Instr Sci (2015) 43:365–379
DOI 10.1007/s11251-014-9334-5

Child predictors of learning to control variables via instruction or self-discovery

Barbara Wagensveld · Eliane Segers · Tijs Kleemans · Ludo Verhoeven

Received: 18 April 2014 / Accepted: 8 December 2014 / Published online: 17 December 2014

Abstract We examined the role child factors on the acquisition and transfer of learning the control of variables strategy (CVS) via instruction or self-discovery. Seventy-six fourth graders and 43 sixth graders were randomly assigned to a group receiving direct CVS instruction or a discovery learning group. Prior to the intervention, cognitive, scientific, and linguistic child factors were assessed. Acquisition and transfer of CVS knowledge were measured before and after intervention. Outcomes showed that CVS acquisition and transfer learning gains were higher for the instruction group than for the discovery group. Variation in CVS acquisition learning gains was explained by CVS pretest, verbal reasoning skills, and reading comprehension in the discovery group only. Variation in transfer gains were explained by vocabulary in the instruction group and by CVS prestest, verbal reasoning, vocabulary, and reading comprehension in the discovery group. It can be concluded that instruction provides a more effective method for learning CVS since individual differences in the instruction group played a marginal role explaining CVS posttest. However, since individual differences in linguistic ability account for variation in CVS learning gains discovery learning, these should be taken into consideration when teaching CVS in a more discovery-based manner.

Keywords Control of variables strategy · Deductive reasoning · Discovery learning · Primary school · Experimentation · Language

Introduction

Learning to control variables in an experimental setting has been widely recognized as a principle component in science education (e.g., Klahr 2009; Zimmerman 2000, 2007).

B. Wagensveld (✉) · E. Segers · T. Kleemans · L. Verhoeven
Behavioural Science Institute, Radboud University Nijmegen, Montessorilaan 3, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands
e-mail: b.wagensveld@pwo.ru.nl

Although there is evidence of systematic experimentation in kindergartners (Croker and Buchanan 2011), older children can show difficulties in systematically controlling variables and applying the so-called *control of variables strategy* (CVS; Inhelder and Piaget 1958) during experimentation. Over the last few decades, CVS research has mainly focused on determining the most effective method to teach the strategy to children in general. Instructional methods that have been examined vary from explicit instruction (e.g., Klahr and Nigam 2004) to more implicit discovery-based learning methods (e.g., Dean and Kuhn 2007) and intermediate procedures, such as scaffolded instruction (e.g., Lazonder and Egberink 2013). Although these studies contributed greatly to establishing a consensus on how to teach the CVS to the average child, not much is known about the interplay between these teaching methods and individual differences in children (Siler et al. 2010). This leaves the question which type of instruction works best for which child unanswered. In the present study, it was therefore explored how child factors are related to the acquisition of CVS and its transfer effects in different learning contexts.

Both constructivist and instructivist approaches have been studied in teaching children to utilize the CVS. The constructivist view suggests that children acquiring knowledge in a discovery-based learning situation are better at applying and transferring their knowledge to new situations (e.g., Schauble 1996; Stohr-Hunt 1996). Following this view on learning, Dean and Kuhn (2007) showed that CVS knowledge gained by means of practice in a discovery situation, as opposed to direct instruction, is more robust and long-lasting. However, Klahr and colleagues have shown in several studies that, on average, children receiving direct CVS instruction outperformed children experimenting without or with low levels of instructional guidance. They conclude that complex higher-order strategies, such as the CVS, are difficult for children to discover and are therefore best taught by means of instruction (e.g., Chen and Klahr 1999; Klahr and Nigam 2004; Strand-Cary and Klahr 2008; Klahr 2009; Matlen and Klahr 2013). Even studies comparing scaffolded instruction to direct instruction suggest that direct instruction is at the least as beneficiary as step-by-step guidance (e.g., Sao Pedro et al. 2009; Lazonder and Egberink 2013).

Although there has been a strong focus on instructional methods within CVS research, there is evidence that the employed instructional method is no longer relevant once the strategy is acquired. This observation is referred to as the *path independency hypothesis* by Klahr and Nigam (2004). In their study they compared children who were proficient in applying CVS after instruction with children who were evenly proficient at applying CVS after experimenting by themselves in a discovery-based learning situation. Klahr and Nigam found that, despite the difference in learning method, both groups of children with expertise in CVS performed evenly well on a transfer task. However, although the learning method appears to be irrelevant once CVS is learned, Klahr and Nigam still concluded that direct instruction should still be favored over discovery-based learning. They based their argument on efficiency considerations, since they found that many more individual children learned in the direct instruction group than the discovery group.

The finding that fewer children are able to learn the CVS in a discovery-based manner might be related to individual differences in their pre-existing capacities. Veenman et al. (2004) showed that the ability to learn CVS gradually increases from middle childhood into adulthood. Age differences have also been shown to influence the capacity to transfer the acquired CVS knowledge to novel situations. Older children can more easily transfer CVS knowledge to novel situations than younger children (Chen and Klahr 1999). Furthermore, prior knowledge is an important factor. Lorch et al. (2010) compared CVS performance of children from schools scoring high or low on the science section of an achievement level test following three types of CVS teaching methods (instruction, self-discovery, and instruction combined with self-discovery). They found that students from the combined group performed best, followed by the instruction and discovery group, respectively. Also, children from the high achieving schools showed greater learning gains on all three teaching methods than students from the low achieving schools.

Research on the role of individual child factors in CVS acquisition is scarce, especially when related to the instructional method used for teaching CVS. Siler et al. (2010) examined relations between child factors and CVS performance after instruction in fifth and sixth graders from low socio-economic backgrounds. They found that CVS posttest score correlated with CVS pretests performance and scores on several subtests of the CTB/ Terra Nova[2] (i.e., a standardized achievement test) namely measures of reading comprehension, science, non-verbal intelligence and verbal reasoning. Further exploration of these correlations led to the observation of two main predictors. First, CVS acquisition was predicted by reading comprehension skills ($r = .47$, $p = .03$). Siler et al. suggested that, since the instruction was presented orally and not in written form, this relationship might not be the result of a reading-related ability. Instead they proposed this finding could be to be the result of a more general comprehension skill underlying both reading comprehension and understanding complex strategies, such as CVS. The second factor with predictive value were the verbal deductive reasoning skills which predicted both immediate ($r = .58$, $p = .006$) and delayed CVS transfer abilities ($r = .49$, $p = .04$). Siler et al. argued that this latter finding was related to the ability to understand the logic behind CVS during the instruction phase. In understanding CVS, three reasoning steps are involved. First, the child first needs to identify the variable of interest (e.g., the steepness of a ramp). Second, the child needs to select the two values of that variable that should be compared (e.g., steep and less steep). And finally, the child needs to understand that all other variables (e.g., surface of the ramp and weight and start position of the ball) should be kept constant (Klahr and Dunbar 1988).

The study of Siler et al. (2010) does provide insights into the child factors related to the acquisition and transfer of CVS following instruction, but they only examined children after having received instruction. Child predictors of discovery-based learning of the CVS remain unexplored. Also the developmental perspective in child predictors of CVS acquisition is missing since the authors did not address age as a factor. In the present study, we therefore examined which child factors play a role in both direct instruction and self-discovery learning of CVS in two different age groups. We included all child factors that were found to correlate with CVS performance by Siler et al. (2010) as pre-intervention measures (i.e., reading comprehension, scientific knowledge, non-verbal reasoning, and verbal reasoning). Furthermore, vocabulary was added as a predictor measure as it has been found in previous studies that understanding key concepts in scientific phenomena highly depends on the words to write and talk about them (Snow 2010).

The present study had two main aims. First, by comparing instruction and discovery-based learning in two grade levels we examined if we could replicate the findings of prior studies (e.g., Chen and Klahr 1999). Our second aim was to examine to what extent CVS acquisition under different instructions could be predicted from child factors with a special focus on linguistic skills. The following research questions were answered:

(1) To what extent do CVS acquisition and transfer learning gains differ for children provided with instruction from children learning in a discovery-based manner and how are these findings related to differences in age?
(2) To what extent do child factors predict CVS learning gains in the two intervention groups?

With respect to the first research question, we expected that instruction produces higher learning gains than discovery-based learning, and that transfer learning gains would be higher for the older children than the younger children. With respect to the second research question, we expected that the following child factors, i.e. pretest measures, verbal reasoning, reading comprehension, and vocabulary measures, would predict CVS acquisition and transfer in the instruction group, and even more so in the discovery group, as children in this group have to rely more on their own skills to learn the concept of CVS. We did not expect the non-linguistic factors, i.e., scientific knowledge and non-verbal reasoning, to play a role in the instruction group since Siler et al. (2010) did not observe any relationships with these factors examining children receiving instruction. However, we hypothesized that these factors might be involved in learning in a discovery-based manner since these children may need to rely more on pre-existing scientific knowledge and reasoning skills.

Methods

Participants

Participants were 76 fourth graders and 43 sixth graders from a primary school in the Netherlands. Education in this school was based on the Jenaplan pedagogy[1] in which fourth, fifth, and sixth graders are joined in class. Children from both grades under study were randomly divided over intervention types resulting in 60 children in the instruction group (40 fourth graders, 20 sixth graders) and 59 children in the discovery group (36 fourth graders, 23 sixth graders). Following Strand-Cary and Klahr (2008), we excluded so-called 'natural experts' from the analyses. Natural experts were defined as children reaching the maximum score on one of the pretest CVS measures which resulted in the exclusion of 18 children, the remaining children scored on average 2.92 out of 8 points on the pretest (SD 1.99).

As a result, 38 fourth and 14 sixth grade participants were included in the instruction group and 35 fourth and 14 sixth graders were included in the discovery group. The average age of the fourth graders was 10;0 years (SD 0;6) and the average age of the sixth graders was 12;1 years (SD 0;5). Socio-economic status was defined in terms of the highest completed education level of the parent(s) which was categorized from 1 (primary education only), to 2 (middle secondary education), 3 (higher secondary education), and 4 (academic education) and was on average 2.68 (SD .94) for fourth graders and 2.54 (SD .92) for sixth graders. Intervention groups did not differ in age, $t(99) = .083$, $p = .934$, $d = 0$; socio-economic status, $t(99) = .216$, $p = .227$, $d = .24$, or gender $t(99) = .714$, $p = .477$, $d = .14$. Children were asked to participate in the study, and parents gave passive consent for participation of their child.

[1] The Jenaplan pedagogy is centered on the idea that the school is a community in which children learn, play and interact with not only their peers but also younger and older children, their teachers and parents. Therefore, children are grouped over different achievement levels, resembling a home-situation were older and younger siblings are learning and playing together (Petersen 1974).

Materials

Child factors

Nonverbal reasoning *Raven's Standard Progressive Matrices* (Raven 1976) was used to assess non-verbal reasoning skills. The task consists of five sets (A–E) of 12 puzzles, equaling 60 puzzles in total. The puzzles in set A consist of a patterned figure from which a part is missing. The child was asked to select the missing part by choosing one of six alternatives. In sections B–E, the puzzles consist of a pattern of four (B) to nine (C–E) figures of which the last figure was missing. The child was asked to select the missing figure by choosing from six (B) to eight (C–E) alternatives. Scoring was based on the number of correct answers. Reliability was assessed using Cronbach's alpha and was sufficient for this sample (.79)

Verbal reasoning Deductive verbal reasoning skills were assessed using a task based on a measure designed by Schröder et al. (2000). In this paper-and-pencil task the children were presented with 15 premises (e.g., *All children who get a bike from their parents feel cheerful)* and asked to answer one to four questions per premise up to a total of 36 questions (e.g., *Laura feels cheerful. Did she get a bike from her parents?*). Children were given three multiple choice options per question: *yes*, *no* or *maybe*. Scoring was based on the number of correct answers. Cronbach's alpha was calculated for the current sample and was acceptable (.63)

Vocabulary knowledge Vocabulary knowledge was assessed by means of the subtest *Zinsleeswoordenschat* [Sentence Reading Vocabulary] of the *Taaltoets Alle Kinderen* [Language Test for All Children] (Verhoeven and Vermeer 1993). The test consists of fifty sentences containing one target word which is underlined. Children were asked to select the correct meaning of the target word from four possible alternatives. Scoring was based on the number of correct answers. Cronbach's alpha of the current sample was good (.85).

Scientific knowledge Scientific knowledge was measured using the subtest *Wereldoriëntatie deel 1: Natuuronderwijs* [World orientation part 1: Nature Education] of the established and standardized *Eindtoets Basisonderwijs* [Final Test Primary Education] of the Dutch National Institute for Measurement in Education (CITO 2010). The test consists of 30 multiple choice questions, each question having four possible answers. Questions cover scientific domains such as biology, chemistry, physics and technology. Scores were based on the number of correct answers. The reliability of the test in the current sample was good (.72).

Reading comprehension Reading comprehension scores were received from the school which administers nationally standardized CITO reading comprehension tests each half year (for sixth graders: Feenstra et al. 2010; for eight graders: Weekers et al. 2011). The tests have two versions, for skilled and less skilled readers. For our analyses, we used the ability scores which are corrected for the version administered. Reliability values (MAcc) reported for these test are good, .84 for the sixth grade version (Feenstra et al. 2010) and .88 for the eighth grade version (Weekers et al. 2011).

CVS intervention

To allow for optimal comparison with prior studies the CVS intervention materials were based on the materials used by Klahr and colleagues in their "TED" (Training in Experimental Design) tutor (e.g., Siler et al. 2010, for an overview) and consisted of replica's of the materials for experimentation (i.e., wooden ramps) and translations of the direct instruction protocol, the CVS measure of acquisition (i.e., ramp test) and the paper-and-pencil measure of general CVS transfer (i.e., story test).

Materials for experimentation (ramps) During the intervention phase, two wooden ramps were used for experimentation. The ramps consist of a slope and a landing platform which is slightly steeped to prevent the balls from rolling of the ramp onto the floor. The slope is supported by an L-shaped object which can be turned so that either the longer or shorter part of the L supports the slope of the ramp, allowing the steepness of the slope to be changed from steep to less steep. Starting fences can be placed on the slope in two designated slots, at the top and middle of the slope, to control the starting position of the balls. Two surfaces of different texture can be placed on the slope and each ramp is accompanied by two balls which differ in weight. As a result, the ramps have four binary variables: steepness (steep or less steep), starting position (top or middle), surface (rough or smooth), and ball (heavy or light).

CVS worksheet The worksheet was used for the discovery group only and contained three elements. First, children we asked to write down their research question. Second, a table was presented containing three columns, one for the variables and the others for indicating how these variables were set-up two for each of the ramps. For example, for the variable 'steepness' children were directed to select 'steep' or 'less steep' for each of the two ramps. The third element was a space were children could write down their findings.

CVS acquisition (ramp test) CVS skills were measured by means of a hands-on CVS test (also referred to as the *ramp test* by Klahr and colleagues in e.g., Siler et al. 2010) in which children were asked to design unconfounded experiments in the context of the intervention materials, so by using the ramps. There are four pairs of questions, one pair for each of the variables (i.e., steepness, starting position, surface, ball). The first question of a pair focuses on the design of an unconfounded experiment for one of the four variables (e.g., *Can you design an experiment to test if starting position affects how far the ball rolls?*). The second question of a pair examines the rationale behind the designed set-up (i.e., *Why did you design the experiment like this?*). After each question, the child was given as much time as needed to complete the set-up of the ramps or provide an answer orally. There was no feedback given during the assessment.

Designs in which only the variable of interest was set-up differently for the two ramps were scored as correct, and designs containing differences in more than one variable were scored as incorrect. Answers to the second questions were only scored correct if the use of CVS was made explicit. One point was given per correct answer, resulting in a score range of 0–8 points. Cronbach's alpha indicated that reliability for this sample was good (.82).

CVS transfer (story test) Transfer of CSV knowledge was assessed by means of a paper-and-pencil CVS task (also referred to as the *story task* by Klahr and colleagues in e.g., Siler et al. 2010). Children were presented with three sets of questions which were formulated around three stories about selling lemonade, flying a model rocket, and baking cookies. Each set starts with a question asking children to design an experiment (i.e., *Can you design an experiment to test if the time of day affects the amount of lemonade that is sold?*). The second question asks them to explain their design. The third question in the set presents children with a predesigned experiment, and asks them to evaluate the experiment as either well or badly designed. Then, children are asked to explain their choice (question four) and redesign if they thought that was necessary (question five).

Answers to the second and fourth question were only scored correct if the use of CVS was made explicit. An example of an answer which was scored as correct is: *By only changing start position, I can find out if it's faster to start at the top or middle*. An example of an incorrect answer is: *Because this is the best way to measure*. For all questions, one point was given per correct answer, resulting in a score range of 0–15 points. Cronbach's alpha (.74) indicated reliability of this test was good.

Procedure

Prior to the start of the intervention, all children were assessed on the child factors in two separate 45 min classroom sessions. During the first session, measures of scientific knowledge, verbal syllogistic reasoning skills, and vocabulary were taken. In the second session, non-verbal reasoning skills and reading fluency were assessed. Reading comprehension scores were received from the school and were taken approximately 3 months prior to the experiment. There were maximally 2 weeks between the two pretest sessions. The intervention took place between several hours to 7 days after session two.

During the intervention, children were placed individually in a quiet room within the school. The intervention was done by one of six experimenters (all trained undergraduate students in Educational Science). Three were assigned to the Instruction intervention and three were assigned to the Discovery intervention. The experimenters were trained by the first author and followed a completely scripted protocol. The intervention phase had a pretest-intervention-posttest design.

During the *pretest phase*, children first received the CVS transfer measure to establish their prior knowledge of CVS. Next, all children were provided with an introduction on experimenting in which the experimenter explained that experimenting is about making comparisons and illustrated this with examples (e.g., 'we can examine which type of fertilizer makes plants grow better, fertilizer from brand a or b). After this, the ramp and its variables were introduced and the experimenter showed how the ramp could be set up and altered for all four of the variables (i.e., slope, surface, start position and weight of the ball). Children received the CVS acquisition measure (i.e., ramp test) after the explanation of the ramp.

During the *intervention phase*, children in the *discovery* group were encouraged to design and test their own experimental questions with the ramps. Children could design as much experiments as they wanted. Children were not explicitly instructed to discover CVS but were implicitly guided to do so by presenting them with the CVS worksheet on which they were asked to make notes of their experimental questions, provide the set-ups of their experiment in the table, and describe the outcomes of their experiments. The children did not receive feedback on the experiments they designed. To children in the *instruction* group, the concept of CVS was carefully explained by the experimenter and illustrated with examples of unconfounded and confounded set-ups using the ramp. The child was instructed to listen and observe carefully while the experimenter built the ramp set-ups. After the set-up was completed, the child was asked to evaluate the design and, in case of a confounded set-up, was asked what could be change to meet CVS requirements. Despite the child's ability to correct the set-up, the experimenter always provided the right answers and set-up in the end. At the end of the instruction, all observations were summarized and made explicit to the child. Note that in the instruction group, the children had no hands-on experience with the wooden ramps. Both the discovery and instruction intervention took approximately 20 min per child.

The intervention was ended with the posttest phase in which the child again received the CVS acquisition and transfer measure respectively.

Results

Grade differences and instruction effects

The first research question focused on examining which instruction type was most effective for CVS acquisition within each age group. Table 1 presents the mean scores on the CVS pre- and posttest per task, grade and intervention type.

The raw scores on the acquisition and transfer CVS measure were first transformed into Z-scores ($M = 0$; $SD = 1$) to allow for comparison between the two measures with different scaling. We then performed a GLM repeated measures analyses with CVS knowledge (Acquisition, Transfer) and Time (Pretest, Posttest) as within-subject factors, and Intervention (Instruction, Discovery) and Grade (Fourth, Sixth) as between-subject factors. The results showed main effects for Intervention ($F(1, 97) = 7.91$, $p = .006$, $\eta_p^2 = .08$) and Grade ($F(1, 97) = 27.03$, $p < .001$, $\eta_p^2 = .22$), as well as Time × Intervention ($F(1, 97) = 48.48$, $p < .001$, $\eta_p^2 = .33$) and Time × Task × Grade ($F(1, 97) = 7.09$, $p = .009$, $\eta_p^2 = .07$) interaction effects.

First, the interaction between Time and Intervention was further examined using *t*-tests for independent samples. The results indicated that the score on the combined posttests of the instruction group was significantly higher than the combined posttests score of the discovery group ($t(99) = 6.33$, $p < .001$, $d = 1.27$), while the two groups did not differ at

Table 1 Means scores, standard deviations and p-values for the CVS pre- and POSTTESTS and learning gains per grade and intervention group

Intervention	Fourth grade				Sixth grade			
	Instruction		Discovery		Instruction		Discovery	
	($n = 38$)		($n = 35$)		($n = 14$)		($n = 14$)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Acquisition (max. 8)								
Pretest	2.13	2.15	1.31	1.80	3.43	2.38	4.79	1.63
Posttest	6.21	2.80	2.94	2.27	7.79	0.58	5.36	1.74
Learning gain	4.08	2.81	1.63	1.78	4.36	2.17	0.57	1.91
Transfer (max. 15)								
Pretest	1.45	1.88	1.34	1.19	1.64	2.71	2.50	3.25
Posttest	6.24	4.77	2.17	2.11	9.64	3.75	5.93	4.98
Learning gain	4.79	4.70	0.83	2.01	8.00	3.88	3.43	3.27

Table 2 Mean scores, standard deviations and p-values for the child factors per grade and intervention group

Intervention	Fourth grade					Sixth grade				*p*
	Instruction (*n* = 38)		Discovery (*n* = 35)			Instruction (*n* = 14)		Discovery (*n* = 14)		
	M	*SD*	*M*	*SD*	*p*	*M*	*SD*	*M*	*SD*	
Non-verbal reasoning (max. 60)	40.92	4.79	41.2	3.86	.784	42.29	5.34	39.86	10.68	.456
Verbal reasoning (max. 36)	21.00	3.97	21.11	3.16	.892	23.93	5.26	23.93	5.70	1.000
Vocabulary (max. 50)	35.05	6.20	34.94	5.95	.939	39.00	9.78	33.36	9.62	.136
Scientific knowledge (max. 30)	15.76	4.69	15.66	4.24	.919	18.86	4.26	19.71	6.01	.667
Reading comprehension (max. 50[a]/55[b])	38.76	10.43	38.20	12.02	.832	50.71	14.96	49.57	14.83	.841

[a] Fourth grade

[b] Sixth grade

pretest ($t(99) = -.03$, $p = .980$, $d = 0$). Paired sample t-tests indicated that, although the instruction group outperformed the discovery group at posttest, both groups showed significant learning gains when pre- and posttests were compared, $t_{Instr}(51) = 4.05$, $p < .001$, $d = .59$; $t_{Disc}(48) = 6.52$, $p < .001$, $d = .65$. For the discovery group, we examined if this learning gain was related to the experiments designed in the discovery-phase. On average, the children designed 3.20 (SD 1.29) experiments of which 1.65 (SD 1.42) were experiments in which CVS was applied. The amount of CVS experiments and the score on the CVS posttest were marginally significantly correlated for the acquisition posttest (r = .281, p = .051) and significantly correlated for the transfer posttest (r = .443, p = .001), indicating that practice during the discovery phase was related to the posttest score.

Second, to further investigate the three-way interaction between Time, Task, and Group, learning gains were calculated by subtracting the pretest score from the posttest score for each of the CVS tasks per grade. Differences in learning gains were explored using independent samples *t*-tests which showed that learning gains of fourth and sixth graders did not differ on the CVS acquisition measure, $t(99) = .734$, $p = .465$, $d = .16$. Regarding transfer of CVS knowledge, we found that learning gains were larger for children from sixth grade than children from fourth grade, $t(99) = 3.02$, $p = .004$, $d = .68$.

Child Factors of CVS Acquisition

The second research question focused on the extent to which the CVS acquisition in both the instruction and discovery group could be predicted from child factors. Table 2 displays the mean scores of each child factor per grade and intervention group. Paired samples t-tests showed that within each grade level, the instruction and discovery group did not differ on the child factors.

To gain insight into the predictive value of these child factors, separate hierarchical regression analyses were undertaken for each of the two CVS measures per intervention

Table 3 Correlations Between CVS Posttest Measures and Child factors for the Instruction Group (below the diagonal) and the Discovery Group (above the diagonal)

	1	2	3	4	5	6	7	8	9
1. Non-verbal reasoning	–	.258	.371**	.418**	.358*	−.009	.269	−.096	.191
2. Verbal reasoning	.362**	–	.164	.370**	.530**	.354*	.371**	.401**	.553**
3. Vocabulary	.373**	.210	–	.534**	.384**	.149	−.019	.095	−.073
4. Scientific knowledge	.503**	.317*	.501**	–	.696**	.561**	.336*	.368**	.467**
5. Reading comprehension	.420**	.561**	.435**	.487**	–	.568**	.476**	.282	.639**
6. CVS acquisition pretest	.297*	.170	.320*	.421**	.475**	–	.331*	.690**	.522**
7. CVS transfer pretest	.074	.254	.091	.252	.481**	.207	–	.262	.672**
8. CVS acquisition posttest	.195	.080	.162	.268	.278*	.391**	.158	–	.452**
9. CVS transfer posttest	.337*	.215	.589**	.505**	.524**	.436**	.247	.444**	–

* $p < .05$, ** $p < .01$

group. To avoid multicollinearity issues related to interaction terms, all predictor values were centered prior to the regression analysis. Correlations for both groups on each of the predictor and criterion measures are presented in Table 3.

In Step 1 of the regression analyses, all child factors (i.e., non-verbal reasoning, scientific knowledge, syllogistic reasoning, vocabulary, reading fluency, reading comprehension) were entered, including a Group variable (Fourth grade = 0, Sixth grade = 1). In Step 2, the interaction terms between Group and the child factors were added to examine whether the predictive value of each of the child factors differed per grade level. Adjusted R^2 statistics were used to correct for the growing number of predictors in the model (Voeten and Van den Bercken 2003).

Outcomes of the regression analyses are presented in Table 4. The results in Step 1 showed differential effects of the various child factors in predicting CVS acquisition. For the instruction group, none of the child factors predicted the learning gain on the acquisition measure, but the learning gain on the transfer task was positively predicted by vocabulary ($p = .007$). For the discovery group, the learning gain on the CVS acquisition task was positively predicted by CVS pretest ($p < .001$) and verbal reasoning ($p = .012$). And the learning gain on the transfer task was positively predicted by CVS pretest ($p < .001$), verbal reasoning skills ($p = .038$), and reading comprehension ($p = .013$), and negatively predicted by vocabulary ($p = .013$). In Step 2, no additional interaction effects were found for both groups on both task types, indicating that the observed effects were of the same strength in grade four and grade six.

Discussion and conclusions

The aim of this study was (1) to examine the role of instruction and grade level in the acquisition of the control-of-variables strategy and (2) determine how the learning gains

Table 4 Hierarchical multiple regression analysis predicting CVS posttests from CVS pretests, child characteristics and grade level for the instruction and discovery group

Predictor	Instruction ($n = 52$)				Discovery ($n = 49$)			
	Acquisition CVS		Transfer CVS		Acquisition CVS		Transfer CVS	
	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β
Step 1	.205		.475***		.572***		.710***	
CVS pretest[a]		.287		.073		.727***		.356***
Non-verbal reasoning		.072		.032		−.123		−.029
Verbal reasoning		−.107		−.119		.318*		.211*
Vocabulary		−.052		.377**		.034		−.305**
Scientific knowledge		.078		.164		.133		.192
Reading comprehension		.078		.254		−.341*		.313*
Grade		.187		.106		−.054		.100
Step 2	.025		.040		.071		.036	
Grade × CVS pretest		−.897		−1.132		−2.122		− .479
Grade × non-verbal reasoning		−.024		−.029		.047		.067
Grade × verbal reasoning		.110		.069		.209		−.135
Grade × vocabulary		−.045		−.256		−.319		−.094
Grade × scientific knowledge		.040		.011		.046		−.295
Grade × reading comprehension		−.128		.265		.196		.098
Total R^2_{adj}	−.033		.349***		.511***		.653***	

[a] CVS pretest refers to the acquisition pretest for the columns displaying the outcomes of the regression analysis for the acquisition CVS posttest and to the transfer pretest for the columns showing the regression analysis values for the transfer CVS posttest

* $p < .05$. ** $p < .01$. *** $p < .001$

for each intervention group were related to child factors. Regarding the role of instruction and grade level, we found that overall learning gains were significantly higher in the instruction group than in the discovery group and that children from sixth grade were better at applying CVS in new situations than younger fourth graders after both types of interventions. Both observations are replications of findings in earlier studies on the role of instruction and age in CVS acquisition (e.g. Klahr and Nigam 2004; Strand-Cary and Klahr 2008; Lorch et al. 2010), presenting us with a reliable dataset to address the second and novel part of our research aim, i.e., examining the predictive value of child factors under different types of instruction.

The present study is the first to systematically explore the relationships between CVS acquisition under different levels of instruction and a range of child factors, i.e. non-verbal and verbal reasoning, vocabulary, scientific knowledge, and reading comprehension. Contrary to our expectations and observations of Siler et al. (2010), we did not observe any child factors predicting CVS acquisition in the instruction group and relations between CVS transfer abilities and child factors were limited to the vocabulary measure. The discrepancy between our findings and the observations of Siler et al. may lie in the difference in socio-economic background of the children. Siler et al. only examined children from a low SES group whereas children from the present study had a middle-high socio-economic status. Considering that low SES children are in general more challenged learners, it seems plausible that as a result pre-existing capacities were more influential as is also evidenced by Klahr and Li (2005), who found that instruction is less effective in low SES populations. The present outcomes are in line with the ample studies showing instruction to be a highly effective method for teaching the CVS in middle to high SES populations (e.g., Chen and Klahr 1999; Klahr and Nigam 2004; Strand-Cary and Klahr 2008), indicating that in these populations the instruction is effectively diminishing whatever differences may exist between children.

With regard to the discovery learning group, we—as expected- found child factors that predicted post-test scores, indicating that children who were able to learn CVS knowledge in a constructivist manner relied on pre-existing knowledge and skills. Outcomes showed that, in addition to CVS pretest, variance of both the acquisition and transfer measure of CVS in the discovery group was explained by reading comprehension and verbal reasoning skills. These are the same factors that were found predictive in the study of Siler et al. examining low SES children.

Regarding the verbal reasoning skills, the data seem to suggest that children can learn to discover CVS on their own if they have well-developed verbal reasoning skills. We examined this hypothesis by undertaking additional *t*-tests for independent samples splitting the group on the mean which made clear that children with higher verbal reasoning skills (M 25.33, SD 3.03) indeed performed higher on both the CVS acquisition measure ($p < .045$) as the CVS transfer measure ($p < .003$) than children with lower verbal reasoning skills (M 18.58, SD 2.35). These outcomes, showing that verbal reasoning skills are an important predictor for CVS learning, can be explained in light of the logic that underlies CVS understanding. Since CVS understanding involves a reasoning rationale (Klahr 2009) it was expected that reasoning skills would play an important role for children experimenting in a discovery-based manner.

The relations with reading comprehension showed an interesting discrepancy; reading comprehension was negatively related to CVS acquisition and positively to CVS transfer scores. Children with weaker reading comprehension had higher direct learning gains, whereas children with higher reading comprehension had higher transfer learning gains. The strong correlations between reading comprehension and pretest CVS measures may explain the results. Children with higher reading comprehension had less to learn with regard to CVS acquisition and more with regard to CVS transfer. The CVS transfer task, on the other hand, may have been a step too far for the children with lower reading comprehension.

Finally, transfer performance in the discovery group was also predicted by vocabulary but contrary to the observation in the instruction group these factors were negatively related which might be the result of suppression in the regression analyses since reading comprehension and vocabulary scores were highly interrelated (Nation 2009).

Interestingly, the child factors related to CVS learning were all linguistic in nature. We did not observe any relations with our measure of scientific knowledge or nonverbal reasoning. These findings, showing that CVS acquisition and transfer is related to linguistic capacities, provide more support for the hypothesis that language is highly important for science learning (e.g., Mercer et al. 1999; Snow 2010; Osborne 2002; Wellington and Osborne 2001).

Regarding (practical) implications, the present observations do contribute to the debate on the necessity of instruction for teaching higher order strategies. The data showed that learning gains for the instruction group were larger and less influenced by pre-existing child factors than learning gains of the discovery group supporting the theory that complex higher order strategies are best learned by means of instruction (e.g., Mayer 2004; Kirschner et al. 2006; Mayer 2009; Klahr and Nigam 2004; Alferi et al. 2011). However, children in the discovery group also showed a significant increase in CVS knowledge after experimenting which was predicted by prior CVS knowledge and verbal reasoning and reading comprehension. Since these factors were also found to be predictive of CVS learning gains in a group of low SES children (Siler et al. 2010) prior knowledge, verbal reasoning abilities and reading comprehension skills should be taken into consideration when working with more challenged children or when teaching CVS in a constructivist manner.

Though the observed relations showed that linguistic child factors are important predictors for CVS learning, the present study did not examine direct causal relationships. Future studies should make clear if developing linguistic capacities in general, and verbal reasoning skills specific, directly affect the acquisition of CVS by assessing if CVS performance increases after linguistic interventions. An important point to consider is including a measure of academic vocabulary (Snow 2010). The vocabulary measure used in the present study was a general and not a task- or science-specific measure. It would be interesting to examine if academic vocabulary provides a stronger predictor of CVS learning. Finally, the present study did not include a group of children who did not receive any intervention making it difficult to disentangle learning gains related to the intervention and taking the same test twice. Another possible solution to this problem would be to create multiple versions of the CVS acquisition and transfer test the order of which could be counterbalanced over participants.

To conclude, the present study examined the role of child factors in acquiring CVS. The data showed that instruction led to higher CVS learning gains than discovery based learning and transfer effects increased with age. Relations between child factors and learning gains were more prominent in the discovery-learning group than the instruction group and influential factors were, besides prior CVS knowledge, all linguistic in nature. Language capacities thus play an important role in scientific experimentation and need to be taken into account, especially when teaching systematic experimentation to children in a discovery-based manner.

Acknowledgments This research was funded by a Grant from the Dutch Curious Minds research program. The authors would like to thank David Klahr and his colleagues for kindly providing us with the CVS tests and sketches of the materials and Huub Hennissen for building the ramps. We thank Janneke van Druenen, Marly van Gemert, Syrthe Niessen, Helena Porteners, Femke Snelting, and Eline Vrenken for their help with collecting and digitizing the data.

References

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

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*, 1098–1120.

Cito [Dutch National Institute for Measurement in Education] .(2010). Wereldoriëntatie deel 1: Natuuronderwijs [World orientation part 1: Nature education]. Arnhem: Cito.

Croker, S., & Buchanan, H. (2011). Scientific reasoning in a real-world context: The effect of prior belief and outcome on children's hypothesis-testing strategies. *British Journal of Developmental Psychology, 29*, 409–424.

Dean, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education, 91*, 384–397.

Feenstra, H., Kleintjes, F., Kamphuis, F., & Krom, R. (2010). Leerling- en onderwijsvolgsysteem. Begrijpend lezen. Groep 3 t/m 6.[Student- and education observation system. Reading comprehension. Grade 3–6] Arnhem: Cito.

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.

Klahr, D. (2009). ''To every thing there is a season, and a time to every purpose under the heavens'': What about direct instruction? In S. Tobias & T. M. Duffy (Eds.), *Constructivist theory applied to instruction: Success or failure?* (pp. 291–310). New York: Taylor & Francis.

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

Klahr, D., & Li, J. (2005). Cognitive research and elementary science instruction: From the laboratory, to the classroom, and back. *Journal of Science Education and Technology, 14*, 217–238.

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

Lazonder, A. W., & Egberink, A. (2013). Children's acquisition and use of the control-of-variables strategy: effects of explicit and implicit instructional guidance. *Instructional Science*, 1–14.

Lorch, R. F., Lorch, E. P., Calderhead, W. J., Dunlap, E. E., Hodell, E. C., & Freer, B. D. (2010). Learning the control of variables strategy in higher and lower achieving classrooms: Contributions of explicit instruction and experimentation. *Journal of Educational Psychology, 102*, 90–101.

Matlen, B. J., & Klahr, D. (2013). Sequential effects of high and low instructional guidance on children's acquisition of experimentation skills: Is it all in the timing? *Instructional Science, 41*, 621–634.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59*, 14–19.

Mayer, R. E. (2009). Constructivism as a theory of learning versus constructivism as a prescription for instruction. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction: Success or failure* (pp. 184–200). New York: Routledge.

Mercer, N., Wegerif, R., & Dawes, L. (1999). Children's talk and the development of reasoning in the classroom. *British Educational Research Journal, 25*(1), 95–111.

Nation, K. (2009). Reading comprehension and vocabulary: What's the connection? In R. K. Wagner, C. Schatschneider, & C. Phythian-Sence (Eds.), *Beyond decoding: The behavioral and biological foundations of reading comprehension* (pp. 176–194). New York: Guilford Press.

Osborne, J. (2002). Science without literacy: A ship without a sail? *Cambridge Journal of Education, 32*(2), 203–218.

Petersen, P. (1974). *Der kleine Jena-Plan*. Weinheim: Beltz.

Raven, J. C. (1976). *Standard progressive matrices: Sets A, B, C, D & E*. Oxford: Oxford Psychologists Press.

Sao Pedro, M. A., Gobert, J. D., Heffernan, N. T., & Beck, J. E. (2009). Comparing pedagogical approaches for teaching the control of variables strategy. In N. A. Van Taatgen & H. Van Rijn (Eds.), *Proceedings of the 31st annual meeting of the cognitive science society* (pp. 1294–1299). Austin, TX: Cognitive Science Society.

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

Schröder, E., Bödeker, K., & Edelstein, W. (2000). *The development of syllogistic reasoning: A manual including measurement procedures and descriptive analyses*. Berlin: Max Planck Institute for Human Development.

Siler, S., Klahr, D., Magaro, C., Willows, K., & Mowery, D. (2010). Predictors of transfer of experimental design skills in elementary and middle school children. In *intelligent tutoring systems* (pp. 198–208). Berlin, Heidelberg: Springer.

Snow, C. E. (2010). Academic language and the challenge of reading. *Science, 328*(5977), 450–452.

Stohr-Hunt, P. M. (1996). An analysis of frequency of hands-on experience and science achievement. *Journal of research in Science Teaching, 33*(1), 101–109.

Strand-Cary, M., & Klahr, D. (2008). Developing elementary science skills: Instructional effectiveness and path independence. *Cognitive Development, 23*, 488–511.

Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction, 14*, 89–109.

Verhoeven, L., & Vermeer, A. (1993). *Taaltoets alle Kinderen. Handleiding* [Language test for all children. Manual]. Arnhem, The Netherlands: Cito.

Voeten, M.J.M., & Van den Bercken, J.H.L. (2003). *Lineaire regressieanalyse* [Linear regression analysis]. Groningen, The Netherlands: Stenfert-Kroese.

Weekers, A., Groenen, I., Kleintjes, F., & Feenstra. H. (2011). Cito Volgsysteem primair onderwijs LOVS. Begrijpend lezen Groep 7 en 8. [Cito Observation primary education LOVS. Reading comprehension Grade 5 and 6] Arnhem: Cito.

Wellington, J. J., & Osborne, J. (2001). *Language and literacy in science education*. Buckingham: Open University Press.

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review, 20*(1), 99–149.

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