Direct Instruction vs. Task Structuring

Are they equally effective to teach the control of variables strategy?

Lucian Walther

University of Twente, Department of Instructional Technology

Enschede, Netherlands

2012
Abstract

Teaching children effectively to design meaningful experiments and to draw conclusions from the outcomes is essential in early science education. Prompted by suggestions from Kuhn and Dean (2005) and Lazonder and Kamp (2011), the present study investigated whether structuring a task in a learning environment is as effective as direct instruction to acquire an effective experimentation strategy. Fifty-two 5th graders engaged in a simulation-based inquiry task. One group ($n = 19$) received a direct instruction about why and how to apply the control of variables strategy and were provided with one broad research question about all the variables in the simulation. Another group ($n = 17$) did not receive any information about how to set up experiments but their research question was divided into several successive simpler questions, each addressing only one variable. The third group ($n = 16$) worked on the broad research question but did not receive any instruction about how to approach the experimentation. The results did not show a significant benefit of neither the direct instruction nor the structured task in comparison to the control condition. Based on this result, several issues are discussed on how to improve direct instruction and task structuring in future research.

*Keywords:* scientific reasoning, inquiry learning, direct instruction, task structuring, control of variables, computer simulations.
Introduction

During school time, every child will encounter science education. Not only domain knowledge about, for example biology or physics, but also science skills have to be learned. Instead of simply identifying that a phenomenon occurred, scientists try to explain phenomena by determining the conditions, causes and consequences of the observed event. Kuhn (2002) described scientific activity as involving three stages: inquiry, analysis, and inference. During the inquiry stage, investigators formulate precise questions that can be tested experimentally. The analysis stage involves designing and running informative experiments. Finally, in the inference stage, investigators modify their theories on the basis of evidence from their experiments.

While investigating, it is essential for children to recognize that to conduct a real test of the effect of one variable all other variables must be held constant, so that the effects of these other variables do not influence the outcome (Kuhn, Black, Keselman & Kaplan, 2000). This strategy is called control of variables strategy (CVS). Tschirgi (1980) found that when children get a negative outcome from an experiment and they want to find out the reason why, they often use CVS effectively. In case of a positive result they mainly use the technique of just holding the variable of interest constant. It seems that children usually use the CVS to falsify unconvincing hypotheses. These findings suggest that even young children have some knowledge of how an experiment should be conducted. They mainly need support in how to apply their existing knowledge in their investigations (Lazonder & Kamp, 2011).

There are different methods to assist students in applying the CVS during inquiry tasks. Although a vast amount of research is done on this topic, the findings are not always consistent. The next section will discuss the pros and cons of direct instruction and inquiry learning in learning to apply the CVS.

Klahr (2005) suggests that, for a domain general science process skill such as CVS, it may be more efficient to instruct the students explicitly and directly so that the students can engage more meaningfully in experimentation and discovery across domains. In another study (Klahr, 2009) he states that in teaching the CVS, explicit instruction or at least directive guidance is optimal. He suggests that it might be more efficient to instruct learners directly on how to implement CVS and then to give them enough opportunities to practice. According to information processing theories such as cognitive load theory, another advantage of direct instruction is that, contrary to discovery conditions, direct instruction is less likely to overload working-memory capacity (Sweller, 1988).
While some would argue that with direct instruction the students role would shift from an active discovering investigator to a rather passive participator in the learning process, Tweed (2004) found that in the context of teaching and learning CVS, direct instruction does not necessarily induce passivity. It surely depends on the design of the direct instruction, the way it is given and how long and detailed it is.

Klahr and Nigam (2004) have presented evidence that brief direct instruction effectively teaches the CVS, and on this basis they advocate such instruction as the most efficient, desirable method of developing inquiry skills. According to Fletcher (2009), more direct forms of instruction work better when learners have little prior knowledge, which applies to primary school students. Alfieri, Brooks, Aldrich, and Tenenbaum (2011) made similar findings and stated: “Adolescents were found to benefit significantly more from explicit instruction than did adults.” (p. 11). Chen and Klahr (1999) compared explicit training combined with probe questions with providing probes without direct instruction. They found that if children receive a direct instruction about the theory behind controlling variables in science experiments, and additionally examples of unconfounded comparisons, this resulted in improved use of CVS and transfer of it to other domains, relative to students who did not receive the direct instruction.

It is not clearly identified how children continue to use the CVS after the instructional context is withdrawn. Dean and Kuhn (2007) found that the immediate effects of direct instruction fade away after three months of practice. However, Lorch et al. (2010) showed sustained benefits of direct instruction even after three years and without any efforts to consolidate and maintain acquired skills. Case (1974) stated that a problem of the direct instruction method is that it typically fails to develop transfer outside of the domain of instruction. Chen and Klahr (1999) showed that only if the main task and the transfer tasks are set in well-structured environments with clearly defined research hypotheses provided for the subjects, transfer is achieved with the direct instruction method.

On the other hand, contrary to the ideas of direct instruction, constructivist literature suggests that knowledge constructed by the learners themselves is more robust than that which is simply “told” to the learner, especially in challenging and complex real-world tasks (Elkind, 2001; Piaget, 1970). Slamecka and Graf (1978) stated that memory is enhanced when learning material is generated by the learner in some way, which is commonly referred to as the generation effect. Keselman (2003) suggests that as long as sufficient scaffolding is provided to prevent the emergence of misconceptions, inquiry learning is likely to be beneficial even for students whose experimentation skills are less than perfect. A meta-
analysis by Alfieri et al. (2011) confirms this idea by showing that unassisted discovery does not benefit the learners, whereas feedback, worked examples, scaffolding and elicited explanations do. The need for such scaffolding has been widely recognized, given studies in which ‘free’ exploration, that is offering computer simulations without any support, has been shown not to benefit learners (Klahr & Nigam, 2004; Mayer, 2004), whereas supported discovery learning with simulations has been shown to be an effective mode of learning (Ketelhut, 2006). Alfieri et al. (2011) also found that the construction of explanations or participation in guided discovery is better for learners than being provided with an explanation or explicitly taught how to succeed on a task. These ideas are already known for a longtime from the theory of the zone of proximal development by Vygotsky. As McNeill, Lizotte, Krajcik, and Ronald (2006) explain:

“The zone of proximal development (ZPD) defines the area between a child’s independent problem solving capabilities and the level of potential problem solving capabilities with the guidance of people or tools (Vygotsky, 1978). Stone (1993) argues that scaffolds allow students to achieve a higher level of understanding within their zone of proximal development. In order for a scaffold to promote student understanding, it needs to reside within a students’ current ZPD. If a scaffold provides too much information, the student will not be challenged to learn more. The scaffold should provide just enough information that the learner may make progress on his/her own (Hogan & Pressley, 1997).” (p. 12)

Taken the pros and cons of direct instruction and guided inquiry learning, the question arises which one works better to support learning CVS and if a combination of both methods is more effective than applying only one of them. For example, Klahr (2009) suggests that direct instruction as a basis for subsequent discovery can give structure to inquiry-learning tasks and therefore safe cognitive capacities of the learner. It facilitates constructivist learning by reducing task ambiguities and learning time while improving process comprehension and potential generalization.

But there are other ways to save cognitive capacities and to provide the circumstances which are beneficial for inquiry learning. Instead of using direct instruction, teachers can make important aspects of the inquiry learning environment visible through questions that scaffold children’s learning by modeling, coaching, and eventually fading some of their
support (Hmelo-Silver, Duncan & Chinn, 2007). Scaffolding can also be used to guide instruction and decrease cognitive load by structuring a task and let the students focus on relevant aspects of the task (Hmelo-Silver, 2006). In accordance with these findings, Lazonder (2011) suggests an alternative way to support students’ learning with inquiry tasks:

“Studies in the field of software training demonstrated the educational advantages of a restricted user interface that makes the more advanced commands available only after the basic ones are mastered (Bannert, 2000; Carroll & Carrithers, 1984). Likewise, studies on model progression revealed that gradually introducing learners to increasingly more sophisticated or comprehensive subject matter enhances their inquiry performance (Mulder, Lazonder, & De Jong, 2011). Together these studies substantiate the present conclusion that more clear-cut task goals lead to more effective learning activities and outcomes.” (p. 8)

The idea is to reduce the complexity of the task and the learning environment and focus the attention of the children on the relevant task features step by step. Lazonder and Kamp (2011) analyzed whether such task structuring can be used to support inquiry learning. They demonstrated that when children investigate a series of narrow inquiry questions, this facilitates the CVS and promotes inference performance and conceptual understanding in comparison to the children in the control condition who received one broad inquiry question. This shows that splitting up an inquiry task into simple, single-focused research questions can help children to apply the CVS in setting up systematic comparisons. The important issue of transfer of the learned CVS skills to other domains has not been addressed in their study. Kuhn and Dean (2005) showed that children who receive a hint at the start of a lesson have more goal-oriented experimentation plans and make more valid inferences than children in a control condition who engage in self-directed practice without hints. For the transfer of the CVS to a similar domain they found a nearly significant difference between the two conditions in favor of the condition that received the guiding hints.

The findings about structuring the inquiry task in the way described by Lazonder and Kamp (2011) in order to prompt children to select one variable as the focus of investigation lead to the assumption that this method might be as effective to teach CVS as direct instruction in combination with unguided inquiry. If this would be the case and the task structuring method would be optimized the direct instruction phase might become redundant.
The students could construct the ideas themselves, benefit from the generation effect and by this learning time and teaching resources could be saved.

The purpose of this study was to investigate whether task structuring is equally effective as direct instruction to promote systematic experimentation in an inquiry-based learning environment and if both methods are more effective than unguided discovery. Research activities of children who engaged in a simulation-based inquiry task were examined. The study used a between-subject design in which the type of instructional support was manipulated across three conditions. Children in the direct instruction condition received direct instruction about why and how to use the CVS, directly before they started working with the simulation. Children in this condition investigated one broad research question about all four simulation variables. As the children in the narrow question condition in the study by Lazonder (2011), Children in the structured task condition were given four successive single-faceted research questions, one about each variable in the simulation. They did not receive any kind of instruction about CVS. Children in the control condition investigated the same single broad research question about all four simulation variables as children in the direct instruction condition.

The first hypothesis was that children in the structured task condition would apply the CVS equally effectively as children in the direct instruction condition, and more effectively than children in the control condition. This hypothesis was based on the ideas of Lazonder and Kamp (2011) and Kuhn and Dean (2005) that the formulation of question(s) organizes and gives meaning to inquiry activities, which might lead the children in the structured task condition to learn the CVS intuitively. They might get a sense of which variables are relevant for their investigations and which are not. It would prevent them from varying other variables across two-instance comparisons.

The second hypothesis was that, because of equally effective application of CVS, no difference would be found between the direct instruction and structured task condition concerning valid inferences from their inquiry and valid final beliefs. Furthermore it was expected that, because of the first hypothesis, both experimental conditions would make more valid inferences and have more valid final beliefs than the control condition.
Method

Participants

The sample consisted of 52 German fifth graders (28 boys, 24 girls) with a mean age of 10.66 (SD = 0.56). The participants were randomly assigned to three conditions: the direct instruction condition (n = 19), the structured task condition (n = 17) or the control condition (n = 16). Two children refused to participate before they were assigned to the conditions.

Materials

Learning environment. During the experiment the children had to interact with a computer simulation of a gong. Their task was to find out whether and how four variables (with either two or three possible values) influenced reverberation time of the gong. These variables were sound pitch (low, medium, or high), distance of the listening person to the gong (near, medium, or far), direction the listening person is facing (right or left), and the furnishings of the room (empty or full). Pitch had a positive effect on reverberation time: the higher the tone the longer it took for the sound to decay. The distance variable had a non-linear effect: a short distance (near) increased the time one could hear the gong, the values medium and far had the same effect. When the room was empty the reverberation time was longer while a furnished room shortened the reverberation time. The direction the listener is facing had no effect.

The children’s actual task was to design and conduct experiments by choosing a value for each of the variables and running the simulation by pressing the start button. Doing so enabled them to test the effects of the different variables. Depending on the chosen values for the variables, the gong in the simulation oscillated for a shorter or longer time and a matching sound of a gong was played. A stopwatch in the simulation measured the precise time of reverberation and showed it on the screen. Every action of the children in the simulations was recorded in a log file.

Two different versions of the gong-simulation were used: one presented a single broad research question that addressed all variables in the simulation at the same time; the other presented four narrow questions, each addressing only one of the variables. The questions were shown above the simulation during the children’s investigations. The direct instruction condition and the control condition received the broad research question (“What is the influence of sound pitch, distance, furnishing, and facing direction on the time you can hear the gong echoing?”) while the structured task condition received the four narrow questions (e.g., “What is the influence of sound pitch on the time you can hear the gong echoing?”).
These questions were presented in successive order and the children had to click on a “continue” button when finished working on one question to see the next one.

Direct instruction. The researcher gave a short but precise explanation to the direct instruction condition about why and how to use CVS. To get their attention and prepare them for active learning, the children were informed that they would have to work like real scientists. The instruction was designed to promote the acquisition and transfer of recurrent constituent skills. This means that in order to avoid cognitive overload, the given information was restricted to what was needed to learn the skill: what is CVS, why it is important and how should it be used. Furthermore, the information was presented just in time, right before the children started their inquiry with the simulation, to make sure that as much cognitive resources as possible were available to understand and learn the CVS. Furthermore, the general applicability of the CVS was explicitly mentioned in order to help the children in generating a cognitive schema of the CVS, independent of the context in which it was learned. To support understanding and show the general applicability, a simple example problem situation (a pen rolling down a ramp) was used to demonstrate the application of CVS. The different variables that might influence the speed of the pen (length, thickness, shape, color and material of the pen and angle and material of the ramp) were explained and examples were shown of good and bad experiments.

Beliefs questionnaire. A four-item questionnaire was used to assess children’s beliefs about reverberation before and after their inquiry activities. Pre- and post-test contained identical items addressing every variable in the simulation. These were semi-open questions that asked the children to indicate whether they thought a particular variable would influence the reverberation time. If they chose “yes”, they were asked to explain what they thought this influence would be. Semi-open questions were used because the children had no prior knowledge of this topic (it had not been taught in class), but might have intuitive ideas about the variables they would encounter during their work with the simulation. The children’s answers were checked against the simulation’s underlying model to determine whether they were true or false.

Answer sheet. An unstructured answer sheet was handed out to the children with the instruction to take notes during their experimentation and write down the conclusions they drew from their investigations.
Procedure

The study took place during regular lessons and was conducted by the researcher with help of the children's teachers. First, they had to sign an informed consent form. Then each child randomly received a paper indicating their condition (A, B or C) and subject number and told to keep it. Next they had to answer the beliefs questionnaire (pre-test). The experimenter exemplified the concept of reverberation and read aloud the questions, one at a time. Next, the author took the children from the direct instruction condition to the computer room and gave the direct instruction. Children from the structured task condition and the control condition remained in the classroom and read a short text about the history of the gong under surveillance of their teacher which took about 10 minutes. This text was used to compensate for the time that was needed for the direct instruction and did not contain any helpful information for the inquiry task. When they finished reading, their teacher brought them to the computer room and they had to wait outside. When the direct instruction was finished, the children from the structured task and control condition entered the computer room. The children of all three conditions gathered at one computer and the experimenter showed the learning environment and demonstrated how to set up and conduct an experiment. An example of how an inference from the conducted experiment can be drawn and formulated was given orally. The children were informed about the two versions of the learning environment to make sure they understand that they were either work on a single or on multiple questions. They were instructed to start the program on the desktop that was named after their experimental condition (A, B or C) and follow the directions that would appear on screen. The unstructured answer sheet was handed out to each child with the instruction to take notes during their experimentation and write down the conclusions they drew from their investigations. Then the children logged in to the learning environment and started their investigations. They conducted individual experiments with the simulation for 25 minutes. The experimenter kept the children focused on their task, and reminded them to work on the research question(s) displayed in the simulation interface. Children received assistance on computer technical issues only. Immediately after their investigations, the children once again filled out the beliefs questionnaire (post-test) to assess whether their gained new insights on the topic of reverberation.

Scoring and Data analysis

Children’s beliefs were assessed from their answers to the two questionnaires and their inferences from their notes and conclusions from the answer sheet. All other data was
obtained from the simulation’s log files. A trial was called one execution of an experiment, thus choosing values for the different simulation variables and pressing the start button. The mean number of value changes of the different variables from one trial to the next was used as a measure of systematic experimentation. The closer this number was to 1, the better was the use of CVS. To check if the children’s beliefs and inferences matched the simulation’s underlying model, the number of valid beliefs was assessed from the answers to the questionnaires that were administered before and after their investigations. To score their answers, one point was awarded for each correctly drawn inference (the variable has an effect or not) and an additional point was given if the direction of the effect was correct. The children’s conclusions from the answer sheet which they filled in during their investigations were scored in the same way. The number of correct conclusions was called valid inferences. Furthermore, the number of variables and relations between variables they mentioned in their conclusions were counted.

Every answer on open questions was rated by two raters. The Cohen’s kappa inter-rater reliability for the pre-test was 0.87, for the post test 0.92 and for the inferences from the answer sheet it was 0.84. The quantitative data were analyzed by PASW Statistics 18 (Predictive Analysis Software). To test the research hypotheses univariate analysis of variance (ANOVA), post-hoc analysis with Bonferroni correction and partial correlations were used. The alpha level used as criterion was $p < .05$.

**Results**

The three conditions did not differ significantly in their prior domain knowledge that had been assessed by the questionnaire about their initial beliefs, $F(2,49) = 2.90, p = .06$. Table 1 shows the mean scores of the pre- and posttest and the interaction measures with the simulation for the direct instruction, the structured task, and the control condition. Univariate analysis of variance (ANOVA) was used to analyze between-group differences. In the simulation, children from the three conditions conducted a mean number of 14.6 experiments (SD = 5.44). No differences between conditions were found ($F(2,44) = 0.76, p = .47$). To check the comprehensiveness of the children’s investigations, the numbers of variables and values they investigated in the simulation were compared. Children in all three conditions scored nearly the maximum in these categories which indicates that most children from all conditions equally tried to find out about the effects of the variables, $F(2,44) = 1.18, p = .311$, and values $F(2,44) = 1.61, p = .211$. 
Table 1

*Means and Standard Deviations of Children’s Performance.*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Direct instruction ( (n = 19) )</th>
<th>Structured task ( (n = 17) )</th>
<th>Control ( (n = 16) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belief questionnaire</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(max. 7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid initial beliefs</td>
<td>2.87 (1.71)</td>
<td>3.35 (0.70)</td>
<td>3.94 (1.24)</td>
</tr>
<tr>
<td>Valid final beliefs</td>
<td>4.05 (1.38)</td>
<td>4.50 (1.26)</td>
<td>4.81 (1.49)</td>
</tr>
<tr>
<td>Experiments (^a)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trials</td>
<td>15.78 (7.26)</td>
<td>13.63 (4.24)</td>
<td>13.69 (4.82)</td>
</tr>
<tr>
<td>Variables investigated</td>
<td>3.78 (0.65)</td>
<td>4.00 (0.00)</td>
<td>3.92 (0.28)</td>
</tr>
<tr>
<td>(max. 4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Values investigated</td>
<td>9.22 (1.96)</td>
<td>9.49 (0.25)</td>
<td>9.85 (0.56)</td>
</tr>
<tr>
<td>(max. 10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean changes per trial (^b)</td>
<td>2.06 (0.54)</td>
<td>1.90 (0.62)</td>
<td>2.33 (0.44)</td>
</tr>
<tr>
<td>Conclusions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variables addressed</td>
<td>3.82 (0.38)</td>
<td>3.79 (0.59)</td>
<td>3.50 (1.21)</td>
</tr>
<tr>
<td>(max. 4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relations addressed</td>
<td>2.16 (1.63)</td>
<td>1.03 (1.19)</td>
<td>1.53 (1.49)</td>
</tr>
<tr>
<td>(max. 4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid inference score</td>
<td>3.42 (2.29)</td>
<td>1.77 (1.80)</td>
<td>1.94 (2.07)</td>
</tr>
<tr>
<td>(max. 7)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) The log files of five children have not been saved

\(^b\) Lower scores indicate better use of CVS

The mean number of value changes per trial was used as a measure for the use of CVS, with lower scores indicating more controlled experimentation. The scores in Table 1 show that the lowest CVS-scores were observed in the structured task condition. However, an ANOVA showed that these scores were not significantly lower than those in the other two conditions \( F(2,44) = 2.21, p = .121 \). Analysis of the log files revealed that across all three conditions only eight students used the CVS relatively consistently which explains the fact that the average number of mean changes per trial was high in all conditions \( M = 2.10, SD = \)
Children in all three conditions addressed a mean of at least 3.5 of the four variables in their inferences on the answer sheet and no significant difference between the conditions was found, $F(2,49) = 0.84, p = .436$. ANOVA indicated that there was a trend towards difference between the conditions concerning the number of relations that had been addressed in the concluding answers, $F(1,49) = 2.72, p = .076$. A post-hoc analysis of this trend with Bonferroni correction showed that the difference between the direct instruction ($M = 3.82, SD = 0.38$) and the structured task condition ($M = 3.79, SD = 0.59$) was not significant, $p = .74$.

Valid inference scores were used as a measure for comprehensive experimentation. ANOVA showed that children from the different conditions differed significantly in the valid inferences they made during their investigations, $F(2,49) = 3.51, p = .038$. However, post-hoc analysis with Bonferroni correction revealed that the differences between the conditions were not significant; between the direct instruction condition ($M = 3.42, SD = 2.29$) and the structured task condition ($M = 1.77, SD = 1.80$) the $p$ value was .061.

Valid final beliefs were used to assess whether the children improved their domain knowledge by working with the simulation. An ANCOVA with condition as between-subjects factor and pre-test as covariate revealed no main effects of condition, $F(2, 46) = 0.058, p = .994$, or pre-test, $F(1, 46) = 0.418, p = .521$. Furthermore, no interaction between condition and pre-test was found, $F(2, 46) = 0.057, p = .945$, which indicates that the conditions did not differ in domain knowledge improvement.

While there was no difference between the three conditions concerning the use of CVS, partial correlations, controlling for initial beliefs, showed a correlation between the mean number of variable changes and valid final beliefs, $r = .393, p = .007$. This means that students using CVS as intended (low mean number of variable changes) had less valid final beliefs. However the mean number of variable changes did not correlate with any other measure for systematic experimentation. Its correlation with the number of investigated variables was $r = .258, p = .083$, with the number of investigated values, $r = .250, p = .093$, with the number of variables addressed $r = .111, p = .462$ and with the number of relations addressed, $r = -.190, p = .205$. Furthermore, no correlation between valid inferences and valid final beliefs cold be found, $r = .017, p = .912$.

Conclusions and Discussion

The aim of this study was to examine whether a structured inquiry task is as efficient as direct instruction followed by unguided inquiry for teaching the control of variables.
strategy to children using an inquiry task simulation. If the findings would support these hypotheses task structuring would be recommended because time and teaching resources could be saved. The first hypothesis was that, according to the suggestions of Kuhn and Dean (2005) and Lazander and Kamp (2011), children in the structured task condition would apply the CVS as effectively as children in the direct instruction condition, and more effectively than children from the control condition. No differences between the conditions were found regarding systematic experimentation. The second hypothesis stated that no difference would be found between the direct instruction and structured task condition concerning valid inferences from their inquiry and valid final beliefs. Furthermore it was expected that both experimental conditions would make more valid inferences and have more valid final beliefs than the control condition. The results do not support this expectation, no difference between the three conditions regarding valid inferences and valid final beliefs were found. Furthermore, the mean number of variable changes correlated positively with valid final beliefs on the post-test. This means that children who used the CVS more had less valid final beliefs.

The finding stands in marked contrast with the literature cited in the introduction, which illustrated that direct instruction and task structuring have advantages over unguided discovery. Why did children in the control condition, who did not receive any support, improve their domain knowledge as well as children who did receive support? One explanation for their good performance might be the trend towards a difference between the conditions that existed in the pre-test. However the improvement (gained domain knowledge) from pre- to post-test was the same for all three conditions. Another reason for the missing difference between the conditions concerning the gain of domain knowledge could be that both of the supports in the experimental conditions did not have any supportive effect. How can this be explained?

The analysis of the log files showed that, although the children in the structured task condition had the lowest number of mean value changes per trial, which would indicate a good use of CVS, most of them did not use the simulation as intended. Oftentimes they skipped questions without doing experiments and thus did not work according to the specified structure that was built in the simulation to support them by organizing their investigations. Several explanations for this are possible, either the instruction on how to work with the simulation was not clear enough or they might not have paid attention while it was explained. Another explanation could be that they were not interested in working on the research questions, which would not be an uncommon behavior in school for children of this age. As
Van Joolingen, de Jong, and Dimitrakopoulou (2007) suggested, it might be possible that ‘completing the learning environment’ instead of answering the questions becomes the main goal of the children. This leads to “gaming the system”, which can be defined as mainly performing actions that take you through the system with a minimum of effort. The fact that they did not use the simulation as intended implicates that no comparison between the effectiveness of the direct instruction method and the structured task method can be drawn.

The question remains why the results from this study did not replicate other studies in which direct instruction was clearly superior to unguided discovery learning in facilitating children’s acquisition of CVS (Chen & Klahr, 1999). One explanation is that, although the children seemed interested and focused on the demonstration, the given direct instruction might still have been too short, not clear enough, or that the children did not pay attention to it, which would all lead to insufficient understanding of CVS and result in a performance similar to the one in the control condition with unguided discovery. Another explanation is that the children need more practice than half an hour to learn the concept and application of CVS. As Klahr (2005) states, a brief teacher demonstration is sufficient to grasp the principle idea of the CVS, however mastery of this strategy can be reached through repeated practice on gradually more comprehensive inquiry tasks. Klahr and Nigam (2004) demonstrated the benefit of a short direct instruction in comparison to discovery learning. However, although the time that was given to the children to conduct experiments is not explicitly mentioned in the article, from their description and the fact that their experiment spread over two days it seems that the time on task was longer than in the present study. This again would support the idea that brief direct instruction is sufficient to teach CVS, but only when enough time to practice is given. The results of the present study indicate that 25 minutes, of which the first time is used to familiarize with the simulation, seems not to be enough time to practice to get a beneficial effect of the direct instruction.

How can the unexpected findings concerning hypotheses two be explained? Why could no differences be found between the three conditions regarding valid inferences and valid final beliefs? Could it be that the belief-questionnaire did not assess their domain knowledge as intended? Although it addressed every variable involved in the simulation with the direction of its effect, it might be possible that the semi-open question format was inappropriate for children of this age and they had difficulties in expressing their ideas about the variables in their own words. While several children could answer the questions adequately, for the others it might be the case that if they only had vague ideas about the
effects of the different variables, it might have been easier for them to express their ideas in a completely closed rather than semi-open question format. This could also explain the question, why no correlation could be found between valid final beliefs or mean number of variable changes and valid inferences. A closer look at the answers given by the children on the answer sheet that was filled in during their investigations, shows that some children did not understand the questions correctly. They only wrote down the time they could hear the gong for the different settings without drawing conclusions about the effects of the individual variables. Some of them did so due to a lack of understanding of their task while some had a lack of time, as they expressed when their work was collected. The pace with which they were working differed between the children. Maybe some of them first wrote down all their findings and before they could draw conclusions from it the time on task was over. If they had more time, which is one way suggested by Mulder et al. (2011) to enhance task performance, it might be possible that they could draw inferences from their findings, increase their understanding and make more valid inferences. The negative correlation between application of CVS and valid final beliefs remains hard to explain.

In order to explain why the data does not support hypothesis 1 and 2 and most of the children’s performance was mediocre, one has to think about those parts of the experiment that were the same across all the three conditions. It might be the case that the instruction about the functionality of the simulation and their objective was not clear enough. This is a serious problem because it has a strong influence on the outcome. While during the direct instruction the children seemed to show interest and pay attention, during the instruction about the functionality of the simulation their attention and interest was low. This might have to do with the group size which was for the direct instruction about 10 and for the simulation instruction about 25 children. Furthermore, a difficulty was that because no projector device was available and the time that was allocated by the school was a bit short, all the children had to watch a short demonstration of how to use the simulation (set up and execute experiments and draw conclusions) on a single computer screen at the same time and then try it out on their individual computers with some further instruction. Because in the simulation no introduction phase exists to try out how to use it, as for example in the study by Kuhn and Dean (2005) who provided a whole initial session for this purpose, it can be assumed that at least the first few experiments that were set up by each child were used to familiarize with the learning environment. Although this was the same across all three conditions, it reduced the time they had to set up and execute meaningful experiments from which they could draw conclusions and by this reduced the time the children in the direct instruction condition had to
profit from their advantage. It might be the case that their advantage would have become visible if they would have engaged in further experimentation tasks, which was not possible because only 1 hour per class was provided by the school. Due to this practical limitation the transfer of the CVS skills to other domains also could not be measured. It can only be assumed that with a little more practice the children from the direct instruction condition would have outperformed the other two conditions in a transfer task.

Further research is needed to verify if task structuring could be used as effectively as direct instruction to teach the basics of scientific working. A larger sample should be used in order to receive more meaningful results. It can be recommended to take enough time to instruct the children about their objective and the functionality of the simulation and to make sure that every child pays attention, so that they will use the learning environment in the intended way. Furthermore it should be considered to include an orientation phase at the beginning of the simulation in which the children can familiarize with its functionality. It might also be helpful to have fewer children at the same time participating, because children in this age sometimes have difficulties working independently on a new task. By doing so, problems that occurred in this study, for example children rushing through the simulation and starting it a second time or logging off from the operating system and by this deleting the log files, could have been avoided.

In order to improve its guiding and structuring function, the design of the learning simulation for the structured task condition should be reconsidered because it has serious impact on how the children interact with it. As proposed by Mulder et al. (2011), it should be considered to prevent the children in the structured task condition from progressing to the next question without really dealing with the previous one. This could be achieved by, instead of using an answer sheet, providing one answer section per question inside the simulation, as was the case in the study by Lazonder and Kamp (2011). To make the structure even more explicit, it might be useful to make sure that only if the children filled in something, for example an answer that includes some specific terms, they would be able to proceed to the next question. Furthermore, to make sure they understand what kind of answers are expected, additionally to the example of an inference given during the instruction of the simulation, it could be considered to design the answer section for the first question in a sentence completing format. For example: “The . . . (higher/lower) the pitch of the sound, the . . . (shorter/longer) is the time one can hear the gong”. In this way, it would become clear what to fill in at the following answer sections which should be open questions/answers in order to
prevent the children from guessing. However, Mulder et al. (2011) stated that restricting the progress through the simulation might be in conflict with the iterative nature of the inquiry process and suggest a learning environment that allows the children to navigate freely through the simulation forwards and backwards. An additional way to make the structure clearer, it might be very helpful to indicate the state of progress in the simulation. Before each question is investigated a short notification could appear on the screen, for example: “Get ready for question 3 of 4!”. This would surely help the children to understand the structure of the simulation and to stay aware of their progress. A less obvious way would be to write the number of the current research question next to the question. It is up to future research to find the right balance between not enough and too much guiding structure in simulation based learning environments.
References


Lazonder, A.W. & Kamp, E. (2011): Splitting up the inquiry question to promote children’s scientific reasoning. University of Twente, Department of Instructional Technology


Appendix

1. Beliefs questionnaire (pre- and post-test)

2. Answer sheet
1.

Fragebogen I
Gruppe: 
Nummer: 


Bitte beantworte die untenstehenden Fragen so gut du kannst indem du Antwort a) ODER b) wählst.

1. Die TONHÖHE beeinflusst die Dauer die man den Gong hören kann.

a) Stimmt! (Erkläre wie die Tonhöhe die Dauer die man den Gong hören kann beeinflusst)

b) Stimmt nicht! Man kann den Gong bei jeder Tonhöhe gleichlange hören.

2. Die ENTFERNUNG beeinflusst die Dauer die man den Gong hören kann.

a) Stimmt! (Erkläre wie die Entfernung die Dauer die man den Gong hören kann beeinflusst)
b) **Stimmt nicht!** Man kann den Gong bei jeder Entfernung gleichlange hören.

---

**3.** Die **BLICKRICHTUNG** beeinflusst die Dauer die man den Gong hören kann.

a) **Stimmt!** (Erkläre wie die Blickrichtung die Dauer die man den Gong hören kann beeinflusst)

b) **Stimmt nicht!** Man kann den Gong bei jeder Blickrichtung gleichlange hören.

---

**4.** Die **EINRICHTUNG des Raums** beeinflusst die Dauer die man den Gong hören kann.

a) **Stimmt!** (Erkläre wie die Einrichtung die Dauer die man den Gong hören kann beeinflusst)

b) **Stimmt nicht!** Man kann den Gong bei jeder Einrichtung gleichlange hören.
2.

Fragebogen 2

<table>
<thead>
<tr>
<th>Gruppe:</th>
<th>Nummer:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Auf diesem Blatt kannst du während du mit dem Computer arbeitest Notizen machen. Wenn du Antworten auf die Frage(n) gefunden hast schreib sie bitte auch auf.

_________________________________________________________________
_________________________________________________________________
_________________________________________________________________
_________________________________________________________________
_________________________________________________________________
_________________________________________________________________
_________________________________________________________________
_________________________________________________________________