Gaining domain knowledge during a hands-on versus simulation task: can the executive functions explain inter-individual differences?

Leonie Brummer
University of Twente
Department: Instructional Technology

Abstract
The influence of executive functions (inhibition, working memory, and fluency) on learning during a balance scale task was studied in 6th grade children (mean age 11.2 years). The balance scale task was considered a guided inquiry learning task with two conditions: (a) a hands-on and (b) a simulation condition. Inhibition, working memory, and fluency were measured with the Color-Word Interference Test, WISC subtest Digit Span, Word Fluency Test, and the Figure Fluency Test. Domain knowledge was assessed using a pre- and post-test. Results revealed a negative correlation between fluency and learning (indicated as post- minus pre-test scores). Findings are discussed as well as limitations and areas for future research.

Keywords: inquiry learning, hands-on vs. simulation, executive functions.

Introduction

Inquiry learning

In inquiry learning, or scientific inquiry, students carry out an investigation to gain (domain) knowledge (Atkin & Black, 2003; In: Hume & Coll, 2010) by observing the world, proposing ideas, and explaining and justifying statements upon gathered evidence (Hofstein & Lunetta, 2003). An example of a widely-used inquiry task is the balance scale task (e.g., Schultz & Takane, 2007; Van der Maas & Jansen, 2003). The core concepts of inquiry learning consist of: (a) developing, formulating and evaluating explanations and (b) communicating and justifying explanations (Sampson & Blancard, 2012; Zimmerman, 2007). Recently, inquiry learning is often embedded in simulated environments (Smetana & Bell, 2012).

Simulation versus hand-on tasks in inquiry learning

Whether simulations rather than hands-on, sometimes referred to as physical, tasks are more suitable to learn science is debated in the literature (e.g., Klahr, Triona, & Williams, 2007; Smetana & Bell, 2012). Zacharia, Loizou, and Papaevripidou (2012) found that physicality was advantageous compared to its virtual counterpart in their study. Participants in the physicality condition learned more about the balance beam (e.g., which aspects influence the balance, such as position and weight), than participants in the virtual condition. Learning, in the study from Zacharia, Loizou, and Papaevripidou (2012), was determined in a clinical interview. Physicality was defined as ‘’the actual and active touch of concrete material and apparatus’’ (Zacharia, Loizou, & Papaevripidou, 2012, p. 448), which is similar to the definition of hands-on used in the current study. On the contrary, Klahr, Triona, and Williams (2007) concluded that no difference in performance existed between the hands-on and a virtual condition (simulation). According to Smetana & Bell’s (2012) literature review of 61 articles, simulations can be as effective (or more effective) as (than) traditional methods (i.e., lecture-, textbook- based or physical hands-on) if used appropriately. Many studies directly compared simulations and more traditional methods (including hands-on materials). Results showed at least equal effectiveness of both methods as well as advantages favouring simulations over traditional methods. However, Smetana and Bell (2012) concluded that inconclusive findings in the studies were probably due to methodological concerns. Thus, it is unclear which approach (hands-on or simulation) is more suitable to gain domain knowledge.

Even though the suitability of hands-on or simulation approaches to learn science is debated, hands-on environments are considered richer learning environments (Grabinger & Dunlap, 1995; Wilson, 1995). Hands-on environments have more “cues” (information) that can influence conceptual change or contribute to domain knowledge. Rich learning environments are also discussed by Grabinger and Dunlap (1995). According to them, Rich Environments for Active Learning (REALs) are characterized by authenticity, knowledge-builders, and are contributors to high-level thinking processes. The information stemming from these environments are considered rich. Additionally, the information facilitates learning (e.g., gaining domain knowledge). For example, in a balance scale task real, tangible objects (“hands-on”) draw to gravity and are considered an additional informational cue that could contribute to domain knowledge. The “gravity” component is absent in the simulation condition. Thus, it is assumed that adding an extra modality, such as tactile input, to the visual (and auditory) input can lead to gaining domain knowledge *if* relevant information is selected (and irrelevant information is ignored). The ability to select relevant information (and ignoring irrelevant information) is considered an inter-individual difference that could explain difference between groups (e.g., hands-on and simulation), which is visible in learning (difference between pre- and post-test scores). Selectively attending to the relevant information (while ignoring irrelevant information) is called inhibition (Diamond, 2013). Students, who are not able to inhibit irrelevant information stemming from the learning environment, have a greater risk of confusing information or being overwhelmed by that information. This can result in gaining less domain knowledge. *If* irrelevant information is not inhibited, more cognitive load is evoked that hinders learning. For example, during a balance scale task, the student must decide whether to focus on specific characteristics (such as the weight, the size, or even the colour of the objects) or to ignore (some of) these characteristics. The cognitive load theory (CLT) assumes that learners can be overwhelmed (or confused) by the amount of interactive informational elements that needs to be handled simultaneously before learning can occur (Paas, Van Gog, & Sweller, 2010). This means that information needs to be selected (inhibited) to reduce the amount of cognitive load, especially in rich learning environments (such as hands-on environments).

Cognitive load is categorized into three categories: (a) intrinsic load, (b) extraneous load, and (c) germane load (Paas, Van Gog, & Sweller, 2010). Intrinsic load is evoked by the complexity of the materials, whereas extraneous load is evoked by poorly designed materials that interfere with schema construction. Germane load is considered the load composed of effective processing that result in learning. Intrinsic and extraneous load are additive, which could lead to exceeding the maximum one’s (cognitive) capacity. It reduces germane load and, therefore, hinders learning (Paas, Van Gog, & Sweller, 2010). In the current study, intrinsic load is not differentiated from extraneous load. When speaking of cognitive load, either both or the addition of intrinsic and extraneous load is meant.

In summary, information evoked by touch and visual aspects means handling more information in order to gain domain knowledge. As a result, a student can be so overwhelmed or confused by that information (more cognitive load) that it eventually interferes with learning. Focussing on specific information (while ignoring other information) refers to inhibition, a component of the executive functions.

Executive Functions

Executive functions (EFs) are responsible for the execution of social and goal-directed behaviour (Cooper-Kahn & Dietzel, 2011; Jurado & Rosselli, 2007; Smidts & Huizinga, 2011). Additionally, regulating behaviours is enabled by EFs (Cooper-Kahn & Dietzel, 2011; Smidts & Huizinga, 2011) that is necessary in learning situations (Neuenschwander, Röthlisberger, Cimeli, & Roebers, 2012). For example, in an inquiry learning task in science. The goal of science is to reach conceptual change (Zimmerman, 2007) and in order for the conceptual change to occur, goals have to be achieved. For example, by completing a workbook and carrying out an investigation in a science-related domain.

The CLT is connected to at least three separately measurable components of EFs, these include (a) the ability to focus without being distracted (inhibition), (b) the active manipulation of information in the memory for later use (working memory; Cooper-Kahn & Dietzel, 2011; Jurado & Rosselli, 2007; Smidts & Huizinga, 2011) and (c) the generation of ideas (fluency; Stockholm, Jørgensen, & Vogel, 2013) or (effective) strategy use (Hughes & Bryan, 2002).

One often-studied EF is inhibition, or the ability to focus without being distracted (Diamond, 2013; Miyake et al., 2000). Sometimes it is defined as the ability to stop and think prior to acting upon those thoughts (Cooper-Kahn & Dietzel, 2011; Diamond, 2013). In order to reduce the cognitive load in learning environments, one must selectively attend to information by selecting relevant information while ignoring other (irrelevant) information. For example, inhibiting irrelevant responses (e.g., thinking that the size of the weights is influential) during a balance scale task.

Working memory (WM) refers to *actively* processing (thus manipulating) information that is necessary moments later (Smidts & Huizinga, 2011). WM is limited in capacity and duration (Paas, Van Gog, & Sweller, 2010; Smidts & Huizinga, 2011). WM is one of the core concepts of the CLT that interacts with an unlimited long-term memory (Paas, Van Gog, & Sweller, 2010). Due to limitations in capacity and duration, the WM can be easily overloaded by information stemming from the learning environment. An example of WM during a balance scale task is memorizing outcomes from experiments and converting the outcomes with information from one's long-term memory to a written explanation.

Fluency refers to the generation of ideas (Stockholm, Jørgensen, & Vogel, 2013) and is involved in strategy use (Hughes & Bryan, 2002). Handling information by using effective strategies can reduce cognitive load, because the cognitive capacity is not exceeded by choosing one effective strategy. Therefore, there is still capacity left for learning. If a student must "search" for effective strategies, more cognitive capacity is used for "searching" and less for actual learning the material.

EFs and age

At the age of 12, students are able to understand the balance scale task (Boom, Hoijtink, & Kunnen, 2001). Additionally, relating the development of EFs to age appears difficult due to inconsistent definitions and operationalizations of EFs (Catroppa & Anderson, 2009). However, Catroppa and Anderson (2009) summarize the development of several EFs (e.g., fluency) with regard to the age at which these EFs presumably develop. They conclude that EFs are visible in late childhood (ages 9-12; Catroppa & Anderson, 2009). Thus, at the age of 11 or 12, the EFs are sufficiently developed to indicate inter-individual differences between students. Therefore, it was decided to choose 6th graders as participants of the current study, with an approximate age of 11 or 12.

Research Questions and Hypothesis

Previous research about hands-on and simulation tasks revealed inconsistent results for educational purposes (e.g., Klahr, Triona, & Williams, 2007; Smetana & Bell, 2012). Therefore, the first research question is: Does learning differ between an inquiry hands-on and an inquiry simulation task?

Rich learning environments can enhance learning; however, the limitation of such learning environments (e.g., environments requiring multiple senses, such as hands-on learning environments) is that an extra dimension of information (tactile input) is added to the information already evoked by other modalities (e.g., visual aspects such as size and color tags). As a result, confusion about what information is relevant for the task at hand, could be created. According to the cognitive load account, more information can result in more cognitive load. More cognitive load can prevent learning to occur. A learning situation with more informational cues relies on the ability to ignore irrelevant information and to focus on relevant information (inhibition) in order to gain domain knowledge. Not being able to ignore irrelevant information causes confusion or the feeling being overwhelmed, which in turn interferes with gaining domain knowledge. Inhibition skills are necessary in learning environments that evoke more information (e.g., tactile input such as the "gravity" component in a hands-on environment) than in learning environments with no extra dimensions (e.g., simulation environment). Inhibition, among working memory and fluency, are often studied together. Therefore, the second research question is: Can the executive functions (inhibition, working memory, and fluency) provide an explanation for differences in learning during an inquiry hands-on and an inquiry simulation balance scale task? Students vary in their ability to ignore irrelevant information (inhibition). In hands-on learning environments; inhibition is an ability that has to be used in order to gain domain knowledge. Therefore, it was hypothesized that children with better inhibition skills learn more on the inquiry hands-on task than children with lower inhibition skills. WM and fluency are also studied in the current study, because both EFs also draw to cognitive load and there is consensus about the measurability of these EFs.

In the current study, the research questions can be summarized as followed:
1. Does learning differ between an inquiry hands-on and an inquiry simulation balance scale task?
2. Can the executive functions (inhibition, working memory, and fluency) provide an explanation for differences in learning during an inquiry hands-on and an inquiry simulation balance scale task?

Hypothesis: Children with better inhibition skills learn more on the inquiry hands-on task than children with lower inhibition skills.

Method

Participants

The participants in the current study consisted of 72 6th graders from two Dutch elementary schools (39 girls and 33 boys). The average age was 11.2 years (SD= .57). Participants were randomly assigned to the hands-on or the simulation condition. Therefore, in total 70 participants were included in the study. Two participants were excluded from the sample. One participant was hospitalized after the second test session and the other moved elsewhere after completing the first test session.

Materials

EFs were measured with the Color-Word Interference Task (inhibition), WISC-III subtest Digit Span (WM), and the Word Fluency Test (WFT; fluency)/ Figure Fluency Test (FFT; fluency). During the hands-on task and the simulation task, specially designed materials were used (workbook and pre- and post-test).

Color-Word Interference Test. The Color-Word Interference Test is inspired on the Stroop-task (Stroop, 1935, In: Delis, Kaplan, & Kramer, 2007) and consisted of four conditions, in which three colours recurred: (a) red, (b) blue, and (c) green. Condition 1 involved naming coloured squares and condition 2 involved reading words (names of colours) aloud. Condition 3 involved naming the colour of the ink while ignoring the written word. For example, the word ‘green’ is printed in red ink. The correct answer in this case was ‘red’; not ‘green’. During condition 4, one has to switch between reading the word and naming the colour of the ink. Each condition had two practice rows, with a total of 10 items. The 50 items were presented in five rows of 10 items each. Conditions 1 and 2 took approximately 90 seconds, whereas condition 3 and 4 took about 180 seconds to complete. As an EF-measure, only condition 3 was taken into account.

WISC-III subtest Digit Span. Digit Span consisted of a number string, that had to be read aloud by the researcher and the participant had to replicate the numbers is a specific sequence (forward or backwards). Digit Span forward involved the replication of a number string forwards (indicative for short-term memory) and Digit Span backwards was a comparable task, only the string of numbers had to be replicated aloud backwards (indicative for WM). Both tasks had 12 items and increased in difficulty. If two following strings of the same length were incorrectly replicated, the test had to be cut off. Each correct item received a score of 1. The maximum score (for WM) was 12.

Word Fluency Test (WFT)/Figure Fluency Test (FFT). The WFT consisted of two parts. Part one consisted of naming as many animals as possible within one minute. Part two involves a similar task regarding professions. Subcategories are allowed, such as biology teachers, history teachers, and so on. Each correctly named animal or profession was scored with one point. The amount of perseverations was noted as well as errors.

The FFT had five parts consisting of three practice examples and one test page. Part one involved drawing as many unique patterns as one can possibly think of using five dots in a circular configuration. Part two and three were similar to part one, with respectively rhombic distracters and lines added to the five dots. Part four and five consisted on dots in two different figurations without any distracters. Each part took approximately 60 seconds to complete.

As an EF-measure, the number of unique words/figures (separately) was taken into account.

Balance scale task.

Hands-on task. The hands-on task consists of a balance beam and 18 weights. The distance from position 3 to the fulcrum is 45 cm, from position 2 to the fulcrum 30 cm, and the distance from position 1 tot the fulcrum is 15 cm. Each position was marked with a pin that was faced upwards.

There are two baskets of each weight, so that similar weights can be placed on both sides of the fulcrum. The weights with the grey tape are 450 grams, red corresponds with 225 grams, and the yellow tape stands for 150 grams. The setting was visible in Figure 1.

Figure 1. Photo of the hands-on materials with the balance beam and the weights.

Simulation. The simulation consisted of a task in which students could place weights on the balance beam presented on a computer screen. The balance beam was visible in the upper part on the screen. Below this balance beam, switches with a slide bar were placed to vary the position (45, 30 or 15 cm to the fulcrum), the weight (150, 225 or 450 grams), and the size (12, 14 or 16 cm^3). This accounts for the left side as well as the right side of the balance beam. Figure 2 shows a screenshot of the program.

Workbook. The workbook contained questions that had to be answered during the session. It consisted of three parts, namely: (a) a pre-test, (b) a hypothesizing and experimenting phase, and (c) a post-test. An hour was scheduled to complete the whole workbook. It was not allowed to ask questions regarding the content, although it was allowed to ask procedural questions.

The pre-test was a prior knowledge test to capture the pre-knowledge about balance and which factors (weight, volume, and position) are influential. Besides 'practicing' with the format, the practice questions involved questions regarding the differences between left and right. Appendix A displays the design (in Dutch) of these two practice questions. This knowledge was essential during the following questions, because students had to indicate which side of the displayed balance beam (with weights on either side) would go down. The remaining pages consisted of the prior knowledge test. See appendix C (Dutch version) for the design of the questions. Each question had a similar format: a picture of the balance scale with a weight on each side of the scale. Next to the picture was a question regarding what will happen to the balance. This question could be answered with: (a) the right side goes down, (b) there is a balance, or (c) the left side goes down. Only one answer was correct. It total, participants were given 15 minutes to complete 20 questions.

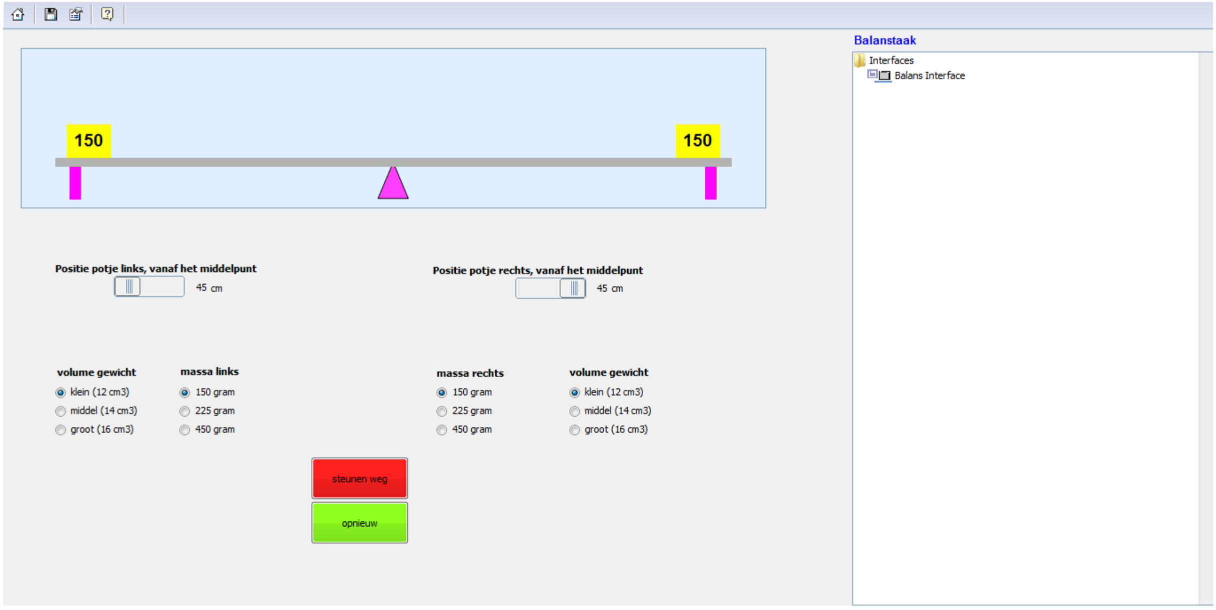


Figure 2. Screenshot of the balance task simulation.

The hypothesizing and experimenting phase included the use of the balance scale task (hands-on materials) or the simulation. Both the children in the hands-on as the simulation condition received part two of the workbook. The first pages provided instruction (written and oral) about how to answer the questions and what could be expected next. Depending on the condition (hands-on or simulation), the oral instruction consisted of explaining that the weights can be placed on the balance beam by either clicking with the mouse or placing the weights on the pins. For both conditions the same instruction was used, although modified for the specific condition. More instruction consisted of question one (appendix C), with arrows to indicate what is meant in the question, and more specific explanations what is expected from the children. Each question had the same format: one page was dedicated to a hypothesizing phase, whereas the other page involved actual experimentation.

Similar sub questions were asked for each of the six questions (see appendix D). First, a situation is given (a picture with weights on each side). For this situation, the sub questions were as follows: (a) what happens to the balance, (b) how certain are you, and (c) why do you think that will happen. Second, the experiment had to be conducted on either the computer or with the hands-on materials. Similar to the first set of questions, the participant had to answer: (a) what happens to the balance; (b) is the weight influential, if so, how, (c) is the size influential, if so, how, (d) is the position influential, if so, how, and (e) how certain are you. After completing the sixth question, two yes/no-questions were asked: (a) do you know which factors influence the balance, and (b) do you want to conduct another experiment. If the latter question was answered with “yes”, the participant had to turn the page and had to go through the process of experimentation again. These questions were slightly different, because the students had to choose which experiment they want to conduct and why they choose that particular experiment (see appendix E). If the question was answered with “no”, the participant was directed to the next question that began with hypothesizing. Participants had 30 minutes to complete part two.

Part three consisted of a post-test and was similar in format as part one. It contained the same questions. Within 15 minutes all 20 questions had to be completed.

Not all children could complete all questions in the given time.

Procedure
The current study was divided into two test sessions. The first session consisted of the measures for the three EFs: the Color-Word Interference Test, Digit Span, and WFT/ FFT (in this order) and was a one-to-one session with the researcher.

The final session consisted of working with either the hands-on materials (condition 1) or the simulation (condition 2). Beforehand, all participants were randomly assigned to one of the two conditions. During this session, the researcher took groups of participants (6 -10 students) from the classroom to complete a prior knowledge test. For part two of the session, the participant received the hands-on materials or its simulated equivalent. Both conditions received part two of the workbook.

After reading the first instructions, the researcher gave oral instructions, suited for either the hands-on or the simulation condition. The oral instruction was followed by written instructions in the workbook. The written instructions were specified by highlighting essential parts and placing arrows. Participants went through a hypothesizing and experimentation phase that alternated. The last past consisted of the post-test. After completion of all parts, the participants returned to the classroom and other participants had to complete the workbook. They received a small gift (pencil or key chain) after completion.

For each new group of participants, the balance beam and the baskets were placed in a similar configuration to secure standardized procedures. The weights were covered when the student was completing the pre-test.

Additionally, the debriefing consisted of explaining all children that only the weight and the position of the weights are influential.

Results

Table 1 summarizes the descriptive statistics (mean score, of EFs measures and pre- and post-test scores. Table 2 displays the descriptive statistics (*n*, mean age (SD), knowledge test scores (SD), minimum, and maximum scores) of the hands-on and simulation condition.

Table 1
Descriptive statistics of EFs measures, NDT, and the pre- and post-test scores (N=70)

Measure	Test	Mean (SD)	Minimum	Maximum
Inhibition	C-WIT	66.87 (13.25)	30	100
Working Memory	Digit Span	4.70 (1.59)	2	10
Fluency	WFT	29.31 (7.24)	14	46
	FFT	52.53 (19.51)	13	108
Performance on knowledge test	Pre-test	12.33 (3.33)	4	20
	Post-test	12.29 (3.22)	4	19

Note. All scores are raw scores. C-WIT= Color-Word Interference Test. WFT= Word Fluency Test. FFT= Figure Fluency Test. NDT= Nederlandse Differentiatie Test.

Table 2
Descriptive statistics of the hands-on and simulation condition

Condition	n	Mean age (SD)	Knowledge test	Mean score (SD)	Minimum	Maximum
Hands-on	35	11.11 (.47)	Pre-test	12.34 (3.67)	4	20
			Post-test	11.86 (3.21)	4	19
Simulation	35	11.34 (.64)	Pre-test	12.31 (2.99)	7	18
			Post-test	12.71 (3.21)	4	18

An independent samples t-test was used to investigate whether the learning (pre- and post-test scores) differed for the two conditions. Levene's test for equality of variances was not significant, thus equal variances were assumed. Learning differed not significantly ($t(68) = 1.30$, $p > .10$).

No significant difference was found between scores of the WFT, FFT, and Color-Word Interference Test between the two conditions. Scores on Digit Span differed significantly between the two conditions ($t(68)= 2.98$, $p < .01$). Levene's test was not significant.

With a one-way repeated measure analysis, the assumption of sphericity is met (due to having only two levels, namely the pre- and post-test). An ANCOVA with repeated measures (with scores on Digit Span Backwards as a covariate), the mean scores on the pre- and post-test did not differ significantly ($F(1,68)= .08$, $p > .75$. Controlling for the condition in which the participants were placed, an ANCOVA with repeated measures did not differ ($F(1,67)= .44$, $p > .51$). The effect size (η_p^2) was .01 and is considered small (Field, 2005).

To investigate whether learning (pre- and post-test scores) differed for the two conditions, an independent samples t-test was used. Levene's test for equality of variances was not significant, thus equal variances were assumed. The performance differed not significantly ($t(68) = 1.30$, $p > .10$).

Normality was assumed when the value of the skewness and kurtosis ranges from -1 to +1 (Plooij, 2011). The values are within this range, thus for all EFs measures a normal distribution is assumed as well as for the scores on the knowledge test (pre- and post-test). As a result, Pearson correlation was used. Table 3 summarizes the correlations between EFs measures and learning for all participants. Learning was derived from post-test minus pre-test scores.

Table 3

Correlations (and SDs) between EFs measures and learning (N=70)

			EFs				**Learning**
			1	**2**	**3**		
			C-WIT	**Digit Span**	**WFT**	**FFT**	
EFs	**1**	**C-WIT**					
	2	**Digit Span**	-.20 (.10)				
	3	**WFT**	-.40** (.00)	.30* (.01)			
		FFT	-.22* (.03)	.27* (.02)	.13 (.14)		
Learning			.04 (.73)	-.04 (.74)	-.07 (.57)	-.24* (.04)	

Note. One-tailed correlations. * $p < .05$; ** $p < .01$. [a]*n*=71. [b]*n*=69. 1=Inhibition. 2=Working Memory. 3=Fluency. C-WIT=Color-Word Interference Test. WFT= Word Fluency Test. FFT= Figure Fluency Test.

Additional analysis consisted of correlations between the EFs and learning for the hands-on and simulation condition separately (see Table 4 and 5).

Table 4

Correlations (and SDs) between EFs measures and learning in the hands-on condition (n=35)

			EFs				**Learning**
			1	**2**	**3**		
			C-WIT	**Digit Span**	**WFT**	**FFT**	
EFs	**1**	**C-WIT**					
	2	**Digit Span**	-.27 (.06)				
	3	**WFT**	-.26 (.06)	.39* (.01)			
		FFT	-.31* (.03)	.21* (.12)	.13 (.22)		
Learning			.01 (.48)	-26 (.07)	-.16 (.18)	-.39* (.01)	

Note. One-tailed correlations. * $p < .05$; ** $p < .01$. [a]*n*=34. ** Significant at a .01-level (2-tailed). C-WIT= Color-Word Interference Test. DS=Digit Span. WFT= Word Fluency Test. FFT= Figure Fluency Test.

Table 5
Correlations (and SDs) between EFs measures and learning in the simulation condition (n=35)

			EFs				**Learning**
			1	**2**	**3**		
			C-WIT	**Digit Span**	**WFT**	**FFT**	
EFs	**1**	**C-WIT**					
	2	**Digit Span**	-.16(.18)				
	3	**WFT**	-.52** (.00)	.29* (.05)			
		FFT	-.16 (.19)	.40** (.01)	.12 (.25)		
Learning			.08 (.32)	.06 (.38)	.01 (.47)	-.11 (.26)	

Note. One-tailed correlations. * $p < .05$; ** $p < .01$. C-WIT= Color-Word Interference Test. DS=Digit Span. WFT= Word Fluency Test. FFT= Figure Fluency Test.

Conclusion and discussion

The current study investigated whether differences existed in learning between a hands-on and a simulation balance scale task and whether the executive function inhibition could provide an explanation for differences in learning during two versions of the balance scale task (hands-on and simulation).

No difference was found in learning between the hands-on and simulation condition and the effect size was considered small (Field, 2005). In other words, the difference in learning between the hands-on and simulation condition was too small to provide a substantial explanation. This means that the difference in learning was not completely due to the condition (hands-on or simulation). This finding corresponds with the conclusions drawn by Smetana and Bell (2012). They concluded that learning depends on methodological considerations rather than the learning environment (hands-on or simulation) itself.

It is remarkable that the students did not learn from the learning task. This could be due to relative high pre-test scores. As a result, little room existed for improvement during the post-test. While completing the post-test, the students already worked roughly 45 minutes. This could make them tired or less motivated and, therefore, the post-test scores were lower than expected. Another explanation could be found in the chance of guessing the right answer. For each question, a student had 33% of guessing the right answer. The pre- and post-test discriminated insufficiently. Finally, it was also possible that both the hands-on materials and the simulation caused insecurity that resulted in non-learning. After the learning task, students were aware of the complexity of the influencing factors (e.g., interaction weight and distance). Emotions, such as insecurity, affect academic performance and the use of learning strategies among other things (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011). Izard (1984, In: Chen & Sun, 2012) even claimed that cognitive activities suffer from negative emotions (e.g., feelings of insecurity). Besides feelings of insecurity, feelings of confusion can be the result of both the hands-on materials and the simulation. The complexity of the material (interaction of weight and distance) increased the intrinsic load and minimized the germane load. Therefore, learning can be hindered.

Regarding the second research question, no correlation existed between inhibition and learning (indicated by post- minus pre-test scores). Therefore, the hypothesis was rejected. However, a negative correlation was found between fluency (Figure Fluency) and learning. This indicated that less fluent children had learned more in the hands-on condition for Figure Fluency only. An explanation can be found in the cognitive load evoked when dealing with multiple sensory input. In the hands-on condition, information is evoked by sight and touch; whereas in the simulation condition, information is only evoked by sight. Therefore, the hands-on condition is a richer learning environment wherein more information needs to be handled by the learner. As a result, in order to gain domain knowledge children must choose effective strategies and fluency draws to (effective) strategy use (Hughes & Bryan, 2002). Apparently, less fluent children are better able to select those strategies that are effective to handle a richer learning environment (hands-on environment). Maybe these children are limited to a couple strategies that appear effective in many situations and, as a result, less cognitive load is evoked by choosing a strategy. Additionally, fluent children can be confused or overwhelmed by so many informational cues that choosing an effective strategy appears more difficult, because more cognitive load is evoked by "searching" for the most effective strategy.

The correlation between fluency and learning in the hands-on condition only concerned the FFT. With the FFT, the participants could see which figures they already had drawn on the page, which is considered an external representation of information. The visual component is apparent in FFT as well as during the balance scale task. When the students draw only a few patterns (corresponds with lower FFT-scores), external representations could result in managing cognitive load (De Jong, 2010). New patterns can be checked externally ("searching" patterns on the paper); not internally ("searching" in one's head for new patterns). A few patterns are easier to check (low(er) FFT-scores) than more patterns (high(er) FFT-scores). With the WFT, the participants could not check whether they mentioned an animal or profession, which stresses the load by "searching" new animals or professions internally. Therefore, a correlation concerning the visual version of fluency test, that reduces cognitive load by displaying the given answer, was not surprising.

Limitations

The current study had several limitations. First of all, the sample size was too small to pose robust conclusions. Second, the workbook (used during the simulation and hands-on task) was not piloted, which means that it was unclear whether the materials could facilitate learning. Third, lower scores on the pre- and post-test could be due to carelessness while completing the test. As a result, a page could be skipped. Skipped questions were scored as errors. Additionally, the chance of guessing was substantial and high(er) pre-test scores left little room for improvement on the post-test scores. Fourth, due to technical problems, the log files from the simulation condition were not registered completely. Only a few log files were useable, but due to the limited number of log files, it was decided not to use the data. The lack of comparison material from the simulation condition resulted in the inability to compare them with the observations from the hands-on condition. Beforehand, it was decided to compare experimentation behaviours from the simulation and hands-on condition; however, after gathering the data, no reliably comparison could be made. Fifth, although the WISC-III is sufficiently reliable and valid, the *Commissie Testaangelegenheden Nederland* (COTAN) advice against interpretation of results subtests individually (COTAN, 2012). Additionally, according to the COTAN, both the WFT and FFT are considered reliable and sufficiently valid. It has to be noted that criterion validity was evaluated as insufficient, due to the lack of research with both tests (COTAN, 2012).

Future research

The current study investigated whether the executive functions could provide an explanation for the performances during an inquiry hands-on and an inquiry simulation balance scale task. Only fluency appeared to influence learning in the hands-on condition. This was probably due to the richness (more informational cues, such as feeling how heavy the weight was) of the hands-on learning environment. Future research should involve a larger sample size to replicate the findings in the current study. Due to the relative small sample size, scores on the pre- and post-test could be based on coincidence. Second, the possible relationship between cognitive load and executive functions (combined with hands-on and simulation materials) is also interesting to address. Cognitive load could explain why the students did not learn from the learning task. After the learning task, the students could be aware of the complexity of the interaction of the weight and distance which results in insecurity or confusion. This addresses the cognitive load theory; however, no cognitive load measures were included in the current study. The concept was only used to explain the findings. Future research should include cognitive load measures, for example, self-reports, physiological measures and/or dual tasks. It should be noted that no direct measures of cognitive load exist. No reliability measures were used for the learning materials; however, this should be addressed in future studies. Fifth, research should at least involve different experimental conditions with varying degrees of cognitive load evoked by the design of the materials as well as the complexity of the materials to explore the role of cognitive load in combination with executive functions and learning. Other explanations (about the fact that students did not learn) should also be considered. Future studies should minimize the chance of guessing as much as possible by, for example, providing four- or five-answer questions or by increasing the number of questions in the pre- and post-test (or a combination of both). The duration of the experiment should also be taken into account to avoid tiredness or other non-motivating feelings and behaviours.

Additionally, the current study investigated a link between executive functions and learning during two inquiry tasks. The link between executive functions and learning is moderated by learning behaviours (for example, experimentation behaviour or method of working). To gain more insight in why executive functions contribute to learning, one has to expose the learning behaviours. The current study originally focused on the role of these behaviours; however, due to technical problems with the log files, questions concerning learning behaviours could not answered and were deleted from the study. Future research could address those learning behaviours.

Acknowledgement

I want to thank Marjolein van Klink for all her time and effort she put in mentoring this study and providing feedback in order to improve of my thesis. She did an amazing job at “teaching” to be a better researcher. I also want to thank Hans van der Meij for his feedback as well as Elise Brummer, and Marloes van Schaik. Additionally, I am grateful for the cooperation of the schools.

References

Boom, J., Hoijtink, H., & Kunnen, S. (2001). Rules in the balance. Classes, strategies, or rules for the balance scale task? *Cognitive Development, 16*, 717-735.

Catroppa, C., & Anderson, V. (2009). Neurodevelopmental outcomes of pediatric traumatic brain injury. Retrieved from http://www.medscape.com/viewarticle/713315_4 on 11 June 2013.

Chen, C.-M., & Y.-C. (2012). Assessing the effects of different multimedia materials on emotions and learning performance for visual and verbal style learners. *Computers & Education, 59*(4), 1273-1285. Doi: 10.1016/j.compedu.2012.05.006.

Cooper-Kahn, J., & Dietzel, L. (2011). *Vergeten, kwijt en afgeleid. Opvoedwijzer om executieve functies bij kinderen te versterken*. Amsterdam: Hogrefe.

COTAN (2012). Figuur-fluency test, FFT, 2006. Retrieved from http://www.cotandocumentatie.nl.proxy.ubn.ru.nl/test_details.php?id=569 on 10 November 2012.

COTAN (2012). Weshler Intelligence Scale for Children – 3rd edition – Nederlandstalige uitgave, WISC-III-NL, 2002-2005. Retrieved from http://www.cotandocumentatie.nl.proxy.ubn.ru.nl/test_details.php?id=99 on 10 November 2012.

COTAN (2012). Woord-Fluency Test, WFT, 2006. Retrieved from http://www.cotandocumentatie.nl.proxy.ubn.ru.nl/test_details.php?id=568 on 10 November 2012.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2007). *Handleiding Color-Word Interference Test*. America, Harcourt Assessment, Inc.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135-168. Doi: 10.1146/annurev-psych-113011-143750.

Field, A. (2005). Effect sizes. Retrieved from http://www.statisticshell.com/docs/effectsizes.pdf on 18 May 2013.

Grabinger, R. S., & Dunlap, J. C. (1995). Rich environments for active learning: a definition. *The Journal of the Asssociation for Learning Technology, 3*(2), 1-30. Doi: 10.1080/0968776950030202

Hofstein, A., & Lunetta, V. N. (2003). The laboratory in science education: Foundations for the twent-first century. *Science Education, 88*(1), 28-54.

Hughes, D. L., & Bryan, J. (2002). Adult age differences in strategy use during verbal fluency performance. *Journal of Clinical and Experimental Neurpsychology, 24*(5). 642-654. Doi: 10.1076/jcen.24.5.642.1002.

Hume, A., & Coll, R. (2010). Authentic student inquiry: the mismatch between the intended curriculum and the student-experienced curriculum. *Research in Science & Technological Education, 28*(1), 43-62. Doi: 10.1080/02635140903513565.

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsycholical review, 17*, 213-233. Doi: 10.1007/s11065-007-9040-z.

Jong, T. De. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Sciences, 38*, 105-134. Doi: 10.1007/s11251-009-9110-0.

Klahr, D., Triona, L. M., & Williams, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. *Journal of Research in Science Teaching, 44*(1), 183-203. Doi: 10.1002/tea.20152.

Maas, H. L. J. Van der., & Jansen, B. R. J. (2003). What response times tell of children's on the balance scale task. *Journal of Experimental Child Psychology, 85*, 141-177. Doi: 10.1016/S0022-0965(03)00058-4.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex ''frontal lobe'' tasks: a latent variable analysis. *Cognitive Psychology, 41*, 49-100. Doi: 10.100/cogp.19999.0734.

Mulder, J. L., Dekker, P. H., & Dekker, R. (2006). *Handleiding Woord-Fluency Test/ Figuur-Fluency Test*. Nederland, Leiden: PITS.

Neuenschwander, R., Cimeli, P., Röthlisberger, M., Roebers, C. M. (2012). Personality factors in elementary school children: Contributions to academic performance over and above executive functions? *Learning and Individual Differences, 25*, 118-125. Doi: 10.1016/j.lindif.2012.12.006.

Paas, F., Gog, van. T., & Sweller, J. (2010). Cognitive Load Theory: new conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review, 22*(2), 115-121. Doi: 10.1007/s10648-010-9133-8.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students’ learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology, 36*, 36-48. Doi: 10.1016/j.cedpsych.2010.10.002.

Plooij, F. (2011). *Onderzoek doen. Een praktische inleiding in onderzoeksvaardigheden*. Nederlands, Hilversum: Pearson Education Benelux.

Sampson, V., & Blanchard, M. R. (2012). Science teachers and scientific argumentation: Trends in views and practice. *Journal of Research in Science Teaching, 49*(9), 1122-1148. Doi: 10.1002/tea.21037.

Schultz, T. R., & Takane, Y. (2007). Rule following and rule use in the balance-scale task. *Cognition, 103*, 460-472. Doi: 10.1016/j.cognition.2006.12.004.

Smetana, L. K., & Bell, R. L. (2012). Computer simulations to support science instruction and learning: A critical review of the literature. *International Journal of Science Education, 34*(9), 1337-1370. Doi: 10.1080/09500693.2011.605182.

Smidts, D., & Huizinga, M. (2011). *Gedrag in uitvoering. Over executieve functies bij kinderen en pubers.* Amsterdam: Nieuwezijds.

Stockholm, J., Jørgensen, K., & Vogel, A. (2013). Performance on five verbal fluency tests in a healthy, elderly Danish sample. *Aging, Neuropsychology, and Cognition, 20*(1), 22-33. Doi: 10.1080/13825585.2012.656576.

Wilson, B. G. (1995). Methaphors for instruction: why we talk about learning environments. Retrieved from http://carbon.ucdenver.edu/~bwilson/wils95 on 17 May 2013.

Zacharia, Z. C., Loizou, E., & Papaevripidou, M. (2012). Is physicality an important aspect of learning through science experimentation among kindergarten students? *Early Childhood Research Quarterly, 27*, 447-457. Doi: 10.1016/j.ecresq.2012.02.004.

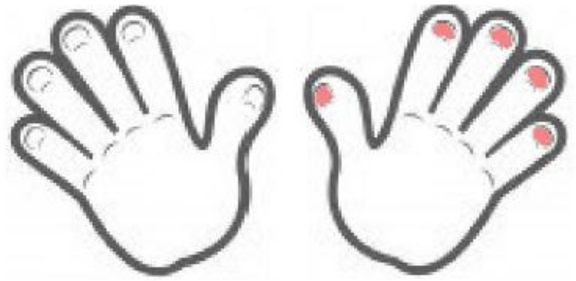
Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27*, 171-223. Doi: 10.1016/j.dr.2006.12.001.

Appendix A

Control questions (left/right) in the pre-test (workbook part 1).

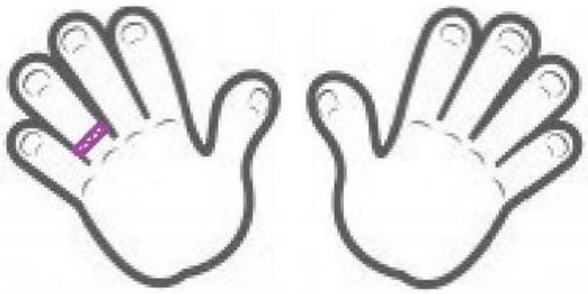
Oefening:

Welke hand (links of rechts) heeft nagellak op de nagels?
Kleur het juiste rondje in.



o links
o rechts

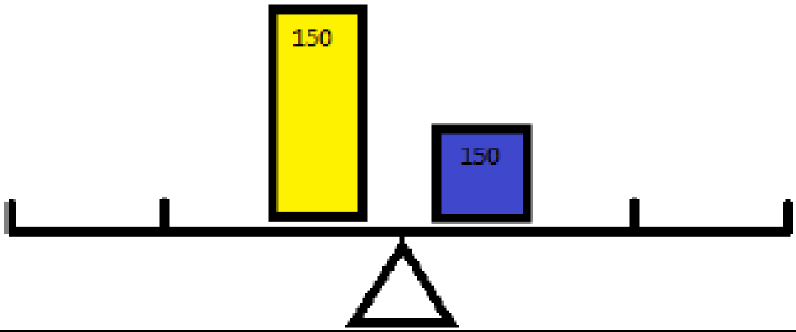
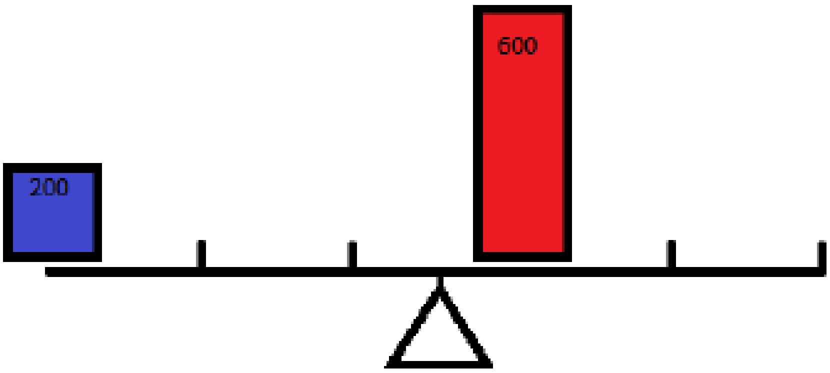
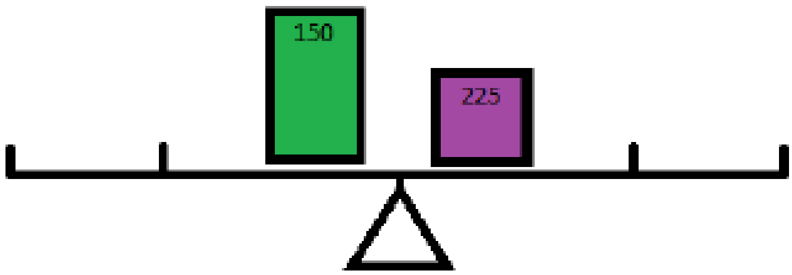
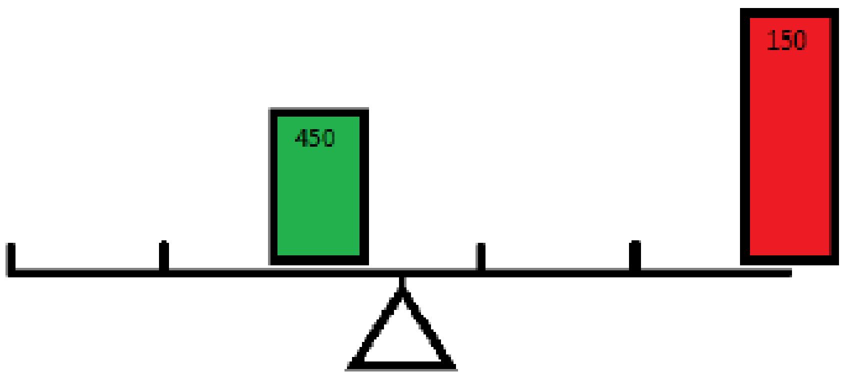
Welke hand (links of rechts) heeft een ring om de ringvinger?
Kleur het juiste rondje in.



o Links
o Rechts

Appendix B

Examples of the design of the questions in the pre- and post-test (workbook part 1 and 3).

1 	Bij deze balans: O gaat de rechterkant omlaag O is er een evenwicht O gaat de linkerkant omlaag
2 	Bij deze balans: O gaat de rechterkant omlaag O is er een evenwicht O gaat de linkerkant omlaag
3 	Bij deze balans: O gaat de rechterkant omlaag O is er een evenwicht O gaat de linkerkant omlaag
4 	Bij deze balans: O gaat de rechterkant omlaag O is er een evenwicht O gaat de linkerkant omlaag

Appendix C

Design of the hypothesizing and experimenting pages, question 1 with inserted instructional cues for the simulation condition (workbook part 2).

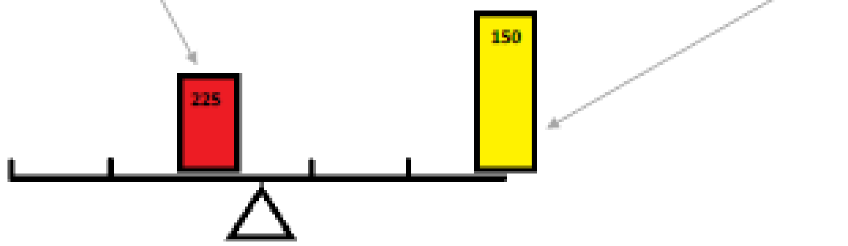
De Balans

Opgave 1:

Stel je de volgende situatie voor:

Links van het middelpunt staat een middelgroot potje van 225 gram op positie 1.

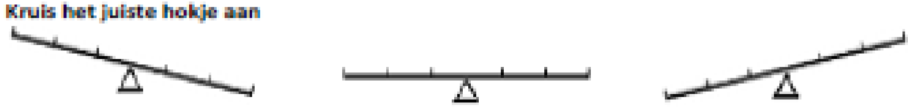
Rechts van het middelpunt staat een groot potje van 150 gram op positie 3.



(a). Wat denk je dat er gebeurt als je deze potjes op de balans zet?

GEBRUIK HIERBIJ GEEN COMPUTER!!!

Kruis het juiste hokje aan



☐ Rechts gaat naar beneden ☐ In evenwicht ☐ Links gaat naar beneden

(b). Waarom denk je dat?

____

____

____

____

(c). Geef een rapportcijfer over hoe zeker je bent van je antwoord bij opgave (a).

Als je erg zeker van je antwoord bent, geef je een hoog cijfer (bijvoorbeeld een 9).

Als je heel onzeker bent, geef je een laag cijfer (bijvoorbeeld een 2).

Kruis het hokje aan bij het cijfer dat je kiest.

(heel onzeker) ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 (heel zeker)

7

De Balans

Voer nu het proefje uit op de computer!

Wat gebeurde er?

____

____

____

Maakt het gewicht van de potjes uit? Zo ja, hoe dan?

____

____

____

Maakt het volume van de potjes uit? Zo ja, hoe dan?

____

____

____

Maakt de positie van de potjes uit? Zo ja, hoe dan?

____

____

____

Geef een rapportcijfer over hoe zeker je bent van je antwoord.

(heel onzeker) □ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □ 8 □ 9 □ 10 (heel zeker)

Attachment D

Design of the hypothesizing and experimenting pages, question 2 for the simulation condition (workbook part 2).

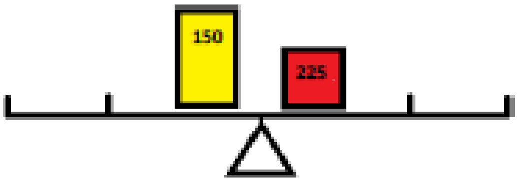
De Balans

Opgave 2:

Stel je de volgende situatie voor:

Links van het middelpunt staat een middelgroot potje van 150 gram op positie 1.

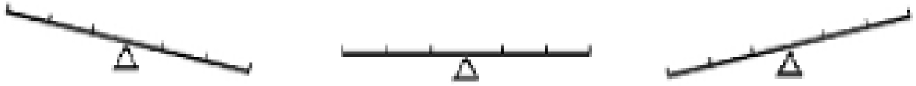
Rechts van het middelpunt staat een klein potje van 225 gram op positie 1.



(a). Wat denk je dat er gebeurt als je deze potjes op de balans zet?

GEBRUIK HIERBIJ GEEN COMPUTER!!!

Kruis het juiste hokje aan



☐ Rechts gaat naar beneden ☐ In evenwicht ☐ Links gaat naar beneden

(b). Waarom denk je dat?

____

____

____

____

(c). Geef een rapportcijfer over hoe zeker je bent van je antwoord bij opgave (a).

(heel onzeker) ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 (heel zeker)

De Balans

Voer nu het proefje uit op de computer!

Wat gebeurde er?

____

____

____

Maakt het gewicht van de potjes uit? Zo ja, hoe dan?

____

____

____

Maakt het volume van de potjes uit? Zo ja, hoe dan?

____

____

____

Maakt de positie van de potjes uit? Zo ja, hoe dan?

____

____

____

Geef een rapportcijfer over hoe zeker je bent van je antwoord.

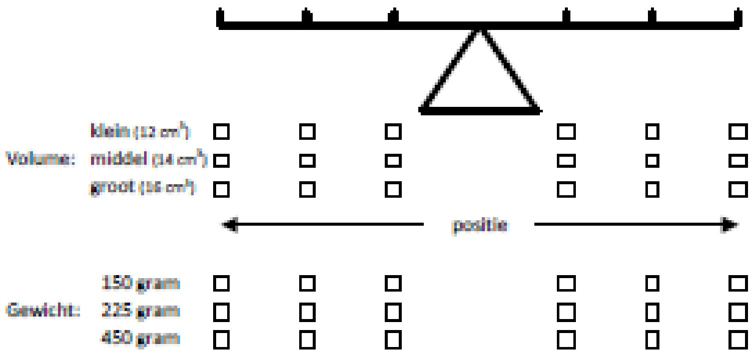
(heel onzeker) ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 (heel zeker)

Attachment E

Design of the hypothesizing and experimenting page after completing the first (obligatory) six questions (workbook part 2).

De Balans

(a) Welk proefje wil je nu uitvoeren?



(b) Waarom is dit proefje nu handig?

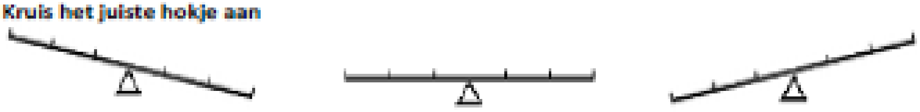
____

____

____

(c). Wat denk je dat er gebeurt als je deze potjes op de balans zet?
<u>GEBRUIK NU GEEN COMPUTER!!!</u>

Kruis het juiste hokje aan



☐ Rechts gaat naar beneden ☐ In evenwicht ☐ Links gaat naar beneden

(d). Waarom denk je dat?

____

____

____

(e) Geef een rapportcijfer over hoe zeker je bent van je antwoord bij opgave (a).

(heel onzeker) ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 (heel zeker)

De Balans

Voer nu het proefje uit op de computer!

Wat gebeurde er?

____

____

____

Maakt het gewicht van de potjes uit? Zo ja, hoe dan?

____

____

____

Maakt het volume van de potjes uit? Zo ja, hoe dan?

____

____

____

Maakt de positie van de potjes uit? Zo ja, hoe dan?

____

____

____

Wil je nog een proefje doen? Kleur het hokje in.

□ Ja: vul een nieuwe gekleurde kaart in.

□ Nee: doe je boekje dicht en steek je vinger op.