Effectiveness of concept cartoons and self-explanations to promote sixth-graders' data-reading and theory-revision skills.

Robbert van Hooij (s0214078)

March 2013

University of Twente, Enschede

Bachelor Thesis Psychology

Instructional Technology (ILO)

First attending: Dr. P. Wilhelm

Second attending: Dr. A.W. Lazonder

De effectiviteit van concept cartoons en self-explanations voor het bevorderen van vaardigheden in data interpretatie en theorie-revisie bij leerlingen uit groep acht.

Samenvatting:

Deze these heeft als doel om te onderzoeken of elf jaar oude bassischool leerlingen kunnen profiteren van concept cartoons, ondersteund door self-explanation instructie. Er wordt onderzocht of de combinatie van deze twee instructie methoden nuttig kan zijn voor het leren begrijpen van data interpretatie en theorie revisie De concept cartoon-methode is een afbeelding waarin stripfiguren elk een andere opvatting hebben over een (natuurkundig) verschijnsel. De afwijkende opvattingen in de cartoons zijn zo ontworpen om een cognitief conflict te creëren, waarbij de participant verleid wordt om na te denken en eventueel een nieuwe theorie te vormen Echter, uit onderzoek blijkt dat deze methode het best kan worden ondersteund met prompts en het stellen van open vragen om effectief te zijn in het onderwijs (Stephenson & Warwick, 2002). Deze aanvulling kan worden gevonden in self-explanations. De constructieve aard, met aanmoediging tot integratie van nieuwe kennis en het continue verloop waarin het wordt uitgevoerd, kan een geschikte aanvulling zijn op de concept cartoon methode. Voor deze these is een natuurkundige taak ontworpen, waarin kinderen experimenteerden met verschillende ballen en moesten bepalen welke factoren (gewicht, grootte en kleur) van invloed waren op de valtijd. Deze taak is gebaseerd op twee principes: de algemene misvatting dat gewicht invloed heeft op de valtijd, en wetenschappelijke literatuur die beschrijft dat kinderen problemen ondervinden om resultaten te interpreteren die niet overeenkomen met hun verwachting. Deze studie onderzocht of de combinatie van concept cartoons en self-explanations effectief zijn om de kinderen te leren de resultaten correct te interpreteren en vervolgens een conclusie te trekken op basis van de resultaten in plaats van op hun oorspronkelijke verwachting. Drie condities werden gebruikt; een controleconditie (N = 14), een concept cartoon interventie (N = 15), en een concept cartoon met self-explanation instructie (N = 16). Drie variabelen werden onderzocht: cognitief conflict, data interpretatie, en theorie revisie. De resultaten lieten geen significante verschillen zien tussen de condities op het voorkomen van cognitieve conflicten of het reviseren van theorie. Ook was er geen significant effect van de interventies op correcte data interpretatie, behalve dat de combinatie van de concept cartoon en self-explanation er voor zorgden dat kinderen efficiënter gebruik maakten van de resultaten; ze hadden minder vergelijkingen nodig om tot een conclusie te komen, ongeacht de correctheid van deze conclusie.

Effectiveness of concept cartoons and self-explanations to promote sixth-graders' data-reading and theory-revision skills.

Abstract:

This thesis aimed to investigate whether eleven-year olds can benefit from concept cartoons supported and self-explanation prompts. The basic assumption was that the combination of these two support measures would enhance children's scientific reasoning, in particular their ability to differentiate between tentative hypothesis and solid experimental data.

Concept cartoons aspire to elicit conceptual change through the visual and textual presentation of conflicting but apparently plausible scientific ideas. The conflicting ideas in the cartoons are designed to elicit a cognitive conflict, whereby the participant is tempted to think about the ideas and maybe form a new theory. However, according to Stephenson & Warwick (2002), it needs to be supported with encouragement and open-ended questions for it to be effective in teaching. This support could be found in self-explanations. Its constructive nature, with encouragement of integrating new knowledge into existing knowledge and the continuous, on-going and piecemeal fashion in which it is used, could be a suitable complement to the concept cartoon method.

A physics task was used in which children had to experiment with different balls to figure out which factors influence drop speed (weight, size and colour). This task was built upon two principles: the common misconception that weight has influence on drop speed, and literature showing that children have problems using data that contradict their initial hypothesis. This study investigated if the combination of concept cartoons and self-explanation could enhance children's ability to correctly interpret data and draw valid conclusions based on that data rather than on their initial beliefs. Three conditions were used; a control condition (N = 14), a concept cartoon condition (N = 15), and a concept cartoon plus self-explanation prompts condition (N = 16). Three variables were investigated: cognitive conflict, data reading, and theory revision.

Results showed no significant difference between the three conditions with regard to the elicitation of cognitive conflict and theory revision. There was also no significant effect on correct data reading. However, the combination of concept cartoons and self-explanation prompts did elicit a more efficient use of data: children in this condition needed less comparisons to draw a conclusion than children from the other two conditions, regardless of correctness.

Introduction

When we consider the contents of education of this century, it becomes clear that children are expected to familiarize themselves with scientific thinking at an early stage in education (American Association for the Advancement of Science, 1993; Greven & Letschert, 2006; Duschl, Schweingruber, Shouse, 2007). In the Netherlands, this is reflected in one of the attainment targets of primary education, which reads: “students learn to study materials and inquire about scientific phenomena such as light, sound, electricity, force, magnetism and temperature.” (Greven & Letschert, 2006). The highest level in secondary education in the Netherlands is called “VWO”, an abbreviation which roughly translates to ‘pre-university education’. The highest aim for an elementary school-curriculum is to prepare children for this science-oriented education. Even though not every child will pursue a career in science, the thinking skills used in scientific inquiry can be beneficial for everyone (Ford & Forman, 2006; Metz, 2004; ONeill & Polman, 2004). Especially eight-grades (sixth-graders in American education) would benefit from closing the gap between elementary education and middle or high-school scientific-oriented education. Currently, this gap is believed to exist and according to Metz (2008), much larger than necessary. The current curriculum, where question, method and data analysis tools are given is still very remotely related to the practices in which knowledge is advanced. According to Metz, big ideas need to be central in the practices of the science classroom to stimulate prediction, observation, interpretation and explanation building. Her conclusion is that the elementary school classroom can and should more adequately reflect the robust goal structure of science as discovery and understanding.

Recent research, documented by Gopnik (2012) has also made some challenging objections to current educational policy. According to her perspective, children’s spontaneous exploratory and pretend play is designed to help them learn. In contrast to developing cognitive skills, early childhood should be more about socio-emotional development. Activities as encouragement of play, presenting anomalies and asking for explanations seem to prompt scientific thinking more effectively than direct instruction (Gopnik, 2012). She also states that policy-makers may acknowledge the importance of the socio-emotional aspect, but consistently seem to underestimate the intellectual capabilities of young children and preschoolers. Educational policy seems to focus on the development of cognitive skills, but play and social development have become shallow and underestimated. While aiming to cognitively develop children into competent and economically viable individuals, children’s natural and play-like learning abilities should not be ignored. In its most ideal form, elementary science education should involve the presentation of interesting anomalies in a playful way. This should be sufficient to activate the child’s spontaneous and exploratory approach which they use to explain anomalies around them every day. In short, the gap between elementary and high school science education, as Metz (2008) argued, can be bridged by utilizing the children’s spontaneous approach and molding this into a way of reasoning that is more in line with the scientific approach.

To achieve this, science educators should first consider what children are already capable of, and with which aspects of scientific reasoning they have difficulties. Zimmerman (2007) has documented the many different aspects of scientific reasoning with which children have problems. She concludes that one of the major problems for children is to coordinate theory and evidence (Tschirgi, 1980; Schauble, 1988; Kuhn, 1993; Kuhn, Garcia-Mila, Zohar and Anderson, 1995; Zimmerman, 2007). When results do not match their hypothesis, children still have a natural tendency to draw conclusions based on their hypothesis. They also tend to ignore contradicting data or interpret it differently to make it fit better with their hypothesis (Kuhn et al, 1995). One can argue that confirmation bias explains this phenomenon, as Mynat, Doherty and Tweney (1977) concluded in earlier work. They found that when subjects design experiments, they try to confirm a hypothesis, instead of disconfirming it. Research by Schauble (1988) supports these findings: she found that children have difficulties discerning between their hypothesis and the results from an experiment, because they think the goal of an experiment is to produce an expected outcome instead of testing this expectation for truth. Whether problems with understanding the discrepancy between solid proof and a grounded (but not set-in-stone) hypothesis can be fully attributed to confirmation bias is questionable, and not the purpose of this study. The purpose of this study is to investigate if it is possible to support children in understanding that an hypothesis has an aspect of uncertainty, and that results from experiments are of a more certain nature. This could be achieved by developing an instructional intervention that assists children in discovering this for themselves, to help give them the insight that a systematic experiment will produce results, which should be used to reflect on a hypothesis to draw a ‘scientific’ conclusion. This thesis argues for a novel method of instruction that utilizes a combination of two potentially powerful interventions; concept cartoons and self-explanation prompts.

Concept cartoons in elementary science education

Concept cartoons are a way of presenting multiple views on a problem. They depict three or more children who propose different explanations for a given (science) problem. One of these explanations is closely related to the accepted scientific view about the problem, the other explanations represent plausible misconceptions. Concept cartoons have been employed in various educational approaches and were initially designed in 1992 in an attempt to 'develop an innovative teaching and learning strategy which took account of constructivist views of learning in science' (Keogh and Naylor, 1999a, p.431). This approach advocates the active engagement of children in the development of their own ideas, guided and supported by the teacher (Driver & Oldham, 1986; Driver, 1989; Gilbert & Watts, 1983).

In an evaluation of concept cartoons in the context of teaching and learning science, Keogh and Naylor (1999) performed two case studies involving teachers, student teachers, and pupils across the primary and secondary age ranges. Their impression was that 'learners were positively engaged in discussion about the concept cartoons, with a firm desire to explore their understandings through investigation and research.' (p.442) They concluded that learners appear to experience a cognitive conflict after having studied the contradictory but apparently plausible ideas depicted in the cartoons. They argued that in this respect, similarities exist between concept cartoons and other active learning techniques that aim to promote conceptual change (Roth & Anderson, 1988; White & Gunstone, 1992; Chinn & Brewer, 1993). All in all, there is evidence that concept cartoons elicit discussion and appear to encourage children to engage in debate (Schauble, 1988; Wittrock, 1994).

Another, more recent study by Stephenson and Warwick (2002) investigated the use of concept cartoons in a scientific context; the understanding of light. They argued that learning should be viewed as a social process with the concept cartoons acting as a tool for both developing and expressing conceptual understanding. In their work, the students also showed encouragement to discuss appropriate responses to the scientific anomalies presented through the cartoons. However, Stephenson and Warwick emphasized that 'the success of the learning experience was equally dependent on the confidence of the teacher' (p.140). For the concept cartoons to work as an instructional tool, they argued that it is crucial for the teacher to actively encourage the student's engagement with the cartoons and the associated investigation by asking open-ended questions and support (Stephenson & Warwick, 2002).

While taking a step back and considering the goal of this thesis, it can be concluded that the concept cartoon method could be beneficial to assist children in learning science. Presenting a scientific problem through a cartoon could, according to the studies described above, be successful in activating children's exploratory approach through the elicitation of cognitive conflict. If we consider this in light with the aforementioned problems with scientific reasoning, concept cartoons appear to be a suitable method to elicit cognitive conflict. However, concept cartoons alone will not be sufficient to stimulate conceptual change, as Stephenson (2002) emphasized, an efficient form of instruction should be included. A form of instruction that provides the support and open-ended questions that are necessary to stimulate discussion, solve the conflict and revise the theory subsequently. Ideally, an instructional intervention that connects with both the constructivist nature of the concept cartoons, and the constructive knowledge building that scientific reasoning entails.

The self-explanation effect in elementary science education

The self-explanation effect appears to have a long history in scientific literature. It can be argued that the roots of this line of research start with the discovery of children's self-directed speech by Piaget (1926). He stated this was an egocentric, non-directed 'talk for self', in which the child verbalizes his or her thoughts in whatever form they occur. Vygotsky (1986) however, postulated that this self-directed form of speech had a much greater significance in the child's cognitive development. According to Vygotsky (1986), language helps children think about mental activities, behavior and to select courses of action. He viewed language as the foundation for all higher cognitive processes, including problem solving and abstract reasoning. He explained this 'self-directed talk' as a way in which children guide themselves while solving a problem and called it 'private speech' (Vygotsky, 1986). Most current studies about this form of self-directed speech support this view (Berk & Harris, 2003). One interesting finding is that when tasks become more challenging, or when children make errors or are confused, they use more private speech (Fernyhough & Fradley, 2005). Furthermore, evidence suggests that children, who use more self-guiding private speech, perform better on tasks than children who use less (Al-Namlah, Fernyhough & Meins, 2006; Berk & Spuhl, 1995; Fernyhough & Fradley, 2005; Winsler, Naglieri, & Manfra, 2006). Interestingly, this private speech tends to be internalized with age, changing into whispers and silent lip movements (Patrick & Abravanel, 2000; Winsler & Naglieri, 2003). It is possible that this phenomenon still exists among older children. If this is the case, it could very well be this internalized speech as a feature that is elicited and strengthened by self-explaining. Therefore, private speech as a feature of self-explanation may be a valuable tool in supporting children with the scientific reasoning process.

It is well documented in the cognitive-developmental literature that when children produce explanations while solving problems, they learn considerably more effective (Chi, Bassok, Reimann & Glaser, 1989; Russell & Kelley, 1991; Chi, De Leeuw, Chio & Lavancher, 1994; Renkl, Stark, Gruber & Mandl, 1995; Siegler, 2002; Calin-Jageman & Ratner, 2005; Williams & Lombrozo, 2009). This phenomenon is called the 'self-explanation effect' by Chi et al (1989). They showed that when asking students to explain overtly what they did (and why) while solving a worked-out physics example, they proceduralize their declarative knowledge of physics and become better at solving similar problems. This conversion of declarative into procedural understanding is considered a critical aspect of learning and comprehension (Anderson, 1987). Placing this in the context of learning scientific discovery and reasoning, it is important that one learns to convert declarative knowledge on 'how to solve a problem' into the procedural understanding of it. That is, when we give students scientific problems to solve, the ultimate goal should be that they learn to comprehend the hypothetical-deductive procedure of science. For this reason, the self-explanation effect could be effective because it seem to fit within the constructive context of scientific discovery.

This association with the constructive context of science becomes more apparent in a deeper analysis of the self-explanation effect by Chi, de Leeuw, Chio and Lavancher (1994). They described three processing characteristics of self-explaining to prove its worth as an effective learning activity. The first characteristic is that self-explaining is a constructive activity. This statement is preceded by work of Simon (1979) who postulated that students learn both by being taught and by self-instruction. He based this on his finding that learning is a constructive process in which a student converts words and examples generated by a teacher or presented in a text into usable skills, such as problem solving skills. This process of conversion is essentially a form of constructive self-instruction (Simon, 1979). Research by Chi et al. (1994) also shows that declarative knowledge is constructed when self-explanation is used. However, as argued before, in the context of scientific reasoning one ought to be more interested in the construction of procedural knowledge of the process. Anderson (1987) argued that when learning occurs, the effortful process lies in the conversion of declarative knowledge into the procedural knowledge. This conversion process is arguably the key element in teaching children to use the scientific reasoning method. Taken together, it seems that the constructive nature of self-explanation plays a big part in the process of children's knowledge building.

The second characteristic of self-explaining is that it encourages the integration of newly learned materials with existing knowledge. Data from Chi et al. (1994) shows that at least 30% of the self-explanations are produced from integrating new information with old knowledge. This does not necessarily mean that the new information always fits with the old knowledge, for example in the case of incorrect prior knowledge. Interestingly, not only do correct self-explanations lead to better learning, but also incorrect ones seem to facilitate a learning experience. One interpretation that Siegler (2002) used to explain this, is that creating an incorrect self-explanation merely objectifies that piece of knowledge and readies it for examination. This can create a conflict that the learner might resolve and it is known that much learning results from these conflict situations (Kuhn, 1972; Doise, Mugny & Perret-Clermont, 1975; VanLehn, 1988).

The third characteristic is that self-explaining is carried out in a continuous, ongoing and piecemeal fashion (Chi et al., 1994). This often results in incomplete and fragmented self-explanations. Chi et al. use an analogy where one thinks of self-explaining as the process of creating or revising a mental structure, i.e. a mental model of a system. While reading example statements about the system, there are many opportunities whereby "what is read contradicts what is being created or exists a priori in one's mental structure" (Chi et al, 1994). This could very well be in line with how Siegler explained learning as a multiple strategy theory (Siegler, 1996, 2000). Similar to his theory, it can be argued that during self-explanation, one has to solve multiple conflicts between one's evolving mental structure and the particular instruction that is given. The continuous and ongoing fashion that self-explaining asks from children, is nearly inescapable as they have to verbalize their thoughts while quickly solving the conflicts and reconstruct their theory on-the-go.

In short, these three characteristics of self-explaining seem to support learning significantly and fit in the constructive scientific context. As Siegler (2002) stated, self-explanations are inferences about causal connections among objects and events, inferences concerning 'how' and 'why' events happen. Interestingly, it can be argued that the 'how' and 'why' questions are the basic questions in science, therefore it seems valuable to explore the relation between self-explanation and learning to reason scientifically. When considering the strengths and weaknesses of the aforementioned concept cartoon method, and combining it with an instruction that elicits self-explanation, two constructive instructional methods come together into one. The effects of these two methods combined could be enough to teach children the importance of experimental results, and learn to discern between these results and their hypothesis. Subsequently, it could help in teaching children proper data-reading which should initiate theory revision about a scientific problem.

This study

This thesis aims to investigate whether sixth grade elementary students can benefit from concept cartoons with added self-explanations to learn to interpret data as a representation of 'what really happened' as opposed to what they initially 'belief that will happen'. This coordination between the hypothesis-space and the experiment-space, as documented by Zimmerman (2007), appears to be problematic for children.

The task constructed for this study is a simple physics experiment in which students have to figure out which of three factors influence the drop speed of small balls. The experiment is built upon two principles: the common misconception that weight has influence on falling time, and studies that show children have problems using data that disconfirms their initial hypothesis (Tschirgi, 1980; Schauble, 1988; Kuhn, 1993; Kuhn, Garcia-Mila, Zohar and Anderson, 1995; Zimmerman, 2007). The reason why this misconception was used, is to ensure that all participants had a common belief about something that is actually wrong and can be proven wrong by a simple experiment. The task will engage students in an experiment that will produce disconfirming evidence and is thus designed to elicit an internal conflict.

This study investigated two aspects of scientific reasoning: data reading and theory revision. Three conditions were used to measure the effects of the concept cartoon and the self-explanation on data reading and theory revision. Two interventions; one with concept cartoons and one with concept cartoons plus a self-explanation prompt. Finally a control group was used for comparison. It was expected there would be differences between the three conditions on three variables; the occurrence of cognitive conflict, data reading, and theory revision.

Cognitive conflict was expected to occur when the children would be confronted with disconfirming evidence. This internal conflict would be visible in a longer experiment time, more experiments performed, and more factors explored that were not initially hypothesized. In the control condition, we expected a minimal amount of time spent and small number of comparisons made. Because without the interventions, children would not be able to accept the disconfirming evidence, firmly holding onto their belief in the misconception Furthermore, we expected participants to mostly make comparisons with the initially hypothesized factor, with minimal effort in investigating the factors that were left out of their hypothesis. This expectation is based on studies that show children have a tendency to confirm their beliefs instead of trying to disconfirm them (Mynat, Doherty and Tweney, 1977; Schauble, 1988). In the cartoon condition we expected more time spent and more total comparisons made by the participants, while they compare more factors that were not initially hypothesized. The concept cartoon would elicit a cognitive conflict for the children when they study the contradictory but apparently plausible ideas depicted in the cartoons (Schauble, 1988; Wittrock, 1994, Keogh & Naylor, 1999). This could be visible in more thorough and time consuming investigation, and more with factors that were not hypothesized initially, as they could show interest in the new ideas presented in the cartoon. This was expected even more in the cartoon-SE condition, as the self-explaining would support elaboration on the cognitive conflict that occurs from the cartoon. This elaboration could increase understanding of how to resolve this conflict (Chi et al. (1994).

Data reading was defined as the ability to draw a conclusion based on the results from the experiments performed with the data-cards. Our expectations were that in the control condition, participants would not be able to draw a conclusion based on the data, and would most likely draw conclusions that contained weight as a factor. According to the aforementioned scientific literature, children would have difficulty accepting the data (Mynat, Doherty & Tweney, 1977; Zimmerman, 2007). In the cartoon condition, it was expected that more participants would draw conclusions based on the data, with less conclusions based on weight (initially expected to be all children's misconception). The second cartoon depicted an example of how to use data to confirm the theories in the first cartoon, even without further prompted elaboration this could still inspire the children to pay more attention to the data. Hence, a small number of children could have gained the understanding that they had to use the data to draw conclusions. For the cartoon-SE condition, we expected participants to draw even more (or close to all) conclusions based on data and with even less based on weight. This is because the self-explaining is expected to elicit discussion and elaboration on the ideas depicted in the cartoons (Kuhn, 1972; Doise, Mugny & Perret-Clermont, 1975; VanLehn, 1988, Chi et al, 1994). Especially after studying the second cartoon, we expect them to understand that they need the data to form a conclusion. Therefore, in this condition we also expect the conclusions to be more in line with the data that was used. The increased understanding of data reading is expected to lead to more efficient usage of it to draw conclusions.

Theory revision was defined as the change in the participants' conceptual knowledge of the subject. As they would be confronted with the disconfirming evidence and acknowledged this, we expected this would elicit conceptual change. This could result in a conclusion that differed from their initial hypothesis and if data reading was valid, it should be a correct conclusion. We expected there would be differences between the conditions in successful theory revision. In the control condition, we did not expect any theory revision for two reasons. First, children would have problems with 'moving' from the hypothesis-space to the experimental-space, which leads them to stay in the hypothesis-space and make conclusions based on the hypotheses. (Tschirgi, 1980; Schauble, 1988; Kuhn, 1993; Kuhn, Garcia-Mila, Zohar and Anderson, 1995; Zimmerman, 2007). Secondly, the aforementioned weight misconception would be too powerful for the participants to show any concept change. In the cartoon condition, we expected that more participants would be in conflict about their initial theory. However, because we did not expect them to correctly read the data, this conflict would not elicit successful theory revision. In the cartoon-SE condition, we expected most or all of the participants to experience conflict about their initial theory and subsequently accept the data, leading to conceptual change, i.e. valid data reading and successful theory revision.

Method

Participants

Participants were drawn from two sixth-grade classes in an elementary school in the west part of the Netherlands. Participants were randomly assigned to one of three conditions: Control (n=15, 6 girls), Cartoon (n=15, 6 girls), Cartoon-SE (n=16, 7 girls). The participants had an average age of 11 years and 8 months (SD: .46, range 11 – 13 years).

Materials

Physics experiment. The complex and systematic nature of experimentation to produce scientifically valid and reliable results, was reason for a simplification of the experimental procedure to be comprehensible by eleven year old students. First, the physics task itself was simplified and they were not asked to come up with the factors that may influence the falling time. Instead, they were given three factors, including a non-causal factor: weight, size and colour (the non-causal factor). In the experiment-phase, children received verbal guidance by the experimenter to do the example experiment.

The experiment consisted of a table setup on which a set of eight different balls was presented (Table 1). The balls were paired with stacks of data-cards. The data-cards were numbered (identical to the balls) and contained the descriptive properties of the corresponding balls. The backsides of these cards depicted the same information but with the (pre-calculated) falling time in seconds. To eliminate the balls' bouncy nature, a plastic bin filled with sand was used for the participants to drop the balls in.

Table 1: *Properties of experimental balls*

Object	**Diameter**	**Mass**	**Colour**	**Time**
Ball 1	4.5 cm	5 gram	Red	0.50 sec
Ball 2	4.5 cm	5 gram	Blue	0.50 sec
Ball 3	6.5 cm	5 gram	Red	0.75 sec
Ball 4	6.5 cm	5 gram	Blue	0.75 sec
Ball 5	4.5 cm	50 gram	Red	0.50 sec
Ball 6	4.5 cm	50 gram	Blue	0.50 sec
Ball 7	6.5 cm	50 gram	Red	0.75 sec
Ball 8	6.5 cm	50 gram	Blue	0.75 sec

Concept cartoons. For the two experimental conditions, two different concept cartoons were used (Figure 1 and 2). Both cartoons contained the same four imaginary students. On the first concept cartoon, these students were depicted with a text-balloon in which an hypothesis was written about which factor(s) influence(s) the falling time (Table 2). On the second concept cartoon, these students were depicted with a text-balloon containing a conclusion related to their hypothesis, with a thought-balloon above depicting two data-cards corresponding to their conclusion (Table 3).

Figure 1: *Concept cartoon A*

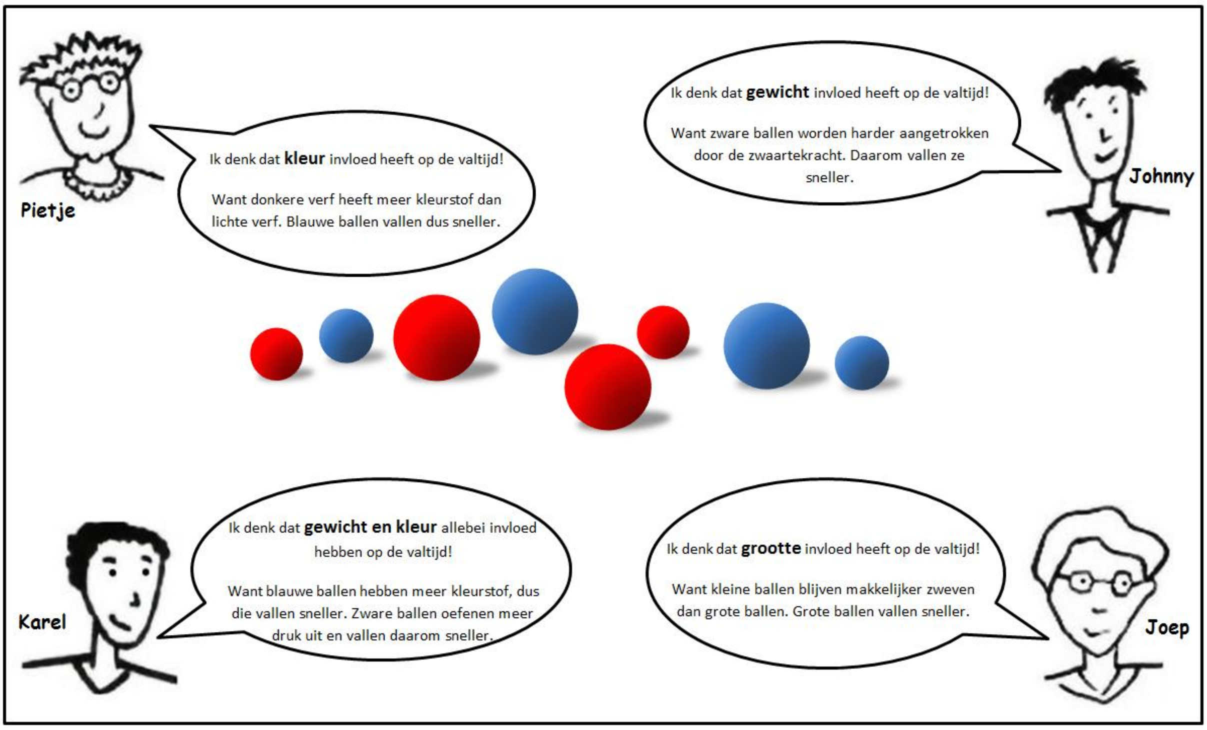


Table 2: *Concept cartoon A: Text balloons – Hypotheses*

Person	Hypothesis	Explanation
Pietje	I think color has influence on the falling time	Because dark paint consists of more pigments then light paint, blue balls must drop faster
Johnny	I think weight has influence on the falling time	Because heavy balls get pulled down more by gravity, so heavy balls must drop faster
Karel	I think weight and color influence the falling time	Because blue balls have more pigments, thus drop faster. Heavy balls exert more pressure so they drop faster.
Joep	I think size has influence on the falling time	Because small balls float more easily, big balls drop faster.

Figure 2: *Concept cartoon B*

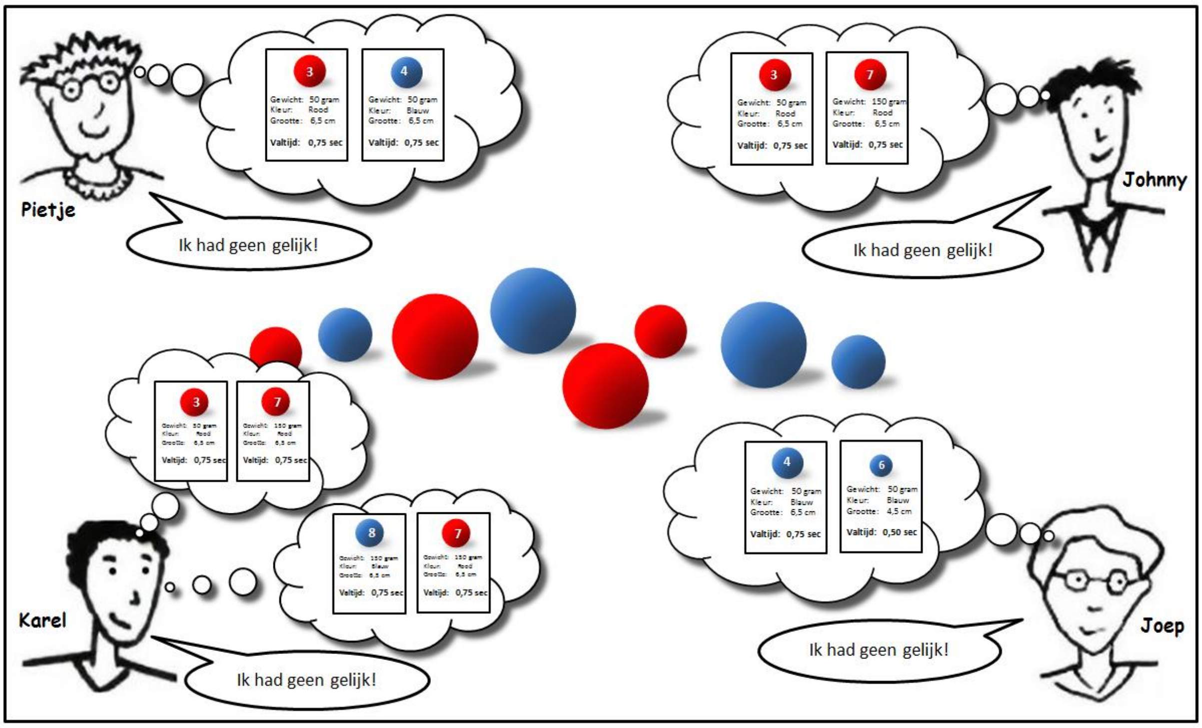
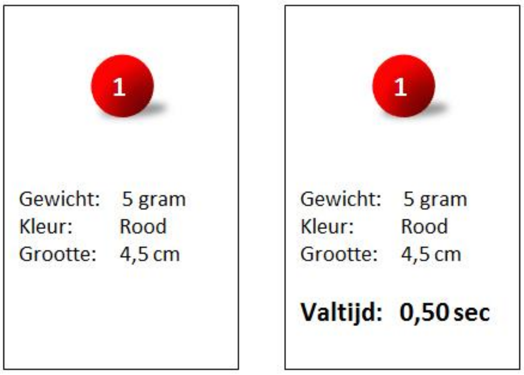


Table 3: *Concept cartoon B: Text-Balloons & Thought-Balloons – Conclusions*

Person	**Conclusion**	**Thought**
Pietje	I was not right	Balls 3 and 4. Same falling time
Johnny	I was not right	Balls 3 and 7. Same falling time
Karel	I was not right	Balls 3 and 7. Balls 8 and 7. Same falling time
Joep	I was not right	Balls 4 and 6. Different falling time

Supporting materials. To support the experiment, a paper placemat and an information card (Figure 3) were used. The paper placemat consists of eight empty spots with the same size and width of the data-cards. This serves the purpose of supporting the participants in making pairs of data-cards for comparison and to encourage them to put the cards in a conveniently arranged manner for a good overview. The information card shows the variable 'falling time' and the three factors influencing it: size, colour and weight. These are complemented by a small illustration explaining what they mean. This information serves the purpose to eliminate any confusion about the three factors.

Figure 3: *Data-cards front and backside.*



Equipment. For measurement purposes, a photo camera and an audio-recorder were used. The photo camera was used to record the number and contents of the comparisons between balls that the participants made with the data cards. The audio-recorder was used to record verbal utterances participants produced after the self-explanations were prompted. The purpose of these recordings was to validate if the prompts actually elicited self-explanations.

Procedure

To investigate whether children were able to draw a valid conclusion based on the presented data, a control condition was included. To investigate the effect of concept cartoons and self-explanation, two experimental conditions were used.

Control condition. In the control condition, participants were seated behind a table with the materials placed on top. The eight balls were presented with the data-cards, in the middle of the table, the information-card at the right side and the bin to drop the balls in on the left side. Participants received the instruction that all balls were different from each other and that they were allowed to inspect them. At this point, participants were asked which of the factors had influence on the falling time, with the specific instruction that there might be more than one. Participants were asked to elaborate on their answer, i.e. when they would say weight, they were asked if they thought a heavier ball would drop faster or slower than a light one. This part of the task will be referred to as the *hypothesis-phase* in the rest of this article.

Next, in the *experiment-phase,* the participants received the instruction: “To see which ball drops faster, we can compare the falling times of the balls”, with the additional instruction that it is very hard to see which ball drops first and that because of that they could use the data-cards. The first experiment was an example, guided by the experimenter. The participant received strict verbal instructions to select two balls that only differ in weight and were asked to first take the corresponding data-cards and place them on the layout. They were then instructed to drop the balls from one meter height at the same time, as precise as they were able to. They could repeat this three times maximum . When they had done this, they were instructed to flip the data-cards to read the exact falling times. At this point, the final instruction was to continue making similar comparisons until they thought they had enough information to form a conclusion. At any point, when they did not show any intent to make another comparison, we repeated this: “Do you think you know how it works now, can you make a conclusion?”

Finally, in the *conclusion-phase*, the participants were asked which factors influenced falling time. Subsequently, they were asked to elaborate on whatever factors they concluded on, specifically: “Does a heavy ball drop faster than a light ball, or the other way around?”, or: “Does a small ball drop faster than a big ball?”, and at last: “How did you come up with this answer?”

Cartoon condition. The first experimental condition was procedurally identical to the control condition, except for the presentation of the two concept cartoons. Concept cartoon A was presented at the start of the experiment-phase, with the specific instruction: “On this cartoon you see students who have thought about this as well. They all think differently about it. Take a good look at what they say.” The participant was allowed to study the cartoon for a maximum of 1 minute before the protocol was continued as described in the control condition procedure. They were then instructed that the falling time of the balls in the experiment was described on the backside of the data-cards, with the explanation that it was impossible to see that small difference in milliseconds without advanced equipment. The second concept cartoon (B) was also presented in the experiment-phase, right after this instruction. This was done in this particular order because we assumed this instruction alone would not be sufficient for the children to understand the link between the experiment and the data. To test for differences in effect between the conditions and the control group, the second cartoon in particular was expected to elicit this understanding. After this cartoon was presented, participants received the following specific instruction: “The students from the last cartoon have looked at the data-cards as well, and they now say the following about it. Take a good look.”. Again they were allowed to study the cartoon for a maximum of 1 minute. Hereafter, both the cartoons were removed from the table and the protocol continued as described in the control condition procedure.

Cartoon-SE condition. The second experimental condition was procedurally identical to the cartoon condition, except for an added self-explanation prompts. This intervention was added subsequently to the presentation of concept cartoon B. At the point where the participants received the cartoon with the complementary instruction, they were asked to explain the cartoon to the experimenter. Specifically: “This student thought … had influence on the falling time. Can you explain to me why you think he now says he has been wrong?” When the participant did not answer or understand, the question was rephrased more specifically: “This student has looked at these data-cards, how does he come to the conclusion that he has been wrong? Can you explain this to me?”. In any case if the participant still was unable to explain, no more questions would be asked about this particular student in the cartoon. Successively, the same questions were asked for the second and the third student in the cartoon. In the end, the cartoons were removed from the table and the protocol was continued as described in the control condition procedure.

Scoring and Data analysis

Weight misconception. The common misconception that weight is of a major influence on the falling time of objects has been the starting point of this study. To be sure that participants indeed had this misconception, they were asked to formulate a hypothesis and every instance where they answered with weight as one of factors, we scored this as (1) yes. The hypotheses without weight were scored as (2) no.

Additionally, we investigated whether there were any structural differences in types of hypotheses between the conditions prior to the experimental procedure with a chi-square analysis. We did this to ensure that the three groups did not have extreme differences in theories about what influences the falling time, prior to measuring theory revision.

Self-explanations. To validate whether the self-explanation prompts actually elicited the self-explaining, recordings of the utterances were used. These utterances were classified qualitatively as a self-explanation when the participant verbally explained both the concept cartoon-figure’s hypothesis and the figure’s motivation.

Conflict. The number of comparisons (pairs) made while experimenting is illustrative of the effort invested in trying to figure out which factors were of influence. This could indicate that the participant might have had doubts about the initial theory and thus could assumed to be in conflict. This number of comparisons is likely to be related to the amount of time used (in seconds) to make comparisons, however, it may be informative to compare these to see if there is indeed a relation (and what kind) between time taken and number of comparisons. For example, a participant could be experimenting slow and thus take a long time. Both time taken and the number of comparisons made were compared across the conditions with analysis of variance.

Time spent experimenting and the number of comparisons alone do not guarantee that participants were actually experiencing a conflict. They may have taken more time and made more comparisons but this could be for many reasons (just to have fun dropping the balls, the ball rolled off the table, the participant did not comprehend or was just slow, etc.). We needed information on why they took longer to experiment or why they made more comparisons before reaching a conclusion. This was measured by looking at which type of comparisons they made in relation to their initial hypothesis. For this purpose, photos of the compared data-cards for each participant were taken to score three variables: (1) number of colour-comparisons, (2) number of weight-comparisons and , (3) number of size-comparisons. We used these variables together with the initial hypotheses to score the number of comparisons made in line and those not in line with hypotheses. A minimal number of comparisons in line with initially hypothesized factors combined with a high number of comparisons not in line with hypotheses, should indicate that the participant has rejected their hypothesis and thus assumed to be in conflict about their theory. However, a high number of comparisons not in line, and a high number of comparisons in line with hypotheses could also be illustrative of a conflict. The participant may not believe the initial test results and thus looks for more confirmation before testing another theory. These scores were compared across the three conditions with a chi-square analysis, to see if any significant changes occurred.

Data reading. To investigate if the experimental conditions had any effect on the participants ability to base their conclusions on data, the performance in correct data reading was measured. First, we asked the participants in the conclusion-phase how they came to their conclusion. When they addressed the data or data-cards in their answers, this was scored as (1) data reference, when they did not address the data or data-cards, this was scored as (0) no data reference. Chi-square analysis was used to test for any differences between the conditions. Secondly, we recorded the number of conclusions that contained weight as a factor and compared this across the conditions with a chi-square analysis. This was done to verify if they had read the data; if a conclusion was drawn based on weight it indicates they had not, as these data would show differences in weight had no influence.

To investigate if the answers from the participants were valid, we compared them with the information on the data-cards selected in the experiments. We then checked if any of the conclusions could actually have been drawn from the used data-cards. We counted how many comparisons they made containing the factors formulated in the conclusion. For instance if a conclusion was drawn that size was of influence, but the data-cards did not represent this factor, it was concluded that they did not use the data to formulate a conclusion. The number of compared pairs that contained the same factors as in the conclusion were then compared across the conditions with chi-square analysis.

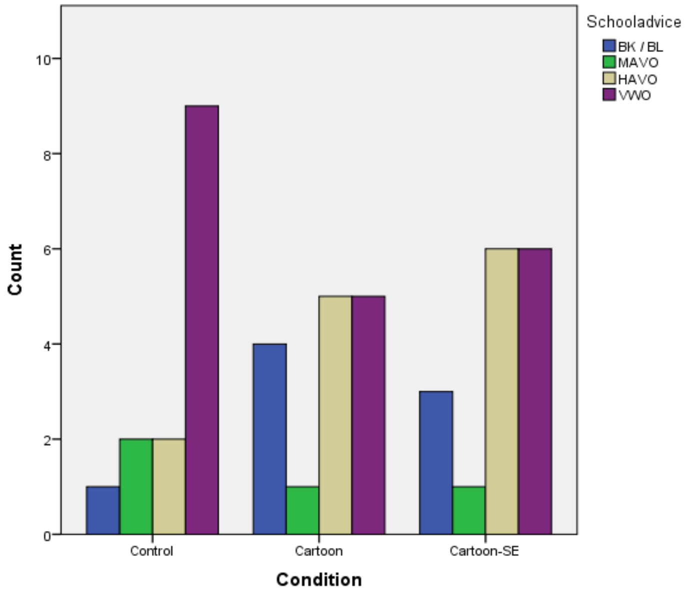
Theory revision. We first checked if the participants drew correct conclusions or not, this would indicate theory change, as all participants were assumed to initially have the weight misconception. A conclusion that described size as the factor of influence, with the explanation that smaller size is faster, was classified as a correct conclusion. Any other conclusions were classified as incorrect. The number of correct conclusions were compared across conditions with a chi-square analysis.

To investigate theory revision regardless of drawing a correct conclusion, we compared the prior hypotheses about the effects of colour, weight and size with the final conclusions. We measured what kind of hypotheses and conclusions they formulated (colour, weight or size). We then compared these to see if concept change occurred. If the conclusion differed from the hypothesis, theory revision did occur. The number of participants that showed theory revision was then compared across the three conditions with a chi-square analysis.

Results

Intellectual capabilities. After our initial data gathering, we considered the intellectual capabilities of our participants as they showed very good understanding of the task, even without interventions. Because of this, we have inquired afterwards about the high-school advice that these participants have received subsequently to an aptitude test (CITO, 2012). We checked how the different levels of advised future education (from low to high) were divided among the conditions, as shown in Figure 4. The two highest levels of education: HAVO and VWO are equally divided among the conditions (11, 10 and 12 respectively). However, compared to the experimental conditions almost twice as many participants in the control group received the highest grade of middle-school advice (9, 5 and 6 respectively).

Figure 4: *School advice compared per condition.*



Weight misconception. Our assumption that nearly all participants would have a weight misconception proved to be correct. 45 participants (98%) thought weight was of influence, of which 43 (94%) thought that a heavy ball would be faster than a light ball. Two participants thought a light ball would drop faster than a heavy one (4%). The one person who did not think weight was of influence, hypothesized size as the factor that influences falling time, we assumed that this person had prior knowledge of the subject. This assumption is based on the fact that she did not test any factors other than size, had the shortest experiment time of all participants and formed a correct conclusion. For this reason we have excluded this one person for further analysis.

None of the participants chose colour as influence on the falling time. All participants hypothesized weight as influence on the falling time. Six participants hypothesized size in the control condition, five in the cartoon condition and four in the cartoon-SE condition. The between-condition differences in hypotheses for weight (X^2 (4, N = 46) = 3.8, p = 0.5) and size (X^2 (4, N = 46) = 4.2, p = 0.4) were not significant.

Self-explanations. To validate if the prompts in our protocol did actually elicit self-explanations, the utterances were analysed. Our results indicate that of the 16 participants that were prompted, 14 (88%) explained the cartoons verbally to the experimenter. No differences between the conditions could be investigated, as these prompts only occurred in the third condition.

Conflict. We performed an analysis of variance on the differences between conditions in the amount of time taken to make comparisons. This analysis did not result in any significant effects of any of the conditions ($F (2, 42) = 1.06$, $p = 0.36$). Also, analysis of variance for the differences between the conditions in the total number of comparisons made, did not show any significant effects of any of the conditions ($F (2, 42) = 1.01$, $p = 0.37$). Furthermore, we expected there would be a relation between time taken and the number of comparisons made. Our analysis did indeed show this relation ($r = 0.45$, $p < 0.005$). We also tested for the number of comparisons made in line with the initial hypotheses. Analysis of variance did not result in any significant differences in between the conditions ($F (2, 42) = .073$, $p = 0.49$). Neither did we find differences between the number of comparisons made that were not in line with hypotheses ($F (2, 42) = 0.07$, $p = 0.99$).

Data reading. The number of participants that referred to the data in response to our questions, was compared across conditions. Chi square analysis did not produce any significant differences between the conditions ($X^2 (2, N = 45) = 0.29$, $p = 0.87$). Initially, all participants hypothesized weight as a factor. A quick glance at the number of weight conclusions per condition (Table 4) did show a small increase in conclusions that reject weight (14 vs. 10), however results show no significant differences between the conditions ($X^2 (2, N = 45) = 2.00$, $p = 0.37$).

Table 4: *Number of conclusions with the factor weight*

	Condition		
Conclusion	**Control**	**Cartoon**	**Cartoon-SE**
Not weight	10 (22.2%)	10 (22%)	14 (31.1%)
Weight	4 (8.9%)	5 (11.1%)	2 (4.4%)

Note: percentages of total number of participants

We also validated if the participants *could* actually have based their conclusion on the data they collected. The number of compared pairs which shared the same factors as the factors used in the conclusion was compared across the conditions with an analysis of variance. Table 5 shows means and SD's for the three conditions. There appeared to be significant differences between the conditions ($F (2, 42) = 3.208$, $p = 0.05$). Further analysis with 95%-Bonferroni confidence intervals showed a significant difference between the control condition and the cartoon-SE condition (Table 6). In the cartoon-SE condition, participants needed less pairs to form a conclusion, regardless of this conclusion being correct or not. No differences were found between the control and cartoon condition, nor between the cartoon and cartoon-SE condition.

Table 5: *Number of tested pairs in line with conclusion per condition*

Condition	**N**	**Mean**	**SD**
Control	14	2.29	1.44
Cartoon	15	1.87	1.23
Cartoon-SE	16	1.25	0.78

Table 6: *Confidence intervals for differences between conditions*

Conditions	***p***	**Interval**
Control - Cartoon	0.97	-0.63 – 1.47
Control - Cartoon-SE	0.05	0.00 – 2.07
Cartoon - Cartoon-SE	0.41	-0.40 – 1.63

Theory revision. We analysed the number of correct conclusions to investigate if the interventions had any effect on theory change. Correct conclusions were assumed to be indicative of theory revision, as all participants were assumed to initially have the weight misconception. We did not find any significant differences between the conditions (X^2 (2, N = 45) = 0.06, p = 0.97). Secondly, a chi-square analysis was used to investigate the effects of the conditions on concept change. Concept change occurred if the conclusion differed from the hypothesis. Table 7 shows how many participants showed concept change per condition. We did not find any significant differences between the conditions in effect on concept change (X^2 (2, N = 45) = 1.28, p = 0.53).

Table 7: *Concept change per condition*

	Condition		
Concept change	**Control**	**Cartoon**	**Cartoon-SE**
Yes	12 (26.7%)	12 (26.7%)	15 (33.3%)
No	2 (4.4%)	3 (6.7%)	1 (2.2%)

Note: percentages of total number of participants

Conclusion

Conflict. We expected an increase among the conditions for the amount of time spent and the number of comparisons that were made. In the control condition we expected a minimal amount of time and a small number of comparisons; in the cartoon-SE condition we expected a significant increase for these two variables. However, our analyses did not show any significant differences. We also expected that in the experimental conditions, fewer comparisons would be made with factors initially hypothesized and more with the factors that were not. Our analysis however did not result in any significant differences between the conditions.

Data reading. We expected an increase in the ability to draw a conclusion based on the data for the cartoon and cartoon-SE conditions. This expectation proved to be wrong; we did not find any significant effects of the experimental conditions on the participant's subjective reference to data. We also expected a decrease in the number of weight conclusions for the experimental conditions. When we investigated the number of conclusions made that were based on weight (the misconception), we did see a slight increase in the number of participants that rejected weight in the cartoon-SE condition (14 vs. 10 in control and cartoon condition) but this did not prove to be a significant difference.

To validate if the participants *could* actually have used the data from their experiments to draw any conclusion (correct or incorrect), we compared the number of compared pairs that contained the same factors as in the conclusion across the conditions. These results show that participants needed less pairs compared to the control condition. From these results we can conclude that the self-explanation effect did not necessarily lead to better understanding, however it did lead to a more efficient use of data. This could indicate a better comprehension of the important connection between reading the data and forming a conclusion based on that, this will be discussed later.

Theory revision. Against our expectations, there was no significant increase found in the number of correct conclusions among the conditions, thus theory revision can not be concluded to be the result of our interventions. We also expected there would be differences in effect of the conditions on concept change. Our results did only show a slight increase in concept change for the cartoon-SE condition (15 vs. 12 in control and cartoon condition). Nevertheless, these results did not show any significant effect from any of the conditions on this increase.

Discussion

Self-explanation effect. In this thesis, we argued for the positive effect of self-explanation on teaching children to understand data reading and subsequently form new conceptual knowledge. Our results did not entirely support this; we did not see any increase in the number of correct conclusions across the conditions. However, when we looked closer at how they actually came to their conclusion from the used data cards, we did see that the self-explanation condition elicited more efficient use of data reading. These results indicate children needed less data to form a conclusion after the self-explanation intervention. The self-explanation intervention did have an effect, as the children needed less information (only one or two comparisons) to draw a conclusion. This indicates that the prompted self-explaining of the concept cartoon did lead to a better understanding of what we attempted to instruct with the cartoon; the connection between reading data and drawing a conclusion. This result needs a critical note however, as this was only visible when we compared the number of data cards that actually contained results that could lead to a right conclusion. When we compared the total number of comparisons made, this difference between the control condition and the cartoon-SE condition did not occur; there were not systematically less comparisons made in the cartoon-SE condition. On the whole, while needing less experiments to come to a conclusion indicates a better understanding of the data-reading, it does not necessarily mean that they experiment ‘better’. In scientific practice it is a given that more test results means more reliable results; to test if results repeatedly occur in multiple occasions to eliminate simple chance. Nonetheless, the purpose of our design was not to teach children how to experiment in a reliable way, it is aimed at teaching them the connection between data as proof to test a theory. The self-explaining combined with the concept cartoon did promote a more efficient usage of data, but not necessarily a more correct usage.

When we consider conceptual change, we did not find any notable results to implicate that self-explanation had any effect on the conceptual change in children’s knowledge of the subject. As the children in the control group showed conceptual change without any intervention; we were preaching to the choir in this respect. It might be that the instructional intervention, regardless of the cartoons or the self-explanation prompt was already strong enough to elicit concept change, e.g. we confronted every participant with the disconfirming proof. Perhaps if we would not have done this, we could possibly have seen an effect.

Physical task. In our control group, we saw conceptual change occur among the participants, and we also saw successful data reading and many correct answers. Due to the amount of structure and guidance we added to the physical experimentation task, children actually demonstrated increased understanding of the subject. They did not have this understanding before, thus we can assume that our procedure, including the physical task, had this effect on the children, even without the instructional interventions. To answer the question why our results did not show any significant differences between the conditions, we have to consider the task difficulty and the participants' capabilities.

At first, we did not expect our task to be too easy for the children, as other studies have proven that children around this age have indeed difficulties with accepting disconfirming evidence (Mynat, Doherty & Tweney, 1977; Zimmerman, 2007) and have problems understanding the discrepancy between the hypothesis-space and the experimental-space (Tschirgi, 1980; Schauble, 1988; Kuhn, 1993; Kuhn, Garcia-Mila, Zohar and Anderson, 1995; Zimmerman, 2007). For these reasons, we invented some simplifications to make the experimental procedure less difficult to grasp. Considering our results we may have overdone this, giving too much structure and making the link between the data and the experiment too obvious for our experimental conditions to have any extra effect. Reflecting upon our work, we conclude that for a next iteration of this research, it would be worthwhile to do a pilot study of the actual task, before studying the experimental conditions. This to ensure that the task alone indeed shows difficulties for the children, before testing if any experimental intervention can relieve these difficulties.

Secondly, when considering the intellectual capabilities of our participants, they may indeed have been too smart for our interventions to have any extra effect on their understanding of the subject. Our assumptions considering their intellect turned out to be correct, we have inquired about the high-school advice that these participants have received subsequently to an aptitude test (CITO, 2012). Compared to the experimental conditions, almost twice as many participants in the control group received the highest grade of middle-school advice. This could account for an increased performance in the control group; if they would have been divided equally, we might have seen more differences between the effects of the control group and the experimental groups.

Implications for future research. Self-explanation combined with concept cartoons as instructional method in learning aspects of scientific reasoning should be studied more exhaustively, to get more robust results about its combined effect. It would be worthwhile to investigate whether the effect of self-explaining on the understanding of ideas presented in concept cartoons, could be stronger in an experiment in which the learning task is less structured and more ambiguous about the usage of the data. It would also be interesting to investigate the differences between the combinations of self-explanation with concept cartoons, and self-explanation with an explanatory text (Chi, de Leeuw, Chio & Lavancher, 1994) or a worked out example (Crowley & Siegler, 1999). While our current results did not show any significant effects of the concept cartoon in itself, it does seem to be an effective 'handle' to teach a concept. The implicit nature of the concept cartoon in itself does not seem to elicit enough understanding, but when combined with a self-explanation, it becomes more effective. The combination of a concept cartoon and the self-explanation in this case prompts the children to think more thoroughly about the idea's actually means. Future research about the connection between self-explaining and concept cartoons could provide valuable results for anyone who is interested in a better science-learning environment for children.

References

Al-Namlah, A.S., Fernyhough, C., & Meins, E. (2006). Sociocultural influences on the development of verbal mediation: Private speech and phonological recoding in Saudi Arabian and British samples. *Development Psychology, 42,* 117-131.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy.* Washington, DC: American Association for the Advancement of Science.

Anderson, J.R. (1986). Problem solving and learning. *American Psychologist, 48,* 35-44.

Berk, L.E., Harris, S. (2003). Vygotsky, Lev. In L.Nadel (Ed.), *Encyclopedia of cognitive science.* London: Macmillan.

Berk, L.E., Spuhl, S.T. (1995). Maternal interaction, private speech, and task performance in preschool children. *Early Childhood Research Quarterly, 10,* 145-169.

Calin-Jageman, R.J., Ratner, H.H. (2005). The role of encoding in the self-explanation effect. *Cognition and Instruction, 23,* 523-543.

Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P., Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145-182.

Chi, M.T.H., De Leeuw, N., Chiu, M., LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive science, 18,* 439-477.

Chinn, C., Brewer, W. (1993). The role of anomalous data in knowledge acquisition: a theoretical framework and implications for science instruction. *Review of Educational Research, 63,* 1-49.

Doise, W., Mugny, G. & Perret-Clermont, A.N. (1975). Social interaction and the development of cognitive operations. *European Journal of Social Psychology, 5,* 367-383.

Driver, R., Oldham, V. (1986). A constructivist approach to curriculum development. *Studies in Science Education, 13,* 105-122.

Driver, R. (1989). Changing conception. *Adolescent Development and School Science, 6,* 79-104.

Duschl, R., Schweingruber, H.A., Shouse, A. (2007) *Taking science to school: Learning and teaching in grades K-8*. Washington, DC: National academy press.

Fernyhough, C., Fradley, E. (2005). Private speech on an executive task: Relations between task difficulty and task performance. *Cognitive Development, 20,* 103-120.

Ford, M.J., Forman, E.A. (2006). Redefining disciplinary teaching in classroom contexts. In J.

Green & A. Luke (Eds.), Review of research in education, Vol.30.

Gilbert, J., Watts, D. (1983). Concepts, misconceptions and alternative conceptions: changing conceptions in science education. *Studies of Science Education, 10,* 61-98.

Greven, J., Letschert, J. (2006). *Kerndoelen Primair Onderwijs.* Den Haag: DeltaHage, Publicatie van het ministerie van Onderwijs, Cultuur & Wetenschap.

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research and policy implications. *Science, 337,* 1623-1627.

Keogh, B., Naylor, S. (1999). Concept cartoons, teaching and learning in science: an evaluation. *International Journal of Science Education. 21,* 431-446.

Kuhn, D. (1972). Mechanisms of change in the development of cognitive structures. *Child Development, 43,* 833-842.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education, 77,* 319–337.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Anderson, C. (1995) Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development, 60* (4, Serial, nr. 245).

Metz, K.E. (2004). Children’s understanding of scientific enquiry: Their conceptualization of uncertainty in investigations of their own design. *Cognition and Instruction, 22,* 219-290.

Metz, K.E. (2008) Narrowing the gulf between the practices of science and the elementary school science classroom. *The Elementary School Journal, 109,* 138-161.

Mynat, C.R., Doherty, M.M., Tweney, R.D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology, 29,* 85-95.

ONeill, D.K., Polman, J.L. (2004). Why educate “little scientists?” Examining the potential of practice-based scientific literacy. *Journal of Research in Science Teaching, 41,* 234-266.

Patrick, E., Abravanel, E. (2000). The self-regulatory nature of preschool children’s private speech in a naturalistic setting. *Applied Psycholinguistics*, *21,* 45-61.

Piaget, J. (1926) *The language and thought of the child.* New York: Harcourt, Brace & World.

Renkl, A., Stark, R., Gruber, H., Mandl, H. (1998). Learning from worked-out examples: the effects of example variability and elicited self-explanations.

Roth, K., Anderson, C. (1988). Promoting conceptual change learning from science textbooks. In P. Ramsden (ed.), *Improved Learning: New Perspective.* London: Kogan Page. pp. 109-141.

Russell, D.M., Kelley, L. (1991). *Using IDE in instructional design: Encouraging reflective*

instruction design through automated design tools. Paper presented at Annual Conference of American Educational Research Association.
Schauble, (1988). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49,* 31-57.
Siegler, R.S. (1996) *Emerging minds: The process of change in children's thinking.* New York; Oxford University Press.
Siegler, R.S. (2000) The rebirth of children's learning. *Child Development, 7,* 26-35.
Siegler, R.S. (2002). Microgenetic studies of self-explanation. Microdevelopment; Transition processes in development and learning. 31-58.
Simon, H.A. (1979). Problem solving and education. In D. Tuma & F.Reif (eds), *Problem solving and education: Issues in teaching and research.* Hillsdale, NJ: Erlbaum.
Stephenson, P., Warwick, P. (2002). Using concept cartoons to support progression in students' understanding of light. *Physics Education, 37(2),* 135-141
Tschirgi, J.E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development, 51*, 1–10.
VanLehn, K.J. (1988). Student modeling. In M. Polson & J. Richardson (Eds.), *Foundations of intelligent tutoring systems.* Hillsdale, NJ: Erlbaum.
Vygotsky, L.S. (1986). *Thought and language.* MIT press.
Williams, J.J., Lombrozo, T. (2009). The role of explanation in discovery and generalization: evidence from category learning. *Cognitive science, 34,* 776-806.
Winsler, A., Naglieri, J. (2003). Overt and covert verbal problem-solving strategies: Developmental trends in use, awareness, and relations with task performance in children aged 5 to 17. *Child Development, 74,* 659-678.
Winsler, A., Naglieri, J., & Manfra, L. (2006). Children's search strategies and accompanying verbal and motor strategic behavior: Developmental trends and relations with task performance on children age 5 to 17. *Cognitive Development*, *21*, 232-248.
White, R., Gunstone, R. (1992). *Probing Understanding.* London: Falmer.
Wittrock, M. (1994). Generative science teaching. In P. Fensham, R. Gunstone and R. White (eds), *The Content of Science.* London: Falmer, pp. 29-38.
Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27,* 172-223.