The Influence of Goal-Specificity on the Discovery of Main Effects and Interaction Effects in a Semi-Complex System

Marian Teigler
University of Twente
Faculty of Behavioral, Management and Social Sciences
Educational Sciences

Abstract[1]

This research investigated the influence of goal-specificity on student's inquiry learning performance in an experimental classroom setting. Sixty-two junior high school students between the ages of 10 and 12 worked on the Plant Growing Task in an inquiry learning environment constructed with the program FILE. This task involved the manipulation of five independent variables which have different effects on the dependent variable 'plant growth'. The participants were assigned to two experimental conditions, in which they either they received a non-specific goal that required them to find out how the five independent variables affect plant growth or the specific goal of achieving a particular outcome. Performance in both conditions was assessed by a written test; possible differences in strategy use, for different problem-solving strategies and rule-inducing strategies, were identified from log files. Results showed that students who received a non-specific goal acquired as much knowledge about the task as students who received a specific goal. Regarding strategy use, neither the hill-climbing strategy nor the history-cued strategy was observed in any of the participants. The control-of-variables strategy did occur, but its application did not differ across conditions. In addition, some students tried to reach their own goal which was to reach the highest possible outcome. The results suggest that neither students' knowledge acquisition about an inquiry task nor their strategy use is affected by goal-specificity.

[1] With special thanks to Prof. Dr. A. W. Lazonder and J. ter Vrugte MSc. from the University of Twente for the supervision of this research.

The Influence of Goal-Specificity on the Discovery of Main Effects and Interaction Effects in a Semi-Complex System

This study will attempt to provide evidence on the impact of goal-specificity on the discovery of main effects and interaction effects in an inquiry learning setting. Furthermore it will investigate which strategies students use when confronted with an inquiry learning task.

Previous studies have consistently emphasized the importance of implementing inquiry learning in schools. Inquiry learning is a form of learning which gives students the opportunity to take a more active role in their learning process (Klahr & Dunbar, 1988). An inquiry learning task enables students to discover information on their own, and make decisions on how to deal with that information (Klahr & Dunbar, 1988; Simon & Lea, 1974; Vollmeyer, 1996). This can for example take place in a computer simulation in which the opportunity is provided do experiments and evaluations in the field of physics, biology etc. . One advantage of this mode of learning is that it complies well with children's innate inquisitiveness (Zion & Zadeh, 2007). Furthermore it appears to reduce the achievement gap for Afro-American students as seen in a study by Kirschner, Sweller, and Clark (2006), which suggests that inquiry learning addresses different learning needs. Another advantage is that inquiry learning supports deep cognitive thinking and scientific reasoning (Zion & Zadeh, 2007). Many researchers argue that this is a desired value of learning that should be supported in education (e.g., Klahr & Dunbar, 1988; Simon & Lea, 1974; Vollmeyer, 1996). In addition, inquiry learning can improve performance as was found in a large study that involved high-stakes standardized tests (Hmelo-Silver et al., 2007). However, others have expressed their concern that inquiry is too often implemented rather poorly or in a way in which learning outcomes are weak (Valanides, Papageorgiou & Angeli, 2014). So in order to take advantage of inquiry learning it is important to ensure that effective learning can take place. The following sections will explain the concepts necessary to understand the theory of inquiry learning.

Students' Development of Inquiry Skills

To successfully implement inquiry learning tasks for students, it is important to know the natural development of inquiry skills. Students develop their approach to experimentation at the beginning of their teenage years. The approaches change within only a few years time. When children enter the first grade, they are able to differentiate between belief demonstration and testing a scientific hypothesis, if they are provided with a concrete task (Koerber, Sodian, Thoermer & Nett, 2005). The ability to think scientifically and create hypotheses develops at the age of 10 (Piaget & Inhelder, 1973). Yet, it is also stated that 10-year olds lack the experience to always do so properly (Piaget & Inhelder, 1973). At this age they still tend to take a belief demonstration or prediction approach when they conduct experiments in the domain of physics (Penner & Klahr, 1996). Around the age of 12, students abandon the goal of belief demonstration and begin with experimentation for inquiry investigation (Penner & Klahr, 1996). Students up to the sixth grade are biased by prior beliefs when making causal judgments (Amsel & Brock, 1996). Fourteen-year-olds are more likely to explicitly test a hypothesis about an attribute without a strong belief or need to demonstrate that belief (Penner & Klahr, 1996). As found in a study by Amsel and Brock (1996) only college-educated adults possess the ability to set prior knowledge aside and make their judgments based solely on data. So, it appears that inquiry skills increase by age and are also influenced by educational level. Yet, these stage-like developmental changes do not apply universally, since exceptions have also been found like preschoolers who show inquiry skills or who lack these skills (Schneider, Knopf & Sodian, 2009). These findings shed light on students' perspectives and intentions when faced with an inquiry task. Apparently ten-years-old students' cognition is ready to learn about scientific hypotheses testing. However, they lack the experience and need several years to put strong believes aside. So they might need some kind of instruction to be able to engage in proper hypothesis testing (Strand-Cary & Klahr, 2008; Lazonder & Kamp, 2012).

Goal-Specificity Effects and Strategy Use

When it comes to the issue of how inquiry learning should be employed to reach good learning outcomes, goal-specificity is often a topic of discussion. According to several researchers goal-non-specificity is more beneficial for learning than goal-specificity (Burns & Vollmeyer, 2002; Lazonder

& Kamp, 2012; Zimmermann, 2007). When students receive a specific goal or think that there is one desired outcome in the task, they will most likely apply a strategy to reach that goal as soon as possible (Burns & Vollmeyer, 2002). The learning effect of this approach is rather low because the student focuses on achieving the goal rather than on investigating how the individual factors affect the outcome. For example, when a student is asked to investigate several variables in a computer simulation in order to make a boat flow fast, there is a fair chance that s/he will try to make the fastest possible boat rather than actually trying to understand which of the single variables are responsible for the boat's speed (Kuhn & Phelps, 1982).

Goal-non-specificity is said to support learning stronger than goal-specificity does (Burns & Vollmeyer, 2002; Lazonder & Kamp, 2012; Zimmermann, 2007). A nonspecific goal in the example above could be to 'examine the different variables that affect flowing speed'. So the specific goal gives students a clear, measurable end state: students know when they have accomplished the task. The non-specific goal, by contrast, just asks to explore the variables in the system. It is left unspecified when exactly this goal has been reached. Many researchers claim that students, when given a non-specific goal, would try to understand the effects of the variables rather than trying to achieve a certain outcome (Bruner et al., 1956; Dewey, 1913; Lazonder & Kamp, 2012). This would lead to stronger learning about the topic of inquiry and would be a more desirable outcome than the understanding how to reach one specific goal (Bjork, 1994; Lazonder & Kamp, 2012; Vollmeyer et al., 1996), because when students are given a specific goal they think they have sufficient knowledge about the system as long as they reach the desired outcome and do not engage in more accurate assessment of the system which would lead to stronger learning (Bjork, 1994).

Simon and Lea's (1974) work described what causes the difference between a condition in which a specific goal is given and a condition in which there is no specific goal given. A specific goal represents a state in an instance space or, as Klahr and Dunbar (1988) would call it, experiment space. The experiment space contains every possible instance, which is mostly represented by every distinct combination of settings of a problem (Klahr & Dunbar, 1988). If a nonspecific goal is given, or no goal at all, it means the absence of a goal state in the instance space. Without a goal that guides the search in the instance space, students might search the rule space, which is also called the hypothesis space, to direct their search in the instance space. This means that students will use what appears to be the most efficient strategy to reach a desired outcome. But if they don't have a goal or desired outcome, they will apply another method. They will use a strategy to investigate the underlying rules of the system and thus, will acquire knowledge about the overall structure.

Considerable effort has been put into identifying the strategies students use when confronted with an inquiry task. A specific goal has mostly been found to evoke the strategy of *focus gambling*, which involves the variation of several variables at once (Vollmeyer et al., 1996). When students use this strategy, they try to come closer to the goal with every variation, yet they do not bother to investigate the actual effects of every single variable. Furthermore a problem-solving strategy that has been observed in the behavior of students, namely *hill-climbing* (Robertson, 2001). Hill-climbing is a strategy which aims to reduce the gap between the current state and the goal state as much as possible (Robertson, 2001). With every step a problem-solver judges the possible solution paths and selects the one which appears to get the closest to the goal state. S/he does so until the remaining moves would increase the difference. The implication of this method is that there is no guarantee to actually reach the goal, because in many problems there is a state from which any move left appears to increase the difference between current state and goal state without having reached the goal. This would force a problem-solver to stop and therefore let the problem remain unsolved or to repeat the procedure from the initial state but taking another path.

The strategies mentioned before apply to tasks in which one or several goal states have to be reached. Different strategies apply when students are given a nonspecific goal. Then strategies are applied which explore the rules and properties of the task. One of these is the *history-cued strategy* (Sweller, 1988). This strategy generates subsequent moves always on the basis of previous results. Consequently, the user of this strategy will make use of records to create new variations of previously found states. It is considered a forward-working strategy that searches possible moves until no further moves are left (Sweller, 1988). Since there is no specific goal to achieve, this strategy aims to reveal the rules underlying the task (Sweller, 1988). There is a strategy, suitable for an inquiry task without a specific goal, which is highly recommended to foster strong learning, namely the *control-of-variables* *strategy* (CVS) (Inhelder & Piaget, 1958). CVS is a strategy in which only the variable of interest is manipulated while keeping all other variables constant. This strategy enables students to discover the single effects each related to one variable, also called the main effects. Although studies do not report frequent observation of the history-cued strategy and the CVS when students try to reach a specific goal, both strategies are also applicable for a task with a specific goal. Then students would explore the task and could possibly use the gained knowledge to reach the specific goal.

The strong learning mentioned before is related to learning about these main effects. Several studies have shown that goal-non-specificity, in comparison with goal-specificity, evokes strategies appropriate for identifying main effects (Lazonder & Kamp, 2012). Yet, it has also been found that students rarely figure this strategy out on their own (Lazonder & Kamp, 2012). So, it is suggested to either teach students the CVS or at least give them the hint to focus on one variable at a time (Lazonder & Kamp, 2012).

While main effects refer to the impact of a single variable, interaction effects are caused by the joint influence of two or more variables. As goal-specificity may prompt students to manipulate several variables at once and because interaction effects are caused by two or more independent variables, this research is made on basis of the assumption that goal-specificity may support the learning about interaction effects more than goal non-specificity does.

So, in order to support students' inquiry learning possibly a specific goal must be provided for the inquiry of interaction effects, while a non-specific goal must be provided for the inquiry of main effects. Therefore it is important to distinctively investigate how goal specificity affects the inquiry of main effects on the one hand and the inquiry of interaction effects on the other hand. Of further interest is which strategies students apply to solve an inquiry task and how goal-specificity affects their strategy use. So, two research questions were investigated in this research:

1. 'What is the influence of goal-specificity on young teenagers' discovery of main effects and interaction effects in a semi-complex system ?'
2. 'What is the influence of goal-specificity on young teenagers' strategy use during an inquiry task in a semi-complex system?'

Method

Participants

Sixty-two junior pupils from a German high school in Münster (Westfalia) participated in this study. Pupils were fifth graders between the ages of 10 and 12. The participants interacted with a computer simulation about plant growth in one of two groups**.** To ensure that both groups were comparable, participants were randomly assigned to one of two equally sized groups. Both groups received a short instruction about CVS, which is described in more detail in the subsection below. Participants in the first group received the nonspecific goal 'Find out how the variables affect the plant length'. The second group received the specific goal 'Increase the plant's length as close as possible to 18 cm'.

Materials

Plant growing task. During the experiment, participants in both conditions worked with a computer simulation of plant growth. This simulation was developed in FILE (flexible inquiry learning environment; Hulshof, Wilhelm, Beishuizen, & Van Rijn, 2005), an authoring platform for designing inquiry learning tasks. With this program the task designer can freely configure the task domain and also the task model, for example the relationship of input and output variables. The task that was used in this study had five discrete input variables with two or three values each, as shown in Table 1. The dependent variable was the length of the plant, which could be either 5, 8, 10, 12,13, 17, 20, 23 or 27 cm.

Table 1
Independent Variables and Levels in the Plant Growing Task

Variables		Giving Water	Insecticide	Dead plant leaves	Location of the plant	Size of flower pot
Levels	1	Once a week	Not using it	Leave them in the pot	Greenhouse	Small
	2	Twice a week	Using it	Remove them	Indoors	Big
	3	-	-	-	Balcony	-

In this model, variables 1 and 5 had a joint effect on the dependent variable 'plant length'. Giving water once a week in a small flower pot would result in a greater length than in a big flower pot. Giving water twice a week in a small flower pot led to a lower length than in a big flower pot. Variables 2 and 3 had no effect on plant length, and variable 4 had a negative main effect: when the plant was indoors, its length was smaller than when the plant was placed on the balcony or in the greenhouse

The simulation recorded participants' actions in a log file that was the main source for data analysis. This feature is explained in more detail in the upcoming sections.

CVS instruction. The CVS instruction involved a short interactive lesson on experimental design that was conducted with a simulation of balls rolling from the back of a truck. The independent variables and their values were load area level (up or down), initial ball position (high or low), and the weight of the ball (heavy or light). The dependent variable was the distance the ball rolls from the truck.

The participants were introduced to the simulation and the research question for the experiment 'determine how these factors influence how far the ball will roll'. The participants were asked to suggest possible variables which affect the distance the ball will roll. The instructor demonstrated how to set values for each variable and explained that the goal of this task was to find out the effect of each of the three variables. The instructor showed two experiments which had two variables changed at a time. The participants were asked to discuss whether these experiments were suitable for finding out single effects of variables. Eventually they were taught that they have to focus on one thing at a time, for example how the weight of the balls influence the distance the ball travels. Thus they had to focus on one variable at a time. The value for this target variable had to be changed while other variables have to remain constant. Furthermore, the participants were taught that they have to run at least two experiments to establish the effect of one variable. Afterwards they were requested to tell how to fix the settings in order to create suitable experiments. After that they received some exercise sheets on which they had to write a research question for each experiment, draw the settings for the experiment which is about to be conducted, and after the instructor had run the experiment they had to write down the answer to their research question.

Domain knowledge test. This test contained six open questions, one about each independent variable in the plant growth simulation, and one additional question about possible interaction effects. The first five items asked participants to indicate the effects of water, insecticide, dead plant leaves, location of the plant, and the size of the flower pot on plant length, respectively. The sixth item asked participants to explain an alternative way on which these five possible causes could affect plant growth. The participants completed the test on paper by writing down a short but detailed answer to each question. The test was administered twice, once as a prior knowledge test and once as a posttest.

Procedure

The researcher served as the instructor for the tests and the trial. Participants first received a briefing. In that briefing the instructor introduced himself as student of educational science at the University of Twente. He explained the purpose of the experiment which is to investigate students' performance in an inquiry learning environment. Participants got informed that their performance will be recorded anonymously and therefore will have no consequences for their grades. After that, the instructor handed out a pretest. It consisted of questions about main effects and interaction effects. This was necessary to determine whether performance may be influenced by prior knowledge and not solely by the independent variable, goal-specificity. The collected the filled in pre-tests as soon the respondents had finished. Then the instructor explained the procedure of the experiment. The instructor told the participants that they first receive a short instruction about the CVS, then start their inquiry of plant growth with a computer simulation and finally take a posttest. The instructor explained the CVS (see section CVS instruction above). Afterwards the participants were introduced to the Plant Growing Task. The instructor explained the topic, Plant Growing, and the possible independent variables which could influence the depended variable, plant size. He instructed the participants how to change the values of the variables and where to look for the output, the plant size in numbers. The goal-non-specificity group received the non-specific goal ‘Find out how the variables affect the plant length’. The second group received the specific goal ‘Increase the plant’s length as close as possible to 18cm’. Here it has to be mentioned that 18 cm. cannot be reached in the Plant Growing Task. To prevent that participants stop their trial without having run a decent amount of experiments, the goal is formulated in a rather ideal way, so participants consider getting closer to 18 cm. The best possible solution would be 17 cm. Both groups were supplied with extra paper and were advised to take notes during the trial. After the trial was finished, the instructor handed out the posttest. With this test it can be determined what the participants learned during the trial about the main effects and interaction effects. The experiment ended when the students had completed this test and the instructor had collected the filled in test sheets. In the end of the experiment the instructor gave a debriefing. He explained that the collected data will be assessed anonymously, will be soon reported at the University of Twente, and will also be acquirable for the participants via e-mail. He also spoke out his thanks for the participation.

Coding and Scoring

The first research question concerned the discovery of main effects and interaction effects in the Plant Growing Task. The given answers for main effects were compared with the model underlying the simulation. For question 2, 3 and 4, which measured knowledge about the main effects, up to 3 points could be earned. No points were awarded if a main effect was not detected. One point was given if the participant had correctly identified a main effect, and one additional point in case the correct direction of effect was specified (e.g., the greenhouse increases the plant length more than the balcony, and the balcony increases plant length more than indoors). If the participant additionally described the magnitude of the effect (e.g., a greenhouse increases the plant length by 7 centimeters) the maximum score of 3 points was granted. For the interaction effect a scoring system, developed by van Leuteren (2014), was used. In this system the given answers for the interaction effect were assigned to 5 different categories.

1. There is no combination of variables described, so no interaction effect was found.
2. Two wrong variables were combined, so the wrong interaction effect was found.
3. An entire experiment, which involves the two correct variables, is described but no clear distinction is made. So, there is an assumption of an interaction effect.
4. The interaction between the two correct variables is described but only on one level, for example: ‘Giving water once increases plant growth more than giving water twice in a small pot.’. Thus a limited interaction effect is described.
5. The interaction of the two correct variables on both levels is described. The respondent has a complete understanding of the interaction effect.

The different categories show the levels of comprehension about the interaction effect. In this system category 1 is the lowest level and category 5 the highest.

For the second research question it must be defined how strategies can be identified from the respondents’ actions. So the following subsections will provide an overview of these definitions.

Hill-climbing. As stated in the introduction, hill-climbing is a strategy which attempts to reduce the gap between the current state and the goal state as much as possible with every move. The strategy of the participant can be analyzed to which degree it matches hill-climbing from the second experiment on. The degree to which the participant’s strategy matches hill-climbing depends on the degree the distance to the goal is reduced with the following experiments. If the outcome of an experiment increases this distance, the participant changes the responsible variable back for the following experiment, if s/he actually uses hill-climbing.

History-cued strategy. This strategy requires the participant to look back to recorded data, so s/he knows which experiments are not conducted yet and have to be explored then. So, the first criterion of this strategy would be that no experiment is repeated. The second criterion is that the participant aims to explore every rule. So, even if the participant's experimentation tempo is too low it is still expected that s/he conducts as many different experiments as possible. So, the performance assessment questionnaire will contain the additional question: 'Did you have sufficient time to conduct all the experiment you wanted?' This question will assess whether the student gave up on finding more rules, thus was not fulfilling the second criterion.

Control-of-variables strategy (CVS). This strategy requires the participant to focus on one variable at a time while keeping the other variables constant. So, when the participant uses this strategy it should be observed that s/he changes the value for only one single variable when conducting an experiment. Next to the variable which gets to be changed all the other variables should remain as they are.

Data analysis

To compare the scores between the two different conditions, goal-specificity and goal-non-specificity an Mann U Whitney test for non-parametric data was conducted via SPSS.

Results

Main Effects

Fifty-seven of the 62 students participated in all four parts of the research session, namely the pre-test, CVS instruction, trial and post-test. Five participants did not complete the post-test and were removed from the sample. As shown in Table 2, the scores regarding main effects on the pre-test and post-test were rather low, which points to a bottom effect. Consequently, the test scores were not normally distributed, because most scores of the students were close or equal to zero. Thus a Mann-Whitney test was chosen, which is appropriate for comparing two means which have no normally distribution. This test was conducted for post-test scores to answer the research question 'What is the influence of goal-specificity on young teenagers' discovery of main effects and interaction effects in a semi-complex system ?'

This test was also conducted for pre-test scores to exclude the possibility that differences between the groups in the post-test can be accounted to prior knowledge. On basis of the assumption that students who received a non-specific goal scored higher on questions about main effects than students who received a specific goal, the scores were tested one-sided with a significance level of .05. The Mann-Whitney test for the pre-test indicated that there was no significant difference between participants in the non-specific goal condition (Mdn= 2) and participants in the specific goal condition (Mdn= 3), U = .514, p = .434. The Mann-Whitney test for the post-test also indicated that there was no significant difference between the scores the non-specific goal condition (Mdn = 10) and the specific goal condition (Mdn = 7), U = 1.567, p = .059. Thus the tests showed no significant difference for finding main effects in the pre-test and the post-test as well.

Table 2
Means and Standard Deviations of Main Effects in Both Groups

	Specific goal group (*n*= 27)		Non- specific goal group (*n*= 29)	
	M	*SD*	*M*	*SD*
Pre-test	.05	.15	.05	.19
Post-test	.27	.31	.05	.55

Interaction Effect

It was found that the interaction effect was found by one single student in the non-goal condition. This participant reached the third level, presumption of an interaction effect. Due to these low numbers, there was no statistical test conducted to compare the scores for interaction effects.

Strategy use

To compare the number of experiments conducted in each condition, an independent-samples t-test for normally distributed data was conducted. The mean number of experiments in the specific goal group (M = 35.59, SD = 9.63) was significantly higher than that in the non-specific goal group (M = 23.72, SD = 11.59), $t(54) = 4.15$, $p < .001$, $d = 1.11$. Furthermore, the logfiles were analyzed to examine whether participants' experimentation matched the goal of their condition, or in other words, whether the effort of students to reach the task was visible in the data. Since participants in the non-specific goal group received the goal to explore every variable they were either expected to use the systematic approach CVS or at least to explore a large portion of all possible combinations of the variables. The exploration of all possible combinations in the system would show effort to investigate every variable. If they tried to explore every possible combination their number of unique experiments should be close to 48, which is the total number of possible unique experiments. In order to prevent that students' attempt to solve their task is not recognized when they missed a small number of experiments, it was decided to use a threshold of 36 unique experiments, which is 75% of the total number of unique experiments. So, participants in the non-specific goal group were labeled as task-solver when they executed at least 36 unique experiments or showed the application of CVS. Participants in the specific goal group who tried to reach a plant length of 18 cm. should have executed a number of experiments which eventually centered around the outputs of 17 cm. and 20 cm. Participants who did so were labeled as task-solvers. Table 3 provides an overview of the number of task-solvers per condition.

Table 3
Number of Task-Solvers and Non-Task-Solvers in Both Groups

	Task-solver	Non-task-solver
Non-specific goal group	8	21
Specific goal group	16	11

Table 3 shows a relatively low number of participants with a non-specific goal who tried to solve their task. A Chi-square test of independence was calculated to compare the frequency of task-solvers among participants in the two groups. A significant effect was found, $\chi^2(1, N = 57) = 5.728$, $p < .05$. The percentage of task-solvers was higher in the specific goal group (59.3%) than in the non-specific goal group (27.6%).

Concerning strategy use, the CVS was applied by 6 participants in the non-specific goal group and 6 participants in the specific goal group. Another Chi-square test of independence was calculated to compare the frequency of CVS users among participants in the two groups. There was no significant effect found, $\chi^2(1, N = 57) = .20$, $p = .889$. The percentage of CVS users was only slightly higher in the specific goal group (22.2%) than in the non-specific goal group (20.7%). CVS users were found among high-achievers and also low-achievers on the post-test. None of the participants made use of the history-cued strategy or the hill-climbing strategy. Thus most participants made use of an undefined strategy. Next to these findings, however, a strategy was found which was not proposed before the experiment. During experimentation some students said that they would increase the output as high as possible. The repetitive effort to increase the plant length to the maximum could also be observed in the logfiles for a number of students in both groups (n = 4).

Conclusion

This study sought to answer two research questions. The first read: 'What is the influence of goal-specificity on young teenagers' discovery of main effects and interaction effects in a semi-complex system?'

Receiving a non-specific goal was predicted to lead to more knowledge of main effects than receiving a specific goal. This hypothesis could not be confirmed by the results because no group scored significantly higher than the other. Receiving a specific goal was predicted to lead to more knowledge of interaction effects than by receiving a non-specific goal. Also this hypothesis was not supported by the results. The only participant who found an interaction effect inquired that effect in the non-specific goal condition. Yet, this result neither proofs the opposite, which would be that receiving a non-specific goal leads to more inquiry of interaction effects. The outcome of this effect was too low for most respondents to make such a statement.

The second research question addressed the strategies students used to investigate plant growth with the simulation. On basis of the literature study, students with a non-specific goal were expected to explore plant growth more profoundly than students with a specific goal would do. Yet, it was found that students who received a specific goal executed more experiments on average than students with a non-specific goal. This finding suggests that students in the latter condition explored less about the Plant Growing Task than students in the former condition. Students in the non-specific goal condition were furthermore expected to apply a strategy appropriate to find main effects. The CVS is such a strategy, and it was expected that these students would adopt this strategy after being introduced to it. Yet, results showed that few students made use of the CVS, which was taught before the trial, during their investigation. Furthermore, the number of CVS users did not differ between the two groups. Some students discarded the CVS and tried to reach their own goal which was to reach an outcome as high as possible. The other strategies, hill-climbing and the history-cued strategy, were not observed during experimentation.

It was also found that the number of students who could be labeled as task-solvers was lower in the non-specific goal condition than in the specific goal condition. These results prompt that less students in the non-specific goal group tried to actually reach their goal than students in the specific goal group.

To conclude, the tests for both research questions do not support the hypotheses, meaning that students in both conditions performed comparably during their investigation of main effects and interaction effects. Another important finding is that some students seemed to abandon the assignment and tried to establish their own goal, which involves reaching the highest possible plant length.

Discussion

When looking at the data gathered during this study, the fact that scores were mostly low stands out. The overview of the data tells that most students scored 0 points at the post-test, which may have occurred because some initial understanding of the topic of investigation is needed (or helpful) for productive experimentation and learning (Zimmerman, 2007). The CVS, which was assumed to support the inquiry of main effects in the non-specific goal condition, was rarely used.

There are several possible issues about the research design which might have been responsible for the low score. Administering the domain knowledge test twice lead to both confusion an boredom. When receiving the post-test, many students paid little attention to it because they either thought they would have already finished this test, or were not enthusiastic to do the same task again. To prevent confusion, the fact that pre-test and post-test were identical was pointed out for the students from the second data collection session on. Yet, still confusion and repetition of the task might both have been plausible explanations for the low scores on the post-test. Furthermore, the short CVS instruction might have been insufficient for the students. According to Strand-Cary and Klahr (2008), sustained practice is an important part of effective CVS instruction. It is said to engage students in applying the strategy. Yet, due to time constraints, an elaborate exercise in which students could try out the strategy on their own, could not been provided. Another reason could be that the CVS instruction used a different task. Students were introduced to a dynamic simulation of a physics topic which has a visual output, a rolling ball. In contrast, the Plant Growing task appears more like a puzzle with biological content and a numerical output. Both systems appear very different in the way variables have to be set and changed. So, it is also possible that students failed to see that they could (or should) transfer the learned CVS to the Plant Growing Task. This assumption is supported by the fact that students did not ask for any advice how to apply CVS during the trial.

Next to the scanty use of CVS, the history-cued strategy and the hill-climbing strategy were not used at all. At least for the latter one, a likely reason might be the kind of task. Hill-climbing has been mostly observed in tasks that offer situations in which a restricted number of paths can be taken with a predictable outcome like in the Tower of Hanoi task or the Cannibal Monk Dilemma

(Robertson, 2003). The Plant Growing task does not lead the problem-solver into situations in which only few paths are possible. Every possible outcome is attainable at any time during the task. Neither are the outcomes predictable. So this might be a reason why hill-climbing was not observed in any of the participants. However, a strategy could be observed which had not been proposed initially. Although students were instructed to solve the task individually some of them competed at pushing the outcome as high as possible. For these students this might have been a likely distraction from the given task.

In addition to the low main effect scores, also the almost complete lack of interaction effect inquiry stands out. It can only be assumed that also the inquiry of this effect was too challenging for the students.

For further research on main effects in semi-complex systems, this study suggests to design the research in a way that the system used for instruction and the system used for the trial have sufficient resemblance for students to apply CVS. To investigate the influence of goal specificity on interaction effects, the research should be further designed in a way that students are more likely to find the interaction effect.

One implication for the practical use of CVS in instruction is to provide a decent amount of training of that principle. Instructors who want to learn in in an inquiry learning environment with a non-specific goal are advised to maintain that students are focusing on their task since such an environment provides the opportunity to establish an own goal or task.

References

Amsel, E. & Brock, S. (1996). The development of evidence evaluation skills. *Cognitive Development*, *11*, 523-550.

Bjork, R. A. (1994). Memory and metamemory consideration in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: knowing about knowing* (pp. 185-205). Cambridge, MA: The MIT Press.

Bruner, J., Goodnow, J., Austin, G. (1956). *A study of thinking.* New York:Wiley.

Burns, B., & Vollmeyer, R. (2002). Goal Specificity effects on hypothesis testing in problem solving. *Quarterly Journal of Experimental Psychology*, *55A*, 241-261. doi:10.1080/02724980143000262

Dewey, J. (1913). *Interest and effort in education.* Boston: Houghton Mifflin.

Hmelo-Silver, C., Duncan, R., & Chinn, C. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99-107. doi:10.1080/00461520701263368

Hulshof, C.D., Wilhelm, P., Beishuizen, J.J., & Van Rijn, H. (2005). FILE: A Tool for the Study of Inquiry Learning. *Computers in Human Behavior*, *21*(6), 945-956.

Inhelder, B., & Piaget, J. (1958). The growth of logical thinking from childhood to adolescence. New York: Basic Books

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41, 75-86. doi:10.1207/s15326985ep4102_1

Klahr, D., & Dunbar, K. (1988) Dual space search during scientific reasoning. *Cognitive Science.* 12, l-55. doi:10.1016/0364-0213(88)90007-9

Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children: Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology,* 64(3), 141-152.

Kuhn, D., & Phelps, E. (1982). The development of problem solving strategies. In H. Reese (Ed.), *Advances in child* development and behavior, 17, 1-44. New York: Academic Press

Lazonder, A. W., & Kamp, E. (2012). Bit by bit or all at once? Splitting up the inquirytask to promote children's scientific reasoning. Learning and Instruction, 22(6), doi:10.1016/j.learninstruc.2012.05.005

Leuteren, van, N. B. (2014). *Finding interaction using animated analogies in inquiry learning*, Unpublished master's thesis, University of Twente, Enschede, Netherlands.

Penner, D. E., & Klahr, D. (1996). When to trust the data: Further investigations of system error in a scientific reasoning task. *Memory & Cognition, 24*, 655-668.

Piaget, J., Inhelder, B. (1973). *Memory and intelligence*. London: Routledge and Kegan Paul.

Robertson, S. I. (2001). *Problem solving*. Hove, UK: Psychology Press.

Schneider, W., Knopf, M., & Sodian, B. (2009). Human development from early childhood to early adulthood: Findings from a 20 year longitudinal study. In W. Schneider & M. Bullock (Eds.), Verbal memory development from early childhood to early adulthood (pp. 63-90). New York: Psychology Press.

Simon, H.A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L.W. Gregg (Ed.), *Knowledge and cognition.* Hillsdale, NJ: Erlbaum.

Strand-Cary, M., & Klahr, D. (2008). Developing elementary science skills: Instructional effectiveness and path independence. *Cognitive Development, 23*(4), 488-511. doi:10.1016/j.cogdev.2008.09.005

Sweller, J. (1988). Cognitive load during problem solving: Effect on learning. *CognitiveScience, 12,* 257-258. doi:10.1016/0364-0213(88)90023-7

Triona, L., & Klahr, D. (2003). Point and click or grab and heft: comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. *Cognition and Instruction*, 21, 149–173. doi:10.1207/S1532690XCI2102_02

Valanides, N., Papageorgiou, M., & Angeli, C. (2014). Scientific Investigations of Elementary School Children. *Journal of Science Education and Technology, 23*(1), 26-36.

Vollmeyer, R., Burns, B. D., & Holyoak, K. J. (1996). The impact of goal specificity and systematicity of strategies on the acquisition of problem structure. *Cognitive Science, 20*, 75-100. doi:10.1207/s15516709cog2001_3

Zacharia, Z. C., & Olympiou, G. (2010). Physical versus virtual manipulative experimentation in physics learning. *Learning and Instruction*. doi:10.1016/j.learninstruc.2010.03.001.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*, 172-223. doi:10.1016/j.dr.2006.12.001

Zion, M., Sadeh, I. (2007). Curiosity and open inquiry learning. *Journal of Biological Education, 41*(4), 162-168. doi: 10.1080/00219266.2007.965