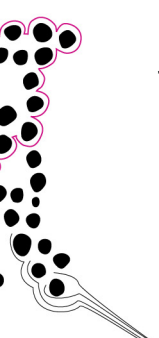
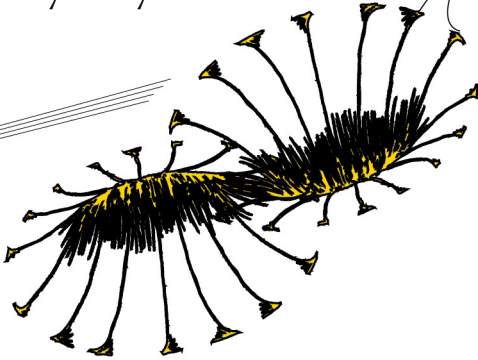




WILL IT FLOAT?



The effect of supporting fourth-grade students in drawing conclusions whilst experimenting in the domain of buoyancy.



Bachelor thesis of

Marit van der Pol

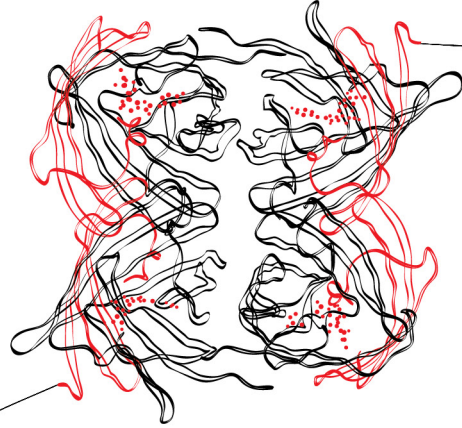
0186716

Psychology

University of Twente

Faculty of Behavioural Sciences

Department of Instructional Technology



Supervised by

Dr. T. H. S. Eysink

Dr. A. H. Gijlers

Enschede, October 7th, 2011



TABLE OF CONTENTS

INTRODUCTION.....	4
<i>Misconceptions regarding floating and sinking</i>	<i>5</i>
<i>Conceptual change: From naïve to scientific</i>	<i>6</i>
<i>Guided inquiry learning: providing students with instructional support</i>	<i>8</i>
<i>Current study: Supporting students in drawing conclusions.....</i>	<i>10</i>
METHOD	11
<i>Participants</i>	<i>11</i>
<i>Materials</i>	<i>12</i>
EXPERIMENTAL SET	12
WORKSHEETS	13
THINKING ALOUD INSTRUCTION.....	14
<i>Inquiry-learning task.....</i>	<i>15</i>
<i>Conditions</i>	<i>16</i>
<i>Questionnaire and tests.....</i>	<i>16</i>
PARTICIPANT CHARACTERISTICS QUESTIONNAIRE.....	16
PRETEST, POSTTEST AND RETENTION TEST	17
<i>Procedure</i>	<i>19</i>
<i>Data Analyses</i>	<i>20</i>
RESULTS.....	21
<i>Reliability of the tests</i>	<i>21</i>
<i>Main effect of condition on the pretest.....</i>	<i>21</i>
MULTIPLE-CHOICE QUESTIONS	22
OPEN QUESTION	22
<i>Main effect of condition on the posttest</i>	<i>22</i>
MULTIPLE-CHOICE QUESTIONS	23
OPEN QUESTION	23

<i>Main effect of condition on the retention test</i>	24
MULTIPLE-CHOICE QUESTIONS	24
OPEN QUESTION	25
<i>Main effect of the inquiry-learning task</i>	25
MULTIPLE-CHOICE QUESTIONS	25
OPEN QUESTION	26
CONCLUSION AND DISCUSSION	28
<i>Hypotheses</i>	28
<i>Comments on the current study</i>	30
<i>Practical implications and suggestions for further research</i>	31
<i>Acknowledgements</i>	32
REFERENCES	33

ABSTRACT

Inquiry learning is a type of learning in which learners have to investigate a domain. Instead of explicitly being told about a domain, the learner has to ask questions, perform experiments, search for explanations, and draw conclusions. By doing so, the learner is supposed to acquire deep knowledge and understanding of the subject matter. However, research has shown that inquiry learning is difficult for learners and support is needed. In this study, 60 children in the fourth grade of elementary education were asked to investigate the domain of buoyancy, whilst either receiving support in drawing conclusions (experimental condition) or not (control condition). They were given a bowl with water and several cubes differing in size, weight, and color and asked to determine which characteristics of the objects influenced the sinking or floating of the objects. All students completed a pretest, posttest and a six-weeks retention test. Results indicate that students in all conditions benefit from the inquiry learning setting; however, students who received support in drawing conclusions achieved higher scores. The question style (open or closed questions) appears to be crucial in testing students' knowledge about the domain. It seems that, in line with former research (Kopitzki & Gijlers, 2011; Zimmerman, 1997), preconceptions and more scientifically sounded concepts remain side-by-side, and when forced to choose, students may prefer to stick to their initial preconception.

Floating

I flattened out my plasticine
And set it in the tank
And do you know the cheeky thing
Turned round and slowly sank.

But when I took it out again
And shaped it like a boat
It wasn't cheeky any more
For it began to float.

(Why did this happen?)

Albert Crawford

INTRODUCTION

Even at a very young age, children will experience the concept of buoyancy when they see objects float or sink in a pool or sea. Because of the previous experiences children have with floating and sinking, they will not enter the classroom as a 'tabula rasa' (blank sheet), but instead have many (naïve) explanations about why things sink or float. Unfortunately, many of these explanations are partially or entirely inadequate from a scientific perspective. This is not surprisingly however, because although buoyancy is a common phenomenon in everyday life, it is a sophisticated science topic. To fully understand why things sink and float, students require complicated knowledge that includes an analysis of forces (buoyancy and gravity) and water pressure. Instead, children will learn about relative density as a simplified explanation for why things sink and float (Pottenger & Young, 1992). Nevertheless, relative density itself is also quite difficult for many children, as density is typically regarded as a higher order concept (Kohn, 1993) involving the ratio of mass to volume (Smith, Snir & Grosslight, 1992) and relative density involves comparing two such ratio variables.

Misconceptions regarding floating and sinking.

Children's inaccurate or incomplete understandings of a scientific phenomenon have been described as preconceptions (Simons, 1999; Ausubel, 1968), alternative frameworks (Driver & Easley, 1978), mental models (Collins & Genter, 1987), intuitive theories (McCloskey & Kargon, 1988), naïve theories (Brewer, 1999) or misconceptions (Chi, 2005) and they tend to be resistant to change (Vosniadou & Brewer, 1992) through instruction (Wandersee, Mintzes & Novak, 1994) This resistance demonstrates how strong our misconceptions can be and how long we can hold on to them. Cognitive theorists such as Jean Piaget would claim that these misconceptions exist in children, because it is beyond the capability of children under the age of nine to fully grasp the concept of buoyancy due to developmental issues (Piaget, 1930). These developmental misconceptions seem to largely center on two ideas.

The first developmental misconception is the confusion between mass and volume. Many students experience the inability to explain volume (size) without referring to mass (weight). When children are asked to explain why a larger item displaces more water than a smaller item, they often persist that it is due to the item's weight, not volume. The second developmental obstacle to understanding density is due to the inability to integrate mass and volume (Kohn, 1993). When asked to address the question of what makes things float, children often focus on one dimension only, as shown by interviews conducted by Smith and colleagues (Smith, Carey & Wiser, 1985; Smith, Maclin, Grosslight & Davis, 1977) with elementary as well as secondary school students. They refer either to the form of an object ("flat objects float"), to its volume ("small things float") or to its mass ("heavy thing sink"). According to Piaget (1930), most children bridge this inconsistency in reasoning by the age of twelve when they reach the formal operational stage and are able to integrate dimensions. Resistance to integrating mass and volume with adult investigators is, however, well documented in Duckworth (1986).

Other misconceptions involve the size-weight illusion, described by Kohn (1993). She found that the greatest number of mistakes in predicting the floating and sinking outcomes of objects arose with those objects that had mass inversely proportional to their volumes. This is mainly due to the fact that most objects that are small tend to be lighter than objects that are large. This confusion is less common in older students than younger ones, but it does persist into adulthood (Duckworth, 1986). Children also have the tendency to provide explanations for individual materials and do not realize that there could be a general explanation

(Biddulph & Osborne, 1984). They often have conceptions that hold true in some situations but not in others. A less prevalent misconception mentioned by both Piaget (1930) and Kohn (1993) involves the amount of water an object is immersed in. Sometimes a student may claim that an object that sinks in a puddle will float in a larger amount of water, such as the sea. This misconception is even held by adults, as Duckworth (1986) documented in her work with eight adult teachers conducting experiments about floating and sinking. It took them more than eight weeks of extensive experimenting with sinking and floating before they began to develop an intuitively grounded concept of density. The idea that a greater amount of water would affect sinking and floating exposes the underlying absent concept of density being a characteristic property of substance. Many children view weight and density as a temporary property of matter (Kohn, 1993; Smith et al., 1985).

Misconceptions on the topic of density can result from the preservation of two parallel explanations of the same concept, a scientific one and a “common sense” one (Kohn, 1993). It seems that many adults and children retain different representations of the same concept within their memory.

Conceptual change: from naïve to scientific concepts.

The process of restructuring underlying misconceptions in order to build scientifically sound conceptions has been labeled conceptual change (Hardy, Jonen, Möller & Stern, 2006; Vosniadou, Ioannides, Dimitrakopoulou & Papademetriou, 2001). Conceptual change research has shown that this process is by no means a sudden shift from a naïve explanation to a scientifically acceptable view but rather a gradual process (Vosniadou et al., 2001) of restructuring interrelated concepts, in the course of which misconceptions may reoccur (Caravita, 2001; Vosniadou et al., 2001; Tyson, Venville, Harrison & Treagust, 1997). This happens because while the shift from misconception through a more coherent pre-conception to the finally scientifically grounded concept occurs, the different concepts remain side by side (Zimmermann, 2007). Dependent on the situation one of the pre-conceptions is chosen as a basis for an explanation of the subject matter.

In the case of the concept of buoyancy, students must simultaneously consider the two dimensions of mass and volume in order to understand the concept of density. To completely apply the notion of density in the context of floating and sinking of objects, an additional conceptual shift needs to come about, which means that students need to

consider the relationship between object and surrounding fluid causal for an object's buoyancy as well. Students need to get confronted with experiences, information, or instruction that is inconsistent with their existing conception of the subject matter in order for them to become dissatisfied with the conceptions they have. This is an essential condition of conceptual change. When students become dissatisfied with the conceptions they have, they experience the need for new explanatory mechanisms and gradually incorporate the new information in their existing explanatory framework. So, in order to challenge the preconception that all floating objects are light, children can be confronted with a comparative light object that sinks (Hardy et al., 2006). Hardy et al. (2006) argue that these kinds of confrontations are crucial for challenging plausible but inapt explanations, and as a result, students may start to consider new ideas.

However, sheer confrontation with experiences that challenge students' naïve ideas is not enough to promote conceptual change (Limón, 2001). The implementation of appropriate scientifically accepted explanations is a constructive process requiring the active cognitive commitment of students (Zimmermann, 2007). Bruner (1961) and Dewey (1938) advocate that learners learn from performing investigations. Observing and explaining counter-evidence to their preconceptions is, however, essential to promote conceptual change (Chinn & Malhotra, 2002) and these direct experiences coupled with instructional support which incorporates discrepant events are essential for helping children change their view of sinking and floating objects (Butts, Hofman & Anderson, 1993). Instructional approaches, that offer students opportunities to discover new principles and stimulate them to engage in meaningful activities and to formulate explanations, are more likely to foster conceptual change (Bransford, Brown & Cocking, 1999). An instructional approach that does just that, is for example (guided) inquiry learning.

Inquiry learning to foster conceptual change.

Inquiry learning is an instructional method based on the constructivistic approach. In this approach a strong emphasis is placed on the learner as an active agent in the process of knowledge acquisition (De Jong & Van Joolingen, 1998). In other words, learners have active experiences with a phenomenon that can lead them to induce characteristics of the domain (De Jong, 2005). The learner must first become oriented to the environment and to prior knowledge related to the phenomenon (De Jong & Van Joolingen, 1998). Considering this

information and supported by guidance in the learning environment, the learner develops hypotheses and tests them by performing one or more experiments (De Jong, 2006). Comparing and relating the collected data makes the underlying ideas and relations between variables salient to the learner. This knowledge can then be integrated into a mental representation and if it becomes clear after reflection that extra information is needed, new experiments can be performed (Klahr & Dunbar, 1988). In short, in a typical inquiry-learning environment, students are confronted with tasks and materials that encourage activities such as exploration, generation of hypotheses, experimentation and reflection. The students can explore and make sense of the material and physical resources in a way that is adjusted to their individual level of prior knowledge and capabilities.

Theories of inquiry learning are usually based on theories of scientific discovery. According to De Jong (2006) the main learning processes in inquiry learning are: *orientation*, *hypothesis generation*, *experimentation*, and *drawing conclusions*. De Jong and Njoo (1992) also make a distinction between regulative (processes that are necessary to control the inquiry learning process) and transformative processes (processes that directly yield knowledge). Essential transformative processes such as designing experiments and interpreting experimental outcomes are of main interest.

Guided inquiry learning: providing students with instructional support

Several studies have indicated that learners may encounter several difficulties in inquiry learning environments (De Jong & Van Joolingen 1998). For example, learners may show “floundering behavior,” that is, performing experiments in an unsystematic way; they may find it hard to state hypotheses (Njoo & De Jong, 1993; Klayman & Ha, 1987) and they often draw conclusions that cannot be drawn from the data (Klahr & Dunbar, 1988). Also, students often tend to focus on observations that are in line with their initial beliefs (Quinn and Alessi, 1994) or they provide explanations that do not question their initial beliefs; this is also known as the confirmation bias. Some students have a strong inclination to search for evidence that support their current hypothesis, even when they are confronted with inconsistent evidence (Dunbar, 1993).

Mayer (2004) pointed out that the translation of principles of constructivist learning into learning environments has largely followed the simple formula “constructivism = hands-on activity”. As a consequence, it may happen that, during an open-ended and thus highly

self-directed learning activity, students do not actually reflect on the relevant concepts, even if the material provided is closely relevant to the phenomenon to be studied, and may perform below expectation. If the inquiry task or setting is too complex and it leaves students too much freedom with respect to directing their own learning, students may be enthusiastic and actively involved, but they might not arrive at the scientific conclusions intended because they do not integrate their sense making into a conceptual framework (Hardy et al., 2006; Kirschner, Sweller & Clark, 2006). Thus, one of the problems might be that students are simply overburdened with the freedom and complexity of unguided inquiry learning settings. Therefore, learners require support for their inquiry learning processes to make it an efficient and effective learning process (De Jong, 2006; De Jong, 2005; Mayer, 2004; De Jong & Van Joolingen, 1998).

Hardy et al. (2006) state that the two key elements of instructional support are the structuring of tasks to allow students to remain focused on important aspects and the support of students' reflection on their insights within a larger context of scientific reasoning (Reiser, 2004). The latter can enable students to make important observations so that they are able to question their prior beliefs (Zimmerman, 2007). Further support can therefore be a matter of making students conscious about contradicting information between their prior knowledge and thoughts, and the real world with its scientific explanations. Students could be stimulated to think about those contradictions by the use of verbal prompts given by, for example, the teacher. This may mean that the teacher provides refocusing statements or challenges assumptions or interpretations made by students (Hogan & Pressley, 1997). Prompting students to evaluate and reflect on their experimental findings might therefore enhance their scientific understanding. Especially prompting for self-explanations promotes understanding (Chi, 1996). Furthermore, De Jong and Van Joolingen (1998) found that successful students take more notes during learning. Inviting students to keep record of their on-going experiments on worksheets by note taking makes it possible to free more memory capacity. This is important since primary school students typically underestimate their memory limitations, when involved in an inquiry-learning task and do not spontaneously keep notes (Zimmermann, 2007).

Several studies suggest that students working in a guided inquiry learning setting achieve a higher degree of conceptual understanding than students in environments of direct instruction (Deslauriers, Schelew & Wieman, 2011; Alfieri, Brooks, Aldrich & Tenenbaum, 2011; Eysink, De Jong, Berthold, Kolloffel, Opfermann & Wouters, 2009; Linn, Lee, Tinker,

Husic & Chiu, 2006; Staub & Stern, 2002; Christianson & Fisher, 1999). We can support students in the orientation phase, the hypothesis generation phase, the experimentation phase and in drawing conclusions. Support in all of these phases will be beneficial for the inquiry learning outcomes, since they are all essential processes in experimenting. In this study, however, we focus on drawing conclusions. From literature we learn that students not only have difficulty drawing conclusions, but they actually fail to draw conclusions from their data (Klahr & Dunbar, 1998). Since experimenting without drawing conclusions is quite useless, supporting students in this phase is crucial for good inquiry-based learning outcomes.

Current study: supporting students in drawing conclusions

The current study wants to contribute to the question of how conceptual change concerning buoyancy may be effectively realized in elementary school's science education using the principle of guided inquiry learning. The key question that this study tries to answer is: What is the effect of supporting elementary school students in drawing conclusions in an inquiry-learning based environment with regard to the domain of floating and sinking on students' learning outcomes?

The learning outcomes of students in two different conditions, who worked in an inquiry-learning environment, have been compared. Both environments were designed to facilitate a high degree of experimentation with relevant material, allowing the child to discover the principles of density and buoyancy. Both conditions were identical, apart from the degree of support students received in drawing conclusions. Students in the so-called control condition were given a small amount of support by means of a worksheet to take notes about predictions and observations of each object they placed into the water. In the experimental condition this worksheet has been expanded with an extra column where students could write down their conclusions. Explicitly requesting students to predict, observe and explain their experiments is an instructional strategy to promote conceptual change (White & Gunstone, 1992). Because of the preconceptions children have, their observations are often inconsistent with their predictions. By creating cognitive dissonance and surprise, predict-observe-explain helps students realize the limitations of their preconceptions and it gets them ready to learn scientific theories (Yin, Tomita & Shavelson, 2008).

According to Koran and Longino (1982) curious students achieve better and deeper understanding than students with lower curiosity levels, probably because curious students explore events and objects longer and intensive and they use many more senses; this is important in inquiry learning since students will have to experiment on their own. They also reported that curious students recall experiences longer. In this study a science curiosity measure is conducted, based on the Children's Science Curiosity Scale that Harty and Beall (1984) introduced.

We compared the learning outcomes of students in two conditions (control and experimental) that worked in an inquiry-learning environment regarding floating and sinking, using a pretest, posttest and six weeks retention test design. The above-mentioned considerations lead to the following hypotheses.

First of all, we expect all students to perform better (i.e., achieve higher scores) on the posttest and retention test than on the pretest, because of the inquiry-learning task, where students will gain knowledge about the field of buoyancy. We also expect that students in both conditions will retain their gained knowledge over time; in other words, we expect to find no differences in scores between the posttest and the retention test. Secondly, we expect that students in the experimental condition will achieve higher scores on the posttest and retention test than students in the control condition, because drawing conclusions will help them to achieve a deeper understanding of the subject matter and it will help them to retain the gained knowledge for longer periods of time. Finally, we expect students with high scores on scientific curiosity to achieve higher scores on the posttest and retention test than students with low scientific curiosity regardless of the condition they are in, based on research by Koran and Longino (1982).

METHOD

Participants.

The data file on which this research is based contained the data of 60 participants (29 male, 31 female). Students who incompletely filled in parts of the test, missed the retention test or whose parents did not send in their parental consent, were omitted from the original sample of 69 students. All participants were in the fourth grade of elementary education (ages eight to eleven years old). The average age of the children was nine years and seven months

(114.80 months), with a standard deviation of six months (6.21 months). All students were randomly assigned to either the experimental condition (30 participants) or the control condition (30 participants). The participants completed the experiment during school time. They did not receive any compensation for their participation.

Materials.

EXPERIMENTAL SET.

The experimental set consisted of 17 cubes differing in size (volume), weight (mass), material, color and density (see Table 1 for an overview of the cubes that students could use to experiment with). The cubes were positioned on a placemat and could be weighted on a scale with 1-gram accuracy.

Cubes	Size (volume)	Weight (mass)	Material	Color	Density (water=1,0)	
A1	Large	137 g	Brass	Gold	8,0 g/ml	Sink
A2	Medium	66 g	Brass	Gold	8,0 g/ml	Sink
A3	Small	8 g	Brass	Gold	8,0 g/ml	Sink
B1	Large	45 g	Aluminum	Silver	2,7 g/ml	Sink
B2	Medium	22 g	Aluminum	Silver	2,7 g/ml	Sink
B3	Small	3 g	Aluminum	Silver	2,7 g/ml	Sink
C1	Large	10 g	Pine	Thick grained	0,35 – 0,60 g/ml	Float
C2	Medium	6 g	Pine	Thick grained	0,35 – 0,60 g/ml	Float
C3	Small	0 g	Pine	Thick grained	0,35 – 0,60 g/ml	Float
D1	Large	13 g	Oak	Thin grained	0,60 – 0,90 g/ml	Float
D2	Medium	7 g	Oak	Thin grained	0,60 – 0,90 g/ml	Float
D3	Small	1 g	Oak	Thin grained	0,60 – 0,90 g/ml	Float
E	Large	20 g	Lignum vitae	Dark wood	1,28 – 1,37 g/ml	Sink
F	Large	21 g	PVC	Grey	1,39 – 1,42 g/ml	Sink
G	Large	14 g	Polypropylene	White	0,85 – 0,95 g/ml	Float
H	Large	18 g	Nylon	Opaque white	1,13 g/ml	Sink
I	Large	20 g	Acrylic	Clear	1,16 – 1,19 g/ml	Sink

Table 1. An overview of the characteristics of the 17 cubes.

WORKSHEET.

Participants had to fill in a worksheet whilst experimenting with the experimental set. See Figure 1 for the worksheet that the students in the control condition received. Through this worksheet, students were supported in the hypothesis generation phase and the experimentation phase. Students had to write down which cube they had chosen, and whether they expected it to float or sink (by ticking the corresponding box). After having put the cube in the water, they had to write down on the worksheet whether the cube floated or sank (by ticking the corresponding box). The weight of the cubes could be written down in the margin.



Item	I guess, that it will 		I see, that it 	
	Float	Sink	Floats	Sinks

Figure 1. Worksheet for students in the control condition.

Besides support in hypothesis generation and experimentation, students in the experimental condition were also supported in the conclusion phase. They received a worksheet with an added column where they could write down their conclusions (see Figure 2). This worksheet followed a predict-observe-explain schema (Yin et al., 2008; White & Gunstone, 1992).



Item	I guess, that it will 		I see, that it 		Now, I know that
	Float	Sink	Floats	Sinks	

Figure 2. Worksheet for students in the experimental condition.

THINKING ALOUD INSTRUCTION.

Students received a think-aloud instruction consisting of a think-aloud example and think-aloud practice. Both involved solving a tangram (a Chinese dissection puzzle of seven pieces) while thinking aloud. The experimenter provided the example by solving a tangram while thinking aloud. The experimenter (a) thought about a possible solution but (b) discovered and indicated that it is not the right solution, (c) told what she thought of the assignment, (d) thought of an action plan for arriving at the solution, and (e) found the correct solution.



Figure 3. The tangram puzzle students had to solve while thinking aloud.

After this example the experimenter discussed it with the participant. After the think-aloud example, the learners had to try to solve a different tangram (see Figure 3) themselves while thinking aloud. Whenever the student fell silent, the experimenter prompted the learner to continue talking. After practicing, the experimenter indicated what went well and why, and how the learner could improve the think-aloud process.

Inquiry-learning task.

The aim of the inquiry-learning task was for students to discover what makes objects float or sink. This was reached by handing participants a bowl full of water and the 17 cubes. Children also received a worksheet and were instructed to choose a cube, write down which cube they chose and whether they thought if the cube would float or sink. Afterwards they placed the cube in the bowl of water and wrote down their findings. Students could also use the scale to weigh the cubes. The experimenter showed them the procedure using a balloon and a spoon. Participants were free to decide in which order they picked a cube and how many cubes they wanted to place in the water. The students were asked to speak aloud about what they were doing and thinking.

Conditions.

Participants were randomly assigned to either the experimental or the control condition. In the control condition, participants could work independently on the experimental task. Students in the experimental condition received support in drawing conclusions through a worksheet with an added column where they could write down their conclusions and through verbal prompts given by the experimenter when they did not write anything down. After placing three to five cubes (depending on whether it was already possible to draw solid conclusions) in the water, the experimenter pointed at the students' worksheet and asked for instance, "Do you know why that cube sank?", "All the A-cubes sank, why would they all sink?" (after all A, B, C or D cubes were used) or "You expected that cube to float because it's small, why did it sink?" (when students expected a different outcome). This was repeated every three to five cubes (depending on whether it was possible to draw solid conclusions, and when the student did not write anything down).

Questionnaire and tests.

PARTICIPANT CHARACTERISTICS QUESTIONNAIRE.

The participant characteristics questionnaire consisted of questions concerning personal data (date of birth, gender, etc.) and scientific curiosity. The latter questions were based on the Children's Science Curiosity Scale (Harty & Beall 1984). The scale consists of 30 Likert-type items. Participants are asked to rate the items on a scale of 5 (strongly agree), 4 (agree), 3 (uncertain), 2 (disagree) and 1 (strongly disagree). The scale is also illustrated by the five "smiley faces" technique for students who have interpretation difficulties and for purpose in the second and third grades. The possible score range runs from low curiosity (30) to high curiosity (150). In order to administer these items to Dutch nine-year olds in a short amount of time only a few items were chosen, translated and slightly adjusted to fit in modern Dutch society (popular informative children's programs as "Klokhuis" and "Jeugdjournaal" and newspapers and scientific magazines targeted at children, such as "Kidsweek" en "Zo zit dat" were integrated in the existing questions). The items that were used in this research are items 1, 2, 6, 8, 11, 15 and 23 (Harty & Beall, 1984).

After gathering the data the internal consistency was calculated. With an *alpha* of 0.53, the internal consistency is considered to be poor. The *alpha* range (with the given item deleted) for the items was 0.41 to 0.59. The item-total correlation coefficients ranged from -

0.01 to 0.49. Shrigley (1983) suggested that an item-total correlation coefficient should be at least 0.30 in order for the discrimination index of an item to be respectable. Based on that, item 11 should be discarded because of an item-total correlation of -0.01. The internal consistency then increases to 0.59, still a poor internal consistency. Item 2 should also be discarded because of an item-total correlation of 0.21, which would increase internal consistency to 0.60. Further discarding would not be beneficial to the internal consistency.


Through factor analysis for all seven items we found three factors; after removing item 11 there were only two factors left. Strikingly only one item loaded on the second factor, namely item 2. Based on both item-total correlation analysis and factor analysis items 2 and 11 were discarded, leaving us with five items with an internal consistency of 0.60, a questionable result. Since reliability of a scale should be at least 0.7, this scale will not be used as an independent variable.

PRETEST, POSTTEST AND RETENTION TEST.

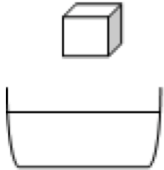
A pretest, a posttest and a retention test were administered to measure the learners' knowledge about floating and sinking. All three tests were based on the pre-conception test by Yin et al. (2008) and consisted of one open question and 25 multiple-choice questions clustered in five different exercises (mc_volume (questions regarding the influence of the size of a cube), mc_mass (questions regarding the influence of the weight of the cube), mc_color (questions regarding the influence of the cube's color), mc_position (questions regarding the influence of the positioning of the cubes relative to the water) and mc_extrapolate (questions regarding the continuity of characteristics)). The open question allowed students to give a more creative and fully self-invented answer. No forced choice and answer directions were given in order to reveal possible hidden knowledge.

In the pretest students started with the open question (to activate students' prior knowledge about floating and sinking), followed by the multiple-choice questions. In both the posttest and retention test the order was reversed; the open question was used as a concluding question about the reason why things float or sink. The pre-test and the post-test took approximately fifteen minutes each, whereas the six-weeks retention test took about five to ten minutes because children did not have to think aloud. The posttest was used as an indicator of immediate effects and the retention test was administered to investigate long-term effects of the interventions.


Mc_volume contained five questions about the influence of the size (volume) of the cube on floating or sinking (see Figure 4) whereas mc_mass asked five questions about the influence of the weight (mass) of the cube on floating and sinking (“This cube is floating/sinking; if this cube would be ten times heavier/lighter, would this cube float or sink?” and “When I know the weight of a cube, I know whether this cube will float or sink. Is this statement true or false?”).

This cube is floating. 

1a) If a different cube, made of the same material, would be 10 times as big, would that cube float or sink?

 That cube would:
 A) float
 B) sink
 C) You can't say, because _____

1b) If a different cube, made of the same material, would be 10 times as small, would that cube float or sink?

 That cube would:
 A) float
 B) sink
 C) You can't say, because _____

3) When I know the size of a cube, I know whether this cube will float or sink. Is this sentence true or false? _____

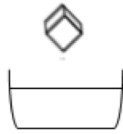
Figure 4. Example of mc_volume: questions concerning the size of the cube.

Mc_color consisted of five questions about whether the color of the cube would influence floating or sinking (“This cube is floating/sinking; if this cube would be of a far darker/lighter color, would this cube float or sink?” and “When I know which color a cube has, I know whether this cube will float or sink. Is this statement true or false?”) and mc_position contained five questions about the influence of the position of the cube relative to the water (see Figure 5). Finally, mc_extrapolate tested the learners’ ability to extend their prior knowledge to a new problem in five different questions (see Figure 6).

This cube is floating.



1a) If I dropped this cube in the water with the edge down, would this cube float or sink?



This cube would:

- A) float
- B) sink
- C) You can't say, because _____

1b) If I dropped this cube in the water with the flat side down, would this cube float or sink?



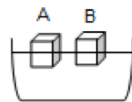
This cube would:

- A) float
- B) sink
- C) You can't say, because _____

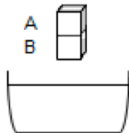
3) When I know in which way a cube falls in the water (with the flat side or with the edge), I know whether this cube will float or sink.
Is this sentence true or false? _____

Figure 5. Example of mc_position: questions concerning the position of the cube compared to the water.

Cube A and cube B are floating.



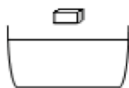
1a) If I glued cube A and cube B firmly together, would the cubes float or sink?



Both cubes would:

- A) float
- B) sink
- C) You can't say, because _____

1b) If I would saw cube A precisely in the middle, would this cube (exactly half of cube A) float or sink?



That cube would:

- A) float
- B) sink
- C) You can't say, because _____

3) When I know whether cubes float or sink, I know whether cubes that are glued together or sawn cubes will float or sink.
Is this sentence true or false? _____

Figure 6. Example of mc_extrapolate: questions concerning the continuity of characteristics.

Procedure.

After the experimenter introduced herself to the class and gave some basic information about the experiment, participants completed the participant characteristics questionnaire in the classroom. In a separate quiet room, participants were then tested individually in one session of 45 – 60 minutes. The session started with a short introduction and a few small talk questions to make the children feel more comfortable. Then the participants were told that they would have to think aloud. They received think-aloud instruction consisting of the think-aloud example and think-aloud practice. Whenever the student fell silent, the experimenter prompted the learner to continue talking. Afterwards, the experimenter indicated what went well and why, and how the learner could improve the think-aloud process. The experimenter then pinned a microphone on the student's clothes and turned on the camera, after explaining why the camera was used.

Subsequently participants received the pretest, which they had to fill in while thinking aloud. After completing the pretest, the students worked with the experimental task. The experimenter explained that the children could use the 17 cubes to experiment and test their own ideas about floating and sinking and demonstrated the procedure with a balloon and a spoon to make sure the participants knew what they had to do. She emphasized that the students had to write down which cube they used and what they expected before they placed the cube in the water. After placing the cube in the water they had to write down their findings. Students used the worksheet (see Figure 1 for the worksheet in the control condition and Figure 2 for the worksheet in the experimental condition) to write down their expectations and findings. When students forgot to write down their expectations or findings, they were prompted to do so ("What do you expect to happen with this cube?" and "What do you see?"). Children in the experimental condition also had to write down their conclusions. When they forgot to do so, they were prompted to draw conclusions (see "Conditions" for a detailed overview).

The children were given the time to experiment with as many cubes as they liked. When students stopped after only three to seven cubes, the experimenter asked them if they had tried everything they wanted to learn (after which most of the students would try some more cubes and others indicated that they had tried everything they wanted). When the students would continue experimenting for more than fifteen minutes, the experimenter asked them whether they already knew why objects float or sink (after which most students

would indicate that they knew why objects float or sink and stopped experimenting, and others would try another one or two cubes or combinations of cubes).

Finally, participants had to fill in the posttest while thinking aloud. The children were then instructed not to communicate with the other students about the content and their findings, after which they could return to their classroom. Six weeks after these individual sessions the experimenter returned to the class and participants filled in the retention test in the classroom without thinking aloud.

Data Analyses

After gathering all data (school, gender, age, condition and answers on the questions of the pretest, posttest and retention test), the raw data was encoded. Whilst encoding the age in months, students born in the same month were considered to be of equal age.

Participants could receive one point for answering each multiple-choice question correctly. The scores on the multiple-choice questions are indicators of the misconceptions the children held in the measured dimensions (i.e., size (volume), weight (mass), color, position relative to the water and continuity of characteristics). The maximum score was 25 points (5 points per measured dimension).

The open question of the test indicates general knowledge about the phenomenon. For this last question, the coding scheme developed by Kopitzki and Gijlers (2011) was used. One point was given for phrases like “weight/size alone is not important”, “it floats, because it contains (more) air” and “it sinks, because it contains (more) metal/stone”. Two points could be received for phrases like “the material is important” and “it sinks, because it is heavy for the water” and three points were awarded for more mature explanations like “the proportion of size and weight is important”. Immature pre-concepts, such as “lighter items float and heavier items sink” or explanations based on individual experiences, like “When I threw a stone into the water it sank” and “My air mattress floats in the sea”, were not rewarded with points. The maximum score was 3 points.

RESULTS

In the following section the results on the tests are presented. Differences between the control condition and the experimental condition have been explored based on the scores of the multiple-choice questions and the open question on the pretest, posttest and retention test. Considering research from Kopitzki and Gijlers (2011), the scores on the open question and the scores on the multiple-choice questions have been explored and tested separately. Furthermore, the multiple-choice questions have been subdivided in the subscales mc_volume, mc_mass, mc_color, mc_position and mc_extraploate and differences between the control condition and experimental condition on the pretest, posttest and retention test have been explored in these subscales as well. Mean scores and standard deviations can be found in Table 2 (multiple-choice questions and open question) and Table 3 (mc_volume, mc_mass, mc_color, mc_position and mc_extraploate). Mean scores have been visualized in Figure 7 (multiple-choice questions and open question), and Figure 8 (mc_volume, mc_mass, mc_color, mc_position and mc_extrapolate).

Reliability of the tests

The reliability of the multiple-choice questions on the pretest, posttest and retention test, as measured with Cronbach's α , was $\alpha = 0.63$, $\alpha = 0.75$, and $\alpha = 0.80$, respectively. Deleting items from the tests did not lead to significantly higher reliabilities. The reliability of the three tests concerning the open question has been calculated through the inter-rater reliability. A second coder rated 20% of the pretest, posttest and retention test. Cohen's Kappa on the tests reached 0.92, which can be considered excellent.

Main effect of condition on the pretest

The mean scores and standard deviations can be found in Table 2. Mean scores have also been visualized in Figure 7 (multiple-choice questions and open question).

Question type	Condition	Mean scores and standard deviations					
		Pretest		Posttest		Retention test	
		Mean	SD	Mean	SD	Mean	SD
Open	Control	0.60	0.77	0.30	0.60	0.27	0.58
	Experimental	0.63	0.72	0.73	0.87	0.80	0.92
Multiple-choice	Control	14.77	3.24	16.20	3.67	16.17	3.70
	Experimental	15.53	3.36	17.73	3.55	17.03	4.66

Table 2. Mean scores and standard deviations on the open question and multiple-choice questions of the pretest, posttest and retention test for the control and experimental condition.

MULTIPLE-CHOICE QUESTIONS

Scores on the pretest were normally distributed for both the control condition (skewness: 0.27; Shapiro-Wilk's test of normality: $W = 0.97$ ($p = 0.54$)) and the experimental condition (skewness: 0.04; Shapiro-Wilk's test of normality: $W = 0.97$ ($p = 0.46$)). An independent samples t-test showed no significant differences between the control and the experimental condition ($t = -0.90$, $p = 0.37$).

OPEN QUESTION

Scores on the pretest were not normally distributed for both the control condition (Shapiro-Wilk's test of normality: $W = 0.72$ ($p < 0.001$)) and the experimental condition (Shapiro-Wilk's test of normality: $W = 0.76$ ($p < 0.001$)). An independent samples Mann-Whitney U test showed no significant differences between the control and experimental condition ($U = 468.50$, $z = 0.30$, $p = 0.76$, $r = 0.04$).

Main effect of condition on the posttest

The mean scores and standard deviations can be found in Table 2 (multiple-choice questions and open question) and Table 3 (mc_volume, mc_mass, mc_color, mc_position and mc_extrapolate). Mean scores have also been visualized in Figure 7 (multiple-choice questions and open question) and Figure 8 (mc_volume, mc_mass, mc_color, mc_position and mc_extrapolate).

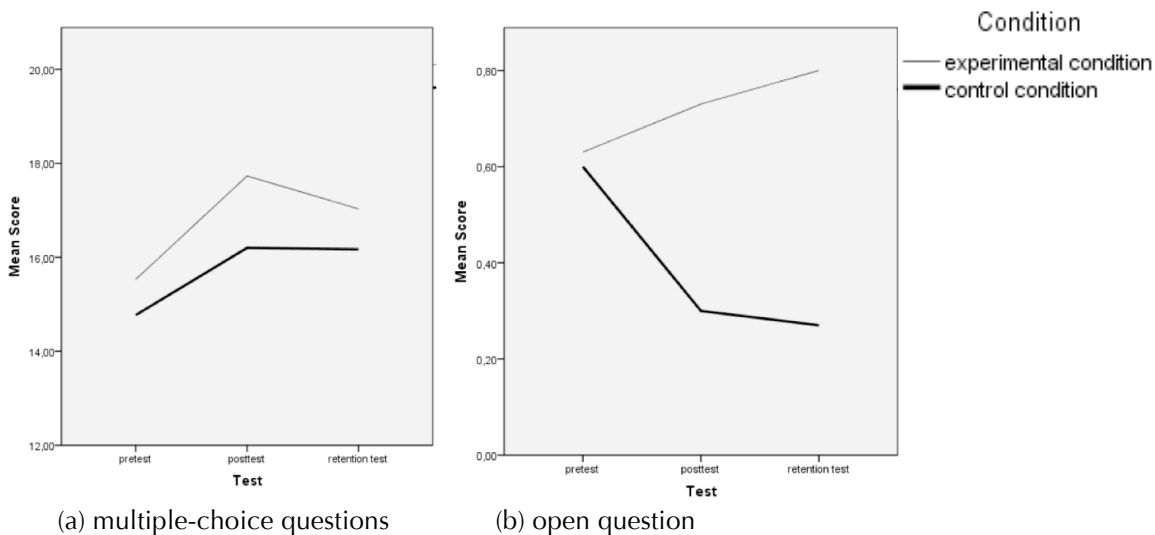


Figure 7. Mean scores on the multiple-choice questions (a) and the open question (b) of the pretest, posttest and retention test for the control and experimental condition.

MULTIPLE-CHOICE QUESTIONS

Scores on the posttest were normally distributed for both the control condition (skewness: -0.17; Shapiro-Wilk's test of normality: $W = 0.96$, $p = 0.26$) and the experimental condition (skewness -0.45; Shapiro-Wilk's test of normality: $W = 0.97$, $p = 0.62$). An independent samples t-test showed no significant differences between the control and experimental condition ($t = -1.64$, $p = 0.11$). When splitting the multiple-choice questions into the five different segments, an independent samples Mann-Whitney U test showed a significant difference between the control and the experimental condition ($U = 606.00$, $z = 2.37$, $p < 0.05$, $r = 0.31$) on *mc_volume*.

OPEN QUESTION

Scores on the posttest were not normally distributed for both the control condition (Shapiro-Wilk's test of normality: $W = 0.56$, $p < 0.001$) and the experimental condition (Shapiro-Wilk's test of normality: $W = 0.73$, $p < 0.001$). An independent samples Mann-Whitney U test showed significant differences between the control and experimental condition ($U = 569.00$, $z = 2.08$, $p < 0.05$, $r = 0.27$).

Question	Condition	Mean scores and standard deviations					
		Pretest		Posttest		Retention test	
		Mean	SD	Mean	SD	Mean	SD
mc_volume	Control	2.37	0.93	2.87	1.31	2.97	1.19
	Experimental	2.73	1.14	3.63	1.27	3.63	1.19
mc_mass	Control	2.13	0.63	1.93	0.64	1.97	0.56
	Experimental	2.17	0.70	2.20	0.85	2.33	0.92
mc_color	Control	3.57	1.43	4.00	1.34	3.90	1.35
	Experimental	3.83	1.34	3.87	1.25	3.77	1.33
mc_position	Control	3.77	1.22	4.07	1.26	3.83	1.21
	Experimental	3.60	1.22	4.23	0.94	3.80	1.47
mc_extrapolate	Control	2.93	1.08	3.33	1.03	3.50	1.01
	Experimental	3.20	0.96	3.80	1.13	3.50	1.14

Table 3. Mean scores and standard deviations on mc_volume, mc_mass, mc_color, mc_position and mc_extrapolate of the pretest, posttest and retention test for the control and experimental condition.

Main effect of condition on the retention test

The mean scores and standard deviations can be found in Table 2 (multiple-choice questions and open question) and Table 3 (mc_volume, mc_mass, mc_color, mc_position and mc_extrapolate). Mean scores have also been visualized in Figure 7 (multiple-choice questions and open question) and Figure 8 (mc_volume, mc_mass, mc_color, mc_position and mc_extrapolate).

MULTIPLE-CHOICE QUESTIONS

Scores on the retention test for the experimental condition are normally distributed (skewness: -0.29; Shapiro-Wilk's test of normality: $W = 0.94$, $p = 0.11$), but scores on the retention test for the control condition are not (Shapiro-Wilk's test of normality: $W = 0.93$, $p < 0.05$). An independent samples Mann-Whitney U test showed no significant differences between the control and experimental condition ($U = 508.00$, $z = 0.86$, $p = 0.39$, $r = 0.11$). When splitting the multiple-choice questions into the five different segments, an independent

samples Mann-Whitney U test showed a significant difference between the control and the experimental condition ($U = 592.00$, $z = 2.19$, $p < 0.05$, $r = 0.28$) on *mc_volume*.

OPEN QUESTION

Scores on the retention test are not normally distributed for both the control condition (Shapiro-Wilk's test of normality: $W = 0.52$, $p < 0.001$) and the experimental condition (Shapiro-Wilk's test of normality: $W = 0.70$, $p < 0.001$). An independent samples Mann-Whitney U test showed significant differences between the control and experimental condition ($U = 586.00$, $z = 2.41$, $p < 0.05$, $r = 0.31$).

Main effect of the inquiry-learning task

To test whether students actually gained knowledge about the domain of floating and sinking, a Repeated Measures ANOVA (with students' scores on the open question and the multiple-choice questions on the pretest, posttest and retention test as within-subjects variable and condition as between-subjects variable) has been executed. The mean scores and standard deviations can be found in Table 2 (multiple-choice questions and open question) and Table 3 (*mc_volume*, *mc_mass*, *mc_color*, *mc_position* and *mc_extrapolate*). Mean scores have also been visualized in Figure 7 (multiple-choice questions and open question) and Figure 8 (*mc_volume*, *mc_mass*, *mc_color*, *mc_position* and *mc_extrapolate*).

MULTIPLE-CHOICE QUESTIONS

A post-hoc Bonferroni comparison showed a significant increase in the test scores between the pretest and the posttest ($p < 0.01$) and the retention test ($p < 0.05$) in the control condition and between the pretest and retention test in the experimental condition ($p < 0.001$). No significant decline has been found between the posttest and retention test in both conditions ($p = 1.00$). When splitting the multiple-choice questions into five different segments (*mc_volume*, *mc_mass*, *mc_color*, *mc_position*, *mc_extrapolate*), an independent samples Mann-Whitney U test revealed a few significant effects.

In both the control condition ($F(2, 58) = 4.37$, $p < 0.05$) and the experimental condition ($F(2, 58) = 7.14$, $p < 0.01$) significant differences have been found on *mc_volume*. A post-hoc Bonferroni comparison showed that students in the control condition score significantly higher on the retention test than on the pretest ($p < 0.05$). In the experimental

condition, students score significantly higher on the posttest ($p < 0.01$) and retention test ($p < 0.01$) than on the pretest. No significant decline has been found between the posttest and the retention test ($p = 1.00$).

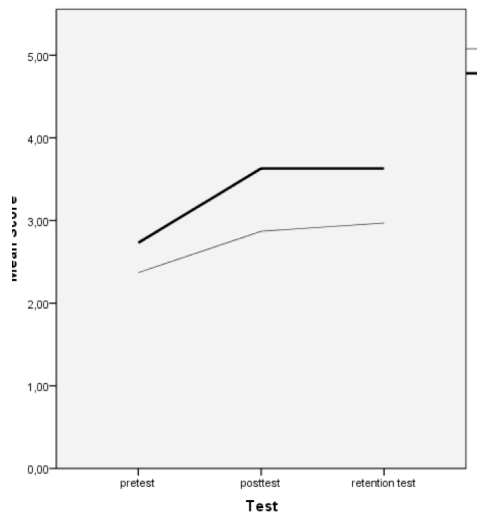
No significant differences have been found on *mc_mass* and *mc_color*. However, interesting trends can be spotted in *mc_mass* (see Figure 8b) and *mc_color* (see Figure 8c). In *mc_mass*, scores of students in the control condition and in *mc_color*, scores of students in the experimental condition, seem to decrease after the inquiry-learning task, whereas all other scores (seem to) increase over time.

On *mc_position* a significant difference has been found in the experimental condition ($F(2, 58) = 4.30, p < 0.05$). A post-hoc Bonferroni comparison showed a significant increase in the scores between the pretest and the posttest ($p < 0.05$). No significant decline has been found between the posttest and the retention test ($p = 0.22$).

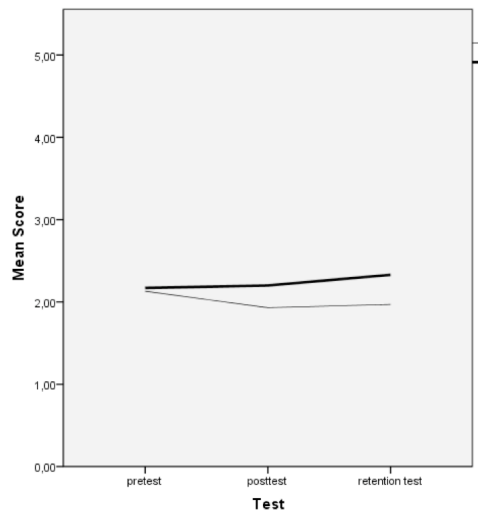
In both the control ($F(2, 58) = 6.63, p < 0.01$) and the experimental condition ($F(2, 58) = 4.53, p < 0.05$) significant differences have been found on *mc_extrapolate*. A post-hoc Bonferroni comparison showed a significant increase in the scores on the pretest and the retention test in both conditions ($p < 0.01$). In the experimental condition a significant increase has been found in the test scores between the pretest and the posttest ($p < 0.01$). No significant decline has been found between the posttest and the retention test ($p = 0.46$).

OPEN QUESTION

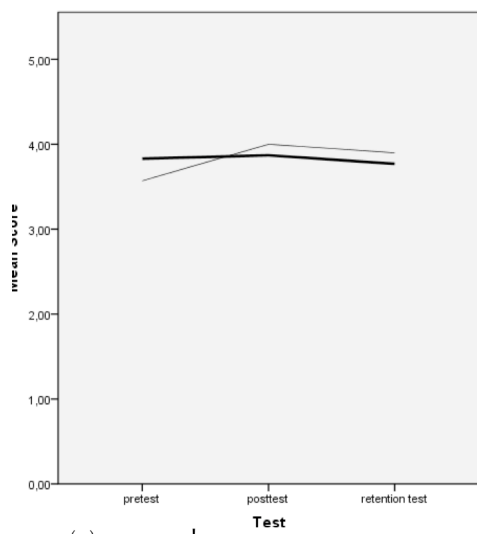
Scores in the control condition decreased over time (a post-hoc Bonferroni comparison showed significant ($p < 0.05$) difference between the pretest and the retention test), whereas the scores on the open question in the experimental condition even seem to increase based on Figure 7b, but this turned out to be non significant ($p = 1.00$).



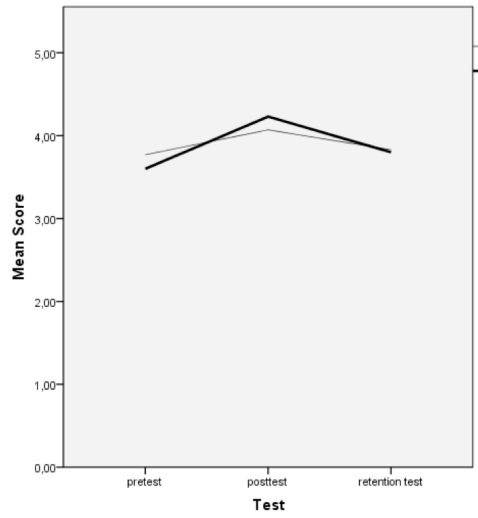
(a) mc_volume



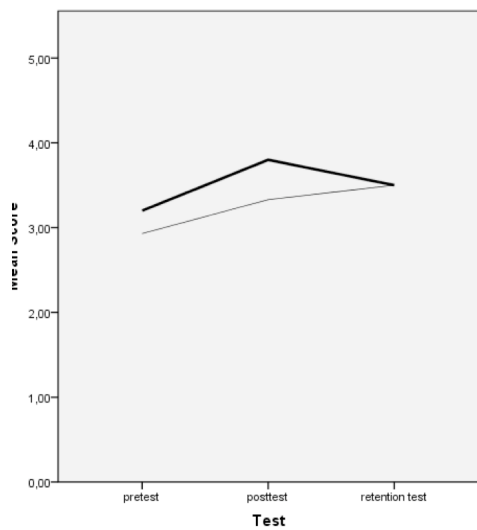
(b) mc_mass



(c) mc_color



(d) mc_position



(e) mc_extrapolate

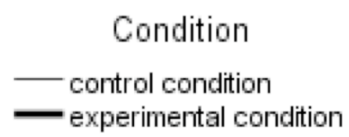


Figure 8. Mean scores on mc_volume (a), mc_mass (b), mc_color (c), mc_position (d) and mc_extrapolate (e) of the pretest, posttest and retention test for the control and experimental condition.

CONCLUSION AND DISCUSSION

The aim of the present study was to investigate the effect of supporting students in drawing conclusions on their learning outcomes whilst working on an inquiry-learning task. We compared the learning outcomes of students who worked on an inquiry-learning task in the domain of floating and sinking in either the control condition or the experimental condition. The control condition contained a worksheet to keep notes about predictions and observations of each object immersed into water. In the experimental condition the worksheet was extended with the request to explain why the object actually sank or floated. Students were also prompted to do so by the experimenter when they did not spontaneously write their conclusions down.

It was hypothesized that (a) all students achieved higher scores on the multiple-choice questions and the open question of the posttest and retention test than on the pretest and that (b) they would retain their gained knowledge over time. We also expected that (c) students in the experimental condition would achieve higher scores on both the open question and the multiple-choice questions of the posttest and retention test than students in the control condition. Finally, we expected that (d) students with high scores on scientific curiosity achieved higher scores on the posttest and retention test than students with low scientific curiosity regardless of the condition they are in.

Hypotheses

First of all, we expected that students in both the control condition and the experimental condition achieved higher scores on the multiple-choice questions and the open question of the posttest and the retention test in comparison to the pretest. Our hypotheses were partially confirmed by our data. In the control condition students did achieve significantly higher scores on the multiple-choice questions of the posttest and the retention test than on the pretest. In the experimental condition however, only the scores on the retention test were significantly higher than on the pretest. In terms of the scores on the open question, students in the experimental condition did not achieve significantly higher scores on the posttest and the retention test than on the pretest; students in the control condition even achieved a significantly lower score on the retention test than on the pretest (see Figure 7). This might be due to confusing experiences the students gained. Students in the control condition did not receive support in drawing conclusions, and this might have prevented them from

experimenting more systematically and integrating their new experiences in their own adjusted theory. Working with the inquiry-learning task only confused them, which led to achieving lower scores on the posttest and retention test than on the pretest.

After dividing the multiple-choice questions into the five different subscales, two significant increases have been found in the control condition. Students achieved higher scores on mc_volume and mc_extrapolate on the retention test than on the pretest. In the experimental condition three significant increases have been found. Students achieved higher scores on mc_volume and mc_extrapolate on both the posttest and the retention test compared with the pretest and they also achieved higher scores on mc_position on the posttest than on the pretest. This might mean that through experimenting students learned that volume alone is not important, that two floating objects still float when you glue them together and that when an object sinks, it also sinks when you saw it in smaller pieces. Because of the support in drawing conclusions, students in the experimental condition also learned that it does not matter how an object is immersed into water; that is, an object that floats will always float, no matter how you throw it into the water.

We also expected that students in both the control condition and the experimental condition retained their gained knowledge over time; that is, we expected a non-significant decline between the posttest and the retention test. Regarding the multiple-choice questions our hypotheses were confirmed by our data. In both conditions a non-significant decline has been found. With respect to the open question, our data did not confirm our hypotheses. In the experimental condition a non-significant increase was found between the posttest and the retention test, whereas in the control condition a non-significant decline has indeed been found, but students in this condition did not actually gain knowledge through the inquiry-learning task; the scores decreased over time, and a significant decrease has been found between the pretest and the retention test. We have also explored the five different subscales within the multiple-choice questions. No significant decline has been found between the posttest and the retention test on mc_volume and mc_extrapolate in both conditions (where scores in the control condition even seemed to increase, see Figure 8) and on mc_position in the experimental condition, suggesting that, in line with our hypothesis, students retained their knowledge over time.

It was hypothesized that students in the experimental condition achieved higher scores on the posttest and the retention test than students in the control condition, because students in the experimental condition were supported in drawing conclusions. In line with research conducted by Kopitzki and Gijlers (2011), we found significant differences between the two conditions on the open question of the posttest and retention test and no significant differences on the multiple-choice questions. However, after splitting the multiple-choice questions into the five subscales (i.e. mc_volume, mc_mass, mc_color, mc_position and mc_extrapolate), a significant difference between the two conditions has been found on mc_volume in both the posttest and the retention test. Children, who were stimulated to draw conclusions whilst experimenting, seem to grasp the idea that volume alone does not matter. A big wooden cube floats whereas a small brass cube sinks.

We also expected that students with high scores on scientific curiosity (as measured with the adjusted items of the Children's Science Curiosity Scale of Harty & Beall (1984)) achieved higher scores on the posttest and retention test than students with low scientific curiosity regardless of the condition they are in. Since the scale turned out to not be reliable ($\alpha = 0.60$), this hypothesis could not be tested.

Comments on the current study

The characteristics of the question style might affect the way children answer them (Kopitzki & Gijlers, 2011). The multiple-choice questions force students to make a decision, and it seems that they then think of a particular situation or object instead of a general rule that holds for several situations and objects. The open questions might trigger students to think about a more coherent concept, because it asks for a general rule and students could use as many words as they wanted to explain floating and sinking. Zimmerman (2007) pointed out that immature preconceptions remain side by side with the new scientifically accepted conceptions and when students feel forced to choose one explanation they may prefer to stick to their initial preconception.

Not only the question style may affect the way students answer them, the answering options might have influence as well. The option "You can't say, because ..." was the right answer alternative in a number of cases, but unfortunately, students often seemed to associate the phrase "You can't say, because ..." with being not able to find the correct

answer. They might also have avoided this specific answer alternative because of the need to explain their decision, which asks for deeper considerations and additional time and effort.

Practical implications and suggestions for further research

Many students fail to or are resistant to integrating more than one variable into one concept to explain a complex idea like density (Hardy et al., 2006). Most kids want to think that ultimately, floating depends on only one thing (usually they say weight). However, even after being confronted with the obvious inadequacies of this explanation, they come back to it time and time again, often forgetting that they experienced the effect of other factors, such as volume, on sinking and floating. Students need to be given many opportunities to investigate both volume independent of mass and mass independent of volume to gain a thorough understanding of the effects of each variable on floating. They also need to be given the time to systematically separate those two variables and then integrate them together (Duckworth, 1986). Elementary schools should therefore give their students many opportunities to experiment with buoyancy. To test whether children gain knowledge, teachers would do well to ask their students open questions. When forced to choose between different answering options, children often fall back on their initial conceptions.

Further research might focus on question style in regard to conceptual change or at least knowledge gain in an inquiry-learning situation. It could provide valuable information for the development of tests for inquiry-learning environments for the classroom as well as for scientifically investigations. It might also be interesting to find out why students, who were supported in drawing conclusions, achieved higher scores than students who did not receive such support on questions regarding the influence of volume on floating and sinking, but not on the other multiple-choice questions. Further research may also explain why students, who did not receive support in drawing conclusions, achieved lower scores on the open question after the inquiry-learning task.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the support of her first mentor dr. T. H. S. Eysink. Thank you for all the valuable comments throughout the process and your patience and faith in me. I could not have done this without your support and guidance. I would also like to thank my second mentor A. H. Gijlers for her advice in the area of data analysis and SPSS, and her constructive feedback and enthusiasm. This thesis would not have been possible without the participation and support of the schools, teachers and students. I would like to thank the students and teachers of the fourth grade of 'Drakensteyn' and 'Glanerbrug-Zuid' in Glanerbrug and the 'Kiezel en Kei' in Borculo for their participation and cooperation. I would also like to thank my grandmother for helping me sort all the paperwork, and my mother for reviewing my work and helping me wherever she could. Many thanks go to J. H. H. Laenen, my friends and my parents for providing me with the much-needed kick in the ass and for their on-going support and love.

Marit van der Pol

REFERENCES

- Alfieri, L., Brooks, P. J., Aldrich, N. J. & Tenenbaum, H. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1): 1-18.
- Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York: Holt, Rinehart and Winston.
- Biddulph, F. & Osborne, R. (1984). Pupils' ideas about floating and sinking. *Research in Science Education, 14*: 114-124.
- Bransford, J. D., Brown, A. L. & Cocking, R. R. (1999). *How people learn: Brain, mind, experience and school*. Washington, DC: National Academy Press.
- Brewer, W. F. (1999). Scientific theories and naïve theories as forms of mental representation: Psychologism revived. *Science and Education, 8*(5): 489-505.
- Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31*(1): 21–32.
- Butts, D. P., Hofman, H. & Anderson, M. (1993). Is hands-on experience enough? A study of young children's views of sinking and floating objects. *Journal of elementary science education, 5*(1): 50-64
- Caravita, S. (2001). A re-framed conceptual change theory? *Learning and Instruction, 11*(4-5): 421-429.
- Chi, M. T. H. (1996). Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology, 10*: 33-49.
- Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences, 14*: 161-199.
- Chinn, C. A., & Malhotra, B. A. (2002). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology, 94*: 327–343.
- Christianson, R. G., & Fisher, K. M. (1999). Comparison of student learning about diffusion and osmosis in constructivist and traditional classrooms. *International Journal of Science Education, 21*: 687–698.
- Collins, A. & Gentner, D. (1987). How people construct mental models. In D. Holland & N. Quinn (Eds.), *Cultural models in language and thought* (pp. 243-265). England: Cambridge University Press.

- De Jong, T. (2005). The guided discovery principle in multimedia learning. In R. E. Mayer, *Cambridge handbook of multimedia learning*, pp.: 215-229. Cambridge: Cambridge University Press.
- De Jong, T. (2006). Scaffolds for computer simulation based scientific discovery learning. In J. Elen and R. E. Clark, *Dealing with complexity in learning environments*, pp.: 107-128. London: Elsevier Science Publishers.
- De Jong, T., Njoo, M. (1992). Learning and instruction with computer simulations: Learning processes involved. In E. de Corte, M. Linn, H. Mandl & L. Verschaffel (Eds.). *Computer-based learning environments and problem solving*, pp.: 411-427. Berlin: Springer.
- De Jong, T., & Van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68(2): 179-201.
- Deslauriers, L., Schelew, E. & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science*, 332(6031): 862-864
- Dewey, J. (1938). *Logic: The Theory of Inquiry*. New York: Holt, Rinehart and Winston.
- Driver, R. & Easley, J. (1978). Pupils and paradigms: A review of literature related to concept development in adolescent science students. *Studies in Science Education*, 5: 61-84.
- Duckworth, E. (1986). Inventing density. Grand Forks: University of North Dakota, Center for Teaching and Learning, North Dakota Study Group on Evaluation.
- Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science*, 17: 397-434.
- Eysink, T. H. S., de Jong, T., Berthold, K., Kolloffel, B., Opfermann, M., & Wouters, P. (2009). Learner performance in multimedia learning arrangements: An analysis across instructional approaches. *American Educational Research Journal*, 46, 1107-1149.
- Hardy, I., Jonen, A., Müller, K., Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "floating and Sinking". *Journal of Educational Psychology*, 98(2): 307-326
- Harty, H. and Beall, D. (1984). Toward the development of a children's science curiosity measure. *Journal of Research in Science Teaching*, 21(4): 425-436.

- Hogan, K. & Pressley, M. (1997). Scaffolding scientific competencies within classroom communities of inquiry. In K. Hogan & M. Pressley (Eds.), *Scaffolding student learning: Instructional approaches & issues* (pp. 74-107). Cambridge, MA: Brookline Books.
- Kirschner, P. A., Sweller, J., Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2): 75–86.
- Klahr, D. & Dunbar, K. (1988). Dual-space search during scientific reasoning. *Cognitive Science*, 12(1): 1-48.
- Klayman, J. & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis-testing. *Psychological Review*, 94(2): 211-228.
- Kohn, A. S. (1993). Preschoolers` reasoning about density: Will it float? *Child development*, 64: 1637-1650.
- Kopitzki, E. & Gijlers, A. H. (2011). Guided inquiry learning: How much support is most effective for children`s learning?
- Koran, J. J. & Longino, S. J. (1982). Curiosity and children's science learning. *Science and Children*, 20: 18-19.
- Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: a critical appraisal. *Learning and Instruction*, 11 (4-5): 357-380
- Linn, M. C., Lee, H. S., Tinker, R., Husic, F. & Chiu, J. L. (2006). Teaching and assessing knowledge integration in science. *Science*, 313: 1049–1050.
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, 59(1): 14-19.
- McCloskey, M. & Kargon, R. (1988). The meaning and use of historical models in the study of intuitive physics. In S. Strauss (Ed.), *Ontogeny, phylogeny and historical development* (pp. 49-67). Norwood, NJ : Ablex.
- 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(8): 821-844.

- Piaget, J. (1930). The level of water. In Paul, K., *The Child's Conception of Physical Causality*, pp.: 164-179. New York: Harcourt Brace & Company.
- Pottenger, F. M. I., & Young, D.B. (1992). The local environment: FAST 1, foundational approaches in science teaching. Honolulu: University of Hawaii Curriculum Research and Development group.
- Quinn, J. & Alessi, S. (1994). The effects of stimulation complexity and hypothesis generation strategy on learning. *Journal of Research on Computing in Education*, 27: 75-91.
- Reiser, B. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences*, 13: 273–304.
- Shrigley, R. L. (1983). Review of the test of science-related attitudes. *Journal of Research in Science Teaching*, 20: 87-89
- Simons, P. R. J. (1999). Transfer of learning: Paradoxes for learners. *International Journal of Educational Research*, 31: 577-589.
- Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition*, 21: 177-237.
- Smith, C., Maclin, D., Grosslight, L., & Davis, H. (1997). Teaching for understanding. A study of students' preinstruction theories of matter and a comparison of the effectiveness of two approaches to teaching about matter and density. *Cognition and Instruction*, 15: 317-393.
- Smith, C., Snir, J., & Grosslight, L. (1992). Using conceptual models to facilitate conceptual change: The case of weight-density differentiation. *Cognition and Instruction*, 9(3): 221-283
- Staub, F., & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology*, 94: 344–355.
- Tyson, L. M., Venville, G. J., Harrision, A. L. & Treagust, D. F. (1997). A multidimensional framework for interpreting conceptual change events in the classroom. *Science Education*, 81: 387-404.
- Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24: 535–585.

- Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction, 15*: 317-419.
- Wandersee, J.H., Mintzes, J.J. & Novak, J.D. (1994). Research on alternative conceptions in science. In: *Handbook of Research on Science Teaching and Learning*, ed. D. Gabel, New York: Simon & Schuster Macmillan, pp.: 177–210.
- White, R. & Gunstone, R. (1992). Probing understanding. *London and New York: The Falmer Press*.
- Yin, Y., Tomita, M. K., Schavelson, R. J. (2008). Diagnosing and dealing with student misconceptions: floating and sinking. *Science Scope 31*(8): 34-39.
- Zimmermann, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27*: 172-223.