Learning and Individual Differences 47 (2016) 145–155

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Ability-related differences in performance of an inquiry task: The added value of prompts

Alieke M. van Dijk *, Tessa H.S. Eysink, Ton de Jong

Department of Instructional Technology, University of Twente, the Netherlands

ARTICLE INFO

Article history:
Received 20 August 2014
Received in revised form 20 January 2016
Accepted 24 January 2016
Available online xxxx

Keywords:
Inquiry learning
Differentiation
Ability level
Instructional support

ABSTRACT

This study investigated how children of different ability levels approached inquiry tasks, whether prompting improved their inquiry process, and whether their inquiry process led to domain knowledge gain. Fifth and sixth graders ($n = 478$) of three different ability levels worked individually with a simulation, either with or without included prompts. Prompts appeared to affect children's inquiry process at all three ability levels. This inquiry process, in turn, was related to their learning outcomes. High ability children, who engaged in more active and effective inquiry than children of lower ability, used the prompts when available. Average and low ability children rarely used the prompts. High and average ability children gained knowledge from pretest to posttest but not from posttest to retention test; low ability children only gained knowledge from posttest to retention test. The results of this study point to a need to find effective ways to support low and average children in inquiry.

1. Introduction

In modern-day elementary education emphasis is increasingly placed on teaching 21st century skills. In the context of the upcoming knowledge society, children should no longer be educated to become passive knowledge-consumers but should actively discover and integrate new knowledge. A well-known instructional approach that enables children to actively gather and process new knowledge is the inquiry method (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011). Even though research has been done to sketch an optimal inquiry process, little is known about differences in inquiry approaches and the effects of inquiry-specific support for children of different ability levels. Optimizing the inquiry learning method for children of different ability levels requires more insight into these differences. The aim of this study was to explore the inquiry approaches of children of different ability levels, whether and how children integrated support that was offered to them into their learning process, and whether children's inquiry approaches affected their learning outcomes and motivation.

1.1. Inquiry learning

Recent studies have shown that inquiry learning, if well-designed, can lead to better results than learning by more direct forms of instruction (see, for example, Eysink & de Jong, 2012; Furtak, Seidel, Iverson, & Briggs, 2012; Smetana & Bell, 2012). This applies to a variety of domains, but inquiry is especially effective for learning in science domains (Arnold, Kremer, & Mayer, 2014). These benefits can mainly be attributed to the fact that in inquiry learning students are expected to actively collect information, process information, and construct knowledge (Alfieri et al., 2011; Mayer, 2003, 2004; Minner, Levy, & Century, 2010). This active engagement in the learning process enhances students' development of knowledge and skills (Manlove, Lazonder, & de Jong, 2006).

When engaging in inquiry, students are expected to learn actively by completing a set of different activities (de Jong, 2006; de Jong & van Joolingen, 1998). The inquiry process often starts with orientation to the domain, which leads to generation of hypotheses concerning the domain. To test the hypotheses, experiments are designed and conducted, after which conclusions are drawn from the experimental outcomes. As a wrap-up activity, the inquiry outcomes and procedure are evaluated (Pedaste et al., 2015). Students are often given considerable freedom in working through these different activities (Mayer, 2004), allowing them to determine their own learning process and learning pace (Minner et al., 2010). The downside of this freedom is that students then experience difficulties with inquiry learning (Mayer, 2004). This is why it has repeatedly been stated that inquiry learning is only effective when it is adequately guided (Alfieri et al., 2011; d'Angelo et al., 2014; Mayer, 2004).

Difficulties students experience with carrying out the different inquiry activities and how to support them in these activities have been addressed in a considerable body of research (e.g., Alfieri et al., 2011; de Jong, 2006; de Jong & van Joolingen, 1998; Rutten, van Joolingen, & van der Veen, 2012). Students are often unsuccessful in generating hypotheses (Gijlers & de Jong, 2009; Njoo & de Jong, 1993), experience difficulties with conducting experiments that go beyond their initial understanding of the variables within a domain (Klahr & Dunbar, 1993),

* Corresponding author at: University of Twente, Faculty of Behavioral Sciences, Department of Instructional Technology, PO Box 217, 7500 AE Enschede, the Netherlands.
E-mail address: a.m.vandijk@utwente.nl (A.M. van Dijk).

http://dx.doi.org/10.1016/j.lindif.2016.01.008
1041-6080/

and find it difficult to draw the right conclusions from the collected data (Klahr & Dunbar, 1988). For younger children, problems with inquiry learning activities are often attributable to difficulties they experience with identifying relevant variables within an inquiry task (Zimmerman, 2007). Identification of these variables is a prerequisite for conducting the right set of experiments to answer a research question, or even for formulating an appropriate research question in the first place.

1.2. Differences between ability levels

Students of different ability levels are expected to differ in how they approach an inquiry task, because they vary in how skillful they are at relating new information to their existing knowledge and determining its relevance and meaning (Wang, Kinzie, McGuire, & Pan, 2010). More specifically, within the context of inquiry learning students are expected to vary in skillfulness at drawing accurate conclusions from experimentation and integrating this knowledge into their existing knowledge schemas (Zimmerman, 2007).

In general, it is assumed that high ability students are skilled at independently figuring out how to solve a problem or complete a task (Diezmann & Watters, 1997; vanTassel-Baska, 2003). They prefer a challenging learning process (Phillips & Lindsay, 2006; Reis & Renzulli, 2010), and favor learning tasks that involve complexity and the possibility of engaging in open-ended discovery (Diezmann & Watters, 1997; vanTassel-Baska, 2003). Challenging and complex tasks align with high ability students' advanced knowledge schemas (Kalyuga, Ayres, Chandler, & Sweller, 2003). In fact, providing high ability students with tasks and support that are too explicit could even be counterproductive. In contrast, low ability students tend to experience more difficulties with navigating through learning tasks than high ability students (Alexander & Schwanenflugel, 1996; Margolis & McCabe, 2003), and most studies conclude that low ability students need more structured tasks to engage in successful learning (e.g., Lou et al., 1996; Wang et al., 2010). Therefore, positive effects of instruction and support seen for children with lower levels of ability might disappear for children with higher levels of ability, as the level of redundancy of the instructional materials might be too high (cf., expertise reversal effect; Kalyuga, 2007; Kalyuga et al., 2003).

Advocates of differentiated instruction maintain that instruction and support should match children's ability level and their specific learning needs (e.g., Tomlinson, 2000; Vygotsky, 1986; Weinert & Helmke, 1998). This should then lead to an efficient learning process with effective learning outcomes (i.e., aptitude-treatment interaction; Cronbach & Snow, 1977), and lead to high motivational levels when working on learning tasks (Lens & Rand, 2000; Margolis & McCabe, 2003). Within the context of inquiry learning, this means that average and low ability students, in particular, need support to engage in effective inquiry learning (Lou et al., 1996; Wang et al., 2010). When they are challenged too much and have to face the learning task and its difficulties on their own, they can become discouraged, demotivated, and even become frustrated (Margolis & McCabe, 2003). High ability students, however, might be able to carry out successful inquiry without much or any additional support, as they prefer challenging, complex, and open-ended learning tasks (Diezmann & Watters, 1997; vanTassel-Baska, 2003). Challenging and complex tasks are an important motivator for high ability students (Lens & Rand, 2000). Providing these students with too much support might decrease their levels of motivation when working on these tasks. Recent research has shown that these children could also benefit from support (Eysink, Gersen, & Gijlers, 2015), but that additional studies should be done to look into the type of support that fits the needs of these children best.

1.3. Differentiated support

A suitable type of support that capitalizes on differences between children could be prompts. First, prompting is a form of support that incorporates autonomous learning (Davis & Linn, 2000). When available upon request, prompts function merely as stepping stones and do not necessarily intrude upon children's inquiry process. Consequently, high ability children should profit from prompts, which still leave them enough freedom to determine their own learning process (Diezmann & Watters, 1997; vanTassel-Baska, 2003). For the same reason, high ability children's motivation should not be negatively affected by offering prompts during their inquiry process (Lens & Rand, 2000). Lower ability children would also benefit from prompts, as the prompts provide them with direct assistance when needed (Margolis & McCabe, 2003) and guide them through difficulties they often experience with their inquiry process. Consequently, the guidance offered by prompts should positively influence lower ability children's level of motivation (Margolis & McCabe, 2003).

Second, prompts have proven to be effective in explaining scientific domains and underlying principles (Davis & Linn, 2000), and can provide students with proper guidance during formulation and conducting of experiments (Zacharia et al., 2015). As stated above, most difficulties that younger children experience with inquiry learning are related to their inability to identify relevant variables within a domain (Zimmerman, 2007).

1.4. Research issues

The literature indicated that children's learning processes and their need for support in general are ability dependent. However, specific differences between children of different ability levels in their learning approach have not yet been investigated in the context of inquiry learning. Studies on inquiry learning show that support is needed for inquiry learning processes to be effective, without making a distinction between children of different ability levels (Alfieri et al., 2011; d'Angelo et al., 2014; Mayer, 2004). This discrepancy gives us reason to further explore differences in children's inquiry processes and the role of support for different ability levels. Consequently, we explored high, average, and low ability children's inquiry approaches in conditions with and without support to structure their inquiry process.

With regard to children's *inquiry learning process*, two main issues were central to this study. First, the abovementioned literature provides clear indications that children of different ability levels differ on how they address problem-solving tasks. This gave us reason to believe that these differences would also manifest themselves in children's inquiry approaches. The current study was designed to give more insight into how these differences express themselves in the experimental activities children of different ability levels undertake, such as performance of unique experiments, and the correctness of their conclusions (Research Question 1).

Second, literature related to children's use of support during their learning process is more ambiguous. Even less information is available on children's tendency to use support in the context of inquiry learning. On the assumption that problems young children experience with inquiry are mainly attributable to identification of relevant variables (Zimmerman, 2007), prompts were offered to help children identify the relevant variables within the domain. A major issue we wanted to explore was high ability children's inclination to use the prompts during their inquiry approach. High ability children might, on the one hand, consider the prompts redundant and disruptive to their learning process (cf., expertise reversal effect; Kalyuga, 2007); they might therefore be inclined not to use the prompts during the inquiry tasks. On the other hand, as all children need to be supported for their inquiry process to be effective (Mayer, 2004), and prompts are considered a way of support that allows children to act autonomously (Davis & Linn, 2000), high ability children might feel they could benefit from the prompts and therefore use them to structure their inquiry process. High ability children's level of *motivation* is likely to coincide with this. In the event they consider the prompts as beneficial, this might enhance their level of motivation. When the prompts are considered disruptive to their learning process, level of motivation might be reduced. Another issue we wanted to gain insight into was average and low ability children's use of prompts. We tried to gain insight into the suitability of prompts to support these children's inquiry process. To check for prompts as a suitable type of support in the context of inquiry learning, children's use of the prompts, whether their use of the prompts positively influenced their inquiry approach, and their level of motivation were explored (Research Question 2).

In order to see the effect of inquiry approach on learning, we additionally explored children's gain in *domain knowledge*. For inquiry learning to be effective, children should be able to conduct the different inquiry activities correctly (de Jong, 2006; Mayer, 2004). Assuming differences in inquiry approaches between children of different ability levels, differences in domain knowledge between these children might occur. As we had reason to believe that the prompts would positively influence children's inquiry approaches, we investigated differences in domain knowledge between children who received the prompts and children who did not receive the prompts (Research Question 3). To gain more insight into coherence between children's inquiry approach and their outcomes, relations between children's inquiry activities and their domain knowledge gain were also examined (Research Question 4).

2. Method

2.1. Participants

The sample consisted of 478 fifth and sixth graders (222 boys, 256 girls; $M_{age} = 11.3$ years, $SD = .72$, ranging from 9 to 13 years) attending seven different elementary schools located in a mid-sized city in the Netherlands. Children participated in this study as an extracurricular activity during school hours. They were categorized according to their ability level (high ability, average ability, and low ability). Categorization was based on scores from the Dutch student monitoring system (CITO, 2012), a standardized scoring system that defines children's level of ability. CITO takes into account both earlier performance by the individual child and the child's position relative to his or her age group. On the basis of these data, CITO generates a learning profile for every student indicating their learning abilities. The scoring system presents scores as Roman numerals on a continuum, varying from 'I' (highest scoring students) to 'V' (lowest scoring students). For the present study, four subjects were selected that best define children's general learning skills: mathematical skills, spelling skills, reading skills, and reading comprehension. Children were categorized as high ability when they scored 'I' on at least three out of the four subjects ($n = 95$; 20%), and as low ability when they scored 'V' on at least two out of the four subjects ($n = 69$; 14%). The remaining children were categorized as average ability children ($n = 315$; 66%). A team of primary school teachers were asked to check the categorization method so that grouping was in line with teacher's intuitive ability-grouping.

Originally, 533 children were to participate in the study. Two children were excluded beforehand, as their parents did not give active consent for their child to participate in the experiment. Children were randomly assigned to the prompted and unprompted conditions. However, 53 children were later excluded from the sample as they did not complete the pretest, simulation tasks, or the posttest, or due to technical difficulties while working with the simulation. The prompted condition consisted of 249 children (high: 23 boys, 29 girls, $M_{age} = 11.3$ years; average: 77 boys, 85 girls, $M_{age} = 11.3$ years; low: 18 boys, 17 girls, $M_{age} = 11.5$ years), and the unprompted condition included 229 children (high: 19 boys, 24 girls, $M_{age} = 11.3$ years; average: 60 boys, 92 girls, $M_{age} = 11.2$ years; low: 25 boys, 9 girls, $M_{age} = 11.5$ years).

2.2. Materials

The goal of the learning task was to study the effects of gravity and air resistance on how fast objects fall. Children had to comprehend the relations of gravity and air resistance for falling objects. More far-reaching information, such as underlying formulas, was not considered. A simulation was developed that allowed children to drop different items of different weights and shapes (familiar objects, such as an apple, a hammer, a feather and a piece of paper) in environments with different gravities, that is, on the earth and on the moon (see Fig. 1). Children could drag the objects to the characters on the earth and the moon. To investigate further the effect of air resistance, children could also vary the presence of an atmosphere (on the earth as well as on the moon). Children could investigate the effects by watching the objects fall, and comparing the time in seconds shown on an indicator that ran while the object was falling (time remained visible till the next simulation).

In order to identify the effects, children could infer patterns by interacting with the simulation. Three tasks were distinguished, covering the following topics: a) the effect of gravity on the speed with which objects fall, b) the effect of air resistance on the speed with which objects fall (related to the presence of an atmosphere), and c) the effect of an object's shape on the speed with which it falls (combining the main effects). For each task, children had to answer two research questions, for which they could develop an answer by working with the simulation (see Fig. 1 for an example of an assignment and the two associated research questions). The simulation was identical in all three tasks. Answers to the questions could be typed in a textbox that was integrated into the learning environment.

Together with the questions concerning the effect of gravity, air resistance, and shape of objects on the speed of falling objects, children in the prompted condition received a set of five prompts per task to help them infer the requested information from the simulation. These prompts were intended to steer children in the right direction regarding identifying the variables they should include in their experiments. The support given in the prompts progressed from merely trying to activate prior knowledge and promote recognition and understanding of the relevant topic area (e.g., 'An ice skater wears a really tight suit to skate as fast as possible.') to suggesting informative simulation runs to answer the questions under investigation (e.g., 'What differences do you see when you drop a paper ball and a piece of paper on the earth and on the moon?'). In order to proceed to the next task, children had to click on all prompts. Children in the unprompted condition did not receive these prompts.

A pilot study was conducted to test whether the children could work with the inquiry tasks and the prompts, and whether consulting the prompts resulted in useful experiments. Think-aloud protocols and observations indicated that the children could work with the set of five prompts per task. A number of linguistic changes were made to better match the language level of the participants' age group.

2.3. Measures

2.3.1. Domain knowledge

A domain knowledge test was developed containing 16 questions (nine essay questions and seven multiple choice questions with three alternatives each) measuring students' knowledge of the effects of gravity, air resistance, and shape of objects when objects fall. The test was designed as a representative test that consisted of four conceptual knowledge questions (de Jong & Ferguson-Hessler, 1996; two essay, two multiple choice), and twelve questions that covered near and far transfer knowledge (Reed, 1999; seven essay, five multiple choice). The conceptual knowledge questions requested children to name and describe the main concepts of the domain (e.g., 'What causes objects that are thrown in the air to come back to the earth eventually?'). Near transfer questions tested children's knowledge in a context comparable to the simulation (e.g., 'There is a glass tube. The air is removed from the tube. Drop a wooden ball and a feather in the air-free tube. Which object hits the bottom first? Explain your answer.'), and far transfer knowledge was measured in a context different from the simulation

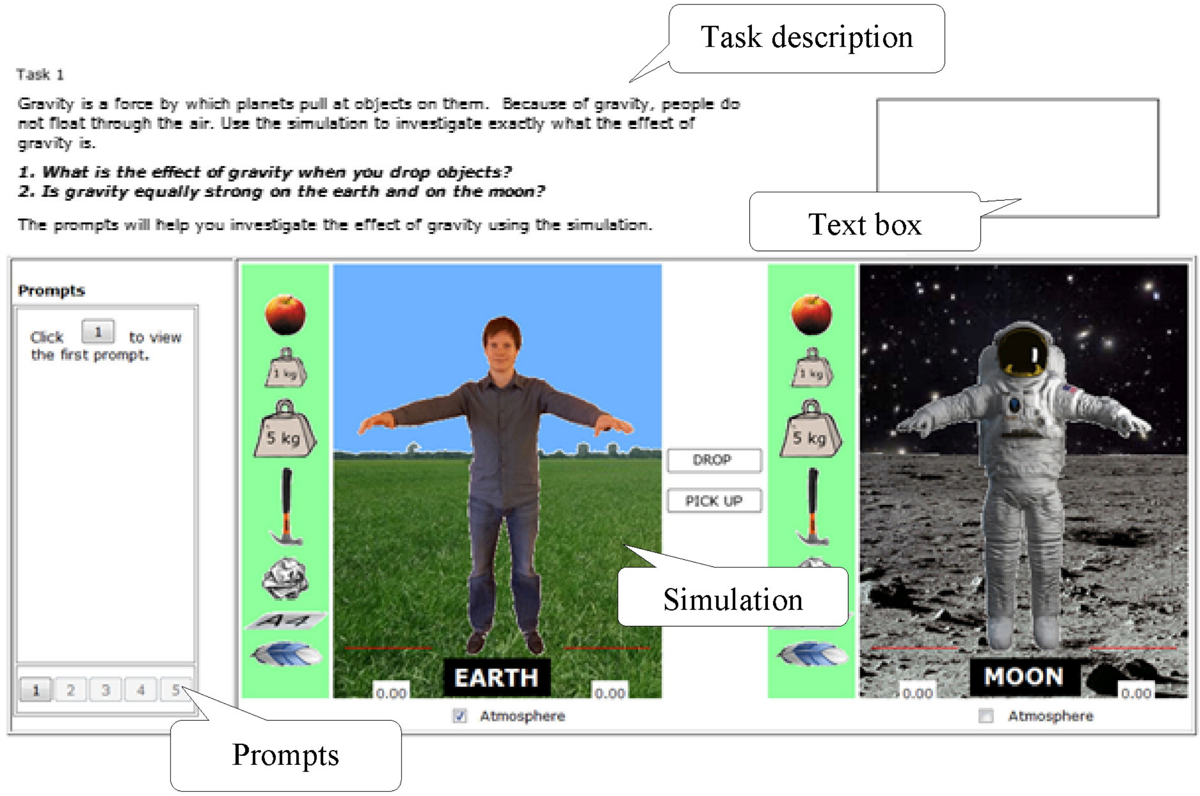

Fig. 1. First simulation task for the prompted condition.

(e.g., 'The earth and the moon pull on one another. How is the gravitational force of the moon visible on earth?').

The final version of the domain knowledge test was tested in a pilot study which indicated that the questions were suitable for the target group. Pretest, posttest, and retention test were identical. The domain knowledge test had adequate reliabilities at the pretest, posttest, and retention test, reaching Cronbach's alpha's of $\alpha = .72$, $\alpha = .71$, and $\alpha = .74$, respectively.

2.3.2. Motivation

Children's motivation was assessed by means of a short questionnaire, consisting of three statements. Motivation was measured as expectation of success and positive valence ('I like the tasks.', 'I will do a good job finishing the tasks.', and 'I know how to do the tasks.'), derived from earlier work by Vollmeyer and Rheinberg (2000). Children could indicate their opinion using a five point Likert scale, ranging from 'I totally disagree' to 'I totally agree'.

The motivation questionnaire was administered three times, in two different versions. A version in future tense was used after they received explanation of the tasks at hand but before starting on the simulation tasks, and after they completed the first simulation task. The version administered after all three simulation tasks had been completed was written in past tense. The motivation questionnaire reached adequate reliabilities for all three testing time points. Cronbach's alpha was $\alpha = .64$, $\alpha = .73$ and $\alpha = .81$, respectively.

2.4. Procedure

Children began by completing the pretest of domain knowledge. They were given 30 min to complete the test. One to four days after they completed the pretest, children took part in the experiment; this session lasted 90 min. Children worked individually with the simulation. They were given an introduction prior to the simulation by means of a screen recording of the simulation, demonstrating the different features of the simulation. Children were told that they were to follow in the footsteps of Galileo and Neil Armstrong, and perform similar experiments themselves with the simulation. Then, the children filled out the first motivation questionnaire. After this, children worked on the inquiry tasks, for which they were given 45 min. Children were instructed to answer all questions, which was checked by the researcher before closing the simulation. After completing the first of the three simulation tasks, children filled out the second motivation questionnaire. Directly after they finished the simulation tasks, children completed the posttest, and filled out the final motivation questionnaire. Children were again given 30 min to complete the domain knowledge test. After approximately three or four weeks (18–29 days), the retention test on domain knowledge was administered. They were again given 30 min to complete the test.

2.5. Data analysis

2.5.1. Log files

To capture children's activities, their actions in the simulation were logged. The following measures were extracted from the log files to examine children's experimental approach: the number of (unique) simulation runs (i.e., 'unique runs in total'), and whether or not students conducted experiments with the most informative variables (i.e., 'most informative runs'). In order to extract the latter measure, for every question a unique run or configuration of runs was identified that was most informative for obtaining the information needed to answer the question being addressed (e.g., the question 'Is the gravity equally strong on the earth and the moon?' could be answered by dropping the same object both on the earth and on the moon). For every 'most informative' simulation run or set of runs conducted for the given question, children received 1 point. Accordingly, children could receive a score varying between 0 to 6 points.

For children in the prompted condition, two measures were extracted to examine children's use of support. First, it was determined whether or not students viewed the prompts *before* answering the questions (i.e., 'views'). Children's views score was set at the point when they typed their final answer to the questions in the simulation. The number of prompts viewed by that point was counted for each of the three tasks, and summed accordingly; views score could thus vary from 0 to 15. Second, it was determined whether or not students conducted the simulation runs suggested in a number of the hints (i.e., 'use'). Suggestions were given in the third, fourth, and fifth hints in all three simulation tasks. Children's use score could thus vary from 0 to 9.

The effectiveness of children's inquiry process was determined by extracting the answers to the questions presented in the three simulation tasks from the log files. These answers were scored to gain insight into the children's understanding of the topics covered in the simulation tasks. The answer to each of the six questions was determined to be either correct (1 point) or incorrect (0 points), on the basis of a previously developed coding scheme. The maximum possible score was 6 points. A second coder scored 10% of the answers to the questions. Neither coder was aware of condition and ability level of the children when scoring the answers. The inter-rater reliability coefficient reached .74 (Cohen's kappa).

2.5.2. Domain knowledge tests

The essay questions on the domain knowledge test were scored following a coding scheme. Answers to essay questions were rated on presence of essential concept(s), and/or explanations of essential process(es). An overview of the concepts and/or processes that were to be described or explained, and how many points should be granted to these concepts and processes, was drafted for each question. An answer including all concepts and mechanisms received a score of 4 points, and scores per question could vary from 0 to 4 points. A second coder scored the answers on the essay questions for 10% of the domain knowledge tests. Neither coder was aware of condition and ability level of the children when scoring the answers. The inter-rater reliability coefficient reached .92 (Cohen's kappa).

The multiple choice questions were treated similarly to the essay questions, since they had similar levels of difficulty, with 4 points given when the answer was correct; an incorrect answer was granted zero points. The maximum score on the test was, therefore, 64 points (16 points for the four conceptual knowledge questions, and 24 points each for the six near transfer knowledge and six far transfer knowledge questions).

2.5.3. Motivation

The motivation questionnaire was administered three times: before the simulation tasks, after the first task, and directly following completion of the three tasks. To create an overview of children's motivational state while working with the simulation tasks, these three measures were combined into one overarching score. Each questionnaire consisted of three items that could be answered using a 5-point Likert scale, ranging from low motivation to high motivation. Consequently, children's motivation could be represented as the sum of scores on the three measures, varying between 9 and 45 points.

3. Results

The main focus of this study was to gain insight into children's inquiry approaches and their use of prompts during inquiry. We first discuss differences in children's learning processes. Thereafter, to explore whether inquiry approach influenced children's learning outcomes, differences in domain knowledge gain were analyzed. These analyses are followed by correlational measures to determine coherence between children's inquiry approach and their knowledge gain. Finally, differences in level of motivation are discussed to find out whether being prompted affected children's motivation.

3.1. Learning process

To gain insight into children's approach to the inquiry task when working with the simulation, their engagement in different activities was analyzed (Research Question 1). Thereafter, children's use of the prompts was examined (Research Question 2). First, differences between ability levels were analyzed to indicate whether and how high, average, and low ability children's inquiry approaches differed with regard to both their inquiry activities and their use of prompts. Second, differences between conditions were assessed for each ability level separately. Children's inquiry approaches were explored in a context in which they were offered prompts to guide their inquiry and a context in which they had to complete the inquiry tasks without prompts. Table 1 shows children's mean scores on the measures concerning their viewing and use of the prompts, simulation runs, and scores on the questions for the simulation tasks.

3.1.1. Differences between ability levels — inquiry activities

Three univariate analyses of variance (ANOVAs) assessed differences between high, average and low ability children concerning the total number of unique runs performed, number of 'most informative' runs performed, and their score on the answers to the simulation task questions. The analyses indicated significant differences between ability levels on the mean number of unique simulation runs ($F(2, 473) = 6.27$, $p = .002$, $\eta_p^2 = .03$), the mean number of 'most informative' simulation runs ($F(2, 473) = 9.14$, $p < .001$, $\eta_p^2 = .04$), and the mean score on the answers on the questions of the simulation tasks ($F(2, 458) = 32.87$, $p < .001$, $\eta_p^2 = .13$). Pairwise comparisons using Bonferroni corrections

Table 1
Mean scores for abilities and conditions on measures of approach to the inquiry task.

	Prompted		Unprompted		Total	
	M	*SD*	*M*	*SD*	*M*	*SD*
High ability						
Prompts						
Views (max = 15)	10.56	4.12	n/a	n/a	n/a	n/a
Use (max = 9)	3.85	2.26	n/a	n/a	n/a	n/a
Simulation runs						
Unique runs in total	14.79	7.47	14.02	7.43	14.44	7.42
Most informative runs (max = 6)	4.63	1.44	2.74	1.03	3.78	1.58
Answers						
Score (max = 6)	3.58	1.55	3.26	1.54	3.43	1.55
Average ability						
Prompts						
Views (max = 15)	9.69	4.15	n/a	n/a	n/a	n/a
Use (max = 9)	2.80	1.97	n/a	n/a	n/a	n/a
Simulation runs						
Unique runs in total	13.43	7.33	11.05	6.15	12.28	6.88
Most informative runs (max = 6)	3.95	1.48	2.42	.90	3.21	1.45
Answers						
Score (max = 6)	2.49	1.39	2.37	1.31	2.43	1.35
Low ability						
Prompts						
Views (max = 15)	8.15	4.86	n/a	n/a	n/a	n/a
Use (max = 9)	1.76	1.76	n/a	n/a	n/a	n/a
Simulation runs						
Unique runs in total	12.77	5.33	8.59	6.19	10.71	6.10
Most informative runs (max = 6)	3.46	1.42	2.21	.54	2.84	1.24
Answers						
Score (max = 6)	1.85	1.37	1.53	1.21	1.69	1.29
Total						
Prompts						
Views (max = 15)	9.66	4.29	n/a	n/a	n/a	n/a
Use (max = 9)	2.88	2.09	n/a	n/a	n/a	n/a
Simulation runs						
Unique runs in total	13.63	7.12	11.24	6.59	12.48	6.96
Most informative runs (max = 6)	4.02	1.50	2.45	.90	3.27	1.48
Answers						
Score (max = 6)	2.63	1.52	2.41	1.43	2.52	1.48

revealed that high ability students conducted more unique simulation runs in total than average ability students ($p = .023$) and low ability students ($p < .001$), conducted more 'most informative' simulation runs than average ability students ($p = .003$) and low ability students ($p < .001$), and showed higher scores than average ability ($p < .001$) and low ability ($p < .001$) students. Average ability children, in turn, demonstrated higher mean scores on the answers for the simulation tasks than low ability students ($p < .001$).

3.1.2. Differences between ability levels — prompts

Children's viewing of prompts before answering the questions and whether they conducted the simulation runs suggested in the prompts were assessed. Two ANOVAs indicated differences between high, average, and low ability children within the prompted condition. The analyses revealed significant differences between ability levels on the number of prompts children looked at before answering the questions for the three tasks ($F(2, 231) = 3.21$, $p = .042$, $\eta_p^2 = .027$), and the number of simulation runs students conducted that followed the suggestions in the prompts ($F(2, 244) = 11.43$, $p < .001$, $\eta_p^2 = .086$). Pairwise comparisons using Bonferroni corrections indicated that high ability children looked at more prompts than low ability children ($p = .036$), and conducted more of the simulation runs suggested in the prompts than average ability children ($p = .004$) and low ability children ($p < .001$). Average ability children, in turn, followed the suggestions in the prompts more often than did the low ability children ($p = .020$).

3.1.3. Differences between conditions

Differences between conditions were analyzed by means of multivariate analyses of variance (MANOVAs) within ability levels and involved the total number of unique runs performed, number of 'most informative' runs performed, and score on the answers to the simulation task questions. Using Wilks' statistic, the MANOVA for high ability children revealed a significant difference between conditions ($\Lambda = .61$, $F(3, 89) = 19.36$, $p < .001$, $\eta_p^2 = .40$). Subsequent ANOVAs indicated that high ability children who received prompts conducted more 'most informative' simulation runs, on average, than the unprompted high ability children ($F(1, 91) = 48.89$, $p < .001$, $\eta_p^2 = .35$).

A second MANOVA revealed significant differences between conditions for the average ability children ($\Lambda = .71$, $F(3, 297) = 40.13$, $p < .001$, $\eta_p^2 = .29$). Subsequent ANOVAs showed that average ability children who received prompts outperformed the unprompted average ability children on the mean total number of unique simulation runs conducted ($F(1, 299) = 9.49$, $p < .001$, $\eta_p^2 = .03$), and on mean number of 'most informative' simulation runs ($F(1, 299) = 114.57$, $p < .001$, $\eta_p^2 = .28$).

A third MANOVA indicated significant differences between conditions for the low ability children ($\Lambda = .69$, $F(3, 63) = 9.26$, $p < .001$, $\eta_p^2 = .31$). Subsequent ANOVAs revealed that low ability children who received prompts conducted more unique simulation runs in total ($F(1, 65) = 8.28$, $p = .005$, $\eta_p^2 = .11$) and conducted more 'most informative' runs ($F(1, 65) = 24.62$, $p < .001$, $\eta_p^2 = .28$), on average, than the unprompted low ability children.

3.2. Knowledge gain

To explore children's domain knowledge gain, ability-related and condition-related differences were assessed (Research Question 3). First, differences in normalized domain knowledge gain between ability levels were examined to gain insight into children's potential for growth when working with the inquiry tasks at hand, which could be informative regarding learning by children of different ability levels when working with an inquiry task. Second, differences in absolute domain knowledge gain between children who were offered prompts and children who did not receive prompts (i.e., by condition) were assessed for each ability level separately by means of repeated measures analyses. The latter analyses explore the learning outcomes with or without prompts for children at each ability level, taken separately. Table 2 shows high ability, average ability, and low ability children's mean domain knowledge test scores on the pretest, posttest, and retention test, and provides an overview of children's normalized learning gain scores, which were used only for comparisons across ability levels.

3.2.1. Differences between ability levels

To assess possible differences in prior knowledge, an ANOVA was conducted. This ANOVA showed that children of different ability levels indeed differed in domain knowledge on the pretest ($F(2, 475) = 43.07$, $p < .001$, $\eta_p^2 = .15$). Pairwise comparisons using Bonferroni corrections indicated that high ability children showed higher prior domain knowledge than average ability ($p < .001$) and low ability children ($p < .001$). Average ability children, in turn, had a higher prior domain knowledge than low ability children ($p < .001$). Differences between conditions and the interaction (Ability level × Condition) were not significant. To take into account these differences in prior knowledge, normalized learning gain scores were calculated (i.e., absolute learning gain divided by the possibility for learning gain based on pretest scores). To determine whether high ability, average ability, and low ability children showed comparable relative learning gains from pretest to posttest, an ANOVA was conducted. This ANOVA indicated significant differences between children of different ability levels ($F(2, 475) = 5.26$, $p = .005$, $\eta_p^2 = .02$). Subsequent pairwise comparisons using Bonferroni corrections revealed that high ability children showed a higher mean normalized learning gain from pretest to posttest than average ability ($p = .045$) and low ability children ($p = .005$). Mean normalized learning gain did not differ significantly between average and low ability children ($p = .334$).

A comparable ANOVA was conducted to determine children's relative learning gain from posttest to retention test. The analysis revealed significant differences between ability levels ($F(2, 450) = 3.30$, $p = .038$, $\eta_p^2 = .01$). Subsequent pairwise comparisons using Bonferroni corrections revealed that high ability children showed a higher mean normalized learning gain from posttest to retention test than average ability children ($p = .032$). Other comparisons were not significant.

3.2.2. Differences between conditions

High ability children's domain knowledge and possible differences between conditions were investigated for the absolute scores on the pretest (Time 1) and the posttest (Time 2) with repeated measures analysis. A significant main effect for the within-subject factor Time ($\Lambda = .89$, $F(1, 93) = 11.24$, $p = .001$, $\eta_p^2 = .11$) was found, indicating

Table 2
Mean domain knowledge test scores and normalized gain scores for the different ability levels.

	High ability		Average ability		Low ability	
	M	*SD*	*M*	*SD*	*M*	*SD*
Absolute test scores[a]						
Pretest						
Prompted	37.47	9.26	28.78	9.18	23.94	8.06
Unprompted	36.63	9.35	28.46	9.25	24.57	8.82
Total	37.10	9.26	28.63	9.20	24.24	8.37
Posttest						
Prompted	39.51	9.88	30.44	9.05	24.75	8.65
Unprompted	38.77	10.20	30.71	8.64	25.50	8.45
Total	38.19	9.98	30.57	8.84	25.11	8.49
Retention test						
Prompted	40.94	9.89	31.09	9.08	25.72	8.40
Unprompted	39.42	9.28	30.90	9.63	27.90	8.29
Total	40.27	9.61	31.00	9.34	26.77	8.35
Normalized gain scores[b]						
Pre to Post	.18	.34	.15	.32	.09	.37
Post to Retention	.09	.28	.03	.32	.08	.34

[a] Maximum score = 64.
[b] Percentage wise gain scores.

an increase in domain knowledge from pretest to posttest. There was no interaction (Time × Condition; $\Lambda = 1.00$, $F(1, 93) = .00$, $p = .980$, $\eta_p^2 = .00$). A similar analysis was conducted to analyze high ability children's knowledge at posttest (Time 2) and retention test (Time 3). There was no significant effect for the within-subject factor Time ($\Lambda = .97$, $F(1, 89) = 2.70$, $p = .104$, $\eta_p^2 = .03$), nor was there a significant interaction (Time × Condition; $\Lambda = .99$, $F(1, 89) = .38$, $p = .539$, $\eta_p^2 = .00$).

Average ability children's domain knowledge at pretest (Time 1) and posttest (Time 2) and possible differences between conditions were similarly analyzed with repeated measures analysis. There was a significant main effect for the within-subject factor Time ($\Lambda = .93$, $F(1, 312) = 23.55$, $p < .001$, $\eta_p^2 = .07$), indicating an increase in domain knowledge from pretest to posttest. There was no interaction (Time × Condition; $\Lambda = .99$, $F(1, 312) = .38$, $p = .537$, $\eta_p^2 = .00$). Analysis of average ability students' knowledge at posttest (Time 2) and retention test (Time 3) showed no significant main effect for Time ($\Lambda = .99$, $F(1, 298) = 1.06$, $p = .305$, $\eta_p^2 = .00$), nor was there a significant interaction (Time × Condition; $\Lambda = .99$, $F(1, 298) = .32$, $p = .573$, $\eta_p^2 = .00$).

Low ability children's domain knowledge and possible differences between conditions were also investigated with repeated measures analysis. From pretest (Time 1) to posttest (Time 2), the main effect for the within-subject factor Time was not significant ($\Lambda = .97$, $F(1, 67) = 1.85$, $p = .178$, $\eta_p^2 = .03$). There was no interaction (Time × Condition; $\Lambda = 1.00$, $F(1, 67) = .00$, $p = .998$, $\eta_p^2 = .00$). Low ability students showed a significant increase in domain knowledge from posttest (Time 2) to retention test (Time 3), indicated by a main effect for the within-subject factor Time ($\Lambda = .91$, $F(1, 60) = 5.68$, $p = .020$, $\eta_p^2 = .09$). There was no interaction (Time × Condition; $\Lambda = .98$, $F(1, 60) = 1.03$, $p = .315$, $\eta_p^2 = .02$).

3.3. Relations between process and knowledge gain measures

To indicate whether the different inquiry activities, as part of children's inquiry approach, were related to effective completion of the inquiry tasks and children's domain knowledge gain, correlational analyses were performed (Research Question 4). Separate correlational analyses were conducted for the three ability levels to reveal the relations between viewing and use of the prompts, the total number of unique runs performed, number of 'most informative' runs performed, score on the answers to the simulation task questions, knowledge gain from pretest to posttest, and knowledge gain from posttest to retention test (see Table 3).

High ability children's capacity to conduct the most informative runs, which was largely related to their viewing and use of the prompts, was positively associated with successful completion of the inquiry tasks. Being able to effectively answer the research questions of the inquiry tasks was, in turn, positively related to high ability children's knowledge gain from the pretest to the posttest.

For average ability children, their level of activity in the inquiry tasks (i.e., the number of simulation runs and of the most informative runs) was positively associated with the effectiveness of the inquiry tasks (i.e., higher scores on the answers to the tasks). Their inquiry process, in turn, was positively influenced by their inspection and use of the prompts. The score that average ability children obtained on their answers to the research questions for the inquiry tasks was positively related to their knowledge gain from the pretest to the posttest.

Low ability children's success on the inquiry tasks (i.e., their score on the answers) was positively related to the total number of simulation runs, and the number of most informative runs performed. However, inspection and use of the prompts were not associated with their ability to answer the research questions of the inquiry tasks correctly. In addition, success on the inquiry tasks did not relate to low ability children's knowledge gain.

3.4. Motivation

Motivational differences between ability levels were assessed to gain insight in the effect of inquiry tasks on high, average, and low ability children's motivation. To explore whether being prompted affected high, average, and low ability children's level of motivation during

Table 3
Correlations between experimental activities and use of prompts by ability level.

	1	2	3	4	5	6	7
High ability							
1. Unique runs in total	–						
2. Most informative runs	.306**	–					
3. Viewing of prompts[a]	.196	.511***	–				
4. Use of prompts[a]	.180	.704***	.459**	–			
5. Score on the answers	.132	.345**	.290*	.002	–		
6. Knowledge gain pre to post	.075	.097	.319*	.153	.258*	–	
7. Knowledge gain post to retention	.060	.149	.129	.236	−.112	−.481***	–
Average ability							
1. Unique runs in total	–						
2. Most informative runs	.319***	–					
3. Viewing of prompts[a]	.262**	.363***	–				
4. Use of prompts[a]	.260**	.722***	.380***	–			
5. Score on the answers	.177**	.282***	.102	.297***	–		
6. Knowledge gain pre to post	.093	.081	.115	.063	.159**	–	
7. Knowledge gain post to retention	−.043	−.036	−.106	.012	−.078	−.410***	–
Low ability							
1. Unique runs in total	–						
2. Most informative runs	.304*	–					
3. Viewing of prompts[a]	.205	.400*	–				
4. Use of prompts[a]	−.013	.591***	.259	–			
5. Score on the answers	.279*	.259*	.050	.158	–		
6. Knowledge gain pre to post	.028	.127	.233	.160	.001	–	
7. Knowledge gain post to retention	.093	−.171	−.083	−.040	.009	−.267*	–

* $p < .05$.
** $p < .01$.
*** $p < .001$.
[a] Prompted condition only.

Table 4
Mean motivation scores by ability level.

	Prompted		Unprompted		Total	
	M	*SD*	*M*	*SD*	*M*	*SD*
High ability	36.84	6.01	37.09	6.01	36.96	5.98
Average ability	34.23	6.96	35.13	5.91	34.67	6.48
Low ability	35.40	6.31	35.26	6.20	35.33	6.21

Note. *Maximum* score = 45.

working with the inquiry tasks, differences between conditions were investigated for each ability level separately (Research Question 2). Table 4 shows high ability, average ability, and low ability children's motivation scores.

3.4.1. Differences between ability levels

An ANOVA was conducted to identify differences between children of different ability levels regarding their motivation while working with the inquiry tasks. Differences between ability levels were significant ($F(2, 468) = 4.71$, $p = .009$, $\eta_p^2 = .02$). Subsequent pairwise comparisons using Bonferroni corrections showed that high ability children were more highly motivated than average ability children ($p = .007$). Other comparisons were not significant.

3.4.2. Differences between conditions

Motivation was assessed for every ability level separately by means of three ANOVAs. The analyses showed no significant difference between conditions for high ability children ($F(1, 92) = .04$, $p = .841$, $\eta_p^2 = .00$), average ability children ($F(1, 306) = 1.51$, $p = .221$, $\eta_p^2 = .01$), and low ability children ($F(1, 67) = .01$, $p = .929$, $\eta_p^2 = .00$).

4. Discussion and conclusions

Ability level is generally considered to play a crucial role in children's potential to learn. This study explored whether and how children of different ability levels varied in the way they approached an inquiry task, whether and how support played an additional role in their inquiry process, whether this affected learning outcomes, and if children's motivation was influenced by the presence of prompts.

4.1. Learning process

Based on the literature on general differences in children's learning skills and preferences, high ability, average ability, and low ability children were expected to show different inquiry approaches. High ability children should know their way around the inquiry tasks at hand as their learning skills seem to match the required skills to engage in inquiry learning best (e.g., Diezmann & Watters, 1997; vanTassel-Baska, 2003). Average ability and low ability children should experience more difficulties navigating through the inquiry tasks and need more support to complete the inquiry activities (Alexander & Schwanenflugel, 1996; Margolis & McCabe, 2003). Results of the present study suggest that high ability children indeed showed a more effective learning process than did the average ability and low ability children. Even though knowledge gains were small, high ability children showed a more effective inquiry process, as they outperformed the others on responding correctly to the research questions accompanying the simulation tasks, and learned relatively more than the average and low ability children, as shown on the knowledge tests. With regard to the efficiency of their inquiry process, results are more ambiguous as high ability children performed more simulation runs than the other children. One might have expected that these children would need fewer simulation runs to solve the problems, as these children's problem-solving skills would have made it easier for them to navigate through the tasks at hand (Diezmann & Watters, 1997; vanTassel-Baska, 2003). However, this result might also be due to differences in prior knowledge between the different ability levels. High ability children's prior knowledge was indeed higher than that of the average and low ability children. Lazonder, Wilhelm, and Hagemans (2008) showed that prior knowledge affects strategy use in the experimentation phase, such that children who start an inquiry task with some domain knowledge are more likely to adopt a theory-driven approach than children with lower levels of prior knowledge. This theory-driven approach comes with a high number of predictions or conjectures, which are then tested by conducting a relatively low number of carefully planned experiments. The present results point to interpretations within this line of reasoning. For example, it could be that high ability children had far more ideas to investigate than average and low ability children, and hence conducted more experiments. It could have increased their ability to 'search the experiment space for a combination of experiments' to answer the research question (as described in the SDDS model; Klahr & Dunbar, 1988). High ability children's motivational level, which showed to be higher than that of the average ability children, might corroborate this line of reasoning. On the one hand, a higher level of motivation might have resulted in a higher task orientation, and, hence, in more simulation runs to complete the task successfully. On the other hand, being able to navigate through the tasks without experiencing difficulty might have positively affected children's level of motivation. However, the correlational analysis showed no relation between the number of experiments performed and the score on the answers on the questions of the inquiry tasks for the high ability children. Future research could be done to gain more insight into the role of prior knowledge in children's experimentation activities.

Considering high ability children's preference for open and challenging learning tasks that are minimally supported (Diezmann & Watters, 1997; Phillips & Lindsay, 2006; Reis & Renzulli, 2010; vanTassel-Baska, 2003), and lower ability children's explicit need for support during challenging tasks (Lou et al., 1996; Wang et al., 2010), one could have expected to see a difference in the use of prompts during the inquiry tasks. High ability children inspected and used the prompts more often during their learning process than the average and low ability children. On the one hand, this result might be surprising as children who need support most used the prompts the least. On the other hand, this result might indicate that the prompts that presented students with further assistance during the experimental phase might have been a better addition to high ability children's existing knowledge of the domain than they were for the average and low ability children. High ability children were consequently perhaps better able to grasp the content of the prompts, which allowed them to follow the line of reasoning presented.

High ability children's motivational levels were not affected by offering prompts during the inquiry tasks. This result, together with the finding that high ability children viewed and used the prompts more during their inquiry process than the other children, suggests that these children did not feel hindered by the support offered during the simulation tasks. In fact, their experimental activity became more efficient when using the prompts; they conducted 'most informative' simulation runs more often in the prompted condition than in the unprompted condition. Since high ability children seem to have benefited from prompts during their inquiry process, it might be interesting to look into the effects on their inquiry process when offered other forms of support, and how much support needs to be given in order to reach the optimal learning process (see also Tomlinson, 2000).

A different pattern emerged for the other ability groups. Both average ability and low ability children were expected to need support to engage in effective inquiry (Lou et al., 1996; Wang et al., 2010) and prompts are often recommended as a good way to offer support during inquiry learning (Arnold et al., 2014; Zacharia et al., 2015). However, the average and low ability children made minimal use of the prompts. An interesting issue here is to gain insight into why the results show that these children responded oppositely to what might have been expected. First, average and low ability children's relatively low level of experimental activity (i.e., viewing and use of prompts, and number of simulation runs) might have been a result of their low prior knowledge. This might have influenced their ability to identify and conduct appropriate experiments during the inquiry tasks (Lazonder et al., 2008). A recent study by Roll, Briseno, Yee, and Welsh (2014) revealed that scaffolding online simulation tasks was beneficial for children with high prior knowledge, and did not benefit students with lower levels of prior knowledge.

Second, the content and implementation of the prompts might have influenced children's capacity and inclination to use the prompts during their inquiry process. The prompts were offered in progressive levels of concreteness, meaning that the first prompts that focused on activating children's prior knowledge and understanding of the domain might have been too abstract for them to convert the information into experimental activities. When the lower ability children were confronted with these first prompts, they might have been discouraged from consulting the other prompts, since they did not feel that these prompts could help them with completing the inquiry tasks.

An alternative, third explanation could be that the lower ability children might not have been aware of their poor ability to answer the simulation task questions, and therefore did not seek the assistance of the prompts in order to do so. Research shows that the discrepancy between children's (prior) domain knowledge and confidence as to how they will perform on certain tasks tends to be very large (de Bruin & van Gog, 2012), which might lead these children to fail to realize that they are not able to complete the task without support and, therefore, not to seek assistance (Ferguson-Hessler & de Jong, 1990). These findings might also explain the result that average and low ability children's motivation, similar to the high ability children, did not differ between conditions.

Prompted children showed a more active and effective inquiry process than their unprompted counterparts. When the children incorporated the prompts in their inquiry process, they generally showed a more effective inquiry process, although an effect of prompting on low ability children's inquiry process was not found. Since working with the inquiry tasks seemed to have lacked effectiveness for the low ability children, even when prompted during these tasks, further investigation is called for regarding how these children should be supported so that they could benefit from inquiry learning activities. In this study, the prompts could be considered an additional challenge for the average and low ability children and might therefore have failed to achieve their objectives. Prompts should perhaps be more prominent and mandatory, or should be offered according to the ‘just-in-time’ principle (Berry & Broadbent, 1987; Hulshof & de Jong, 2006). To make sure that support becomes actually supportive instead of an additional challenge for these children, future studies should look into other forms of support that are better suited in terms of concreteness of the information offered, and into the implementation of these forms of support.

4.2. Domain knowledge

With differences in children's approach to the inquiry tasks in mind, it could be expected that high ability children would also show the largest domain knowledge gain. The results confirmed this expectation, showing that, even though differences were quite small, high ability children indeed showed a higher normalized learning gain than average and low ability children. The size of high ability children's learning gain did not differ between conditions. However, at the same time the results showed that high ability children's experimental activities became more efficient when consulting the prompts, leading to more effective performance on the inquiry tasks.

Average ability children showed a small but significant knowledge gain from pretest to posttest as well. Contrary to what might have been expected, whether they received support did not make a difference. Low ability children did not seem to profit in the short term from working with the simulation, as they did not show a significant knowledge gain from pretest to posttest in either the prompted or the unprompted condition. Overall, even though the analyses showed that prompted children had a more active and effective inquiry process than their unprompted counterparts, this did not lead to higher knowledge gain for the children in the prompted condition. Children – regardless of their ability level – were not able to translate the knowledge learned through working with the inquiry tasks to the domain knowledge tests.

The low ability children were the only ones whose scores increased significantly from the posttest to the retention test, although this gain did not exceed high and average ability children's normalized knowledge gain. This finding is rather peculiar, since the children were not exposed to the learning material in the weeks between posttest and retention test. A possible, highly speculative explanation could be that these children required more time than the other children to grasp the meaning of the different concepts of the domain. When confronted with the same concepts for the third time, responding to the retention test, they might have been better able to integrate their new knowledge into their answers on the questions. Research into the incubation effect might corroborate this assumption. An incubation period might enhance children's ability to solve problems, as a break from the task might distract them from false and fixed assumptions (i.e., fixation; Segal, 2004; Wiley, 1998). A study into the effects of domain knowledge differences on the incubation effect showed that the effect was observed for students with lower knowledge of the domain (Wiley, 1998). However, although results supporting the incubation effect have increased, discussion continues about the mechanisms that cause the effect (Sio & Ormerod, 2015). Future research should therefore be done to corroborate a possible incubation effect for low ability children in the context of inquiry learning.

4.3. Limitations

Two issues that could be considered as limitations to the current study are the grouping method and the domain knowledge tests. Even though the grouping method that divided the children into high ability, average ability, and low ability children was based on conventional performance measures and was in line with the teacher's intuitive division of abilities in their classes, it did not pay specific attention to children with learning problems, such as dyslexia. Future studies could take learning problems into account when grouping children, or gain more insight into the effect of these learning problems on children's inquiry learning activities.

The results of this study revealed that an effective inquiry process seemed to be unrelated to children's domain knowledge gain. This result raises questions regarding the test's suitability for assessing children's domain knowledge that was gained through inquiry learning. Since in most educational settings children must show their capabilities by means of performance testing, it might be relevant to investigate the conditions that need to be met to enable children to translate an effective inquiry learning process into knowledge gain on standardized knowledge tests. Future studies might give more insight into this issue by testing children's knowledge transfer from inquiry learning settings to other types of knowledge tests. Furthermore, the relatively small learning gains could be ascribed to testing effects, especially the posttest-retention test learning gain for the low ability children. However, if a testing effect were the case, these gains should also be present for the high and average ability children — which did not happen. To rule out testing effect as an explanatory variable, parallel knowledge tests should be administered in a replication study.

4.4. Conclusions

The results of this study showed that there were prominent differences in approach to inquiry tasks between children of different ability levels. High ability children, as expected, showed a more active and effective learning process than average and low ability children. However, there were surprising findings regarding the role that could be attributed to the prompts within children's learning process. Although it is often assumed that high ability children do not need and may not welcome additional support to engage in an effective and efficient learning process, offering prompts improved their inquiry process. The absence of this effect for average and low ability children suggests that further research into differentiated support and the suitability of inquiry learning for children of different ability levels is needed. Consequently, in educational practice, teachers should match support to children's ability level when engaging in inquiry learning activities. Special attention should be paid to the fit between the support offered and the children it is intended for, with particular attention to both the type of support that is offered and how it is implemented in the inquiry tasks.

Acknowledgments

The research reported in this article took place in the context of the BE COOL! project (Bevorderen van Excellentie door Coöperatief Onderzoekend en Ontwerpend Leren; Promoting Excellence by Cooperative Inquiry Learning and Learning by Design). We would like to thank OnderwijsBewijs (programmabureau van het Ministerie van Onderwijs, Cultuur en Wetenschap; programme office of the Dutch Ministry of Education, Culture and Science for funding the project. We would also like to thank all other project members: Manon Hulsbeek from the educational consultancy office Expertis, Marga van Amerongen, Mieke van Hecke, and Atteke van Aar from WSNS Lelystad, and Jakob Sikken and Anjo Anjewierden from the University of Twente.

References

Alexander, J.M., & Schwanenflugel, P.J. (1996). Development of metacognitive concepts about thinking in gifted and nongifted children: Recent research. *Learning and Individual Differences*, *8*, 305–325. http://dx.doi.org/10.1016/S1041-6080(96)90021-7.

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. http://dx.doi.org/10.1037/a0021017.

d'Angelo, C., Rutstein, D., Harris, C., Bernard, R., Borokhovski, E., & Haertel, G. (2014). *Simulations for STEM learning: Systematic review and meta-analysis.* Menlo Park, CA: SRI International.

Arnold, J.C., Kremer, K., & Mayer, J. (2014). Understanding students' experiments: What kind of support do they need in inquiry tasks? *International Journal of Science Education*. http://dx.doi.org/10.1080/09500693.2014.930209.

Berry, D.C., & Broadbent, D.E. (1987). Explanation and verbalization in a computer-assisted search task. *The Quarterly Journal of Experimental Psychology*, *39A*, 585–609. http://dx.doi.org/10.1080/14640748708401804.

de Bruin, A.B.H., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction*, *22*, 245–252. http://dx.doi.org/10.1016/j.learninstruc.2012.01.003.

CITO (2012). Toetsscore, vaardigheidsscore… en dan? [Test score, proficiency score… and then?]. www.cito.nl/~/media/cito_nl/Files/Primair%20en%20speciaal%20onderwijs/cito_toetsscore_vaardigheidsscore_en_dan.ashx (Retrieved from www.cito.nl/~/media/cito_nl/Files/Primair%20en%20speciaal%20onderwijs/cito_toetsscore_vaardigheidsscore_en_dan.ashx).

Cronbach, L.J., & Snow, R.E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions.* New York, NY: Irvington Publishers.

Davis, E.A., & Linn, M.C. (2000). Scaffolding students' knowledge integration: Prompts for reflection in KIE. *International Journal of Science Education*, *22*, 819–837. http://dx.doi.org/10.1080/095006900412293.

Diezmann, C.M., & Watters, J.J. (1997). Bright but bored: Optimising the environment for gifted children. *Australian Journal of Early Childhood*, *22*, 17–21.

Eysink, T.H.S., & de Jong, T. (2012). Does instructional approach matter? How elaboration plays a crucial role in multimedia learning. *The Journal of the Learning Sciences*, *21*, 583–625. http://dx.doi.org/10.1080/10508406.2011.611776.

Eysink, T.H.S., Gersen, L., & Gijlers, H. (2015). Inquiry learning for gifted children. *High Ability Studies*. http://dx.doi.org/10.1080/13598139.2015.1038379.

Ferguson-Hessler, M.G.M., & de Jong, T. (1990). Studying physics text: Differences in study processes between good and poor performers. *Cognition and Instruction*, *7*, 41–54. http://dx.doi.org/10.1207/s1532690xci0701_2.

Furtak, E.M., Seidel, T., Iverson, H., & Briggs, D.C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching. *Review of Educational Research*, *82*, 300–329. http://dx.doi.org/10.3102/0034654312457206.

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

Hulshof, C., & de Jong, T. (2006). Using just-in-time information to support scientific discovery learning about geometrical optics in a computer-based simulation. *Interactive Learning Environments*, *14*, 79–94. http://dx.doi.org/10.1080/10494820600769171.

de Jong, T. (2006). Technological advances in inquiry learning. *Science*, *312*, 532–533. http://dx.doi.org/10.1126/science.1127750.

de Jong, T., & Ferguson-Hessler, M.G.M. (1996). Types and qualities of knowledge. *Educational Psychologist*, *31*, 105–113. http://dx.doi.org/10.1207/s15326985ep3102_2.

de Jong, T., & van Joolingen, W.R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, *68*, 179–201. http://dx.doi.org/10.3102/00346543068002179.

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

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*, 23–31. http://dx.doi.org/10.1207/S15326985EP3801_4.

Klahr, D., & Dunbar, F.K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*, 1–48. http://dx.doi.org/10.1207/s15516709cog1201_1.

Klahr, D., & Dunbar, F.K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology*, *25*, 111–146. http://dx.doi.org/10.1006/cogp.1993.1003.

Lazonder, A.W., Wilhelm, P., & Hagemans, M.G. (2008). The influence of domain knowledge on strategy use during simulation-based inquiry learning. *Learning and Instruction*, *18*, 580–592. http://dx.doi.org/10.1016/j.learninstruc.2007.12.001.

Lens, W., & Rand, P. (2000). Motivation and cognition: Their role in the development of giftedness. In K.A. Heller, F.J. Mönks, R.J. Sternberg, & R.F. Subotnik (Eds.), *International handbook of giftedness and talent* (pp. 193–202) (2nd ed.). Kidlington, Oxford: Elsevier Science Ltd.

Lou, Y., Abrami, P.C., Spence, J.C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, *66*, 423–458. http://dx.doi.org/10.3102/00346543066004423.

Manlove, S., Lazonder, A.W., & de Jong, T. (2006). Regulative support for collaborative scientific inquiry learning. *Journal of Computer Assisted Learning*, *22*, 87–98. http://dx.doi.org/10.1111/j.1365-2729.2006.00162.x.

Margolis, H., & McCabe, P.P. (2003). Self-efficacy: A key to improving the motivation of struggling learners. *Preventing School Failure: Alternative Education for Children and Youth*, *47*, 162–169. http://dx.doi.org/10.1080/10459880309603362.

Mayer, R.E. (2003). *Learning and instruction.* Upper Saddle River, NJ: Prentice Hall.

Mayer, R.E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, *59*, 14–19. http://dx.doi.org/10.1037/0003-066X.59.1.14.

Minner, D.D., Levy, A.J., & Century, J. (2010). Inquiry-based science instruction: What is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, *47*, 474–496. http://dx.doi.org/10.1002/tea.20347.

Njoo, M., & de Jong, T. (1993). Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. *Journal of Research in Science Teaching*, *30*, 821–844. http://dx.doi.org/10.1002/tea.3660300803.

Pedaste, M., Mäeots, M., Siiman, L.A., de Jong, T., van Riesen, S.A.N., Kamp, E.T., ... Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47–61. http://dx.doi.org/10.1016/j.edurev.2015.02.003.

Phillips, N., & Lindsay, G. (2006). Motivation in gifted students. *High Ability Studies*, *17*, 57–73. http://dx.doi.org/10.1080/13598130600947119.

Reed, S.K. (1999). Abstracting solutions. *Word problems: Research and curriculum reform* (pp. 101–115). Mahway, NJ: Lawrence Erlbaum Associates, Inc.

Reis, S.M., & Renzulli, J.S. (2010). Is there still a need for gifted education? An examination of current research. *Learning and Individual Differences*, *20*, 308–317. http://dx.doi.org/10.1016/j.lindif.2009.10.012.

Roll, I., Briseno, A., Yee, N., & Welsh, A. (2014). Not a magic bullet: The effect of scaffolding on knowledge and attitudes in online simulations. *Paper presented at the ICLS 2014, Boulder, Colorado*.

Rutten, N., van Joolingen, W.R., & van der Veen, J.T. (2012). The learning effects of computer simulations in science education. *Computers & Education*, *58*, 136–153. http://dx.doi.org/10.1016/j.compedu.2011.07.017.

Segal, E. (2004). Incubation in insight problem solving. *Creativity Research Journal*, *16*, 141–148. http://dx.doi.org/10.1207/s15326934crj1601_13.

Sio, U.N., & Ormerod, T.C. (2015). Incubation and cueing effects in problem solving: Set aside the difficult problems but focus on the easy ones. *Thinking and Reasoning*, *21*, 113–129. http://dx.doi.org/10.1080/13546783.2014.886626.

Smetana, L.K., & Bell, R.L. (2012). Computer simulations to support science instruction and learning: A critical review of the literature. *International Journal of Science Education*, *34*, 1337–1370. http://dx.doi.org/10.1080/09500693.2011.605182.

vanTassel-Baska, J. (2003). What matters in curriculum for gifted learners: Reflections on theory, research, and practice. In N. Colangelo, & G.A. Davis (Eds.), *Handbook of gifted education* (pp. 174–183) (2nd ed.). Boston: Pearson Education, Inc.

Tomlinson, C.A. (2000). *Differentiation of instruction in the elementary grades (Report No. ED 443572).* Champaign, IL: ERIC Clearinghouse on Elementary and Early Childhood Education.

Vollmeyer, R., & Rheinberg, F. (2000). Does motivation affect performance via persistence? *Learning and Instruction*, *10*, 293–309. http://dx.doi.org/10.1016/S0959-4752(99)00031-6.

Vygotsky, L.S. (1986). *Thought and language.* Cambridge, MA: The MIT Press.

Wang, F., Kinzie, M.B., McGuire, P., & Pan, E. (2010). Applying technology to inquiry-based learning in early childhood education. *Early Childhood Education Journal*, *37*, 381–389. http://dx.doi.org/10.1007/s10643-009-0364-6.

Weinert, F.E., & Helmke, A. (1998). The neglected role of individual differences in theoretical models of cognitive development. *Learning and Instruction*, *8*, 309–323. http://dx.doi.org/10.1016/S0959-4752(97)00024-8.

Wiley, J. (1998). Expertise as mental set: The effects of domain knowledge in creative problem solving. *Memory & Cognition*, *26*, 716–730. http://dx.doi.org/10.4758/bf03211392.

Zacharia, Z.C., Manoli, C., Xenofontos, N., de Jong, T., Pedaste, M., van Riesen, S.A.N., ... Tsourlidaki, E.T. (2015). Identifying potential types of guidance for supporting student inquiry when using virtual and remote labs in science: A literature review. *Educational Technology Research and Development*, *63*, 257–302. http://dx.doi.org/10.1007/s11423-015-9370-0.

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