overlap with the *target conceptual model*, since they receive background information about a domain of which they will not have any prior knowledge. The initial amount of domain knowledge of learners in the H condition will be much smaller because they do not receive any domain information, but hypotheses with an unknown truth value. Because of this greater initial knowledge it was predicted that learners in the DI condition would be cognizant of a greater part of the *target conceptual model* than learners in the H condition.

The two types of support were further predicted to promote hypothesis generation equally well. In the DI condition, learners are provided with a rich knowledge base from which subsequent hypotheses can be generated. In the H condition, the *learner hypothesis space* gets filled with a sufficient number of hypotheses, thus enabling the learner to retreive testable hypotheses from the *hypothesis space* a certain number of times. This is expected to speed up the inquiry process enabling learners to acquire new domain knowledge at a faster pace. From this newly gained knowledge subsequent hypotheses can be generated easily and independently. Therefore, it was expected that learners in the H condition would generate as many, and equally specific hypotheses, and that they would conduct equally few exploratory experiments compared with learners in the DI condition. However, they were expected to discover a smaller part of the *target conceptual model* than learners in the DI condition. (Hypothesis 2).

3. Method

3.1 Participants

The sample consisted of 58 university students (37 males). Seventeen participants majored in a technical subject such as applied physics or mechanical engineering, 34 were social science students, and 7 participants followed a different, non-technical study. The participants' mean age was 23.22 years (*SD* = 3.29). Participants were randomly assigned to either the DI condition (*n* = 18, 14 males), the H condition (*n* = 19, 11 males), or the control condition (*n* = 21, 12 males).

3.2. Materials

3.2.1. Inquiry task and learning environment

Students worked with the computer-based scenario SINUS which is a simulation of the Dynamis approach (Funke, 1992). Dynamis simulations are dynamical systems based on linear structural equation systems. Students have to discover the causal structure of the system by changing the values of the input variables and observing their effects on the outcome variables. The cover story of SINUS tells the learner that there are 6 alien species on the planet SINUS and that he or she has to discover how they are related. The input variables were represented by the species *Olschen*, *Mukern* and *Raskeln*, the output variables by the *Gaseln*, *Schmorken* and *Sisen*. Three types of relations had to be discovered: how the input variables affected the output variables; how the output variables influenced each other (side effects), and finally whether and how the output variables changed by themselves over time. Since the task is abstract, learners do not have any prior knowledge about the relations of the variables with each other. The names of the alien species were newly created words without an existing meaning.

Before the start of the task, students read the task description. SINUS is divided into rounds, and each round is divided into a certain number of trials. In each trial, students had to specify the value of the input variables to see the effect of this manipulation in the following trial. At the end of each round, the values of the output variables are set back to the initial starting point and the learner can start again. During the task, the learner could always re-read the task description.

The SINUS program was adjusted to the requirements of this study. Originally it consisted of 4 main effects with positive values, one side effect, and one case of

Eigendynamic, that is an decrease of ten percent in time of the variable Sisen. In order to distribute the relations more evenly across the variables, the variable Gaseln received the *Eigendynamic*. This way each of the output variables was affected by two other variables. Also, the relation between Raskeln en Schmorken got a negative value. See Figure 2 for a diagrammatic discription of the model underlying SINUS. The number of rounds was set to

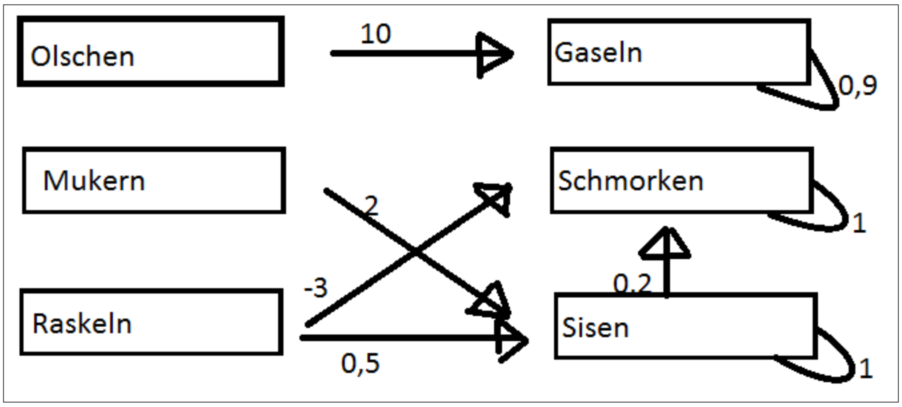

Figure 2. Diagramme of adjusted causal model underlying SINUS. Numbers represent the magnitude of the linear relations.

infinite to make sure that the learner would have sufficient rounds to complete the task.

The SINUS environment also comprised a hypothesis scratchpad in which the student can formulate a hypothesis by specifying the input variable, the dependent variable, the expected relation between these two and the size of effect. The relations that could be chosen were: “has effect”, “has no effect”, “has a positive effect”, “has a negative effect”, “changes by itself”, “decreases by itself”, “increases by itself”, “stays constant” and a “?” in case no hypothesis could be generated. The possibility to state the magniture of effect only appeared when a relation with direction of effect was chosen, such as “has a positive effect”.

Students in the DI condition received additional domain information about each output variable. The information was offered in the form of a booklet. One page was dedicated to one variable and contained each six statements, according to the following structure : (1) its relations to all of the three input variables, (2) its relations to both other output variables (side effects), and (3) whether it changes by itself over time (the variable’s *Eigendynamik*). For example, one statement about the relation between an output variable with one of the three input variables looked like this: “If the population of the Olschen increases, then the population of the Gaseln increases as well”. The magnitudes of the effects (see Figure 2) were left undisclosed and had to be discovered during the task. Before the session started, the learner had sufficient time to read the domain information. During the session, the domain support

laid besides the computer's keyboard, so that the learner could consult the support whenever needed. Two versions of the domain support, that differed concerning the sequence of the statements, were created; their distribution among participants was randomized.

Learners in the H condition received information that equaled the domain information in form and amount, except that 3 out of 6 hypotheses were incorrect. The learners were informed about the fact that hypotheses could either be true or false. The hypotheses were written in the same form as the domain information, for instance "If the population of the Olschen increases, then the population of the Gaseln increases as well". To create false hypotheses, the domain information was taken as base and the direction of effect was either changed from positive to negative, from negative to positive, from non-existent to either a positive or a negative relation and the other way round. For example, the domain information statement "If the population of the Olschen increases, then the population of the Gaseln increases as well" was turned into the false hypothesis "If the population of the Olschen increases, then the population of the Gaseln decreases". Two versions of the hypotheses support that differed concerning the sequence of the statements, were created, whose distribution among participants was randomized.

Participants in the control condition did not receive any support and could start with the task right away. They worked on exactly the same task as learners in the DI and H condition.

3.2.2 Knowledge assessmentt

To assess participants' acquired knowledge of the causal structure underlying SINUS, they were asked to complete a a diagram that resembled the one in Figure 2 except that the relations were missing. Participants were asked to write down all discovered relations, their directions (positive or negative) and the magnitude of their effects. Participants could use the notes they had taken during the task. One point was given for each correct relation found, one additional point if the direction of the effect was correct, and one additional point if the magnitude of the effect was correct. Two raters coded 7 randomly collected diagrams and yielded an interrater agreement of.93 (Cohen's κ).

3.3 Procedure

Students participated in the experiment one at a time. At the beginning of their session the experimenter explained the procedure and the task. After that, students in the control condition were given pen and paper for their notes and a calculator for calculations during the task, and then could begin with the task. Participants in the H and the DI condition received the support information, were told to take their time to read it, and that they would be allowed to use the information during the task. As soon as students had finished reading, the experimenter started the simulation. They were provided with pen and paper for their notes, the support information, and a calculator. From this moment on, the timer in the simulation started. Demographic details about their age, gender, nationality, and study were collected in the simulation. This was important to get an indication of the heterogeneity of the sample. Especially study was important to be able to distinguish between technical and non-technical students. As soon as participants had filled them in, the cover story including the task description was displayed. After this, and before participants started the task, the experimenter explained how the simulation had to be operated. The experimenter used a standard text, and answered participants' questions only when they concerned the operation of the simulation. Participants had 45 minutes to complete the task. During the task, the experimenter observed whether participants in the H and DI condition used the support measures.When 45 minutes were passed, the experimenter interrupted the students, and asked them to fill in the knowledge test. Participants had sufficient time to complete this assessment, requiring in general less than 15 minutes.

3.4. Measures and scoring procedure

The *time* subjects spent on the task was recorded by the experimenter with a stopwatch. The starting point was when the simulation was started. Participants were free to stop before the 45 minutes had been reached, by indicating this to the experimenter.

The number of unique and duplicated *experiments* was logged by the simulation. An experiment was coded as unique if the combination of factor values only appeared once, and it was coded as duplicated when the combination of factor values appeared more than once during the task. Furthermore, each of the unique/duplicated experiments was coded as exploratory when no hypothesis, but a questionmark was filled in in the hypothesis scratchpad before the conduction of the experiment. This way the percentage of exploratory experiments could be assessed.

Participants' hypotheses were logged by the simulation from the hypothesis scratchpad (see Figure 3). The number of hypotheses generated during a session could thus be measured. Hypotheses of different specificity could be filled in, as well as the intention to conduct an exploratory experiment. Hypotheses were classified as either unspecified (predicting that a relation between two variables exists), partly specified (predicting a relation and its direction) or fully specified (predicting a relation, its direction and its magnitude). For each participant the overall mean hypothesis specificity was calculated. In addition, the mean hypothesis specificity per quartile was calculated by dividing the total number of hypotheses into four quartiles and calculating the mean hypothesis specificity for each quartile. This was done in order to detect possible effects of time on the hypothesis specificity.

Participants' score on the knowledge test gave an indication of their *performance succes*. Six relations, including their direction and magnitude of effect, had to be named, and

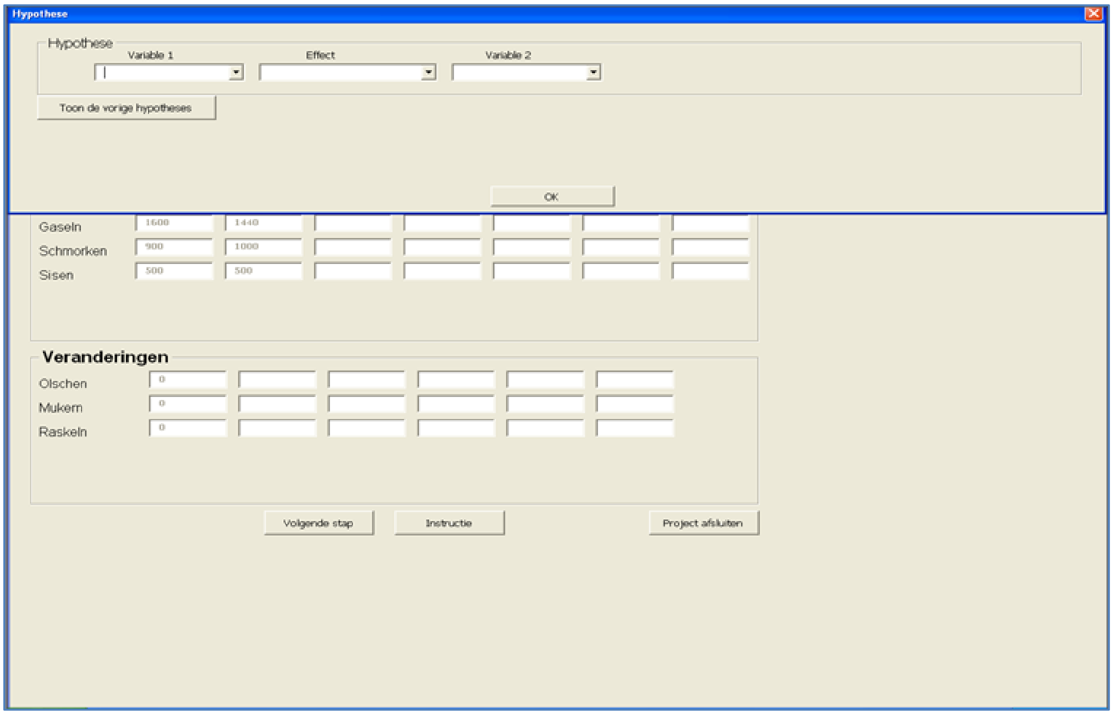

Figure 3. SINUS interface, including hypothesis scratchpad

for each relation a maximum of 3 points could be earned. One point was given for each correct relation found, one additional point if the direction of the effect was correct, and one additional point if the magnitude of the effect was correct.The maximum number of points was 18.

The amount of *support usage* was noted by the experimenter who stayed in the room during the experiment to score how often participants used the support measures. She did that by sitting behind them and noting down everytime a participant looked at the support, which was attached to the table right to the participant. This forced the learner to look right whenever he or she wanted to consult it, enabling the experimenter to recognize when the support was used. Furthermore, at the end of the experimental session, the experimenter asked learners in the H condition the open question: “What did you find of the support?” and noted down their answer.

4. Results

Table 1 summarizes the descriptive statistics for the participants’ performance.

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

	Condition		
	DI (*n* = 18)	H (*n* = 19)	Control (*n* = 21)
Time on task	42.23 (6.42)	41.25 (5.51)	43.00 (6.17)
Number of experiments			
Unique experiments	7.17 (6.54)	7.89 (4.56)	8.52 (5.15)
Duplicated experiments	11.72 (9.93)	18.16 (10.68)	16.33 (9.38)
Exploratory experiments (%)	4.74 (8.97)	11.73 (15.24)	8.67 (15.78)
Number of hypotheses	17.39 (8.78)	22.05 (9.59)	22.10 (9.50)
Mean hypothesis specificity	2.56 (0.40)	2.39 (0.26)	2.35 (0.35)
Hypothesis specificity per quartile			
First quartile (Q1)	2.57 (0.49)	2.49 (0.40)	2.43 (0.35)
Second quartile (Q2)	2.55 (0.48)	2.38 (0.44)	2.29 (0.40)
Third quartile (Q3)	2.52 (0.50)	2.35 (0.44)	2.38 (0.48)
Fourth quartile (Q4)	2.53 (0.43)	2.39 (0.47)	2.42 (0.51)
Support usage	10.00 (6.94)	5.32 (6.20)	-
Performance success	10.67 (3.52)	10.47 (4.53)	9.33 (4.48)

Note. DI = domain information support; H = pre-defined hypotheses support; Control = control condition (no support). Dashes indicate that data was not obtained.

A univariate analysis of variance (ANOVA) was conducted on the data for time on task. This revealed no significant differences between conditions, $F(2, 55) = 0.32$, $p = .73$, indicating that the conditions did not differ in the amount of time needed to complete the task.

A Kolmogorov-Smirnov test indicated that data on number of hypotheses in the H condition was not distributed normally, $D(19) = 0.22$, $p = .02$, with a skewness of 0.82, and a kurtosis of - 0.35. Therefore, a Kruskall-Wallis test was conducted to compare the conditions, $X^2(2, N = 58) = 3.52$, $p = .17$, indicating that participants in all threeconditions did not differ in the number of hypotheses they generated during the task.

The mean specificity of all hypotheses generated during the inquiry task was computed for each condition and compared by using a one-way ANOVA. The results, $F(2, 55) = 1.92$, $p = .16$, did not show any significant differences, expressing that the conditions did not differ in the overal specificity of hypotheses participants generated. However, a *t*-test for two independent samples discovered a significant higher mean hypothesis specificity of the group technical students ($M = 2.66$, $SD = 0.28$) compared to the group non-technical students ($M = 2.34$, $SD = 0.33$) across conditions, $t(56) = 3.52$, $p < .001$, suggesting that technical students generally generated more specific hypotheses than non-technical students. This comparison was drawn to show that a second factor, namely whether students had a technical background or not, was influencing scores on hypothesis specificity.

In addition to the overall analysis of participants' hypotheses, the specificity of hypotheses was also analyzed through time and across conditions. Therefore, the hypotheses of each participant was divided into four quartiles. A mixed-ANOVA was used to examine the effects of time and condition, as well as their interaction. Since the sphericity assumption was violated, $X^2(5) = 16.89$, $p = .01$, and the Greenhouse-Geisser coefficient exceeded .75, the Huyn-Feldt coefficient was used for within subjects analysis. The mixed design ANOVA revealed neither a main effect of time, $F(2.68, 147.18) = 0.76$, $p = .51$, nor a main effect of condition, $F(2, 55) = 1.244$, $p = .30$, nor a significant interaction effect between time and condition, $F(5.35, 147.18) = 0.31$, $p = .92$. This means that the hypothesis specificity did not change over time, and that the specificity of participants' hypotheses within a certain quartile did not depend on the condition they were assigned to.

A multivariate analysis of variance (MANOVA) was used to examine cross-condition differences in the amount of unique and duplicated experiments participants conducted. The

MANOVA revealed no significant differences between conditions, Pillai's trace = 0.10, $F(4, 110) = 1.43$, $p = .23$. This finding implies that participants in all three conditions conducted as many (unique and duplicated) experiments.

The percentage of exploratory experiments was compared across conditions using univariate analysis. Results showed no significant differences between the groups, $F(2, 55) = 1.19$, $p = .31$. However, a Kolmogorov-Smirnov test showed that data on the percentage of exploratory experiments was neither distributed normally in the DI condition, $D(18) = 0.42$, $p = .001$, nor in the H condition, $D(19) = 0.26$, p = .00, nor in the control condition, $D(21) = 0.29$, $p < .001$. Therefore, a Kruskal-Wallis test was performed to compare group means, finding a trend towards a significant difference between groups , $X^2(2, N = 58) = 5.41$, $p = .07$. Further analyses indicated that learners in the H condition conducted a higher percentage of exploratory experiments than learners in the DI condition.

The frequency of support usage was recorded during the experiment by the experimenter, using pen and paper to note down everytime the support was inspected. A two-samples independent *t*-test found a significant difference between the two support groups, $t(35) = 2.17$, $p = .04$, indicating that the domain information was more frequently used than the hypotheses. Fifteen out of 19 learners in the H condition were questioned about what they found of the predefined hypotheses they could consult. Five out of them stated that they found the hypotheses useless or not sufficiently helpful. The reasons for that varied. One participant stated that he knew he would have to conduct his own tests. Another said that the hypotheses were insufficient to solve the problem. A third participant stated that he could generate hypotheses on his own, and that they would distract him from the model of the system he already had in his mind. Four out of 8 participants who regarded them as useful, stated that they felt guided by the hypotheses concerning their own experimentation, and 3 indicated to find them useful because of the overview they supplied of the relations in the system.

Data on performance succes was not distributed normally in the control group, $D(21) = 0.20$, $p = .03$, with a skewness of 0.51, and a kurtosis of -0.68. Therefore, a Kruskall-Wallis test was conducted to compare the group means, $X^2(2, N = 58) = 1.46$, $p = .83$, finding no significant differences between groups, implying that groups did not differ in the scores they obtained on the knowledge test. Interestingly, technical students ($M = 12.94$, $SD = 3.77$) performed significantly better than non-technical students ($M = 8.95$, $SD = 3.81$) on the knowledge test , as was indicated by a univariate analysis of variance, $F(1, 56) = 13.25$, $p < .001$. In order to examine possible interaction effects between condition and whether learners did a technical study or not, a univariate analysis of variance was conducted, not finding any signficant interactions, $F(1, 52) = 2.22$, $p = .12$, suggesting that performance differences between technical and non-technical students did not depend on the condition they were assigned to.

Furthermore, performance succes per relation was computed to give an indication of level of difficulty of each relation seperately, as summarized in Table 2. A one-way repeated measures ANOVA found a significant within-subjects effect of relation, $F(5, 285) = 16.67$, $p < .001$. The lowest mean score was found for the side effect between Sisen and Schmorken (relation 5), and Bonferroni-corrected pairwise comparisons indicated that the mean score on relation 5 differed significantly from the mean score of relation 1, $p < .001$, relation 2, $p < .001$, relation 4, $p < .00$ and relation 6, $p = .01$, but did not differ significantly from relation 3, $p = .13$.

Table 2
Means (and *SD*) for posttest scores for each of the 6 relations in SINUS

Olschen – Gaseln (R 1)	Mukern – Sisen (R 2)	Raskeln – Schmorken (R 3)	Raskeln – Sisen (R 4)	Sisen – Schmorken (R 5)	Gaseln ‚Eigendynamik' (R 6)
1.81(1.03)	2.31(1.17)	1.38(1.14)	2.21(1.20)	0.88 (1.22)	1.53 (1.14)

Note. In order to get an indication of the differential difficulty of the relations in SINUS, performance scores were computed per relation, across all conditions. Maximum score per relation was 3 points. R = relation.

5. Conclusions and Discussion

The aim of the present study was to compare the effectiveness of two types of support within simulation-based inquiry learning: offering domain information and offering pre-defined hypotheses with an unknown truth value. It was expected that learners in both support conditionswould gain more domain knowledge, generate more and more specific hypotheses, and conduct fewer exploratory experiments than learners in the control condition (Hypothesis 1). Furthermore, it was predicted that learners who received pre-defined hypotheses would generate as many and equally specific hypotheses, conduct equally few exploratory experiments, but discover a smaller part of the domain than learners in the DI condition. (Hypothesis 2).

Hypothesis 1 was not supported by the results. Students in the control condition did apparently not gain less knowledge, did not state fewer and less specific hypotheses and did not generate less explorative experiments than learners who either got domain information or hypotheses as support. Hypothesis 2 was partly supported by the results: learners in both support conditions generated as many and equally specific hypotheses. Furthermore, a trend in the data points at the possibility that learners in the H condition may have generated a higher percentage of exploratory experiments. Also contradictory to expectations, learners in the H condition apparently gained as much domain knowledge as learners in the DI condition.

The fact that Hypothesis 1 was not supported by the results is surprising since related research found that learners benefit from additional domain knowledge in a simulation-based problem solving task (Lazonder et al., 2010). There are several differences between the present study and the study by Lazonder et al. (2010) that might explain why the present study did not find benefits of providing domain information during the task. One possible explanation concerns the heterogeneity of the sample. Individual differences in level of systematicity, motivation, strategy use, and general intelligence that are said to contribute to large standard deviations in complex problem solving research (Funke, 2005), may have played a role in the present study, as well. The constitution of the sample - about one third were technical students - probably further increased the level of within-sample variance as compared to the sample in the study by Lazonder et al. (2010) where the sample consisted of first year social science students. Not surprisingly, it was found that technical students outperformed non-technical students on knowledge gain (as indicated by higher knowledge test scores) and that the hypotheses they stated were more specific. The reason for this may be that technical studies are much more mathematically oriented and are designed to teach the usage of proper strategies when faced with complex problem solving (CPS) tasks, as compared to social studies where this is much less the case. Examples of particularly useful strategies within the Dynamis situation are manipulating not more than two variables at a time, selecting large input variables to detect side effects, and leaving all values at zero when trying to detect Eigendynamics (Blech & Funke, 2005). Perhaps technical students made better use of these strategies.

The fact that 4 of the 5 high-achievers in the control condition (with posttest scores of 15 points or more) were technical students may partly explain why the control condition performed as proficiently as both support conditions. In the DI condition, in contrast, only one out of 3 high-achievers had a technical background implying that being a technical student or not was not the only factor that contributed to achieving a high score in the DI condition, but that the support may have had a positive influence on knowledge gain, as well. This was less the case in the H condition where three out of 4 high-achievers were technical students. Although there was no significant interaction effect between treatment and whether or not students followed a technical study, these findings point at a possibile aptitude-treatment interaction (ATI). Aptitude treatment interactions occur when learners with different levels of aptitude are exposed to different kinds of treatments. Relevant aptitudes can be “knowledge, skills, learning styles, [and] personality characteristics” (Kalyuga, 2007, p.527). So, perhaps technical students, because of their higher ability in solving mathematical problem solving tasks, but also because of their prior training in the systematic solving of complex problems, performed better when they were not given any support.

Another interesting finding is that learners in the DI condition did not perform as well as they could have, considering the amount of information they had been given. The relations and their direction had been disclosed to them, enabling them to get at least 12 points on the knowledge test. However, with a mean score of 10.67, the group clearly did not make maximum usage of the domain information they had been supplied with. Further analyses revealed that 11 out of 18 participants in the DI condition obtained a posttest score of less than 12 points. What these learners had in common is that they, out of 6 relations they could have written down on the knowledge test, did not write down one or more entire relations. Many of them did not write down the side effect between Sisen en Schmorken on the knowledge test, although its existence and direction of effect had been disclosed to them by the domain information. When examining the notes they had taken, it was found that they had not made any notes about these effects during the task. Taking proper notes can be counted as a regulative learning proces (de Jong & Njoo, 1992) and is an important task when performing an inquiry learning task. This indicates that poor regulative abilities contributed to the low scores of learners in the DI condition.

The difficulty of the task and the associated intrinsic cognitive load of the CPS task may also partly explain why the present study did not find the assumed superiority of the support. The intrinsic cognitive load of a task is concerned with “the natural complexity of information that must be understood and material that must be learned” (Sweller, 2010, p.124). When comparing the current task to the task in Lazonder et al.’s (2010) study, in which the effects of 5 variables on 1 outcome variable had to be discovered, one could argue that the present task with 3 dependent variables and 3 independent variables, one side effect and one case of *Eigendynamic* was more difficult. Together with a rather short period of time (45 minutes), the task may have burdoned learners with a cognitive load that was too high.

Also, the fact that the mean score on the knowledge test in Lazonder et al. (2010) in the domain information condition was 13.21 points out of 15 points, as compared to a mean score of 10.67 out of 18 points in the DI condition in the present study, suggests that the present task was more difficult. That the task was rather difficult in general is further indicated by the fact that there was one relation in SINUS which was not discovered by 38 of the 58 participants and on which participants achieved significantly lower scores compared to 4 of the other 5 relations: the side effect between the two output variables Sisen en Schmorken. The side effect was characterized by its small magnitude and the fact that it could not be manipulated directly. Prior research on the effects of side effects in Dynamis situations has indeed found that they decrease knowledge acquisition in dynamical systems (Blech & Funke, 2005).

The assumed effectiveness of offering pre-defined hypotheses was notsupported by the results either. The H condition did not outperform the control condition. Interestingly, the hypothesis support was used less often than the domain information support, as became apparent from the notes the experimenter had taken during the experimental sessions about the frequency of usage. Although this registration method is of moderate reliability, it gives an indication that the hypotheses were found less useful than the domain information support. When asked, several learners stated that they did not find the hypotheses useful and therefore hardly used them. One possible explanation might be that a part of these learners were technical students, for whom the support was possibly not necessary because of their trained skills in performing complex problem solving tasks (aptitude treatment interaction). One technical student stated that he, at the beginning of the task, had created an underlying model of the system in his mind, with which the hypotheses interfered because they created doubt about whether his model was correct. He therefore quit using the hypotheses.

Another reason may be that the learners did not believe that the hypotheses would support them, or that they were too focused on the task itself. Possibly, the character of the task was too mathematical and therefore the usefulness of additonal hypotheses was not apparent to many learners. One learner, for example, stated that she was mainly focused on discovering the magnitudes of the relations and therefore hardly consulted the hypotheses. Another learner stated that the hypotheses, because they had an unknown truth value, did not help him in completing the task. And yet another learner said that he found them useless since he had to find out the relations on his own after all.

A third reason may have been that learners had to state their hypotheses during the task in the hypothesis scratchpad. Although it was not meant to be a support measure, the hypothesis scratchpad may also have supported the learners in generating hypotheses because it provided an overview of all possible variables and relations from which hypotheses could be generated, and thus may have competed with the actual support measure, the list of pre-defined hypotheses. Research on the influence of hypothesis scratchpads on the generation of hypotheses has found that hypothesis scratchpads indeed benefit theory-driven experimentation (van Joolingen & de Jong, 1991). So, maybe because of that, the list of pre-defined hypotheses became a redundant support measure. This was not the case with the domain information support -and this was supported by the higher frequency of usage– because the domain informaton support provided background information which was qualitatively different to the information that the hypothesis scratchpad provided.

Hypothesis 2 was partly supported by the results. It was expected that learners in both support conditions would generate equally specific hypotheses, and this was supported by the results. However, what must be taken into account is that, across conditions, technical students significantly stated more specific hypotheses than non-technical students. More interestingly, all of the four students with the highest mean hypothesis specificity (2.6 points or higher) in the H condition were technical students, and none of them was a non-technical student. In the DI condition, by contrast, half of the students with the highest mean hypothesis specificity were technical students, the other half were non-technical students. These findings raise the issue of whether the hypothesis support was equally effective in increasing hypothesis specificity as the domain information support. The trend that was found towards a higher percentage of exploratory experiments in the H condition as compared to the DI conditon further casts doubts as to whether pre-defined hypotheses can support theory-driven inquiry as well as domain information. However, these findings may also suggest that the hypotheses support interacted with the hypothesis scratchpad. Possibly, having to state a hypothesis they had just received from the support in the hypothesis scratchpad was deemed by learners as an unnecessary, possibly time-consuming, additional step, and therefore learners have chosen to not state this hypothesis in the scratchpad, but to skip this step. Therefore, these findings may also suggest that the hypotheses support interacted negatively with the method of measurement of participants' hypotheses: the hypothesis scratchpad.

Hypothesis 2 also predicted that learners in the H condition would gain less domain knowledge than learners in the DI condition. This was not supported by the results, indicating that both supports did not differ in the influence they had on knowledge gain. Factors adding to inner group variance, as discussed above, may have hindered the detection of significant differences here as well. Interestingly, the distribution of high-achievers (15 points or more on the knowledge test) in the H condition was distributed in favor of technical students: 3 out of 4 high-achievers were technical students, as compared to the DI group in which only 1 out of 3 high-achievers was a technical student. Keeping in mind that technical students outperformed non-technical students in knowledge gain, this finding raises some doubts as to whether both support approaches did not differ in influencing knowledge gain.

Future research about the differential effectiveness of pre-defined hypotheses and domain information should consider the following improvements. The large standard deviations found in the present research made it more difficult to find between-group differences. Therefore, in future research the groups should be more homogenous. This may be accomplished by only considering social science students, or by examining the two populations, technical and non-technical students, seperately in two large groups. Snow and Lohman recommend to always rule out possible interaction effects between aptitude and treatment when comparing different instructive methods, in order to find significant main effects of the treatment (Snow & Lohman, 1984). Testing both groups of learners would additionally bring some insights into how learners with different aptitudes react to support measures in CPS tasks. Furthermore, individual differences in problem solving seem to play a role in explaining large standard deviations in CPS research (Blech & Funke, 2005). So, a bigger sample may also benefit statistical analyses.

Because of its mathematical character, the task may not have elicited the usage of hypotheses as support measure. In order to successfully compare the provision of domain information with the provision of hypotheses in simulation-based inquiry tasks, one may want to consider a less mathematical domain. This may be an abstract domain with less mathematical relations to be discovered. This would have as advantage that prior domain knowledge will be controlled for. Alternatively, a more realistic domain may be chosen in order to increase external validity.

A last recommendations concerns the measurement method chosen in the present study to examine participants' hypotheses. It should rather not be done with a hypothesis scratchpad because of its possibly interfering effect with the hypotheses support. Instead, probing questions could be chosen as in Lazonder et al. (2010), asking learners what they want to investigate and what they expect as an outcome.

References

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1-18.

Blech, C., Funke, J. (2005). Dynamis review: An overview about applications of the Dynamis approach in cognitive psychology. Unpublished Manuscript, University of Heidelberg, Germany

de Jong, T. (2006). Scaffolds for Scientific Discovery Learning. In J. Elen & R. E. Clark (Eds.), *Handling Complexity in Learning Environments: Theory and Research* (1 ed., pp. 107-128). Amsterdam Elsevier.

de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*(2), 179-201. doi: 10.2307/1170753

de Jong, T., & Njoo, M. (1992). Learning and instruction with computer simulations: learning processes involved. In E. de Corte, M. Linn, H. Mandl, & L.Verschaffel (Eds), *Computer-based learning environments and problem solving*. Berlin: Springer

Funke, J. (1992). Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung. Berlin: Springer Verlag

Gijlers, H., & de Jong, T. (2009). Sharing and confronting propositions in collaborative inquiry learning. *Cognition and Instruction, 27*(3), 239-268. doi: 10.1080/07370000903014352

Hmelo, C. E., Nagarajan, A., & Day, R. S. (2000). Effects of high and low prior knowledge on construction of a joint problem space. *Journal of Experimental Education, 69*(1), 36-56. doi: 10.1080/00220970009600648

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*(4), 509-539. doi: 10.1007/s10648-007-9054-3

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*(3), 579-588. doi: 10.1037/0022-0663.93.3.579

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

Lazonder, A. W., Hagemans, M. G., & de Jong, T. (2010). Offering and discovering domain information in simulation-based inquiry learning. *Learning and Instruction, 20*(6), 511-520. doi: 10.1016/j.learninstruc.2009.08.001

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*(6), 580-592. doi: 10.1016/j.learninstruc.2007.12.001

Lazonder, A. W., Wilhelm, P., & van Lieburg, E. (2009). Unraveling the influence of domain knowledge during simulation-based inquiry learning. *Instructional Science, 37*(5), 437-451. doi: 10.1007/s11251-008-9055-8

Mayer, R. E. (2004). Should There Be a Three-Strikes Rule Against Pure Discovery Learning? *American Psychologist, 59*(1), 14-19. doi: 10.1037/0003-066x.59.1.14

National Research Council (1996). *National science education standards.* Washington, DC: National Academy Press.
Njoo, M., & de Jong, T. (1993a). Supporting Exploratory Learning by Offering Structured Overviews of Hypotheses. In D. M. Towne, T. De Jong & H. Spada (Eds.), *Simulation-Based Experiential Learning* (pp. 207-223). Berlin New York: Springer-Verlag.
Njoo, M., & de Jong, T. (1993b). Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. *Journal of Research in Science Teaching*, *30*(8), 821-844
Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *Journal of the Learning Sciences*, *1*(2), 201-238. doi: 10.1207/s15327809jls0102_3
Shute, V.L. , Glaser, R. (1990). A large-scale evaluation of an Intelligent discovery world: Smithtown. *Interactive Learning Environments, 1*, 51-77
Snow, R. E., & Lohman, D. F. (1984), Toward a theory of Cognitive Aptitude for Learning From Instruction. *Journal of Educational Psychology*, *76*(3), 347-376
Swaak, J., de Jong, T., & van Joolingen, W. R. (2004). The effects of discovery learning and expository instruction on the acquisition of definitional and intuitive knowledge. *Journal of Computer Assisted Learning*, 20, 225-234. Doi: 10.1111/j.1365-2729.2004.00092.x
Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review*,22, 123-138. Doi: 10.1007/s10648-010-9128-5
Van Joolingen, W. R., de Jong, T. (1991). Supporting hypothesis generation by learners exploring an interactive computer simulation. *Instructional Science*, 20, 389-404
van Joolingen, W. R. (1999). Cognitive tools for discovery learning. *International Journal of Artificial Intelligence in Education*, 10, 385-397.
van Joolingen, W. R., & de Jong, T. (1997). An extended dual search space model of scientific discovery learning. *Instructional Science*, *25*(5), 307-346. doi: 10.1023/a:1002993406499
van Joolingen, W. R., de Jong, T., & Dimitrakopoulou, A. (2007). Issues in computer supported inquiry learning in science. *Journal of Computer Assisted Learning*, *23*(2), 111-119. doi: 10.1111/j.1365-2729.2006.00216.x