Learning by ZAP-ing

Learning Effects of Doing versus Viewing

Masterthesis

May 2011

Matthias Matzner, S0109452

First Supervisor: Dr. Hans van der Meij

Second Supervisor: Dr. Tessa Eysink

Abstract

This article investigates whether learning by doing is more beneficial than learning by viewing. 60 university students followed two ZAPs. The students were randomly assigned to either of two groups, a learning-by-doing group that followed the ZAPs normally and a learning-by-viewing group that were offered a print-out of the ZAPs. Three knowledge measures (factual declarative knowledge, conceptual declarative knowledge and procedural knowledge) were taken. The analysis focuses on comparison between and within groups. Results showed significant difference between both groups on the efficiency measure for ZAP1. All other differences were not statistically significant, suggesting that learning-by-doing and learning-by-viewing with ZAPs are equally effective. Doing as well as viewing seems to invoke cognitive activity and lead to learning. Implications for further research are discussed.

Contents

1 Introduction

1.1 Theoretical Background

Is doing helpful for learning? Any knowledge is supposedly deeper when gained through an active approach (i.e., doing; Jonassen, 1999; Stull & Mayer, 2007). It is claimed that engaging a problem actively instead of passively consuming the solution is the better way of learning. In more general terms, an active approach means acquiring knowledge through one's own means, while a passive approach is acquiring the knowledge by provided means.

The behavioural activity/passivity is thought to lead to cognitive activity/passivity, which then leads to learning (Mayer & Moreno, 2002). One idea that this article builds on is that doing, which is behavioural activity leads to cognitive activity, and viewing, which is behavioural passivity leads to cognitive passivity. When a learner is cognitively active, the subject matter is better learned than when the learner is cognitively passive (Moreno & Mayer, 2002). Thus learning material should be constructed so that the learner has to engage actively with the material instead of consuming it passively.

Basic statistics can serve as an example to illustrate the difference between an active learning approach and a passive learning approach. The knowledge that is to be conveyed is about the statistical parameters the mean and the standard-error. In a passive approach the teacher may give an explanation of these terms and then demonstrate how these terms are calculated. There is no problem that the learner can/has to engage with. The task in the passive approach is remembering and organizing the information into a coherent mental model.

This passive approach can also be described as learning-by-viewing (Stull & Mayer, 2007). Learning by viewing is closely related to the Nurnberg Funnel in which knowledge is poured into the brain of the learner and the learner thus remains mostly inactive (Farkas & Williams, 2002). The passive approach means that the learner uses the provided structures of the learning material only. Nevertheless, learning-by-viewing has its merits since it is effortless opposed to an active learning approach that could create too much cognitive load (Stull & Mayer, 2007), hence disrupting the processing of information (i.e. learning). According to Cognitive Load Theory the brain has limited processing capabilities and too much redundant information and activity hinders the learning process. A well-constructed passive learning approach can optimally utilize the processing capabilities and prevent overload.

An empirical example of well working passive learning is reported by Cheng & Warren (2005). Their article focused on the potential of incorporating peer review into language lessons. Undergraduate Engineering students were asked to assess the language proficiency of their peers as exhibited in seminars, presentations and alike. Both teachers and students found the peer review beneficial for deep language learning. The assessment added to the value of the lessons by creating new, individual guidance through the language learning process, while allowing low cognitive load and learn without doing any more than consuming the offered critique and advice.

On the other side of the spectrum there is the active approach. An example of an active approach on the statistics example would be to describe to another learner how a certain distribution looks like and then find numerical expressions to refine that description. When two distributions are presented, one flat, the other pointed, an easy way to describe them is in terms of "flatness" and "pointy-ness". When the refined task is to describe them more thoroughly, the learner has to come up with a way to find numerical expressions for the "flatness" and "pointy-ness" of the distribution. The learners derive necessity and meaning of the statistical parameters on their own.

This process of acquiring knowledge about a problem while solving the problem itself is called learning-by-doing (Savery & Duffy, 2001). Learning-by-doing may take longer to complete, but is said to be more effective (Eysink et al, 2009). According to Activity Theory, conscious learning can only emerge by activity and cannot happen without it (Jonassen, 1999; Stull & Mayer, 2007). Consciousness is where attention, intention, memory, reasoning, and speech are joined (Vygotsky, 1978). Jonassen (1999, p 65) put it bluntly "learning and doing are inseparable". This means for any learning environment that it has to incorporate an appropriate activity to foster conscious learning.

A few empirical examples of well working active learning are reviewed by Lesgold (2001). His article reviews several learning-by-doing approaches, most notably a project called Sherlok where students and Airforce personnel alike learned complex problem solving skills. Technicians were taught by intelligent tutors of how to repair test stations for aviation electronics in field. Results showed dramatically improved problem solving capabilities for both groups. Another implementation of the learning-by-doing approach of Lesgold at a corporation led to similarly beneficial results. The corporation was initially the most opposed to the learning-by-doing approach, since they have to think in cost-efficient terms.

When it comes to applying the above in educational material the theoretically poles of the two directions dissipate and a continuum spanning from passive to active emerges. There are 4 defined approaches mentionable here: (a) hypermedia learning, where the learner is told how it works, (b) observational learning, where the learner is shown how the subject matter works (c) self-explanation–based learning, lets the learner explain how it works, and (d) inquiry learning, lets the learner investigate how it works (Eysink, de Jong, Berthold, Kolloffel, Opfermann, & Wouters, 2009).

In hypermedia learning, the learner is presented with bits of information which are linked and accessed through hyperlinks. This approach is most close to being passive of the four approaches. The learner can self determine what he wants to learn and the speed at which he wants to learn. The learner's most crucial task with hypermedia learning is to integrate the bits of information into a coherent mental model. However, it seems difficult for learners to identify the information they need to close their knowledge gaps (Eysink et al, 2009). On a sidenote, this also becomes apparent when learners are supposed to learn from viewing texts (Brown, Campione & Day, 1981).

In observational learning, the learner observes an expert performing a task or solving a problem while explaining the procedure and underlying rationale for this procedure. This approach represents a more practical learning approach than hypermedia learning since the expert performs an authentic task, rather than presenting loose bits of information. The subject matter can target actually relevant tasks to enhance motivation in the learner. Further, the learner gains insight into the decision processes and problem solving strategies. Again, there are also some pitfalls for this approach; firstly, the learner may not be able to filter out the relevant information from the irrelevant, and secondly, the learner may passively observe the expert instead of actively and immediately using the expert's information (Eysink et al, 2009).

In self-explanation–based learning, the learner is presented with worked-out examples and encouraged to trace the thoughts in the solution. This approach represents a more practical approach than observational learning since the learner now not only has to perceive, but also to explain what is happening. The learner is not forced to come up with a solution on his own, but rather can spend his mental resources on retracing the solution steps. Missing information becomes apparent through self-explanation. However, it has been shown that learners have difficulties in understanding worked-out examples and the self-explanation process is not spontaneously started (Eysink et al, 2009). Another study also supports that worked-out examples have their pitfalls, most notable for adopting a conventional problem-solving approach (Darabi, Nelson & Palanki, 2007).

Inquiry learning is strongly associated with activity and is based on experiential learning (Kolb, 1984). Kolb's learning cycle is comprised of four stages which are run through consecutively. In inquiry learning the learner is confronted with an information rich environment and observes, reflects, abstracts and generalizes, then observes anew to close the cycle. Since this learning process is cyclical there is no definite starting point from which learning should begin. One stage is where the learner starts with a concrete experience. The learner has to pay attention to the experience to fully absorb the event. The learner then reflects on the made observations in the second stage, relating the observations to existing conceptions. Then the learner abstracts the earlier observations and comes to a generalization in the third stage. The learner then tests his generalization by encountering new situations, closing the circle. In doing this the learner can come to deep conceptual understanding of the subject matter. However, it has been shown that learners are often not sufficiently able to generate hypotheses, perform valid experiments, and draw valid conclusions (Eysink et al, 2009; Kirschner, Sweller, & Clark, 2006).

Besides the learning process, the psychology of learning is concerned also with the outcome of learning. The outcome of any learning process is knowledge, which can be further described by type and quality (de Jong & Ferusson-Hessler, 1996). Knowledge types can be described as situational (knowledge about events and cases), declarative (factual information and concepts; see also Friege & Lind, 2006), procedural (process-based knowledge that enables to act; Anderson, 1982; Lewicki, Czyzewska, & Hoffman, 1987; Bowles & Gintis, 2002), and strategic (analyzing and thinking up more efficient ways to solve a problem). Whereas knowledge qualities can be described as level (surface versus deep), structure (isolated versus structured), automation (declarative versus compiled), modality (verbal versus pictoral), and generality (general versus specific). The knowledge types relevant for the paper at hand are declarative (factual information and concepts) and procedural (process-based knowledge that enables to act). Using the example of statistical parameters again, the knowledge about what the term *standard error* means is declarative factual knowledge. The knowledge about the relation between *standard error* and the *mean* is declarative conceptual knowledge. The knowledge of how to calculate these parameters is procedural knowledge, i.e. how to produce these parameters from the observed values.

When learning leads to the desired outcome this phenomenon is called (positive) transfer. Transfer is the phenomenon that an input moment (i.e. learning) changes subsequent behavior. In more technical terms, the learning process is the transfer source, and the performance or knowledge is the transfer target (Helfenstein, 2005). The domain the learning material stems from is another factor in the equation on learning. The two domains this article focuses on are mathematics and cognitive skills. Both domains seem to have a history of being taught using the passive approach. Education in mathematics revolves around enabling conceptual change of faulty mental models/concepts (Carey, 2000). In this the learners are not active. The teacher tries to find out where the learners err and then demonstrates where and why these concepts are flawed. Although there are researchers that propose an investigative approach for pupils to learn mathematics (Ponte, 2001; Hiebert, Morris and Glass, 2003), now the gross of mathematics education subscribes the learners to a passive role.

Procedural knowledge (how to process and reason with certain information) too is often taught in a passive approach. An interesting observation is that transfer source and target do not have to match. It is possible to learn cognitive skills – reasoning skills to be precise – through viewing. Renken and Nunez (2010) conducted an experiment where participants had to either read (i.e. viewing a text) about an inconsistency in a simple physics experiment or conduct the physics experiment themselves and discover the inconsistency. Viewing showed better results than conducting the experiment in which too much cognitive load may have hindered the learning process of the participants.

Procedural knowledge is often complex and it seems obvious that teaching them is also complex. Contra to this conception, Lee and Anderson (2001) showed that complex procedural knowledge can be decomposed into simple actions which can, in return, be learned faster. The authors used the Kanfer-Ackerman Air Traffic Controller Task as a complex task and decomposed the task into simple sub-tasks. The decomposition showed a speed-up in the learning performance. However promising, decomposition of more complex cognitive procedural knowledge is also laborious (e.g. Brand-Gruwel, Wopereis and Vermetten, 2005). A possibility to acquire knowledge is using ZAPs. ZAP is a Dutch project and stands for *zeer actieve psychologie*, which translates into *very active psychology*. The ZAP project bore a variety of stand-alone applications, called ZAPs. ZAPs can be described as "short, self-contained computer programs that encourage students to experience psychological phenomena in a vivid, self-explanatory way, and that are meant to evoke enthusiasm about psychological topics" (Hulshof, Eysink, Loyens and de Jong, 2005). ZAPs offer the possibility to actively engage in experimenting and discovery (Eysink et al., 2003). ZAPs are generally built up of 3 components: an introduction, an activity and a concluding section. The introduction introduces the phenomenon that the ZAP deals with. Some introductions start with an example from everyday life where the topic of the ZAP applies. The main idea behind the introduction is to disclose enough information about a topic to make the reader interested in the phenomenon. The activity component of any ZAP exemplifies the phenomenon or offers the possibility to discover the phenomenon or to engage in an experiment. The activity itself consists of 3 subcomponents, namely, instructions on how to complete the activity, the activity itself and an evaluation of the data. The instructions try to ensure a smooth user experience. The activity can either involve exploration, experimentation or experiencing. The theory section explains the phenomenon. The student can use this information to put the experience into context and form a broader and deeper understanding of the phenomenon as well as the science behind it.
ZAPs support learning-by-doing (Hulshof, C., Eysink, T., & de Jong, T., (2006); Kolb, 1984). An earlier study assessed two learning outcomes of ZAPs (Hulshof, Eysink, Loyens, & de Jong, 2005). The authors looked at declarative factual knowledge and declarative conceptual knowledge. They observed that the learner gains initially less declarative knowledge than a control group, but retains the information better, through using ZAPs. The authors conclude that ZAPs are well suited to promote better retention and understanding of factual information.

1.2 Research Questions

Which is better: learning-by-viewing or learning-by-doing? One group will engage in learning-by-doing by working with the ZAPs as they are intended (DoZAPs), the other group will engage in learning-by- viewing by working with the ZAPs in which the activity is replaced by a text on the phenomenon (ViewZAPs).

1. Is there a difference in time spent on ZAPs between the groups?

The DoZAPs will probably take longer than the ViewZAPs. The DoZAPs are more immersive and provide a novel activity to follow. It is necessary to determine whether time spent on the ZAPs should be treated as a covariate in the further analysis.

2. Is there a difference on procedural knowledge between the groups?

The difference in knowledge between groups will tell whether learning-by-doing differs from learning by viewing for procedural knowledge. It may be safe to assume that procedural knowledge is learned by DoZAPs (Aleven & Koedinger, 2002), but the question whether the amount learned is equal to ViewZAPs remains (Renken & Nunez, 2010).

3. Is there a difference on declarative knowledge between the groups?

The difference in knowledge between groups will tell whether learning by doing matters for learning declarative knowledge. Research on ZAPs suggests that we should expect better results for DoZAPs (Hulshof, Eysink, Loyens, & de Jong, 2005).

4. Is there a difference in gain of procedural knowledge within the groups?

The difference in learning gain will tell whether each group learns well and consecutively if the groups learn differently.

5. Is there a difference in gain of declarative knowledge within the groups?

The difference in learning effect will tell whether each group learns well and consecutively if the groups learn differently.

2 Method

2.1 Participants

60 students with a mean age of 20 years (SD 1.35 years) from the University of Twente participated in the study. The students were recruited from the psychology and the communication sciences track. The participants were rewarded with 3 points for a mandatory point system of the university. Participants were randomly assigned to conditions. Chi-square checks for randomization on demographic data showed no significant differences between groups.

Three participants worked longer on both ZAPs than the prescribed limit and were excluded on analyses regarding the respective ZAP. Additionally, three participants worked longer than 15 minutes on ZAP1 and were excluded from these analyses; seven participants worked longer than 30 minutes on ZAP2 and were excluded from the analyses for this ZAP. All exclusions due to time took place in the active condition. The number of analysed participants fluctuated between measures due to incomplete data.

2.2 Materials

ZAP1: Confirmation Bias. ZAP1 was about confirmation bias. In this ZAP one has to find out what rule underlies a consecutive order of numbers. The set of 2-4-6 suggests the rule that from left to right the numbers increase by 2. People show the tendency to confirm this assumption by trying other sets of numbers that confirm this (e.g. 8-10-12).

After the presentation of the 2-4-6 sequence with an unknown rule the participant is asked to enter a sequence of three numbers. Then the system responds whether this sequence corresponds to the underlying rule (see Figure 1). The participant can continue testing sequences of three numbers until he is certain that he knows the rule. The participant then has to click on a triangle which leads to a screen with different rules. The computer offers a rule and asks the user to determine whether that rule is correct. When the rule is incorrect the participant is advised to try different sets of numbers. The participant continues until the rule is discovered. When the rule is correct, the participant is provided the theory behind the confirmation bias.

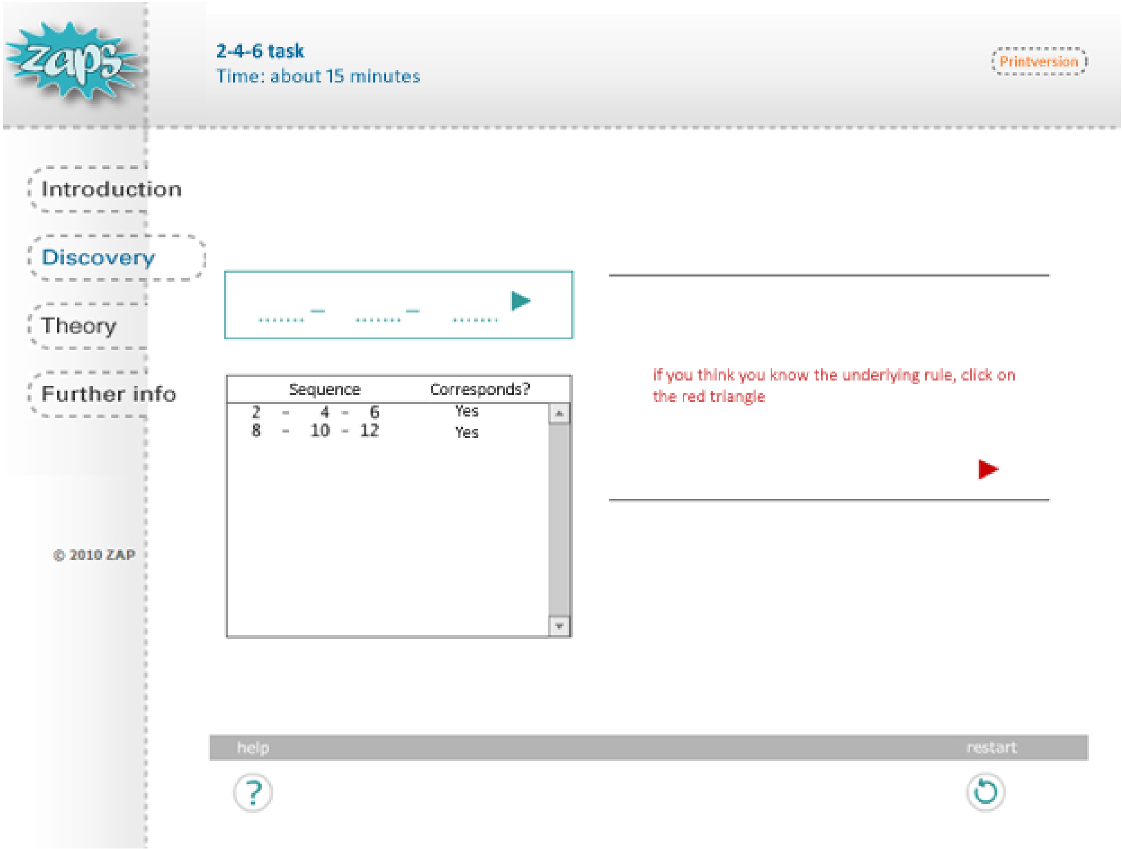


Figure 1. Example Screenshot of the Activity of the 2-4-6 ZAP

ZAP2: Signal Detection. In psychophysics, signal detection tasks determine the ability of people to give the right response to the strength of a real signal (e.g., a telephone ringing) or to a noise. Depending on the signal strength the sensor (human or automatic) gives a response which can be used to determine the ratio of hits (signal present and detected) over misses (signal present but undetected), and correct rejections (signal missing and no response) over false alarms (signal missing but "detected").

In the ZAP, the participant can exercise with different questions and tasks (see Figure 2). The ZAP includes 14 exercises. After he is finished his score is displayed. Each exercise needs to be filled in correctly by the participant to be permitted to continue.

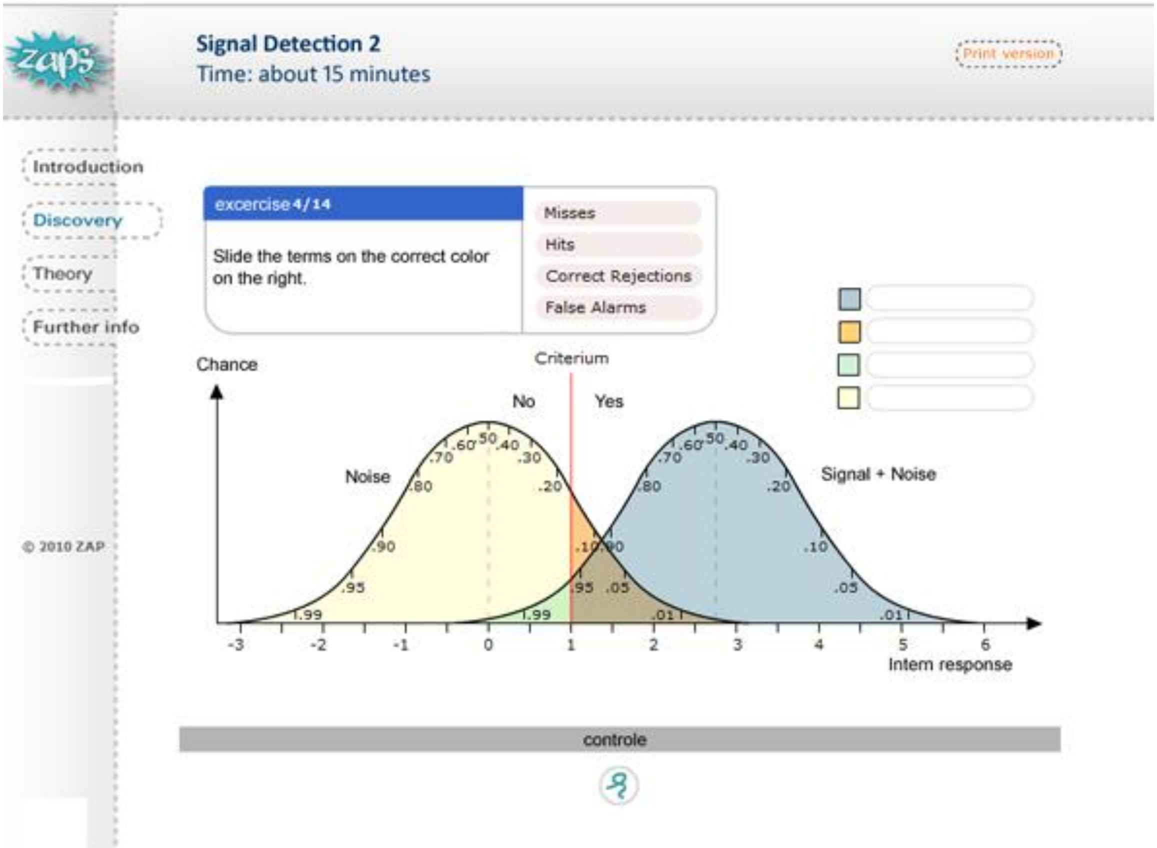


Figure 2. Example Screenshot of the Activity of the Signal Detection 2 ZAP; Dutch task description reads: "slide these terms onto their correct colours to the right."

Tests

Procedural knowledge test. The post- and retention-test each consisted of six tasks. The first task concerned ZAP1, the others ZAP2.

Procedural knowledge of ZAP1 was assessed by testing for confirmation bias. This was done by the activity of the "concept formation" ZAP. In this ZAP, the participant had to find out an underlying rule by clicking on tiles. The rule dictates the properties the tiles have on four dimensions (i.e., color, object, amount of objects and amount of frames). Each dimension can have one of four properties (see Table 1).

Table 1. Possible properties of the rule on "concept formation" ZAP

Dimensions	Property			
Color	Red	Green	Black	Irrelevant
Object	Crosses	Circles	Rectangles	Irrelevant
Amount of Objects	One	Two	Three	Irrelevant
Amount of Frames	One	Two	Three	Irrelevant

Figure 3 displays an example screen. The start screen displayed only the third tile from the left from the left hand side. Every tile that is being chosen is displayed on the lower left portion of the screen, signifying whether that tile corresponds to the underlying rule.

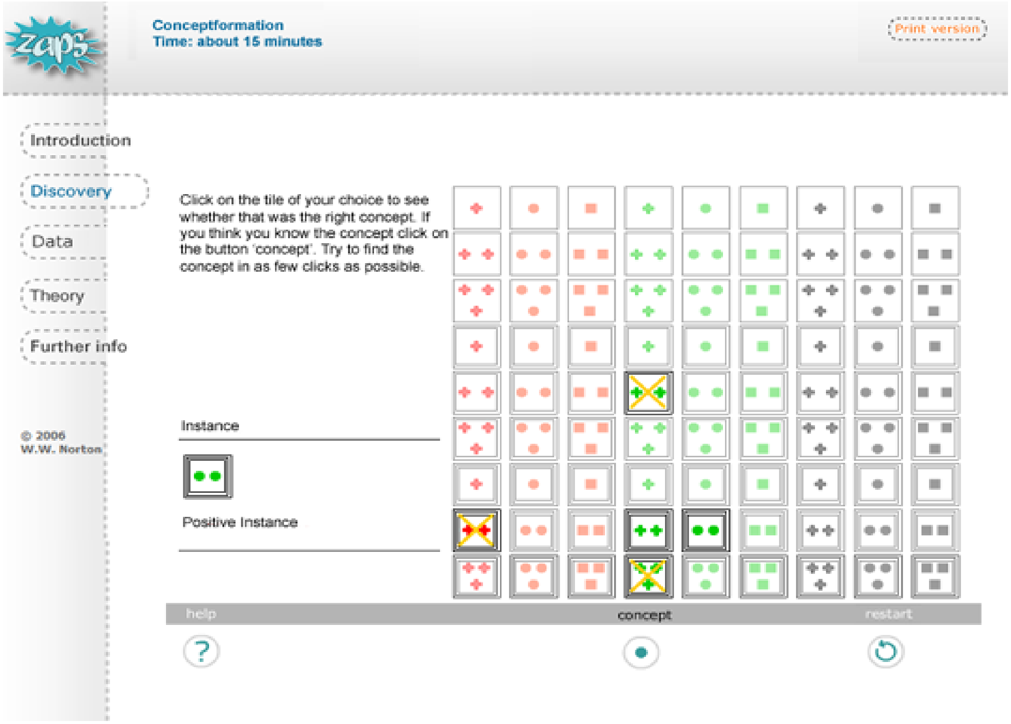


Figure 3. Example of the Task to assess the Procedural Knowledge of the Confirmation Bias Domain.

The participant can then try to fill in the correct rule (see Figure 4). When the rule was not discovered, the participant was returned to the tiles and had to continue to find the rule. When the rule was correct the application displayed the sequence of tiles chosen (see Figure 5).

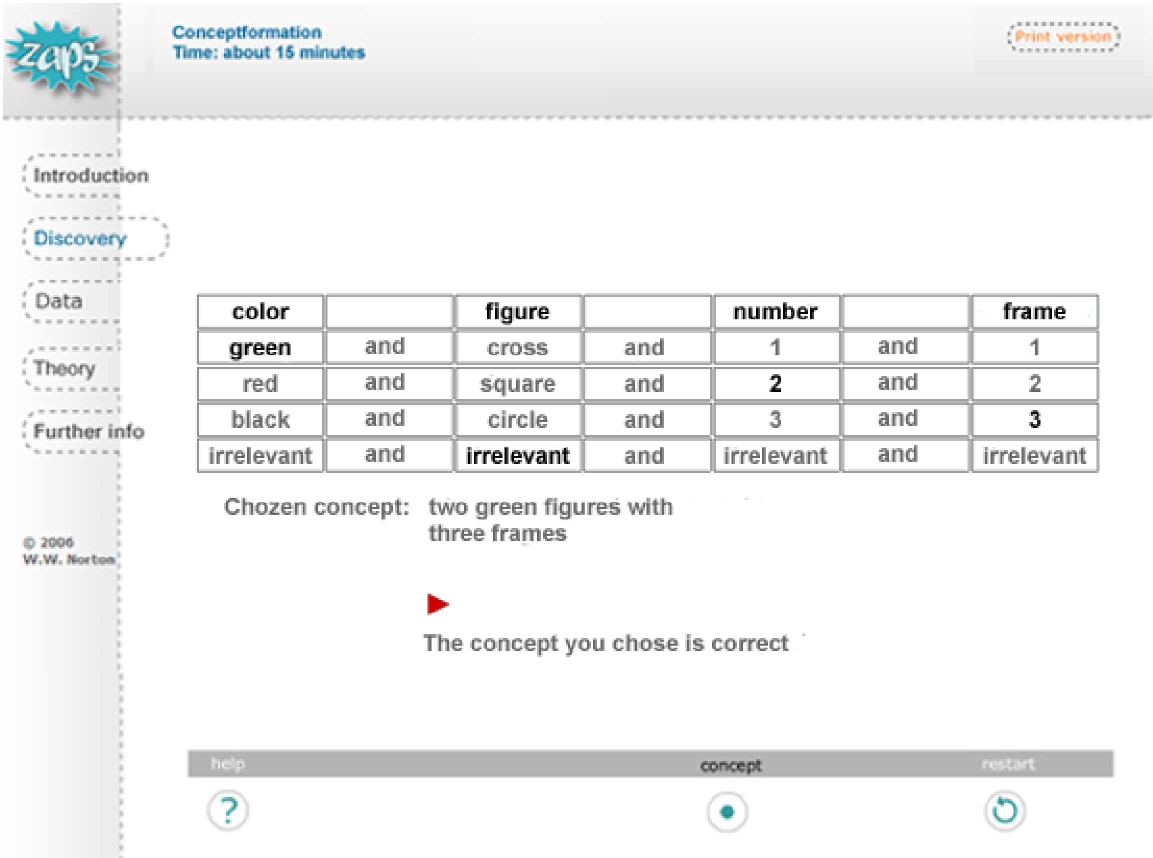


Figure 4. Example screen of an entered rule of the concept formation task

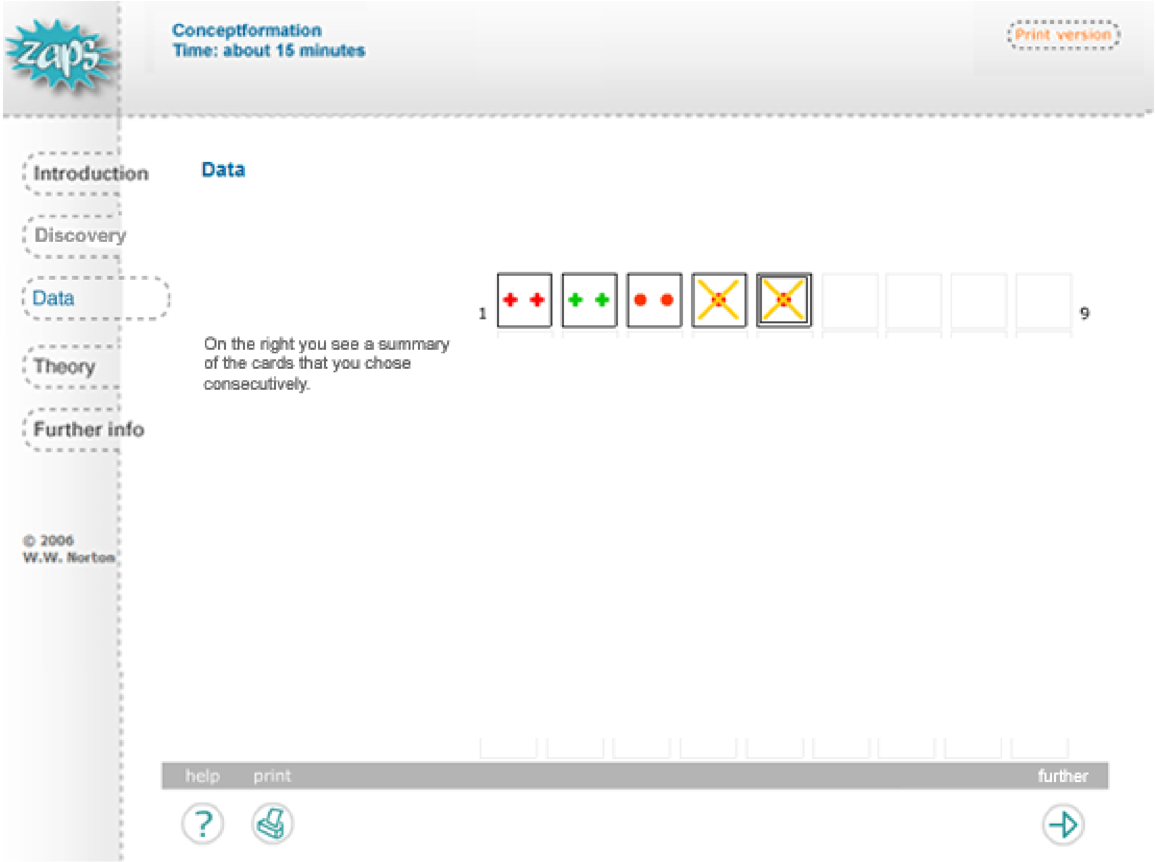


Figure 5. Example tile sequence of the concept formation task

Every tile chosen by the participant was defined as a trial. One measure of effectiveness was the total number of trials. Another measure was the efficiency of the trials. The efficiency score was calculated as the ratio of total number of trials and trials that provided novel information. Through the screen-saves of trials (see Figure 5) it was possible to conduct a sequential analysis.

Every trial was evaluated whether it provided novel information. A trial was defined as not providing novel information when no new conclusion about the underlying rule could be inferred (see Appendix A).

The efficiency measure was calculated with the following formula:

Number of trials providing novel information/Total amount of trials x 100

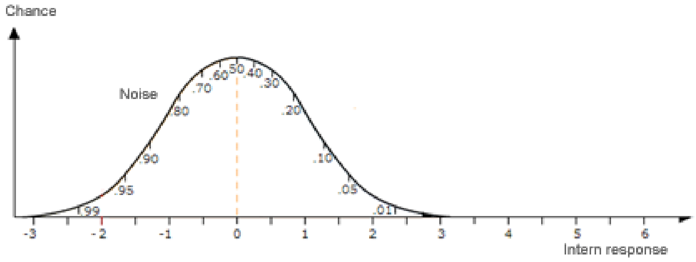
The efficiency measure reflects the participant's tendency to (dis-)confirm previous information. Participants were instructed to complete the task. The task was limited to 80 trials. A pre-test was omitted to avoid testing effects.

There were five tasks about ZAP2. The ZAP2 tasks were story-based signal detection tasks where the participant had to fill in criterium, signal curve and/or the associated chance (see Figure 8). The score was calculated by giving one point for every correct element of the answer. The maximum score for ZAP2 was 15 points.

Prost-and Retentio-test procedural ZAP2

Athletes often pretend to be wounded to get a penalty or similar for their team. Suppose you have to give advice to an official of the sport's union to increase the threshold for when the play will be

interrupted due to an (alleged) injury. To support your suggestion you have to sketch out a diagram about that situation. At the moment 99 percent of all real injuries and 30 percent of all faked injuries lead to an interruption.
Draw the curve and the threshold in the diagram. Pay special attention to the top of the curve, the intersection of the curves and the threshold and give the percentages where they meet.



Declarative knowledge tests. To assess pre-existing factual and conceptual knowledge about ZAP1 items were used from Hulshof & Eysink (2005). The questions about ZAP2 were constructed to be similar. For every correct element of the answer one point was given (see Table 2). The post-test and retention-test items were constructed to be similar and were scored in the same way as the pre-test (see Table 3).

Table 2. Declarative knowledge: Pre-test

Number of items	knowledge aspect and domain	Maximum score
2	Factual Knowledge ZAP1	4
2	Factual Knowledge ZAP2	4
2	Conceptual Knowledge ZAP1	8
2	Conceptual Knowledge ZAP2	4

Table 3. Declarative knowledge: Post- and retention-test

Number of items	knowledge aspect and domain	Maximum score
2	Factual Knowledge ZAP1	4
2	Factual Knowledge ZAP2	4
3	Conceptual Knowledge ZAP1	6
3	Conceptual Knowledge ZAP2	6

Following, example items for each test session are given. Note that the post-test and the retention-test consisted of items which were slightly altered on a surface level, while retaining the structural level.

Pre-test facts ZAP1

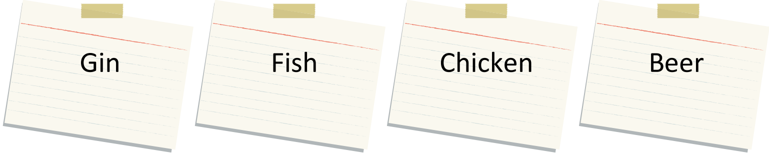
What is confirmation bias?

Pre-test facts ZAP2

What is a response criterium?

Pre-test conceptual ZAP1

Situation: There are 100 quests in a restaurant. Everyone orders a drink and a meal. Every guest orders either a gin or a beer as drink and either chicken or fish as meal. All orders are put on a separate card. The drink is on one, the meal on the other side of the card. The cards of four guests are presented below. The restaurant follows rigid rules when it comes to combination of drinks and meals. Everyone that orders fish has to drink gin with it. Which cards have you have to turn to find out whether the rule is followed by the guests?



Circle the card(s) which you would turn.

Pre-test conceptual ZAP2

As a fire detector goes off too often while there is no fire, how would the sensitivity to be altered to improve this?

Post-test/Retention-test Factual ZAP1

What is the mistake called which is often done in the 2-4-6 task and what does this mistake entail?

Post-test/Retention-test Factual ZAP2

Suppose that somebody believes to have heard a noise while there actually was none. What is the mistake called if this somebody would react to this noise?

Post-test/Retention-test Conceptual ZAP1

The sequence 2-4-6 which corresponds to the underlying rule is given at the beginning of the 2-4-6 task. If your idea is that the rule reads something like: “the sequence consists of three voluntarily placed numbers under 10”, what is the fastest manner to test if this rule is correct?

Post-test/Retention-test Conceptual ZAP2

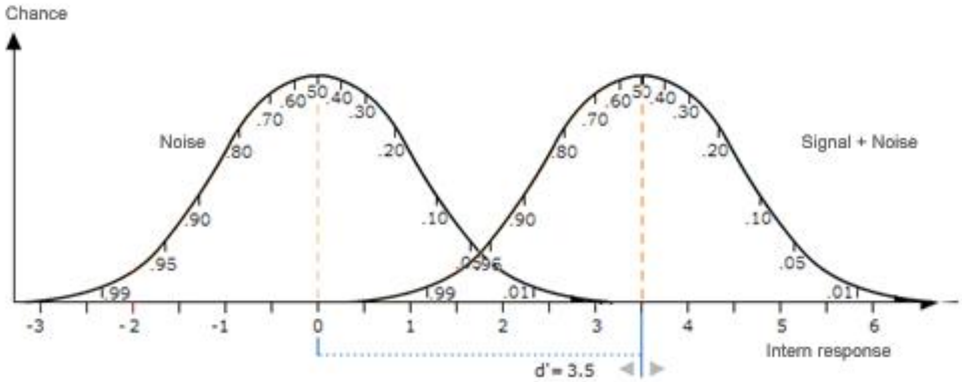
Look at the picture below. You may recognize it from the “Signal Detection” information. Where do you have to place the criterium so that there are as few false alarms as possible and so many hits as possible. Describe your answer in respect to the intern respons (thus in numbers, -3 up to 6):

____

____

____

____



3. Procedure

Participants were welcomed in a study room with pc's and then received a 5-minute description of the experiment. Next, they signed a consent form for participation and answered the demographics questionnaire. Thereafter, they were given 10 minutes to complete the pre-test. Participants were then seated behind a computer and could start working with the ZAPs. Participants had 15 minutes to complete ZAP1 and 30 minutes to complete ZAP2. Immediately after working with the ZAPs participants received a post-test for which they had 30 minutes to complete. The participants received a declarative knowledge questionnaire and then took the procedural knowledge tests. Thereafter, they were thanked for their cooperation and a new appointment for the retention test was set up. The time between the post and retention test was about seven days. Participants had 30 minutes to complete the retention test.

4. Analysis

The program SPSS (Statistical packages for Social Sciences) was used for the analyses. The data were analysed using AN(C)OVA, analysis of (co)variance. IF the time for completing the ZAPs differs significantly between groups it will be taken into account as a covariate for the rest of the analyses. Cohen's d will be used to determine learning effects.

5. Results

5.1 Is there a difference in time spent on ZAPs between the groups?

ZAP1 & 2. A univariate analysis of variance (ANOVA) was performed on the *time* to complete the ZAPs (Table 4). The analysis showed a significant difference between conditions for ZAP1 ($F(1,54)= 7,21$, $p=.010$) and for ZAP2 ($F(1,49)= 55.53$, $p=.00$). Participants in the DoZAP condition needed about 30% longer than those in the ViewZAP condition for completing ZAP1 and about 100% longer for ZAP2. *Time* was thus incorporated as a covariate in the relevant analyses.

Table 4. Time (in absolute minutes) for conditions on ZAPs

Condition		ZAP1			ZAP2	
		M	SD		M	SD
DoZAP	n=26	10.1	4.3	n=22	22.6	8.5
ViewZAP	n=29	7.6	2.6	n=28	9.9	2.8

Maximum for ZAP1 is 15 minutes; maximum for ZAP2 is 30 minutes

5.2 Is there a difference on procedural knowledge between the groups?

ZAP1. A univariate analysis of covariance (ANCOVA) was performed on the *number of trials* on ZAP1 (Table 5). The analysis showed no significant difference between conditions on both the post-test and the retention-test ($F(1,51)=1.52$, $p=0.22$ and $F(3,51)=0.14$, $p=0.71$, respectively).

Another univariate analysis of covariance (ANCOVA) was performed on the *efficiency* on ZAP1 (Table 6). The analysis showed no significant difference between conditions on both the post- and the retention-test ($F(1,52)=2.98$, $p=0.09$ and $F(3,51)=1.89$, $p=0.18$, respectively).

Table 5. Procedural Knowledge for conditions on ZAP1: number of trials

Condition		Post-test		Retention -test	
		M	SD	M	SD
DoZAP	n=25	14.2	15.4	7.1	4.2
ViewZAP	n=27	9.3	8.8	6.9	2.7

Maximum number of trials is 80

Table 6. Procedural Knowledge for conditions on ZAP1: efficiency

Condition		Post-test		Retention -test	
		M	SD	M	SD
DoZAP	n=25	63.0	20.5	80.0	14.6
ViewZAP	n=27	74.2	18.7	80.3	16.2

Maximum efficiency is 100

ZAP2. A univariate analysis of covariance (ANCOVA) was performed on the procedural knowledge about ZAP2 (Table 7). The analysis showed no significant difference between conditions on both the post- and the retention-test ($F(1,49)=0.02$, $p=0.88$ and $F(1,49)=1.95$, $p=0.17$, respectively).

Table 7. Procedural Knowledge for conditions on ZAP2

Condition	N	Post-test		Retention -test	
		M	SD	M	SD
DoZAP	22	4.3	4.3	4.2	3.8
ViewZAP	28	4.1	3.6	3.4	4.0

Maximum is 15

5.3 Is there a difference on declarative knowledge between the groups?

ZAP1. A univariate analysis of covariance (ANCOVA) was performed on the *overall declarative knowledge* over ZAP1 (see Table 8). The analysis showed no significant difference between conditions on the pre-, post-, and retention-test ($F(1,54)=2.48$, $p=0.12$, $F(1,53)=0.50$, $p=0.48$ and $F(1,53)=0.01$, $p=0.92$, respectively).

Table 8. Declarative Knowledge overall for conditions on ZAP1

Condition		Pre-test		Post-test		Retention-test	
		M	SD	M	SD	M	SD
DoZAP	n=26	4.9	1.8	6.8	2.8	5.6	2.4
ViewZAP	n=29	5.7	2.1	6.7	2.6	5.8	2.4

Maximum for the pre-test is 12; the maxima for the post-test and the retention-test are 10

A univariate analysis of covariance (ANCOVA) was performed on the *declarative factual knowledge* about ZAP1 (see Table 9). The analysis showed no significant difference between conditions on the pre-, post-, and retention-test ($F(1,54)=0.04$, $p=0.83$, $F(1,53)=0.96$, $p=0.33$ and $F(1,53)=0.01$, $p=0.93$).

Table 9. Declarative Factual Knowledge for conditions on ZAP1

Condition		Pre-test		Post-test		Retention-test	
		M	SD	M	SD	M	SD
DoZAP	n=25	0.5	0.8	2.6	1.3	1.7	1.3
ViewZAP	n=29	0.6	0.9	3.0	1.3	1.9	1.1

Maximum is 2 for the pre-test, 4 for the post-test and the retention-test

A univariate analysis of covariance (ANCOVA) was performed on the *declarative concepts* over ZAP1 (see Table 10). The analysis showed no significant difference between conditions on the pre-, post-, and retention-test ($F(1,54)=2.79$, $p=0.10$, $F(1,53)=1.89$, $p=0.18$ and $F(1,53)=0.02$, $p=0.89$).

Table 10. Declarative Conceptual Knowledge for conditions on ZAP1

Condition		Pre-test		Post-test		Retention-test	
		M	SD	M	SD	M	SD
DoZAP	n=25	4.4	1.6	4.2	1.8	3.9	1.6
ViewZAP	n=29	5.1	1.9	3.7	1.8	3.9	1.7

Maximum is 8 for the pre-test, 6 for the post-test and the retention-test

ZAP2. A univariate analysis of covariance (ANCOVA) was performed on the *declarative overall knowledge* about ZAP2 (Table 11). The analysis showed no significant difference between conditions on the pre-, post-, and retention-test (F(1,49)=0.54, p=0.47, F(1,49)=0.50, p=0.49 and F(1,49)=0.99, p=0.33).

Table 11. Declarative Knowledge overall for conditions on ZAP2

Condition		Pre-test		Post-test		Retention-test	
		M	SD	M	SD	M	SD
DoZAP	n=22	3.3	1.1	3.2	3.0	3.4	3.0
ViewZAP	n=28	2.9	2.2	3.9	2.6	2.8	2.5

Maximum is 8 for the pre-test, 10 for the post-test and the retention-test

A univariate analysis of covariance (ANCOVA) was performed on the *declarative factual knowledge* about ZAP2 (Table 12). The analysis showed no significant difference between conditions on the pre-, post-, and retention-test (F(1,49)=0.35, p=0.56, F(1,49)=0.03, p=0.86 and F(1,49)=2.4, p=0.13).

Table 12. Declarative Factual Knowledge for conditions on ZAP2

Condition		Pre-test		Post-test		Retention-test	
		M	SD	M	SD	M	SD
DoZAP	n=22	0.7	0.9	1.6	1.2	1.6	1.6
ViewZAP	n=28	0.9	1.1	2.1	1.3	0.7	0.9

Maximum is 4 for the pre-test, 4 for the post-test and the retention-test

A univariate analysis of covariance (ANCOVA) was performed on the *declarative conceptual knowledge* about ZAP2 (see Table 13). The analysis showed no significant difference between conditions on the pre-, post-, and retention-test (F(1,49)=2.16, p=0.15, F(1,49)=1.03, p=0.32 and F(1,49)=0.11, p=0.75).

Table 13. Declarative Conceptual Knowledge for conditions on ZAP2

Condition		Pre-test		Post-test		Retention-test	
		M	SD	M	SD	M	SD
DoZAP	n=22	2.6	1.1	1.6	2.1	1.8	1.9
ViewZAP	n=28	2.0	1.5	1.8	1.8	2.0	2.4

Maximum is 4 for the pre-test, 6 for the post-test and the retention-test

5.4 Is there a difference in gain of procedural knowledge within the groups?

ZAP1 & 2. A univariate analysis of covariance (ANCOVA) was performed to determine which effect the retention delay had on the *number of trials* taken on ZAP1. The analysis showed no significant difference between conditions (F(1,51)=1.66, p=0.20).

Another univariate analysis of covariance (ANCOVA) was performed to determine which effect the retention delay had on the *efficiency* ZAP1. The analysis showed a significant difference between conditions (F(1,50)=5.013, p=0.03). The effect size, Cohen's d, 0.94 and

0.32 are thus statistically significant different. On the efficiency measure the DoZAP outperformed the ViewZAP.

Another univariate analysis of covariance (ANCOVA) was performed on ZAP2 to determine which effect the retention delay had on ZAP2. The analysis showed no significant difference between conditions (F(1,46)=1.26, p=0.27).

Table 14. Effect Size for conditions on procedural knowledge

	ZAP1 trials	ZAP1 efficiency	ZAP2
DoZAP	0.63	0.94	0.02
ViewZAP	0.37	0.32	0.18

5.5 Is there a difference in gain of declarative knowledge within the groups?

ZAP1. A univariate analysis of covariance (ANCOVA) was performed to determine the learning effect for *overall declarative knowledge*. The analysis showed no significant difference between conditions for pre-test to post-test (F(1,53)=1.10, p=0.30) and post-test to retention-test (F(1,53)=0.15, p=0.70).

Another univariate analysis of covariance (ANCOVA) was performed to determine the learning effect for *declarative factual knowledge*. The analysis showed no significant difference between conditions for pre-test to post-test (F(1,53)=0.89, p=0.35) and post-test to retention-test (F(1,53)=0.31, p=0.58).

Another univariate analysis of covariance (ANCOVA) was performed to determine learning effect for *declarative conceptual knowledge*. The analysis showed no significant difference between conditions for pre-test to post-test (F(1,53)=3.46, p=0.07) and post-test to retention-test (F(1,53)=1.02, p=0.32).

Table 15. Effect Size for conditions on declarative knowledge ZAP1

	Pre- to post-test			Post- to Retention-test		
	Overall declarative knowledge	declarative Factual knowledge	declarative Conceptual knowledge	Overall declarative knowledge	declarative Factual knowledge	declarative Conceptual knowledge
DoZAP	0.76	1.85	0.12	0.46	0.69	0.17
ViewZAP	0.42	2.15	0.76	0.36	0.91	0.11

ZAP2. A univariate analysis of covariance (ANCOVA) was performed to determine the learning effect for *overall declarative knowledge*. The analysis showed no significant difference between conditions for pre-test to post-test (F(1,46)=1.14, p=0.29) and post-test to retention-test (F(1,46)=1.09, p=0.30).

Another univariate analysis of covariance (ANCOVA) was performed to determine learning effect for *declarative factual knowledge*. The analysis showed no significant difference between conditions for pre-test to post-test (F(1,46)=0.00, p=0.95) and post-test to retention-test (F(1,46)=1.80, p=0.19).

Another univariate analysis of covariance (ANCOVA) was performed to determine the learning effect for *declarative conceptual knowledge*. The analysis showed no significant difference between conditions for pre-test to post-test (F(1,46)=2.37, p=0.13) and post-test to retention-test (F(1,46)=0.12, p=0.74).

Table 16. Effect Size for conditions on declarative knowledge ZAP2

	Pre- to post-test			Post- to Retention-test		
	Overall declarative knowledge	declarative Factual knowledge	declarative Conceptual knowledge	Overall declarative knowledge	declarative Factual knowledge	declarative Conceptual knowledge
DoZAP	0.04	0.9	0.6	0.07	0.0	0.1
ViewZAP	0.42	1.0	0.1	0.43	1.25	0.1

6. Discussion

This paper aimed to examine whether any differences exist between an active and a passive learning approach. In specific the question was raised what differentiates learning-by-doing from learning-by-viewing. On the one hand, proponents of learning-by-doing claim that any knowledge is supposedly deeper when gained through an active approach (i.e., doing; Jonassen, 1999; Stull & Mayer, 2007). On the other hand, learning-by-viewing is supported by Cognitive Load Theory (Stull & Mayer, 2007). Through low cognitive load, learning-by-viewing may prevent overload and therefore aid the learning process.

The paper compared the learning effects of DoZAPs and ViewZAPs on various knowledge measures (i.e. procedural knowledge, declarative overall knowledge, declarative factual knowledge and declarative conceptual knowledge). The findings show that the DoZAPs rank at least equal to the ViewZAPs on all measures, and are even superior on one measure, the efficiency on ZAP1. Doing the ZAPs is at least as helpful as viewing them. Both methods of learning are equally helpful.

The DoZAP outperformed the ViewZAP on ZAP1 about confirmation bias as the efficiency measure. Cohen’s d is 0.94 for the DoZAPs and 0.32 for the ViewZAPs. This suggests that cognitive skills may be more efficiently learned by behavioural activity. This finding seems to contradict Renken and Nunez’s (2010) findings. However confirmation bias and reasoning skills are related, but are not the same constructs. The findings reported in the article at hand show that the reasoning regarding confirmation bias may be improved better by doing than by viewing. The reason for the different improvement has to lie within the difference between doing and viewing. The participants had direct feedback from the program, which could have boosted the learning process, as it has been found that the learning process regarding reasoning skills can be improved by feedback (Hayes & Devitt, 2008). Future research should clarify whether it is the feedback that the participants received that made the difference between the groups. Is feedback beneficial for learning critical thinking skills?

The research presented in this paper is nonetheless limited to the comparison of two ZAPs. The robustness of the confirmation bias as subject matter may have been too complex and the results may be different for less complex skills. However it may even have been the complexity that brought out the slight advantage of doing over viewing.

7. References

Aleven, V.A.W.M.M., and Koedinger, K.R., (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science* 26, 147–179.

Anderson, J.R., (1982). Acquisition of cognitive skill. *Psychological Review*, 89(4), 369-406.

Anzai, Y., and Simon, H.A., (1979). The theory of learning by doing. *Psychological Review*, 86(2), 124 – 140.

Ashcraft, M.H., (2006). Cognition. *Pearson Education*, Print 4, ISBN 9780131552715, 23-33.

Bowles, S., & Gintis, H., (2002). The inheritance of inequality. *The Journal of Economic Perspectives*, 16(3), 3-30.

Brand-Gruwel, S., Wopereis, I., & Vermetten, Y., (2005). Information problem solving by experts and novices: analysis of a complex cognitive skill. *Computers in Human Behavior*, 21(3), 487-508.

Brown, A.L., Campione, J.C., & Day, J.D., (1981). Learning to learn: on training students to learn from texts. *Educational Researcher*, 10(2), 14-21.

Carey, S., (2000). Science education as conceptual change. *Journal of Applied Developmental Psychology*, 21(1), 13-19.

Cheng, E.W.L., & Ho, D.C.K., (2001). A review of transfer of training studies in the past decade. *Personnel Review*, 30(1), 102.

Cheng, W., & Warren, M., (2005). Peer assessment of language proficiency. *Language Testing*, 22(1), 93-121.

Chiaburu, D.S., & Marinova, S.V., (2005). What predicts skill transfer? An exploratory study of goal orientation, training self-efficacy and organizational supports. *International Journal of Training and Development*, 9(2), 110-123.

Darabi, A.A., Nelson, D.W., & Palanki, S., (2007). Acquisition of troubleshooting skills in a computer. *Computers in Human Behavior*, 23, 1809–1819.

Eysink, T. H. S., Hulshof, C. D., Loyens, S., de Jong, T., Pieters, J. M., & te Winkel, W. (2003). Psychologie interactief ervaren en ontdekken: het ZAP-project. *Onderzoek van Onderwijs*, 32, 64-66.

Eysink, T.H.S., de Jong, T., Hulshof, C.D. & de Jong, M. (2005). ZAP: Interactief ervaren en ontdekken in het psychologieonderwijs. In M. Mirande, J. van der Veen, & M. van der Wende (Eds.), Van trend naar transformatie. *ICT Innovaties in het Hoger Onderwijs* (201-213). Groningen: Wolters Noordhoff.

Eysink, T.H.S., de Jong, T., Berthold, K., Kolloffel, B., Opfermann, M., Wouters, P., (2009). Learner performance in multimedia learning arrangements: an analysis across instructional approaches. *American Educational Research Journal*, 46 (4), 1107-1149.

Farkas, D.K., & Williams, T.R., (2002). John Carroll's the Nurnberg funnel and minimalist documentation. *IEEE Transactions on professional Communication*, 33(4), 182-187.

Friege, G., & Lind, G., (2006). Types and qualities of knowledge and their relations to problem solving in physics. *International Journal of Science and Mathematics Education*, 4, 437-465.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: a research and practice model. *Simulation and Gaming*, 33, 441-467.

Hayes, K.D., & Devitt, A.A., (2008). Classroom discussions with student-led feedback: a useful activity to enhance development of critical thinking skills. *Journal of Food Science Education*, 7(4), 65-68.

Helfenstein, S., (2005). Transfer: review, reconstruction, and resolution. Doctoral dissertation, University of Jyäskylä, Finland. Jyväskylä, *Studies in Computing*, 59.

Hiebert, J., Morris, A.K., & Glass, B., (2003). Learning to learn to teach: an “experiment” model for teaching and teacher preparation in mathematics. *Journal of Mathematics Teacher Education*, 6(3), 201-222.

Hulshof, C.D., Eysink, T. H. S., Loyens, S. & de Jong, T. (2005). ZAPs: using interactive programs for learning psychology. *Interactive Learning Environments*, 13, 39-53.

Hulshof, C., Eysink, T., & de Jong, T., (2006). The ZAP project: designing interactive computer tools for learning psychology. *Innovations in Education and Teaching International*, 43(4), 337–351.

Jonassen, D.H., (1999). Activity theory as a framework for designing constructivist learning environments. *Educational Technology Research and Development*, 47(1), 61-79.

De Jong, T., & Fergusson-Hessler, M.G.M., (1996). Types and qualities of knowledge. *Educational Psychologist*, 31(2), 105-113.

Kirschner, P.A., Sweller, J., & Clark, R.E., (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86.

Kolb, D.A., (1984). *Experiential learning: experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.

Lee, F.J., & Anderson, J.R., (2001). Does learning a complex task have to be complex? A study in learning decomposition. *Cognitive Psychology*, 42(3), 267-316.

Lesgold, AM., (2001). The nature and methods of learning by doing. *The American Psychologist*, 56(11), 964-973.

Lewicki, P., Czyzewska, M., & Hoffman, H., (1987). Unconscious acquisition of complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(4), 523-530.

Mayer, R.E., & Moreno, R., (2002). Aids to computer-based multimedia learning. Learning and Instruction, 12, 107–119.

Moreno, R., & Mayer, R.E., (2002). Learning science in virtual reality multimedia environments: role of methods and media. Journal of Educational Psychology, 94(3), 598–610.

Ponte, J. P. (2001). Investigating in mathematics and in learning to teach mathematics. In F. L. Lin & T. J. Cooney (Orgs.), *Making sense of mathematics teacher education* (53-72). Dordrecht: Kluwer.

Renken, M.D. and Nunez, N., (2010). Evidence for Improved Conclusion Accuracy after Reading about Rather than Conducting a Belief-inconsistent Simple Physics *Experiment. Appl. Cognit. Psychol.*, 24, 792–811.

Savery, J. R., & Duffy, T. M. (2001). Problem based learning: An instructional model and its constructivist framework. *CRLT Technical Report No. 16-01*, Indiana University: Center for Research on Learning and Technology.

Stull, A.T., & Mayer, R.E., (2007). Learning by doing versus learning by viewing: three experimental comparisons of learner-generated versus author-provided graphic organizers. *Journal of Educational Psychology*, 99(4), 808–820.

Vygotsky, L., (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.

Appendix A

Sequential Analysis and Calculation of Efficiency on Concept Formation ZAP

In Figure 6 the first figure is the one that the program provides. The participant's first trial (the second figure, one green cross with 2 frames) provided novel information about whether colour plays a role in the rule. The application returned that the tile did not correspond to the rule and thus it could be inferred that the colour had to be black. This is novel information. All of the participants' answers were analyzed in a detailed sequential analysis (see Table 17).

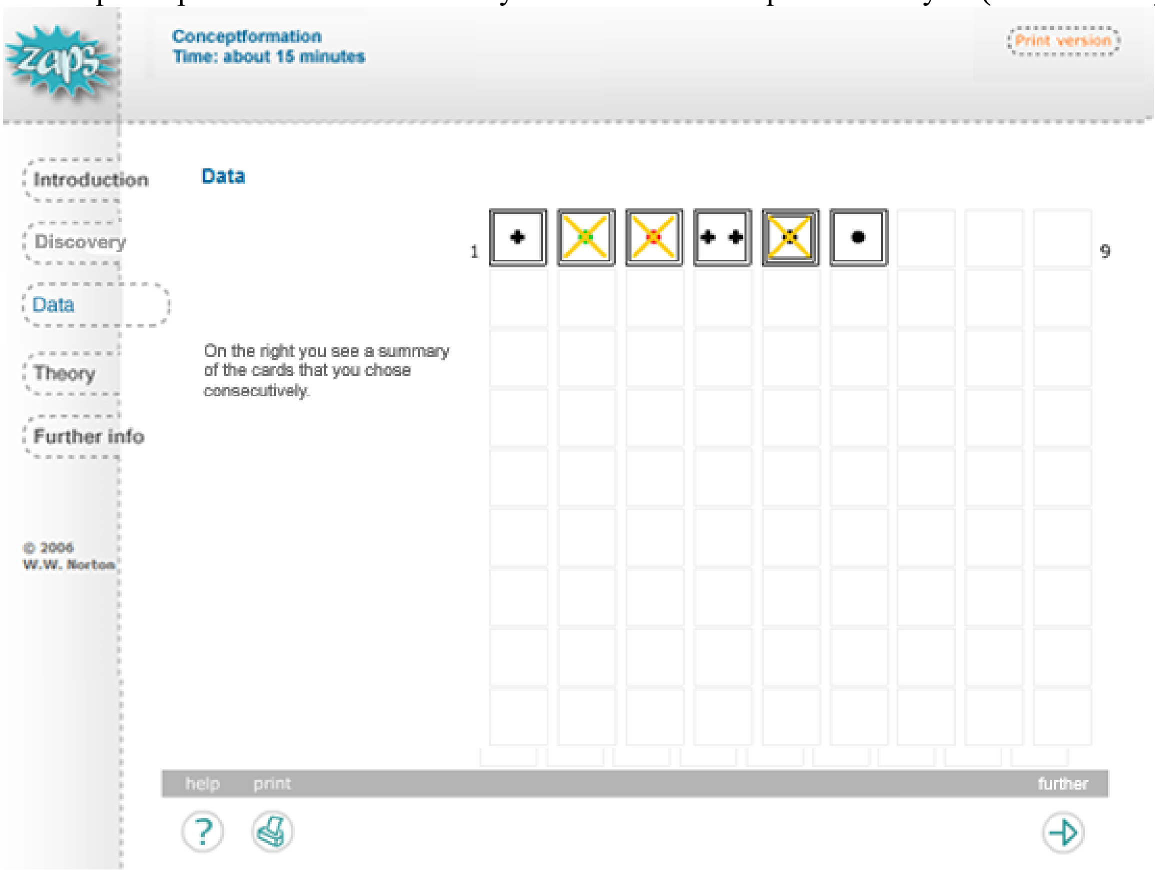


Figure 6. Example sequence concept formation ZAP

Table 17. Sequential Analysis of an example sequence

Number of Trial	Trial	Novel Information	Information gathered
1	One green cross with two frames	Yes	Colour has to be black
2	One red cross with two frames	No	-
3	Two black crosses with two frames	Yes	Number of objects has to be one
4	One black cross with three frames	Yes	Number of frames has to be two
5	One black circle with two frames	Yes	Which kind of object it is irrelevant

Summarized, the rule goes that tiles have to be black with 2 frames.

The efficiency measure was calculated with the following formula:

Number of trials providing novel information/Total amount of trials x 100

Number of trials providing novel information = 4
Total amount of trials = 5
4/5 x 100 = 80% efficiency