Learning and retention of procedural skills in echocardiography

(Bachelorproject)

Evert van der Weit, s2021552, E.W.van.der.Weit@ai.rug.nl,
F. Cnossen* en R. A.Tio†

July 3, 2014

Abstract

In many medical domains simulators are used to learn medical skills. In echocardiography this is usually done by a one-day training session. The aim of this study was to find the best method for optimizing the long-term performance of a procedural skill so that the effect of the training session is maximized. In an attempt to maximize long-term retention the testing effect were used. Two research questions were answered: "Can the testing effect be generalised to procedural skills?" and "Can long-term retention of procedural skills be enhanced when improving declarative knowledge using the testing effect?". Participants in the study received a training session and after 4 weeks they either performed a declarative test, a procedural test or both. 15 weeks after that, participants performed a final declarative and procedural test. Final performance on the procedural test was measured by evaluating screenshots of the echo's participants made. Results show no significant result of the testing effect on procedural skills. Also no effect of declarative knowledge on procedural skill performance is found.

1 Introduction

In the medical world there are many experts on separate medical domains, such as cardiology or anaesthesia. It's not only declarative knowledge that is important in these domains; there are many procedural skills that medical experts have to master. Learning and training of these skills is not always possible on real human patients, because of high risk of the procedure or because the equipment is expensive or always in use. For that reason, learning and training a medicall skill is often done on a simulator. Anderson and Warren (2011) concluded in their research of simulation on neonatology: "Simulation can be used as an educational strategy to enhance the clinical skills necessary for optimal performance in complex, unpredictable, time-critical, high-stakes environments like neonatology."

Training on a simulator has been shown to have a positive effect on complex procedural skills, even after a retention period of one year. For example, Boet, Borges, Naik, Siu, Riem, Chandra, Bould, and Joo (2011) conducted a study in which they trained anaesthetists the cricothyroidotomy skill (emergency airway puncture) on a simulator through a one-day training. Results showed that the skill performance of the subjects was better immediately after the training but also after an interval of 6 or 12 months, compared to the pre-training performance.

When making, for example, an echocardiogram there are two skills that need to be mastered: knowing how to manipulate the probe to get the right image, know how this image is supposed to look and how to interpret this image. There are multiple distinctions in the types of knowledge that humans can master, but one in particular is relevant: knowing what to do, which reflects declarative knowledge, and knowing how to do it, which reflects procedural knowledge. These two types of knowledge have different learning mechanisms (Johnson (2003)). Anderson, Fincham, and Douglass (1997) argue that a complex skill uses both of these learning mechanisms. Initially the student refers to examples while learning a skill, either from instructions or declarative memory. After getting enough practice, rules form to solve this particular problem so, it is no longer necessary to retrieve examples from declarative memory. After enough practice this knowledge transitions from declarative to procedural form.

*University of Groningen, Department of Artificial Intelligence

†UMCG Groningen, Thorax Center

This study will try to use this distinction in skill acquisition by trying to improve either of the two mechanisms. To achieve this, currently available methods will be used in order to optimize the performance of making a heart echocardiogram. Anderson et al. (1997) also noted that there was an advantage when initial study examples were repeated. This suggests that improving the declarative representation of the material influences the final performance. The aim of this study is to see how performance on procedural skills is best retained by using an effect which is commonly known as the 'testing effect'.

Testing effect

We learn something with the goal to remember it after a short period of time, but also to remember it after a few months of even years. One way to improve this long term retention is to use the testing effect. Testing the material gives better long term retention than studying the same material again (Roediger and Karpicke (2006)). This effect has been found for the first time in the early 1900s (Abott (1909); Gates (1917)). Butler and Roediger (2007), for example, performed a study with a recall test, at which a group of participants viewed a series of three lectures and after each lecture either studied a summary, made a multiple choice test or did a short answer test. The group who took a short answer test after the lectures performed significantly better on the final short answer test conducted a month after viewing the lectures. The group who took the short answer test after the lectures got 47% correct on the final test versus 36% correct on the groups who only studied a summary or took a multiple choice test.

To see whether this testing effect also applies to non-verbal learning, Kang (2010) conducted a study which used abstract visuospatial information to see whether the testing effect applied to those stimuli as well. Kang (2010) used a set of Chinese characters as stimuli. Subjects who did not have any prior experience with Chinese characters studied pairings of Chinese symbols with the English words it represents. Afterwards they either restudied the material twice or they had to covertly retrieve the Chinese characters twice. A final test was given after 10 minutes or 24 hours and on both instances the subjects who retrieved the Chinese characters twice were more accurate at drawing the Chinese characters, as opposed to subjects who restudied the material twice.

The testing effect seems to be an effect which affects different types of declarative learning. The question then is, how far does this effect generalize? Can the testing effect also be used to improve procedural learning, directly or by improving declarative knowledge that is needed to perform the skill? Kromann, Jensen, and Ringsted (2009) performed a study which looked at the effects of testing on skills learning. They compared two groups of participants, one group got a 4 hour resuscitation training on a simulator and the second group got a 3.5 hour training and a 30 minutes test immediately after the training. Performance was evaluated after 2 weeks in a simulated scenario using a checklist. Scores for the second group were higher (82.5%) compared to the group who only got 4 hours of training (73.3%). This suggests that the testing effect can also be applied to procedural skills.

Current study

The training in which intensivists learn to perform and interpret a heart echocardiogram currently consists of a one-day training. In this training they attend a theory session, followed by a practical session in which they practice the skill under supervision. Finally a short multiple-choice test is conducted on the anatomy and function of the heart and on interpreting echoes. For the practical session in this training, a simulator is used. Participants in this study will only be asked to make the heart echocardiograms, not to interpret them. This is done because interpreting an heart echocardiogram is mostly a declarative skill, you need to know what to look for and what the meaning is of what you see, which is beyond the purpose of the current study.

The aim of this study is to find the best method for optimizing the long-term performance of a procedural skill in medicine, and this is done by examining the skill of making a heart echocardiogram. An echocardiogram is a method that uses sound waves to create a moving image of the heart. These images can be used by doctors to assess heart function, check for diseases, see locations of blood clots, etc. There are two types of heart echocardiograms: Transthoracic (TTE) and transesophageal (TEE). The first one is a non-invasive method through the thorax and the other one is an invasive method through the oesophagus. For this study the TTE method will be used. In this method a transducer or probe is positioned on the chest and moved and tilted in such a way that the proper image of the heart is visible. Lungs, ribs and tis-

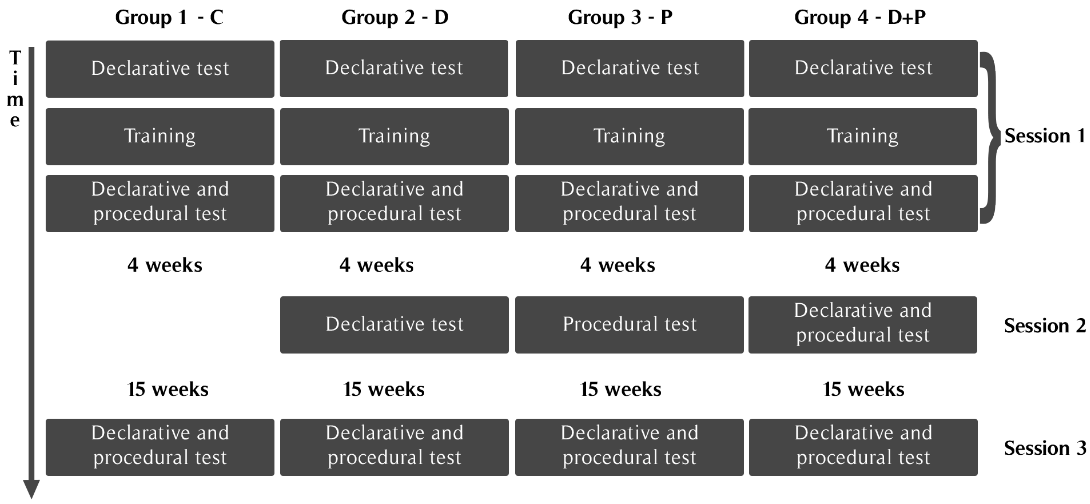


Figure 1: Overview of the test design.

sue obstruct the view and therefore make it hard to find the right position from which the image has to be made.

Research question

Not much has yet been published about the testing effect applied on procedural skills. Therefore more research on this subject is necessary, because it can improve the effectiveness of simulation training and thereby reduce costs. In this study the findings of Kromann et al. (2009) will be extended by adding a second testing moment in the design. The effect of testing will also be measured over a period of 21 weeks to look more closely at the long-term retention of this effect. The precise design of the study will be explained in detail in the methods section. Research questions that will be answered are:

- Can the testing effect be generalised to procedural skills?
- Can long-term retention of procedural skills be enhanced when improving declarative knowledge using the testing effect?

2 Methods

For this study 2nd and 3rd year medical students were recruited on a voluntary basis. A small reward was given after completion of the whole experiment in the form of a book token. A total of 31 students applied for this experiment from which 28 completed all the trials.

Design

All participants first received a training session of approximately 2.5 hours. Training times varied with the amount of people attending the training at once. With 2 participants the training took about 2 hours and with 4 participants it lasted almost 3 hours. This is due to the individual practice time each participant got during the practical training session. The training will be explained in more detail later.

After the training the participants were randomly assigned to 1 of 4 groups. Group 1 took a declarative test 4 weeks after the training, group 2 took a procedural test 4 weeks after the training and group 3 took both a declarative and a procedural test after 4 weeks. The final group did not take any test after 4 weeks.

The final measurement was conducted 15 weeks after the second measurement, and all 4 groups took a declarative and a procedural test. For an overview of the design of this experiment see Figure 1.

Training

Training was done with multiple participants at a time, but with a minimum of 2 and a maximum of 4. The training consisted of 5 parts. First all participants took a declarative test, this was done to compare knowledge of the material previous to starting the experiment. After this test they received a video instruction, which lasted 21 minutes. This consisted of: how to use the equipment, how to position a patient, a instruction on how to position the probe for certain images and an overview of the different images with an explanation on what was seen on these images. Afterwards participants got a live instruction. The instructor was the researcher in this study, who is not a cardiologist. In the instruction they were shown how to make the different views on the simulator and which elements were visible on those views. Also, tips and tricks were given on how to manipulate the probe.

After these instructions participants were allowed to practice the differents views (6 in total) with feedback and further instructions from the instructor. Each participant practiced approximately 10-15 minutes in which practiced every image twice. Participants were allowed to watch each other practice. After each of the participants had had a chance to practice on the simulator they all had to make another declarative test. Finally they took a procedural test in the same order as in which they practiced on the simulator, so the first participant to practice also took the procedural test first.

Declarative test

The declarative test consisted of 20 multiple-choice questions. In 14 of these questions participants watched a short video of a echocardiogram view and they had to name the area under [x], as was shown on the video. The other 6 questions were about how the probe should be manipulated to change from one view to another. The same test was used every time a participant had to make a declarative test but the order of the questions was randomized every time. Participants were not allowed to change an answer to a previous question and only saw one question at a time. Feedback was given after the entire test was completed. Scores on the declarative test range from 0 to 100. For a list of the questions used, see the appendix.

Procedural test

In the procedural test, the participant had to make six different echocardiograms which they had learned in the training session. They always made these six echoes in the same order, the same order in which they also learned to make them. The participant was asked to notify the instructor when they were satisfied with their obtained image. The instructor then stopped the time and made a screenshot of the image. After the six images had been made, they received short feedback on their performance. Participants were not allowed to watch each other during the procedural test. The screenshots were sent to an expert for evaluation. The screenshots were rated on a scale from 0 to 4.

Measures

The scores on all declarative tests were recorded and the scores obtained on the six echos were summed up and divided to get the means for every procedural test. These two variables were used to answer the research questions; other variables measured such as time it took to complete an echo were only used to check for abnormalities in the data. To analyze the data, t-tests were used to determine if there was any difference between the groups. Correlation was used to check if declarative knowledge had an influence on procedural test performance.

Results

For easier reference the groups are tagged. ”C” is group 1 who only did the training and the final measurement, ”D” is group 2 who did a declarative test in between, ”P” is group 3 who did a procedural test in between and finally ”D+P” is group 4 who did both a declarative and a procedural test in between. (see also Figure 1)

Inspection of the data

The boxplot in Figure 2 shows the variance in the declarative test conducted before the training. When comparing group P (M = 44.50, SD = 10.97) and group D+P (M = 55.71, SD = 9.35) no significant difference in means was found,

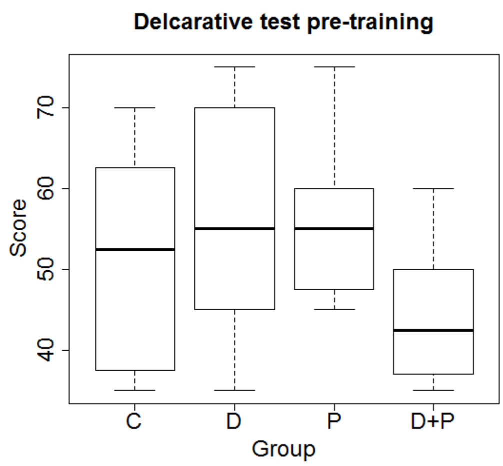
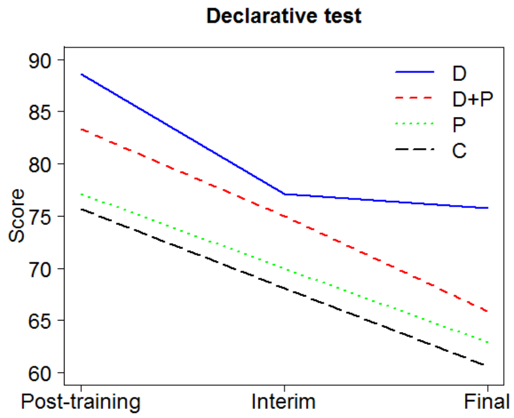


Figure 2: Declarative test scores before the training session.



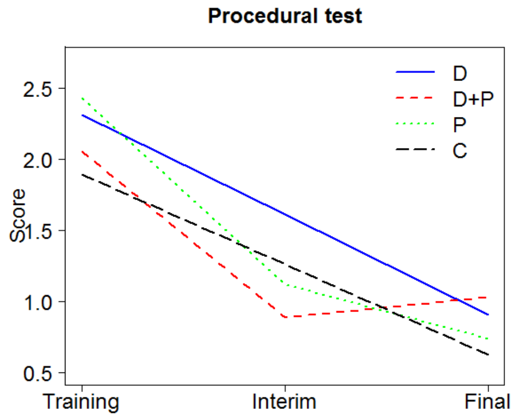


Figure 3: a) Declarative test scores without the pre-training data. b) Procedural test scores

Group	N	Before training	After training	Final test
C	8	51.25	75.62	60.62
D	7	56.43	88.57	75.71
P	7	55.71	77.14	62.86
D+P	6	44.50	83.33	65.83

Table 1: Mean declarative test scores per group

Group	N	Training	Final test	Difference
C	8	1.89	0.62	1.27
D	7	2.31	0.91	1.40
P	7	2.43	0.74	1.69
D+P	6	2.06	1.03	1.03

Table 2: Mean procedural test scores per group

t(11) = 1.96, p = .075. Procedural test performance during training is shown in Table 2. When comparing C (M = 1.89, SD = 0.67) and P (M = 2.43, SD = 0.21) no significant difference in means was found, t(8.59) = -2.15, p = .061. Post-training declarative test performance is shown in Table 1. When comparing C (M = 75.62, SD =) and D (M = 88.57, SD =) no significant difference in means was found, t(10.03) = -2,16, p = .056. Figure 3 shows an overview of declarative and procedural test scores over the course of the experiment. For the precise numbers corresponding to these figures see Tables 1 and 2.

Analysis

For analysis the procedural test performance was examined as shown in Figure 4. To account for different scores after the training, the difference between the training score and the final test score was used for every group. No difference in means was found between C and either of the other groups. Comparing means of C (M = 1.27, SD = 0.78) and P (M = 1.69, SD = 0.52) no significant effect is found, t(12.24) = -.1.23, p = .24. Comparing means of C (M = 1.27, SD = 0.78) and D+P (M = 1.03, SD = 0.95) no significant effect is found, t(9.63) = 0.51, p = .62. When looking at the groups with the biggest difference in means, the groups P (M = 1.69, SD

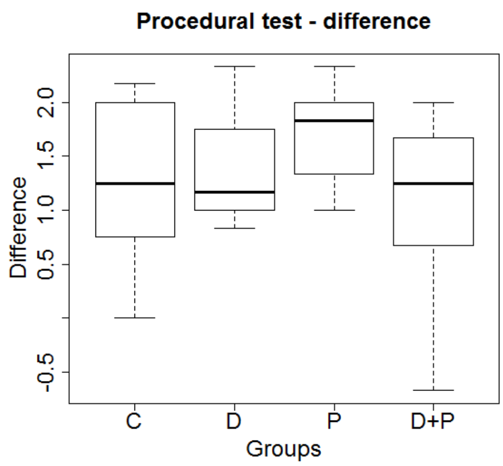


Figure 4: The difference in procedural test performance between the training and the final test.

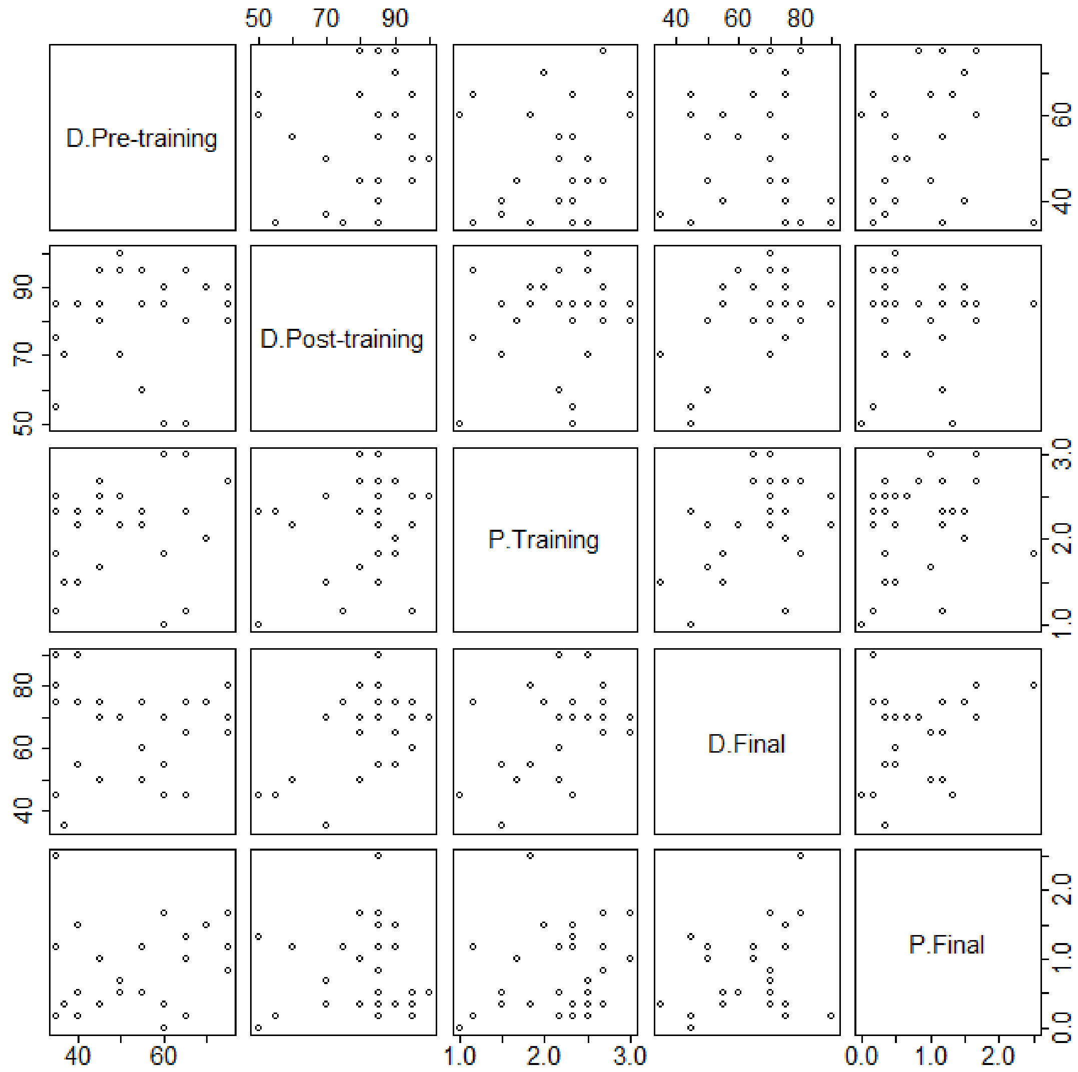


Figure 5: Matrix of scatterplots with declarative and procedural test data from the training and the final measurement. The prefix D in this figure stands for declarative and the prefix P for procedural. Each graph shows a scatterplot of the two variables that are in the same column and the same row as the graph. Each graph is thus shown twice but they are diagonally flipped relative to each other.

= 0.52) and D+P (M= 1.03, SD = 0.95), no significant difference in means was found, t(7.52) = 1.53, p = .168. When combining the groups who did receive an interim test and the groups who did not, still no significant effect is found; t(26) = -0.186, p = 0.85 with the groups D and C (M = 1.33, SD = 0.67) and P and D+P (M = 1.38, SD = 0.79). Finally a matrix of scatterplots was created to investigate possible correlations between declarative test performance and procedural test performance, as seen in Figure 5. The correlation between the final declarative test score (D.Final) and the final procedural test score (P.Final) was not significant, r = .202, p = .30. All the other correlations between declarative tests and procedural tests are of similar strenght and none of those were significant.

Discussion

When comparing the mean declarative test scores there seems to be a difference in means between group the group who took a procedural interim test (group P) and the group who took both a procedural and declarative interim test (group D+P), but this is not significant (p

= .075). Groups were assigned randomly and participants had no experience with echocardiograms. A large portion of the declarative test was based on the anatomy of the heart, which should be something that participants already learned previous to the experiment. It must be a conincidence that there seem to be differences in performance on the declarative pre-test. Something similar seems to be the case with the procedural test performance, a large difference in means between the group who took no interim tests (group C) and group P is seen but this, also, is not significant (p = .061).

Another important thing to mention is that the test design was flawed to correctly answer the first research question. Since there was no group who received an interim training, the effect of the group who received an interim procedural test cannot be interpreted as being more or less benificial than a similar group receiving a training instead of being tested. With this in mind we will look at the research questions. Research questions posed in the introduction were:

- Can the testing effect be generalised to procedural skills?
- Can long-term retention of procedural skills be enhanced when improving declarative knowledge using the testing effect?

First to check wether the testing effect can be generalised to procedural skills, we should look at groups P and D+P who both did an interim procedural test and see if their performance on the final test is better than the other 2 groups who did not have this procedural interim test. Since neither group P or D+P has a significantly higher final procedural test score than the groups who did not take a procedural interim test, we can conclude that there is no beneficial effect from the interim test in the procedural test. If there is no beneficial effect from an interim test there is certainly not a testing effect in which this interim test would be more beneficial then another training.

If long-term retention of procedural skills are enhanced when declarative knowledge is improved using the testing effect we should expect to see participants who scored better at the declarative test also performing better at the procedural test. Figure 5 shows no clear correlation between the score on the final declarative test and on the final procedural test (p = .30). Thus it cannot be concluded that scoring better at the declarative test means also scoring better on the procedural test.

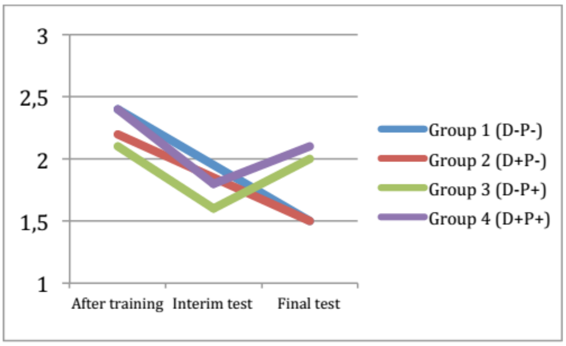


Figure 6: Mean procedural test scores, taken from Mooibroek (2014).

Both research questions have to be answered negatively based on this research. No effect of an interim test is found, and no effect of declarative knowledge on procedural skills.

In general the training seems to be insufficient for good performance on the procedural test post-training. Scores averaging around 2 - 2.5 are seen after training, while the maximum score to be obtained is 4. This indicates that the 3 hour training session is not enough. Knowledge which has not been acquired cannot be mastered can't be strengthened by a test, which would be necessery for the testing effect to work. It has been shown that the degree of overlearning is an important determinant of both skill and knowledge retention (Winfred, Winston, Stanush, and McNelly (1998)). In this study there seems to be no overlearning; the data show that they did not even learn to perform the skill sufficiently. Procedural skill training on a simulator should be given in such a way that after the training the participants master the skill, there should not be a strict time limit to a simulator training.

Interesting is the comparison between the current study and the master's thesis done by Mooibroek (2014). Our study was almost exactly the same as the study performed by Mooibroek (2014), the only difference being the time between the measurements. In this study the time between the training and the interim session was 3 weeks there compared to 4 weeks and the time between the interim session and the final session was 8 weeks compard to 15 weeks in this study. When we compare Figure 6, which shows the data from Mooibroek (2014), and Figure 3b we see that procedural test scores after the training are pretty much similar. Mooibroek (2014) showed a significant effect of the interim test on the procedural test ($p < 0.01$). So, interestingly, the effect that still exists after 8 weeks is no longer present after 15 weeks, as the current study shows. It seems to be that only one interim test is not enough to improve retention after 15 weeks, but it is benificial after a period of 8 weeks. Both studies show a similar decline in procedural test performance over time; 1.33 points in 19 weeks versus 0.8 points in 11 weeks for the groups who did not receive a procedural interim test. This is a decline of 0.07 points per week in both studies. Both studies thus only differ in the interval periods between measurements.

Future research

Testing can enhance procedural skills better in comparison with an equal amount of time spent on training when this test immediately proceeds the training Kromann et al. (2009). It should therefore be likely that a test after a certain interval would improve long-term retention even more. Further research should be conducted to explore the effect of testing on long-term retention. Preferably a procedural skill should be chosen that requires little or no declarative knowledge to perform. The procedural skill that is te be learned should be at maximal performance after the training to maximize the effect of testing.

References

E. E. Abott. On the analysis of the factor of recall in the learning process. *The Psychological Review: Monograph Supplements*, 11(1): 159–177, 11 1909. ID: 2011-13725-005.

J. M. Anderson and J. B. Warren. Using simulation to enhance the acquisition and retention of clinical skills in neonatology. *Seminars in perinatology*, 35(2):59–67, 4 2011.

J. R. Anderson, J. M. Fincham, and S. Douglass. The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(4):932–945, 07 1997.

S. Boet, B. C. R. Borges, V. N. Naik, L. W. Siu, N. Riem, D. Chandra, M. D. Bould, and H. S. Joo. Complex procedural skills are retained for a minimum of 1 yr after a single high-fidelity simulation training session. *British Journal of Anaesthesia*, 107(4):533–539, 2011.

A. C. Butler and H. L. Roediger. Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19(4-5):514–527, 2007.

A. I. Gates. Recitation asafactor in memorizing. *Archives of Psychology*, 6(40), 1917.

A. Johnson. *Procedural Memory and Skill Acquisition*, pages 495–517. Handbook of Psychology. John Wiley & Sons, Inc., 2003. ISBN 9780471264385.

S. K. Kang. Enhancing visuospatial learning: The benefit of retrieval practice. *Memory & cognition*, 38(8):1009–1017, 12/01 2010.

C. B. Kromann, M. L. Jensen, and C. Ringsted. The effect of testing on skills learning. *Medical education*, 43(1):21–27, 2009.

S. Mooibroek. *The testing effect applied to procedural skills.* Master's Thesis. Human Machine Communication, University of Groningen, 2014.

H. L. Roediger and J. D. Karpicke. The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3):181–210, September 01 2006.

A. Winfred, B. Winston, P. L. Stanush, and T. L. McNelly. Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance*, 11(1):57–101, 1998.

Appendix - Declarative test

What structure/area is under the X?

Fourteen videos with moving 2D heart images have been shown, stopping at a certain point with a cross at the structure/area that had to be named.

When does the pulmonary valve close?

A. When pulmonary artery pressure rises above right ventricular pressure.
B. When pulmonary artery pressure dives under right ventricular pressure.
C. When pulmonary artery pressure rises above left atrial pressure.
D. When pulmonary artery pressure dives under left atrial pressure.

When does the aortic valve open?

A. When left ventricular pressure rises above left atrial pressure.
B. When left ventricular pressure rises above aortic pressure.
C. When right ventricular pressure rises above aortic pressure.
D. When right ventricular pressure rises above left atrial pressure.

What is the main purpose of the heart valves?

A. To prevent backflow.
B. To maintain forward flow of blood throughout the cardiac chambers.
C. To guarantee coronary blood flow.
D. To improve myocardial contractibility.

How should you manipulate the probe in order to change the 2-chamber view to a 4-chamber view?

A. Rotate the probe 135 degrees clockwise
B. Rotate the probe 45 degrees clockwise
C. Increase the angle
D. Decrease the angle

How should you manipulate the probe in order to change the short axis view to the long axis view?

A. Increase the angle of the probe
B. The long-axis view is taken from another position
C. Rotate the probe 90 degrees
D. Rotate the probe 90 degrees and adjust the angle slightly

How should you manipulate the probe in order to change the 4-chamber view to the 5-chamber view?

A. Rotate the probe 45 degrees
B. Rotate the probe 90 degrees
C. Increase the angle by moving the tail of the transducer upwards/medially
D. Decrease the angle by moving the tail of the transducer downwards/laterally