Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computers in Human Behavior 21 (2005) 933–943
Computers in Human Behavior
www.elsevier.com/locate/comphumbeh

Studying inquiry learning with FILE

P. Wilhelm [a,*], J.J. Beishuizen [b], H. Van Rijn [c]

[a] *Department of Instructional Technology, Faculty of Behavioral Sciences, University of Twente, P.O.Box 217, 7500 AE Enschede, The Netherlands*
[b] *Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*
[c] *Department of Social Science Informatics, University of Amsterdam, Amsterdam, The Netherlands*

Available online 9 April 2004

Abstract

The FILE program (flexible inquiry learning environment) is a research tool, which allows researchers in inquiry learning to design and administer learning tasks in which task domain and task model (i.e. the relations between input and output variables) can be adjusted independently, while other factors (e.g. interface) are held constant. Its monitoring facilities allow for on-line measurement of learning behavior. This paper offers one example of the possibilities FILE has to offer to researchers. Data are presented which illustrate the sensitivity of FILE to age differences in inquiry learning outcome and processes. Instructional applications of FILE are also discussed.

1. Introduction

The FILE program (flexible inquiry learning environment) is a research tool, which allows researchers in inquiry learning to design and administer learning tasks in which task domain and task model (i.e. the relations between input and output variables) can be configured independently, while other factors (e.g. interface) are held constant. Also, its monitoring facilities allow for on-line measurement of learning behavior. On a general level, Hulshof, P, Beishuizen, and Van Rijn (2002) discussed the principles of FILE. They concluded that FILE is a flexible research

* Corresponding author.
E-mail address: p.wilhelm@gw.utwente.nl (P. Wilhelm).

0747-5632/$ - see front matter
doi:10.1016/j.chb.2004.03.013

tool, which can be used to set up studies in inquiry learning for different age groups in an efficient way, saving expensive programming time. This paper elaborates on the way data gathered with FILE can be analyzed. For this purpose, we will give an example of the application of FILE in research. Empirical data are presented with regard to age differences in inquiry learning outcome and learning processes in two tasks from the domain of biology. It is our conviction that in doing so the value of FILE as a research tool is illustrated best. The empirical findings are part of a larger study, which included more and larger age groups (Wilhelm, 2001).

The development of inquiry learning has been described both as an improvement in general, domain-independent strategies, and as a side effect of a growing knowledge base. On the one hand, there are researchers (e.g. Carey, 1985; Keil, 1981) who emphasize that, differences between children and adults in inquiry learning are, for the most part, due to differences in domain-specific knowledge. The other position, which is held by, for example Case (1992) and Kuhn (1989), stresses that there are qualitative differences between children and adults in inquiry learning. For example, Kuhn (1989) stated that children have difficulty discriminating theory and evidence. They tend to think with theories instead of thinking about them and tend to be merely influenced by evidence instead of thinking about it. Klahr (2000) takes an intermediate position. He stated that both changes in domain-specific knowledge and domain-independent skills are responsible for the age differences found in inquiry learning. Adult superiority stems from a set of domain-independent skills dealing with the coordination between the search in the hypothesis and the experiment space (Klahr & Dunbar, 1988; Hulshof et al., 2002). On the other hand, the plausibility of a particular hypothesis influences the search in both the hypothesis space and the experiment space. This plausibility builds on domain-specific knowledge of which the acquisition is in principle independent from age. In this paper, we focused on domain-independent inquiry skills. FILE was used to confirm the hypothesis that children and adults show clear differences in learning measures pertaining to the coordination between the search in the hypothesis and the experiment space. With respect to the search in the hypothesis space, the experiment space and other measures of learning behavior we gathered (planning activities and data management activities), we had no specific prior hypotheses.

This paper is organized according to the basic applications of FILE (design, administration, analysis). First, we will explain the design of the learning tasks used in the study. Secondly, we will explain how FILE was administered and finally we will discuss the results of the study.

2. Design of the tasks

Two tasks in the domain of biology were designed, the Plant Growing and the Food task. In Fig. 1, the interface of the Plant Growing task is depicted.

The problem posed to the learners in this task was to find out how different input variables affected the maximum height a plant could reach (output variable). The input variables were: (1) giving water, either once or twice a week; (2) usage of an

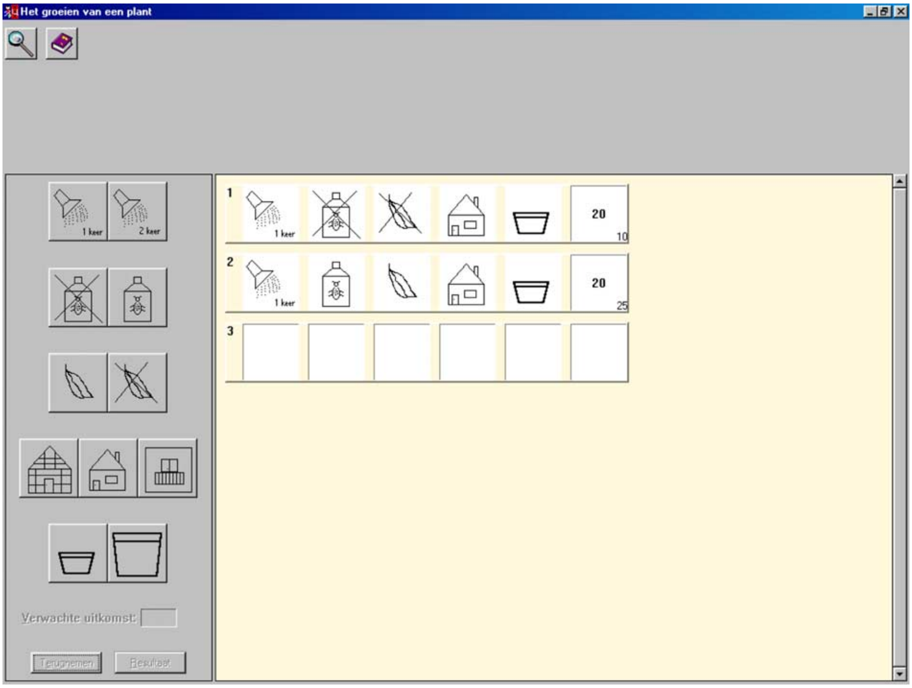

Fig. 1. Interface of the plant growing task.

insecticide to keep away plant louses or not; (3) putting dead plant leaves in the flower pot or not; (4) placing of the plant, either indoors, on a balcony or in a greenhouse and (5) size of the flower pot, either big or small. The levels of the output variable (height of the plant) were: 5, 10, 15, 20 and 25 CMS. In the Food task learners had to find out how usage of different food articles affected the health status of an imaginary person, called Hans. The articles were: (1) carbohydrates, either cornbread, potatoes and pasta or white bread, potatoes and rice; (2) alcohol, a glass of wine a day or not; (3) vitamins, either taking a vitamin supplement or not; (4) snacks, either fried cakes, ice cream or pastry and (5) fat; either steak, filet of chicken and low-fat milk or chops, drumsticks and normal milk. The health status was a five-point scale ranging from 1 (very unhealthy) to 5 (very healthy). The underlying model in both tasks was identical. Input variable 1 and 5 in each task interacted. This interaction was disordinal in nature, meaning that the effect of a particular input variable was reversed under the influence of another. For example, in the Plant Growing task, giving water once a week and using a small flowerpot made the plant grow bigger than giving water once a week in a big flowerpot. This effect was reversed when the plant was given water twice a week. This effect can be explained by the fact that giving to much water in a small flowerpot causes a plant to die from too much water. Input variable 4 had a main effect, one of its three levels resulted in a different effect than the other two levels, which had no effect. For

Table 1
Model plant growing task

	Small flower pot, water once/wk	Small flower pot, water twice/wk	Big flower pot, water once/wk	Big flower pot, water twice/wk
Inside house	20	5	10	10
Greenhouse	25	10	15	15
Balcony	25	10	15	15

Note. once/wk: once per week, twice/wk: twice per week. Usage of an insecticide and putting dead leaves in the flowerpot represent irrelevant input variables and are not included in this scheme. Figures represent CMS.

Table 2
Model Food task

	Carbohydrate+Fat−	Carbohydrate+Fat+	Carbohydrate−Fat−	Carbohydrate−Fat+
Pastry	4	1	2	2
Fried cakes	5	2	3	3
Ice cream	5	2	3	3

Note. Carbohydrate+: corn bead, potatoes and pasta. Carbohydrate−: white bread, potatoes and rice. Fat−: steak, filet of chicken and low-fat milk. Fat+: chops, drumsticks and normal milk. Drinking a glass or wine and taking vitamin pills represent irrelevant input variables and are not included in this scheme. Figures represent the level of the health status ("1" means: very unhealthy, "5" means: very healthy).

example, growing a plant in a house had a negative effect on growth, while growing it on a balcony or in a greenhouse made no difference. Input variable 2 and 3 were irrelevant. For example, in the Food task, drinking a glass of wine with a meal or taking vitamin pills had no effect on health status. The content of the tasks were designed to be equally familiar to all learners and care was taken to ensure that the discovery of the underlying model would be challenging enough for both age groups. Tables 1 and 2 show the model of the Plant Growing task and the Food task.

3. Task administration

Learners performed the learning tasks on a standard PC running Microsoft® Windows 98. Learners conducted experiments by clicking on the pictures representing the levels of the input variables, entering a prediction, and clicking on the "Result" button (see Fig. 1). To promote on-task behavior, a simple and relatively straightforward FILE configuration was chosen. Learners could choose from a limited set of pre-specified prediction values and could temporarily rearrange experiments by means of a selection function. The maximum number of experiments shown on the screen was four. In this way, the information on the screen was limited, but would still allow for the disordinal interaction effect in both tasks to be visualized on eyesight (for this, four experiments are needed). FILE offers various other options to learners and experimenters, but they were not used for this study (for more details about these configurations, see Hulshof et al., 2002). Two groups of learners (sixth-graders and university students, $n = 10$) performed the Plant Growing task and the Food task successively in one individual test session. All learners received the same five-minute instruction on the purpose of the task and how to use the interface. The introductory text to which the learners had access in the learning environment mainly contained information about the nature of the input variables and the output variable. The learners were instructed to think aloud while performing the task. A recording of someone thinking aloud was played as instruction. They were also told that they would be asked questions during task performance. These questions were asked before and after each experiment conducted and pertained to their research plans (''What are you going to find out?''), predictions and hypotheses (''What do you think the outcome of this experiment will be? Why do you think that?''), and inferences (''What did you find out?''). At the end of each task session the learners were interviewed about the effects they had found (''theory interview''). In this interview the learners were questioned about the effects of the input variables on the output variable. For each variable they were asked: ''What difference do you think {input variable} makes?'' A maximum of 35 min was allowed to work on each task. However, most learners finished the task before this time limit was reached (for more details about this procedure, see Wilhelm, 2001). Verbal protocols were recorded on cassette tape (the digital sound recording facility was not available yet at the moment the data were gathered).

4. Data analysis

Verbal protocols and log file data were used to obtain measures of the search in the hypothesis space and the experiment space, the coordination between the search in these spaces (Klahr & Dunbar, 1988), planning activities and data management activities. The answers to the questions posed during and after experimenting were scored by two independent judges, using a standard protocol (see Wilhelm, 2001). In Table 3, measures of the search in the hypothesis space and the experiment space, the coordination between the search in these spaces, planning activities and data-management activities are depicted.

The measure of the search in the hypothesis space was the total number of hypotheses stated. A distinction was made between simple hypotheses, meaning hypotheses with regard to one variable (e.g. ''I think this outcome will be better, because giving water once or twice a week makes a difference'') and complex ones, meaning hypotheses with regard to interactions between two variables (e.g. ''I think this outcome will be worse because giving too much water in a little flower pot is bad for the plant''). More complex hypotheses, for example, hypotheses with regard to three-way interactions, were very infrequent and were ignored. Measures of the search in the experiment space were: percentage of experiment space covered, total number of (unique) experiments, number of experiments duplicated and the number of variables changed per experiment. Percentage of the experiment space covered was the number of unique experiments conducted divided by 48, the maximum number of unique experiments possible in each task. Number of variables changed per experiment is an indication of the extent to which the CVS (control-of-variables

Table 3
Source of measures of inquiry learning processes

	Verbal protocols	Log files
Search hypothesis space		
Total number of hypotheses	*	
Simple hypotheses (one variable)	*	
Complex hypotheses (interactions)	*	
Search experiment space		
Percentage experiment space covered		*
Unique experiments		*
Experiments duplicated		*
Number of variables changed per experiment		*
Number of learners who generated all possible outcomes		*
Coordination hypothesis and experiment space		
Theory interview score	*	
Prediction error		*
Percentage of valid inferences	*	
Number of learners with evidence of the presence of interaction effect in their dataset		*
Number of learners who noticed the presence of evidence for interaction effect in their dataset	*	
Planning activities		
Frequency of plans	*	
Check effect variable	*	
Check interaction	*	
Generate specific outcome	*	
Check effect of variable in different experiment	*	
Predict outcome	*	
Data-management activities		
Scrolling activities		*
Selecting activities		*

Note. Interraterreliability for think aloud data was calculated using Cohen's κ; search hypothesis space: .65, comprehension score: .92, inferences: .70, notice presence interaction effect: .60, planning activities: .70.

strategy; Chen & Klahr, 1999) was used. When a learner uses this strategy he or she changes only one variable at a time, allowing for valid inferences about the effect of the input variables on the output variable. The mean number of variables changed from one experiment to the next was taken as an indicator of the use of this strategy. In each task, five unique outcomes were possible. Since some experiments are more informative than others because their outcomes are less frequent, we recorded the number of unique outcomes generated. Measures of the coordination between the search in the hypothesis space and experiment space were: theory interview scores, prediction error, percentage of valid inferences and measures with respect to the discovery of the interaction effect. The theory interviews were scored in the following way. For each input variable, the learners were asked: "What difference do you think {input variable} makes?" All statements of the learners were compared with nine correct statements, covering all the effects of the input variables on the output variable in the tasks. A score of two points was given if a statement of a learner matched a correct statement, resulting in a maximum score of 18 points (for more details, see Wilhelm, 2001). Prediction error (Kuhn, Garcia-Mila, Zohar, & Anderson, 1995) was calculated by taking the mean of the standardized differences between learner's predictions and the actual outcomes of their experiments. A prediction error rate near to zero indicates a high level of understanding of the effect of the input variables. The answer to the question: "What did you find out?" posed after each experiment conducted could result in an inference made about the effect of an input variable (e.g. "Giving water once or twice a week does not make a difference"). This inference was denoted as valid if the experiments conducted up to that moment contained the correct evidence for that inference to be made. If not, then the inference was denoted as invalid. We also focused on the discovery of the interaction effect in each task. For example, in the Plant Growing task, it was possible that the data set of a learner contained two experiments in which the effect of flowerpot size was tested. This results in a bigger plant when a small flowerpot is used and water is given once a week. When the set of experiments of a learner also contained an experiment in which a small flowerpot was used and water was given twice a week, a learner should notice that a small flowerpot not always results in a bigger plant. The log files were used to check if the set of experiments conducted by a learner contained this type of evidence. Verbal protocols were used to check whether a learner had noticed this evidence (e.g. "There is something wrong with the flowerpots, sometimes a small pot is better and sometimes a big flower pot").

Measures of planning activities were collected using verbal protocols. The answers to the question: "What are you going to find out?" was scored in one of the following categories: (1) check the effect of one variable (e.g. "I am going to find out if giving more water makes a difference"); (2) check interactions between two variables (e.g. "I am going to find out what giving water has to do with the size of the flower pot"); (3) generate a specific outcome (e.g. "I am going to make the biggest plant"); (4) check the effect of a variable in a different experiment (e.g. "I am going to find out if the effect of using an insecticide is the same in this situation"); (5) predict an outcome (e.g. "I am going to see whether I can tell what height the plant will reach in this situation") and (6) unclear (e.g. "I don't know" or "See what happens"). These categories were assessed ad hoc, but their occurrence could be mapped unto planning behavior found in literature (Wilhelm, 2001). Measures of data-management activities were: frequency of scrolling and selection activities. Usage of these functions is indicative of an active attempt to compare experiments, indicating that a learner attempts to coordinate the search in the hypothesis and the experiment space.

5. Results

In Table 4, measures of the search within the hypothesis space and the experiment space are presented. ANOVA's were used to test differences between the two groups of learners. In both tasks no significant differences were found in the number of

Table 4
Measures of inquiry learning processes

	Sixth grade children		University students	
	Plant growing	Food	Plant growing	Food
Search hypothesis space				
Total hypotheses	2.2 (3.9)	2.0 (3.4)	4.6 (4.2)	6.5 (5.4)*
Simple hypotheses	2.2 (3.9)	2.0 (3.4)	3.5 (2.8)	5.3 (4.1)
Complex hypotheses	0.0 (0.0)	0.0 (0.0)	1.1 (2.0)	1.2 (3.5)
Search experiment space				
Percentage experiment space covered	24.6%	21.9%	25.2%	28.1%
Unique experiments	11.8 (4.2)	10.5 (3.1)	12.1 (2.3)	13.5 (4.1)
Experiments duplicated	3.2 (3.2)	1.9 (1.9)	2.7 (2.1)	2.0 (1.8)
Variables changed per experiment	2.6 (.30)*	2.3 (.46)*	1.4 (.32)	1.5 (.21)
Number of learners who generated all possible outcomes	5	4	2	3

Note. Both groups: $n = 10$. Percentage of experiment space: percentage of total number of experiments possible in each task (48).
* $p < .05$.

simple hypotheses and complex hypotheses stated, although the children did not state any complex hypothesis. In the Food task, the students stated more hypotheses altogether than the children did ($F(1, 18) = 5.0$, $p < .05$).

In both tasks, the children varied significantly more variables per experiment than the students did ($F(1, 18) = 69.6$, $p < .001$, for the Plant Growing task and $F(1, 18) = 23.6$, $p < .001$, for the Food task). Also, the children generated more of the five possible outcomes of the experiments than the students did. No significant differences were found between the children and the students in the percentage of the experiment space covered, the number of unique experiments and the number of experiments duplicated.

In Table 5, measures of the coordination between the search in the hypothesis space and the experiment space are depicted. In both tasks the children had significantly lower theory interview scores than the students ($F(1, 18) = 20.0$, $p < .001$,

Table 5
Coordination between the search in the hypothesis space and the experiment space

	Sixth grade children		University students	
	Plant growing	Food	Plant growing	Food
Comprehension score	5.6 (3.4)	8.1 (4.4)	12.7 (3.7)*	13.3 (3.5)*
Prediction error	1.3 (.31)	1.1 (.24)	1.1 (.30)	83 (.29)
Percentage of valid inferences	42%	58%	88%*	95%*
Evidence for interaction effect	5	5	8	8
Notice presence of interaction	0	1	5	7

Note. Both groups: $n = 10$. Maximum comprehension score = 18. Prediction error = mean standardized difference between predicted and actual outcome of experiment.
* $p < .05$.

Table 6
Measures of planning and data management activities

	Sixth grade children		University students	
	Plant growing	Food	Plant growing	Food
Planning activities				
Frequency of plans	5.8 (5.1)	6.6 (5.7)	11.3 (3.9)*	12.8 (3.9)*
Check effect variable	3.7 (3.5)	4.0 (3.9)	6.9 (1.2)*	6.6 (.97)
Check interaction	0.0 (0.0)	.10 (.32)	1.6 (2.9)	1.8 (3.7)
Generate specific outcome	1.8 (3.3)	2.3 (2.1)*	1.2 (1.8)	.70 (.95)
Check effect of one variable in different experiment	0.0 (0.0)	.20 (.63)	1.4 (1.8)*	3.3 (2.4)*
Predict the outcome	.30 (.95)	0.0 (0.0)	.20 (.42)	.40 (.52)*
Data management activities				
Scrolling activities	2.3 (1.8)	1.1 (.99)	4.7 (2.9)*	5.2 (3.7)*
Selecting activities	0.0 (0.0)	0.0 (0.0)	4.6 (6.4)	5.6 (7.0)

Note. Both groups: $n = 10$.
* $p < .05$.

for the Plant Growing task, $F(1, 18) = 8.4$, $p < .05$, for the Food task). Prediction error did not differ, but the students had a higher percentage of valid inferences in both tasks ($F(1, 18) = 23.4$ $p < .001$, for the Plant Growing task and $F(1, 18) = 15.5$, $p < .005$, for the Food task). The data sets of the children less frequently contained enough information to infer the presence of the interaction effect. However, when it was present the children tended to overlook this information more often than the students did.

In Table 6, results with respect to the different research plans stated by the learners and data-management activities are depicted. The students more often had a specific plan ($F(1, 18) = 7.2$, $p < .05$, for the Plant Growing task, $F(1, 18) = 8.0$, $p < .05$, for the Food task). In the Plant Growing task, students more often checked the effect of one variable ($F(1, 18) = 7.5$, $p < .05$) and more often checked the effect of a variable in a different experiment ($F(1, 18) = 6.2$, $p < .05$) than the children. In the Food task, students more often stated checking the effect of a variable in a different experiment ($F(1, 18) = 15.5$, $p < .05$) and predicting the outcome as research plans than the children ($F(1, 18) = 6.0$, $p < .05$). In the Food task, the children more often stated generating a specific outcome as a plan ($F(1, 18) = 4.8$, $p < .05$). None of the children used the selection function (although we observed children using it very proficiently). In both the Plant Growing task and the Food task the students more often used the scrolling function than the children ($F(1, 18) = 4.9$, $p < .05$, $F(1, 18) = 11.6$, $p < .005$).

6. Conclusions

Klahr's position (2000) that the difference between children and adults lays in the coordination between the search in the hypothesis space and the experiment space was confirmed. The theory interview scores and the percentage of valid inferences made, show that the children are less able than students to translate their data into valid statements about the effects of the input variables on the output variable. In addition, the children tended to overlook findings indicative of the presence of an interaction effect in their data sets more often than the students did. Another finding is that students more often made use of the data-management facilities in FILE (scrolling and selecting). Usage of these functions indicates that an active attempt is made to compare experiments, which is necessary for making inferences about the effect of the input variables. Except for evidence revealing the presence of the interaction effect we did not estimate the number of valid inferences the learners could theoretically make on the basis of their data sets. However, we expect the children and the adults in this study not to differ in this respect. Both groups conducted equal numbers of unique experiments and covered equal percentages of the experiment space. Thus, the evidence seems to suggest that children differ from adults in their ability to coordinate the search in the hypothesis space and the experiment space.

An explanation for why children more often overlooked evidence for the presence of an interaction effect might be that children did not presuppose interaction effects, so that contradicting findings posed no problems to them. They also may have overlooked the evidence because they manipulated more variables per experiment, which inevitably leads to data sets that are more difficult to inspect. A different task perspective might also be due to the differences we found. The fact that children generated more of the possible outcomes and the fact that they more often stated the plan of generating a specific outcome might be indicative of an ''engineering'' approach, more than a ''science'' approach (Schauble, Klopfer, & Raghavan, 1991). Students, in turn, more often had a specific plan aimed at testing or checking the effect of specific input variables. Generating a specific effect is an activity that can be situated in the process of the search in the experiment space for which coordination between the searches in both spaces is unnecessary. Thus, there may be other factors that also play a role when age-differences are concerned.

In Hulshof et al. (2002) the use of FILE in designing, administering and analyzing learning tasks was discussed on a general level. In this paper, the data presented illustrate how FILE can be used as a research tool and how data gathered with FILE can be analyzed. Of course, this is only one example. Due to its flexibility, many other applications are conceivable. Researchers can easily alter features of the system to adapt it to their own research purposes, saving expensive programming time. Different models describing the relations between the input and the output variables can be implemented which makes it possible to design tasks of varying complexity. Moreover, the domain of the tasks can be altered independently from the model, which allows studies of the influence of task domain on inquiry learning. Children from grade 4 (8–9 years) are easily made familiar with FILE and have been observed to enjoy working with it, so comparative studies with various age groups are possible, as illustrated in this paper.

The instructional applications of FILE are as yet unexplored. However, it can be imagined that FILE can be used to teach children about the principles of inquiry learning. Due to its transparency (experiments are accessible on eyesight and the levels of the input variables are visually represented), the use of the CVS and the principles of valid inference on the basis of co-variation can be visualized. The program can be used for various age groups throughout the curriculum and the possibility to administer FILE in a group makes classroom use come into reach. In addition, it is easy to learn to work with FILE and teachers can easily adapt FILE to their own wishes. We invite teachers to include FILE in their science classes.

Acknowledgements

The development of FILE was part of the research project "Inductive Learning" supported by a grant from the Dutch National Science Foundation (NWO 575–22.002). All three authors took part in this project. FILE was designed in close collaboration with Jan Wielemaker (Department of Social Science Informatics, University of Amsterdam, the Netherlands) to whom the authors like to express their gratitude. We also like to thank Frans Prins, Marcel Veenman, Ton de Jong, Wouter van Joolingen, Maarten van Someren, Bert Bredeweg and Bob Wielinga for their useful comments. The empirical data presented in this paper stem from the aforementioned research project and are presented in more detail in a dissertation written by the first author.

References

Carey, S. (1985). Are children fundamentally different thinkers than adults?. In S. Chipman, J. Segal, & R. Glaser (Eds.), *Thinking and learning skills* (Vol. 2, pp. 485–517). Hillsdale, NJ: Erlbaum.

Case, R. (1992). *The mind's staircase. Exploring the conceptual underpinnings of children's thought and knowledge*. Hillsdale, NJ: Erlbaum.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*(5), 1098–1120.

Hulshof, C. D., Wilhelm, P., Beishuizen, J. J., & Van Rijn, H. (2002). FILE: A tool for studying inquiry learning. *Computers in Human Behavior.*, doi:10.1016/j.chb.2004.03.014.

Keil, F. C. (1981). Constraints on knowledge and cognitive development. *Psychological Review, 88*, 197–227.

Klahr, D. (2000). *Exploring science. The cognition and development of discovery processes*. Cambridge, MA: The MIT Press.

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

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*(4), 674–689.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Anderson, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development, 60*(4, Serial No. 245).

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

Wilhelm, P. (2001). Knowledge, skills and strategies in self-directed inductive learning. Doctoral dissertation, Leiden University.