UNIVERSITY OF TWENTE

Building a reliable simulator based assessment suite of MIS-performance out of basic Lapsim tasks

Luuk Lenders – s156889

Bachelor thesis Psychology
Enschede, June 2017

First Supervisor: Dr. Martin Schmettow
Second Supervisor: Dr. Marleen Groenier
University of Twente
Cognitive Psychology and Engineering

Contact: l.b.l.lenders@student.utwente.nl

UNIVERSITEIT TWENTE.

Abstract

Objectives: The learning of Minimal Invasive Surgery (MIS) is paired with difficulties. Due to its difference with normal surgery, MIS is seen as a better way to operate, although slightly more dangerous. These risks seem to be more than only ergonomic or cognitive factors. Previous research has mainly focused on finding cognitive predictors of future laparoscopic skill. This study proposes a new model called the Resemblance Spectrum that could be used to explain why the previous research has been looking at the wrong place. Through the use of individual learning curves this research tries to find out if we can find more predictable results using simulators.

Methods: 40 participants enrolled in this study. All participants performed a set number of repetitions of two innate ability tasks, called PicSOr (35 repetitions) and Map planning (2*10 problems in 3 minutes), and two dexterity tasks called Mirror Drawing and Origami (both 20 repetitions each). Next to this they at least did Grasping and Cutting on the LapSim simulator. If interested participants could partake in a second session in which they did two more simulator tasks, namely Clip Applying and Lifting & Grasping. All simulator tasks were repeated 12 times each. For every task, except the innate ability tasks, the time to complete the task (ToT) was measured and used to build individual learning curves. Pearson pairwise correlations between the learning curve parameters were calculated to explore if it is possible to build a reliable assessment suite of only simulator tasks.

Results: A Pearson Pairwise Correlation was used to assess the correlation between the maximum performance parameter per task. The correlation were all positive, but varied between r = .16 up to r = *.68* . The credibility limits were noted to be fairly large.

Discussion: Although the correlations were originally not expected to be so spread out, it seems to be possible to build a reliable simulator only assessment suite for laparoscopic skill. as the correlations are expected to be spread out due to different underlying factors playing a role in the simulator tasks. For future research it is suggested to look at more simulator tasks, but also task to full procedure validation is still needed to prove the last links in the Resemblance Spectrum.

Table of Contents

1 General Introduction

Minimal invasive surgery (MIS) is a fairly recent development in surgery that has radically changed the way how surgery is conducted. One of the earliest forms of MIS is called laparoscopy. In laparoscopy trocars, in principle hollow tubes, are placed into the abdominal wall through which long, endoscopic instruments are inserted towards the operative site (for a more detailed description, see Cao, MacKenzie, & Payandeh 1996). The trocars prevent that the instruments tug at the skin of the patient; instead they give the instruments more free movement. The procedure of laparoscopy is different from open surgery, in which surgeons have a more direct way of interacting with the to be operated tissue. MIS has many advantages like shorter hospital stays, less need of analgesia, faster recovery time and a cosmetic advantage compared to open surgery (Tinelli et al., 2012).

Despite these clear advantages of MIS, there are major variations in learning rates of surgeons in training (Buckley et al., 2014; Grantcharov & Funch-Jensen, 2009). It has also been noted in the study conducted by Grantcharov & Funch-Jensen (2009) that not everyone will reach the proficiency within 10 repetitions, and some people do not seem to show any improvement at all. A more recent study from Louridas et al. (2017) shows the same results, even with the amount of repetitions increased to 40. To try and accommodate for these individual variations, research has been done into what causes these differences and how we can predict performance in the operating room. Much of this research has focussed on innate ability as a predicting measure for performance. An example of one of such innate abilities required for laparoscopy is spatial ability, needed for building a mental representation of anatomy.

However, so far no single aptitude test has been shown to be a reliable predictor of performance (Louridas, Szasz, de Montbrun, Harris, & Grantcharov, 2016), but simulator performance as indicator of psychomotor ability seems to have the highest predictive value for future performance (Kramp et al., 2016). Another important aspect of learning is the rate of skill acquisition. Both maximum performance and rate are part of learning curves. In this study we want to try and find if the learning curves on one task has any predictive power for the learning curve on similar but slightly different tasks.

1.1 Constraints of laparoscopy

Learning the skills necessary for MIS comes with various constraints regarding the ergonomics of the instruments, human factors (Berguer, 1999) and how surgeons are trained. Previously identified problems with workplace ergonomics for MIS include simplified tools, bad placement of screens and problems with keeping good posture while operating (Berguer, 1999).

Along with ergonomic issues, some cognitive factors have also been distinguished which further limit the speed of skill acquisition surgeons in training might have. First of all, there is no direct contact between the tissue and the surgeons, which results in less haptic feedback (Gallagher & Smith, 2003). Secondly, surgeons need to mentally transform a 2D image to a 3D representation of the operative site, and also interpret what is going on at that moment to correctly identify possible complications. An additional perceptual problem is the use of a single, small camera whose 'feed' is shown magnified and possibly rotated on a monitor (Gallagher & Smith, 2003). Furthermore, the instruments can only pivot around the entry point in the body wall, reducing the degrees of freedom that a surgeon has. The pivot point, also called the fulcrum, causes the instruments to move in the opposite directions of the surgeon's movements in both remaining axes that can be effectively used in MIS. This is a well-known effect which is also referred to as the fulcrum effect. It has been shown that the fulcrum effect has major negative influence on acquisition of endoscopic skills (Gallagher, McClure, McGuigan, Ritchie, & Sheehy, 1998).

All of the above problems make laparoscopy not only a difficult task to perform, but also a difficult task to learn. This difficulty in acquiring the necessary skills has caused a lot of research to examine finding reliable estimators for future performance to reduce errors in real, live surgeries. One way of approaching this problem has been through aptitude testing.

In aptitude testing, the main focus lays on testing the natural ability of a person to perform a certain task. The term aptitude is often used interchangeably with innate ability. Aptitude testing has been a major focus in selection processes in all sorts of fields, going as far back as the 1940's for aptitude testing in the Royal Air Force. In surgery however, most of the selection has relied on previous academic performance and interviews (Wanzel, Ward, & Reznick, 2002). The calls for reliable and valid testing measures go even further back (e.g. See Darzi, Smith, & Taffinder, 1999; Gough & Bell, 1989). Advantages for institutions would include reduced expenditure through more reliable (pre-) selection methods before any major time or monetary commitments are made, while at the same time gaining more capable surgeons.

1.2 Previous research on acquiring MIS skills

Because MIS is becoming increasingly versatile in its application, a lot of research has gone into training and selection within this field of surgery. A lot of this research is based around the Stages of Learning model of Fitts & Posner (1967) (see “Skill Acquisition” in Figure 1) combined with the model for skill learning by Ackerman (1988). The model of Fitts & Posner defines three stages of learning and it describes how attentional and cognitive demand reduces as somebody learns how to execute a task. A different model, as proposed by Ackerman (1988), includes the model from Fitts & Posner and applied this to a radex model as proposed by Marshalek, Lohman, & Snow (1983). According to this unified model from Ackerman, depicted in Figure 1, different cognitive abilities are used in varying degrees during the learning of a task. At the same time task performance slowly becomes more automated which could increase the pace of task completion. Based on this model, one would expect to find the relation between the abilities and phases of learning while learning MIS, next to other possible important abilities on which the task may rely.

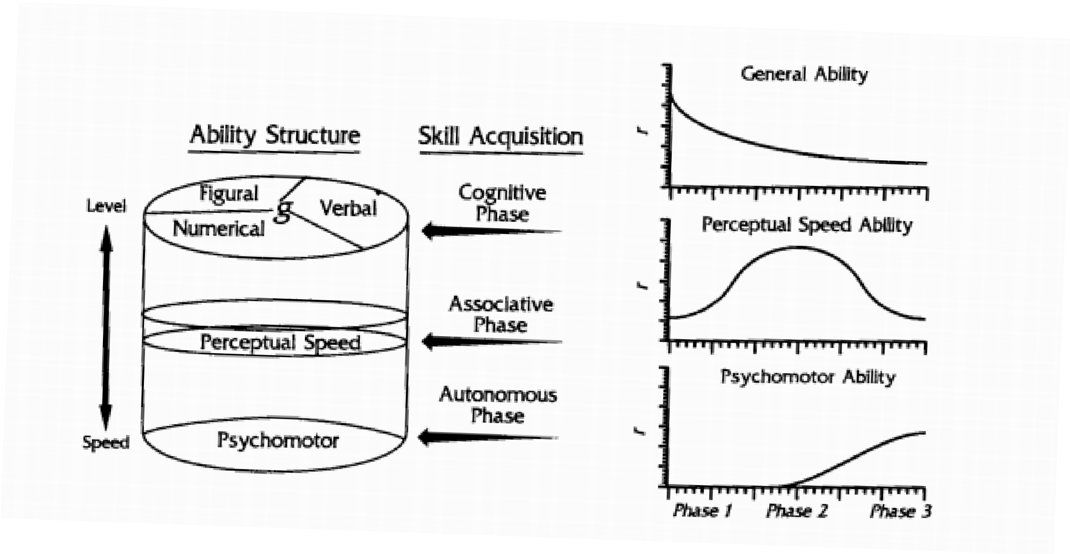


Figure 1 - Ackerman's radex model. As in the original model of Marshalek et al. (1983), complexity is represented by the radius of the cylinder with increasing complexity towards the centre. Reproduced from (Ackerman & Cianciolo, 2000)

For a learning experiment such as the one conducted in this study, this model would predict that we would see a strong relation between the specific abilities and their respective parts on a learning curve. This predication seems to generally hold true as research from Hegarty, Keehner, Cohen, Montello, & Lippa (2007) has proven. However, Hegarty et al. do note that spatial ability correlates highly with performance throughout all phases of skill acquisition. This seems to be mainly due to the content of the task used, which in the case of MIS seems to mainly be spatial ability. Ackerman & Cianciolo (2002) have shown that the content of the task is more important than the consistency of a task. A more in-depth discussion about the importance of spatial cognition in MIS can be found in Keehner (2011), or Hegarty, Keehner, Cohen, Montello, & Lippa (2007) for an even broader discussion on spatial cognition in medicine overall..

The role of psychomotor ability in MIS also been researched, with a study of Stefanidis et al. (2006) finding that a faster acquisition of laparoscopic skill correlates with higher psychomotor ability, contrary to what would be expected based on the model of Ackerman. On the contrary, it has also been found that psychomotor performance positively correlates with performance in the operating room (Kundhal & Grantcharov, 2009). The latter findings are again in line with the proposed model by Ackerman.

However, a more recent systematic review regarding the use of aptitude testing in laparoscopic training shows that separate measures do not give reliable enough results (Louridas et al., 2016). The same review poses the idea that a possible combination of measures of innate ability, their interaction and their relationship with technical skill might be a more reliable predictor. It does also present the possibility of using aptitude testing for more personally tailored training activities. The results of Louridas et al. (2016) are compatible with a recent meta-review from Kramp et al. (2016) which has found that the previously mentioned innate abilities both correlate with performance, but these correlations are not particularly strong. Kramp et al. also noted that simulator based assessment correlated the best (r = .64) with future laparoscopic performance. Due to these weak correlations previously found on using innate ability tasks as a predictor, we are going to try and take a different approach using learning curves.

1.3 Learning curves

In the present study learning curves are going to be used for the purpose of evaluating the possibility of to build a reliable skill assessment suite. Learning curves were first described by Ebbinghaus (1885, transl. 1913) in his famous study about retention of nonsensical syllables. Learning curves are often used to describe the acquisition of proficiency in a task over a longer period of training or trials. A learning curve is plotted with training time on the X-axis, while the Y-axis is used for a measure of learning or proficiency, e.g. Total Damage on a MIS simulator. The start point for a learning curve is dependent on amount of previous experience an individual has. Furthermore, the maximum attainable level of proficiency, or the so called asymptote, also differs per person as can be seen in the study by Grantcharov & Funch-Jensen (2009). The difference between the starting point and asymptote is called the amplitude. The last variable of interest in learning curves is the rate, which determines how steep a learning curve is. All these variables are also presented in Figure 2. Of all of the above mentioned parameters of a learning curve, the maximum performance can be considered the most interesting parameter because of what it describes. The precise model used for the learning curve used in this study will be discussed in the method section.

Learning curves supposedly show compatibility with the previously mentioned model of Fitts & Posner (1967), as the stages of Fitts & Posner's model can be equated to different moments on a learning curve. Learning curves can also help estimate how long one should be trained before it can be deemed safe to allow a surgeon to apply their knowledge in vivo, often this is set to at least 50 cases in training circumstances (The Southern Surgeons Club, Moore, & Bennett, 1995). A more recent study has shown that it might take up to 200 cases until improvement seems to cease (Voitk, Tsao, & Ignatius, 2001).

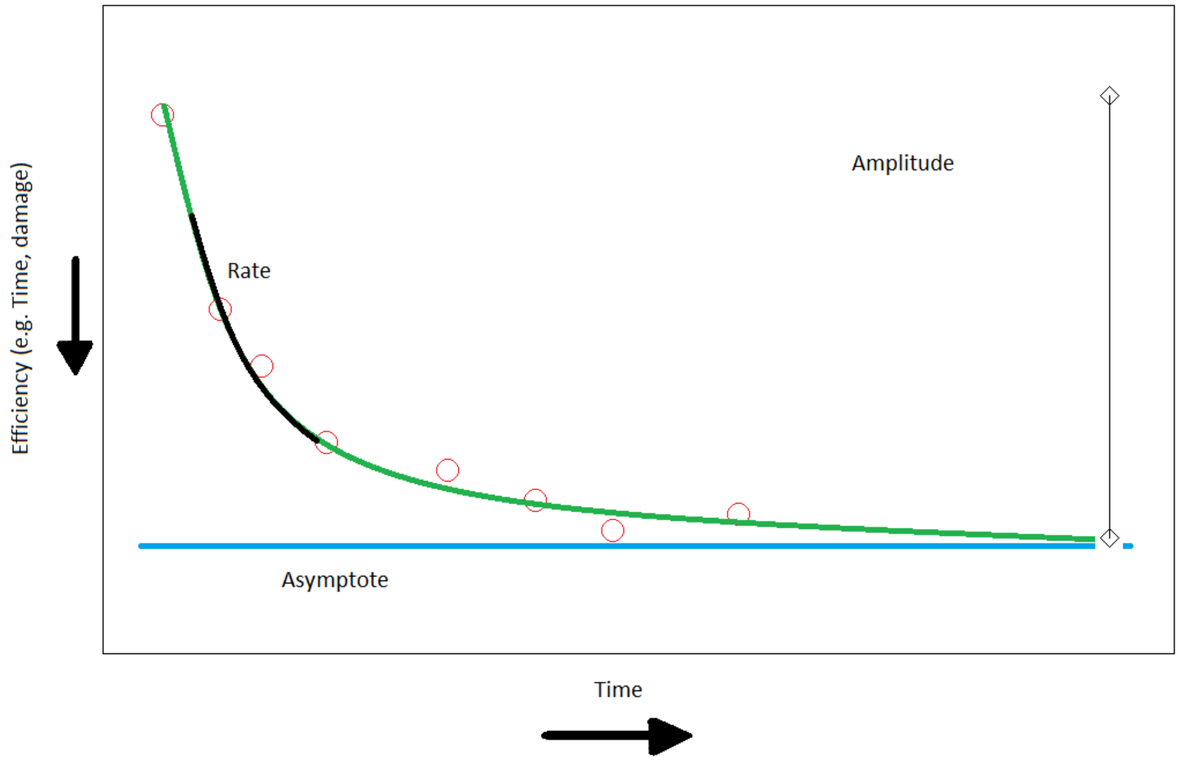


Figure 2 - An example learning curve

1.4 The Resemblance Spectrum

The research by Kramp et al., (2016) and Louridas et al., (2016) gives reasonable doubt to the use of innate ability tests on their own as measure for predicting future laparoscopic skill. In this research we therefore propose a different approach using learning curves and a small battery of tests which test for different cognitive components. We expect that some of the tasks from this test battery have a resemblance to some of the abilities used in laparoscopic tasks. This idea seems to be in line with the ideas posed in Louridas et al. (2016), which states '*It seems that strategies, that assess multiple innate abilities, their interaction, and their relation- ship with technical skill, may be more likely to ultimately serve as reliable predictors of future surgical performance*' (p689). We should note that the paper of Louridas et al. is mainly focused on innate abilities, while our research will focus more on tasks as a whole.

To try and validate our model, which we shall call the *Resemblance Spectrum* which is visualized in Figure 3, we need to define a set of tasks that can be grouped into what we expect will be different constructs. For this, we have made four groups of tasks. First off we have the Innate Ability tests, or IAT. In this group we find the innate ability test used, namely Map Planning task and PicSOr. In the second group we put Origami and Mirror Drawing, which have been shown to correlate in Schmettow, Kaschub, & Groenier (2016). The third group includes preparatory LapSim tasks, such as Grasping and Clip Applying. The fourth and final group that we can test for would be the Simulator Procedure, which would could exist of a full Cholecystectomy in the LapSim.

The Resemblance Spectrum poses a number of questions on the association between the defined constructs. These questions are visualised by the arrows within and between the constructs in Figure 3. We want to know if the tasks have predictive of performance on the other tasks within the same constructs. This would give a measure of reliability on what we are actually measuring within one such construct. Within the same model, we also assume that some constructs to have predictive power for other, more MIS-resembling constructs. With this we want to measure if one construct is a valid measure future performance on another construct, and eventually, a full procedure.

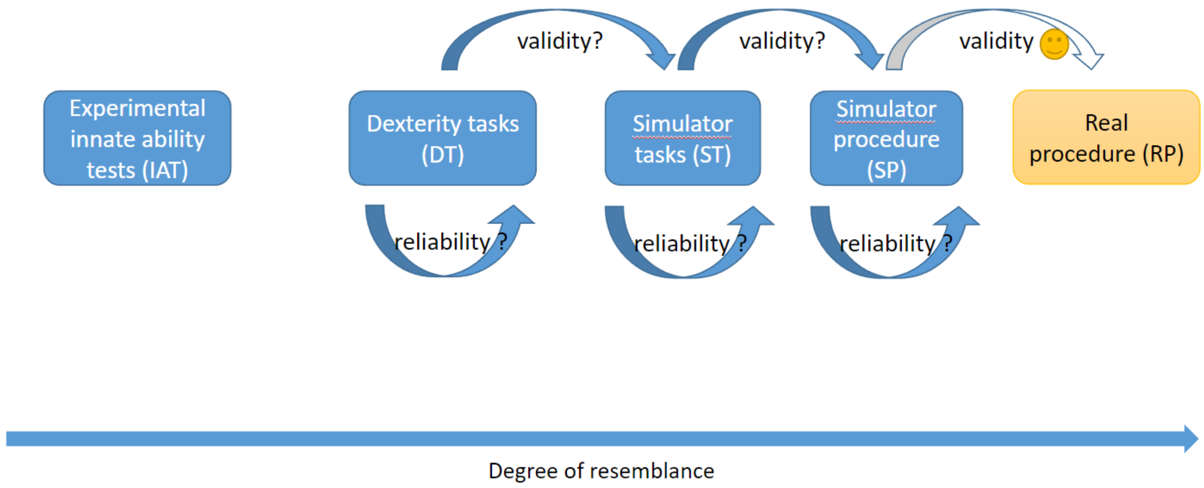


Figure 3 - The resemblance spectrum

1.5 Overview of this study

The aim of this study is to find new ways to find a combination of factors that predict laparoscopic skill acquisition in a valid and reliable way as the previously mentioned aptitude tests do not seem to be giving a consistent and reliable measure. We will try to do this by mapping a person's learning curve on two dexterity based tasks and a minimum of at least two laparoscopic simulator tasks. Two innate ability tasks were also used.

So far, a number of the previously mentioned research has tried to use learning curves to correlate innate abilities with maximum performance or other parts of the learning curve with varying degrees of success. However, close to none of these studies seem to have tried to correlate task specific learning curves with simulator performance or within-simulator task learning curves. We expect that, based on the Resemblance Spectrum, we will find a correlation within the simulator tasks themselves, proving their reliability to some degree. This leads us to the research question that this thesis is trying to answer:

Do individual learning curves of simulator tasks correlate with each other?

Following the results found in the meta-analysis by Kramp et al., (2016), we can assume that a simulator procedure gives a high correlation with future laparoscopic skill. We want to assess how different laparoscopic tasks correlate to each other to build a valid simulator-based assessment suite. We will focus on individual learning curves as this gives us a better overview of personal growth, or lack thereof.

2 Method

2.1 Participants

In this study 40 people participated. All participants were students at the University of Twente and signed up through the faculty of Behavioural, Management and Social sciences test pool and participated for course credits. Two participants were removed (*n* = 38) due to motivational problems and not finishing the session in time. The average age of the participants was 20.6 (*SD* = 2.02, *range* = 18-26). Of these participants, 31 were German, 6 were Dutch and 1 was from a different origin. Of the participants 7 were male and 31 were female. The majority of the participants, 36, reported to be right-handed, the other two were left-handed. Exclusion criteria for participating were physical disabilities regarding hands or legs as both are needed to operate the simulator used. All participants reported having normal or corrected to normal vision. None of the participants had previous experience with a laparoscopic simulator. From all participants 30 participated in a second session.

All participants signed an informed consent; this form can be found in Appendix A. This study has received ethical approval by the Ethics Committee for Behavioural and Management Sciences at the University of Twente (Request No.: 17059)

2.2 Materials

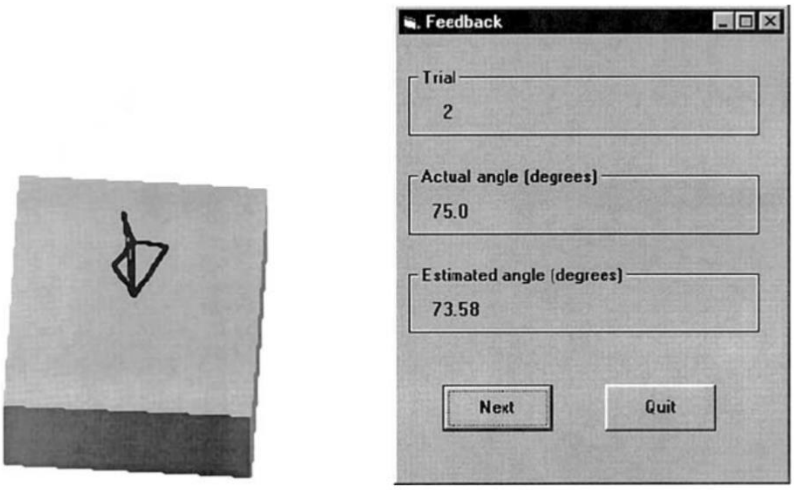
Innate Ability tasks

PicSOr task.

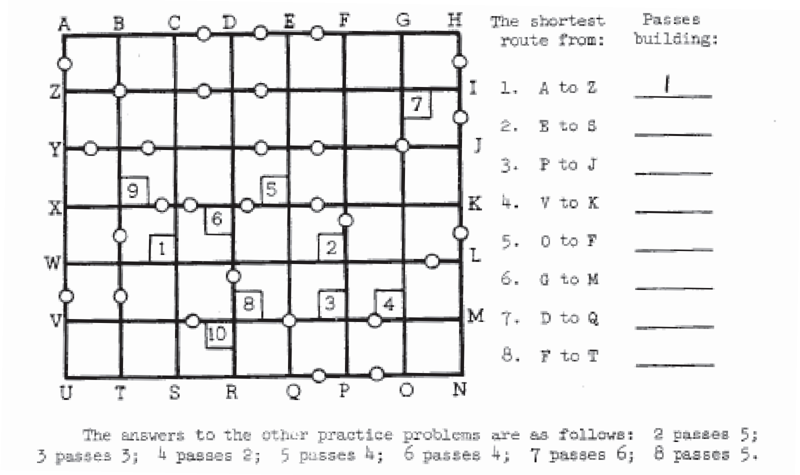
The PicSOr (Pictorial Surface Orientation) task is a computer-based assessment of the perceptual ability related to retrieving depth cues from a two dimensional image. The PicSOr test was validated by Gallagher, Cowie, Crothers, Jordan-Black, & Satava (2003). In the PicSOr task a cube is shown from different angles in each trial, and a spinning arrowhead had to be moved until it was perpendicular to the surface of the cube that the arrow is pointing at. The arrow can be moved by the Up- and Down-arrow keys on the keyboard. The task was presented on a laptop with a 13.5’’ LCD screen, aspect ratio of 16:9. An example of this task can be found in Figure 4a.

Visuospatial Map Planning

The visuospatial Map Planning test is a paper-based assessment by Ekstrom, French, & Harman, (1976). Information about the validity of this task is fairly hard to find, we will therefore use the norm-scores calculated by Henn et al., (2018) from a data-set of 260 participants. The distribution of this set can be found in Appendix B. In the task the participant has to rapidly plan routes. This is done by presenting the participant with a grid with the size of 8 vertical and 7 horizontal lines. This gird is supposedly depicting a city map. In this map white dots depict a road block which cannot be passed, numbered squares touching the intersection of the lines mark a building. The participant had to find the quickest route between two points, marked A up to Z which were located on the edges of the grid including the corners. In every path the participant will pass one and only one building. A test map can be seen in Figure 4b. There is always one and only one fastest route that follows the stated rules. The first page contained instructions and 8 practice routes, with the first one already filled in so the participants understood the concept of this test. The actual task consists of 4 of these grids. One trial consists of two of these grids, and 20 routes have to be determined. Each trial has a time limit of 3 minutes. Three A4 sheets were needed per participant.



a.



b.

Figure 4 - PicSOr and Map Planning task, respectively

Dexterity tasks

Origami.

Twenty white sheets of square paper with sides of 21 cm were used per participant. For every trial a new piece of paper was used to prevent the use of previously folded lines from prior trials. The size of the paper was chosen as modified A4 paper was used. The size was found to be neither too small nor too big, which would both introduce their own difficulties. The figure chosen was the fox, as in the previous study by Schmettow et al. (2016), and should be foldable without any verbal instruction or extra help. The instructions can be seen in Figure 5 and were presented in the same way to the participants. Each step introduces a single new fold towards the end goal. The instructions were printed on an A4 sizes sheet of paper.

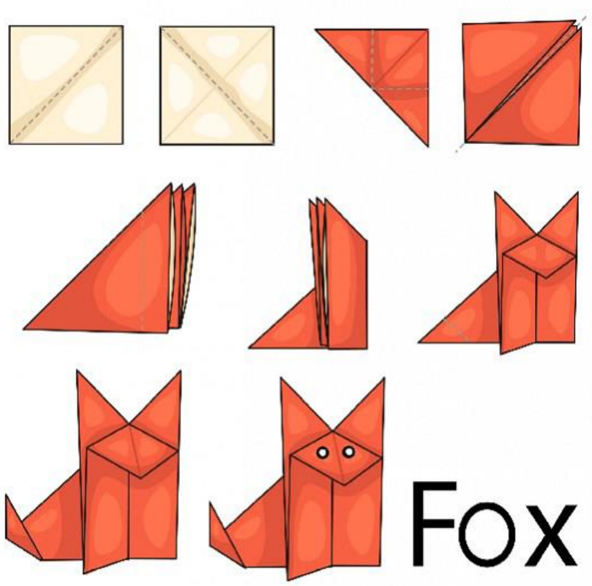


Figure 5 - The instructions given to the participant for the origami task

Mirror Drawing.

For every participant twenty copies were printed of the figure depicted in Figure 6. The space between the two lines was supposed to be traced while only being able to see their movement through a 30x30 centimetre square mirror set as close as possible to being perpendicular to the writing surface. To ensure the participant was only looking through the mirror, they had to put their hand through a cardboard box. This box had a cut-out near the bottom but obstructed the direct view of the paper, pen and hand. The box was coloured plain white to prevent possible distractions that could occur otherwise. The difference with the previously mentioned study by Schmettow et al. (2016) is that the stimulus has been made asymmetrical and the stimuli were turned ninety degrees every trial. The turning was to make sure that the participants did not learn the motion by heart, but instead had to keep on using cognitive abilities to solve the task. These figures were printed with 12.5 cm between the left and right most outer edge.

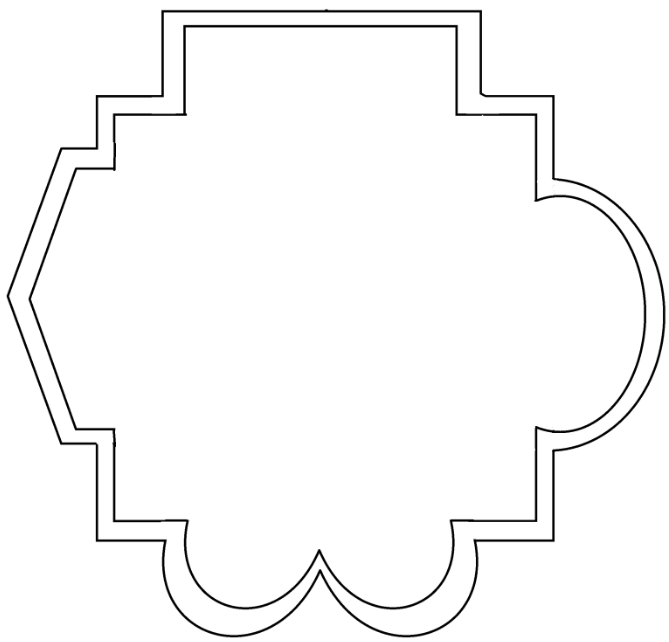


Figure 6 - The asymmetrical mirror drawing stimulus

Lapsim and simulator tasks

The Lapsim is a computer simulator used for the simulation of a variety of laparoscopic procedures and is developed by Surgical Science Sweden AB, Göteborg. The LapSim is combined with the SimBall-module developed by G-coder Systems AB, Göteborg. The version used is the non-haptic simulator, which consisted of two or three SimBall-modules for the insertion of endoscopic tools, a desktop computer capable of running the simulation and a 23” LCD screen with a 16:9 aspect ratio. The endoscopic tools resemble the ones used in an operating room environment, except that their ends do not feature the actual tools used in the task. The SimBall modules register how far the tools are inserted, what angle they are turned towards and force exerted on the handles. The computer translates this into movement of the tools in the digital environment created by the LapSim software

Every task used in this study requires the use of at least one forceps and a possible array of different instruments. If the task requires the participant to switch between different instruments, they have to pull back until a mechanical limit was reached and pinch the instrument handles before inserting the instruments again.

For this study four basic, preparatory tasks of the cholecystectomy module were used, which will shortly be discussed. For all task an easy pre-set was used as all participants could be considered novice in laparoscopy.

Grasping

The Grasping module was the first task participants did on the Simulator, which created the small challenge of getting the forcipes into view as the participants had no previous experience. The task in the LapSim involved grasping a total of six objects in total per trial. The side with which they had to do this alternated between each object with the initial object starting with the right instrument. The objects were to be picked up and stretched until it detached from the abdominal wall and then put in an endoscopic bag. A screenshot of the task can be seen in Figure 7a.

Cutting

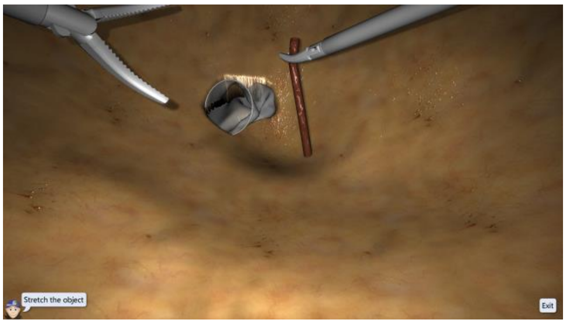
In the Cutting module the participant had to grasp an object much like in the Grasping task. They now also had to stretch the object till a new highlighted area appeared. They then had to apply the ultrasonic-scissors tool on this highlighted area and use the pedals to cut the vessel. The dissected pieces had to be disposed of into the endoscopic bag visible. If the participant stretched the vessel to much it would tear and bleed. To pass this exercise the participant had to repeat this task three times without causing excessive amount of damage. A screenshot of this task can be seen in Figure 7b.

Clip Applying

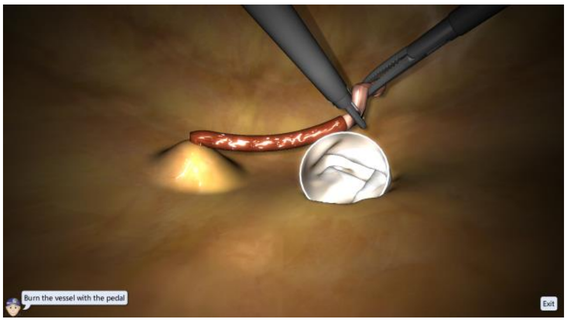
In the Clip Applying module the participants had to use 4 different instruments, requiring multiple switches of instruments. In this module the participant had to grasp a vessel that was connected on either side to the tissue wall and use their other endoscopic instrument as clip applier to place a clip on the highlighted areas. After successfully doing this for both sides of the vessel, the participant can then cut the vessel in the middle with the scissors (called dissectors in the LapSim). If the participant does not apply the clips all the way over the vessel or they rip the vessel apart, they then have to stop the bleeding and clean up the blood with the suction instrument. In case of a torn vessel, the participant won't pass. Each trial consisted of one of these procedures. A screenshot of this task can be seen in Figure 7c.

Lifting and Grasping

In the Lifting & Grasping Module the participant had to use the probe instrument. The probe was used to gently push up a block of tissue in the simulator environment. Underneath this block lay an object which had to be picked up with the other hand using a grasper and be put into an endoscopic bag. The instruments alternated between hands after each object had been picked up. The endoscopic bag was located on the same side as the grasper instrument. One trial consisted of 6 objects to be picked up. A screenshot of this task can be seen in Figure 7d.

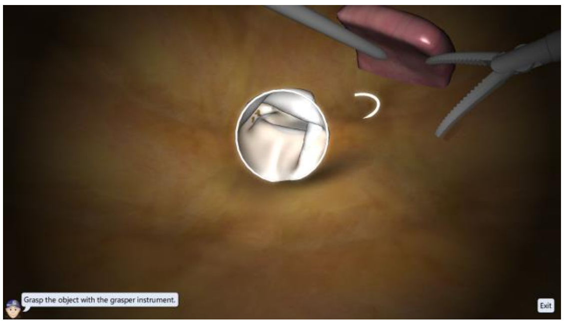


a



b

c



d

Figure 7 - View of the tasks

2.3 Procedure

Location

All experiments took place in the Experimental Centre for Technical Medicine on the campus of the University of Twente. This room was well lit and usually fairly quiet. The LapSim itself was placed in a grey booth, allowing participants to focus on the screen and tools. The non-simulator tests were also conducted in the same room.

Greetings and instructions

After the participant had arrived they were greeted and thanked for their participation. They then received full disclosure on the nature of the research they were about to participate in. If they agreed to continue, they then signed the informed consent form. Every participant got a participant number for any later identification of tasks.

Instructions were given verbally, but have been previously determined and written down in a researcher manual. After every instruction the researcher asked the participant whether they fully understood what they were supposed to do. Standard demographic measurements were recorded together with the handedness of the participant and possible video game experience. This was all recorded through an online questionnaire. The questionnaire and the verbatim instructions can be found in Appendices C and D respectively.

Origami & Mirror drawing

All participants with an even number started with the Origami task, every participant with an odd number started with the mirror drawing task. Both tasks consisted of 20 repetitions. The time, possible errors and observations were recorded in a structured manner by the researcher during and after each repetition. After the tenth trial the researcher told the participant how many trials were left so the participants knew how many trials were left to do. The observation form for both tasks can be found in Appendix E. After each task (20 trials) the participants received a five minute break to prevent possible fatigue.

PicSOr

The PicSOr task was always the third task for the participants. The participants were seated in front of the computer and were instructed on the nature of the task. The participants did four practice trials in which they received feedback from the program. In the practice trials the participants had the chance to get used to the nature of the task and the control of the arrow. After the practice, they would conduct 35 consecutive recorded trials. The participants were instructed to be as fast and accurate as possible. After this task they received another five minute break to prevent possible fatigue.

Map Planning

Map planning was always the fourth and final non-simulator task for the participant. Participants were presented with the practice page of the set and were instructed to read the instructions printed on the page. They then got three minutes to complete the practice items presented on this first page. They could then check their own answers with the correct ones, which were printed underneath the grid, so they truly understood the task. Subsequently, they then got three minutes again for the first trial of twenty items, and another three minutes for the second trial of twenty items. After this task was completed, they received another five minute break before continuing on to the LapSim.

Grasping and Cutting tasks in the LapSim

After completing the two dexterity tasks and the two aptitude tasks, which took one and a half hour on average, the participant was introduced to the LapSim. In the LapSim, the participant received standardised feedback after each trial from the computer, in the form of a number of sliders and a graph that combined these sliders into one score.

The first task was the Grasping task, for which they received instructions from the program. These instructions were presented in text and video. The second task the participants were instructed to complete was the cutting task. The instructions were once again presented through the program in the form of text and video. Data collection for both tasks happened digitally by the LapSim. Both tasks consisted of a total of twelve trials per participant to gather enough data for robust analysis. Between both tasks the participants got a five minute break to prevent fatigue.

Optional advanced tasks in the LapSim – Second Session

After completing the Grasping and Cutting tasks in the Lapsim, participants were asked if they were interested in doing two more advanced tasks in another session. If they decided to participate, a second session would be planned. Instructions were once again given through the LapSim program in form of text and video. Performance was recorded digitally by the LapSim program. Both tasks were repeated twelve times to gather sufficient data for robust analysis. Between both tasks the participant received a five minute break. The second session was planned within four weeks after the first session with at least one day in between to prevent possible fatigue and de-motivational effects. There was one exception to this. This participant did not complete the Lifting & Grasping task due to the, expected, fatigue.

Debriefing

After the participant completed the first or second session, they were asked about their experience with the tasks and if they wanted to receive the results. Following this they got to know their error rates, movement efficiency and completion times. Subsequently they could ask any questions that they had. The participant was thanked for their participation and was given their own copy of the informed consent form which also held contact information for future inquiries.

2.4 Design

For this study a within-subject time series design has been used. All participants received the same, standardised instructions as found in Appendix D, and the same set and repetitions of tasks. The Origami, Mirror Drawing, PicSOr and Map planning tasks were taken into the experiment for further exploration of their possible association with laparoscopy. The Origami and Mirror Drawing task were both repeated a total of twenty times. The PicSOr task consisted of 35 items and Map planning of two trials of 20 items each. The Origami and Mirror Drawing were alternated, based on the participant's number, to control for any possible order effect. This was deemed necessary considering how these two tasks correlated positively in a previous study by Schmettow et al. (2016). The LapSim tasks from the first session were always presented in the same order, first Grasping and then Cutting. Both tasks were repeated twelve times. In the optional second session the participant would do the Clip Applying and Lifting & Grasping task. Both of these tasks were also repeated 12 times. The Grasping task and the Cutting task also served as preparatory tasks for the possible second session. Any possible order effect would only result in an offset of the learning curve, but would not affect the correlations.

2.5 Measurements

Performance on both the Origami and Mirror Drawing tasks was primarily measured by time-on-trial and whether the outcome was a success. For every trial it was noted how many mistakes were made, whether there was change of strategy. The researchers also had space to note down anything they found remarkable. Researchers knew what to focus on through a pilot run of the experiment. Performance for the PicSOr task was determined by the correlation between the chosen angle and actual angle over the course of the 35 items, in line with the paper by Gallagher et al. (2003). These scores were then compared with the norm-scores defined in the study by Henn et al. (2017). The performance on the Map Planning task was defined by the amount of correct choices made in each trial.

While the LapSim is running it registers a lot of data. For this study, the amount of movement, error rate and time-on-task were used, in line with the paper of Luursema, Rovers, Groenier, & van Goor, (2014). From these values, it can be determined what the participants initials performance was, the rate of their learning curve and when they achieved maximum performance.

2.6 Data Analysis

A non-linear mixed effects model with an exponential learning curve as likelihood function is used to build regression models on for the learning curves of the participants. The model described by Heathcote, Brown, & Mewhort (2000) is described by the formula:

$$Y_{ptN} = Asym_{pt} + Ampl_{pt}exp(-Rate_{pt}N)$$

The parameter Asymptote of participants stands for the maximum performance that the participant is expected to reach eventually. The nearing of this performance is reached asymptotically with continued practice. The parameter Amplitude of participant is the strength of the improvement and shows the difference between the initial performance and the Asymptote. The Rate parameter describes the speed of learning and equals the steepness of the learning curve. The word steepness is used as in mathematics, signifying how fast the curve approaches the asymptote. Due to formula used, the problem arises that the amplitude parameter does not have a natural interpretation. Low amplitude can mean either a small improvement due to previous knowledge, or a small improvement due to a low potential for a high maximum performance. It is expected however, that all participants have close to zero previous experience.

Due to this lack of natural interpretation, it has been chosen to use a slightly different parametrisation of the exponential model which takes into account (virtual) previous experience as this is deemed more meaningful than Amplitude. If a participant has a high amount of previous experience, the function needs to be shifted to left on the x-axis so that it appears flatter in the observed range. If no previous experience exists the parameter functions as a measure for initial performance. The formula used for this model is:

$$Y_{ptN} = Asym_{pt}\left(1 + exp\left(-Rate_{pt}(N + Prev_{pt})\right)\right)$$

For both dexterity tasks a learning curve has been plotted. Four learning curves were estimated for the LapSim based on the time-on-task of a trial. For the individual learning curves a multi-level model was used. Each of the learning curves provided us with the three parameters described above. Then the correlation of the three parameters between tasks was assessed by calculating a Pearson pairwise correlation coefficient, as well as a 95% Bayesian credibility limit for each of these correlations. Main focus in this analysis is on the Asymptote, as this represents the maximum performance a participant is able to eventually reach. All data analysis has been conducted with R 3.3.30 (Murdoch, 2016). For syntax, see Appendix F. The reasoning behind the model created will briefly be discussed before the results, and this description has been adapted from Arendt, Schmettow, & Groenier (2017)

Multi-level learning curve model

To estimate the learning curve of the participants in the conducted tasks, a non-linear learning curve model has been built with R 3.3.0. The model is non-linear as we can expect that the learning effect is greater at the start of a task and will then decline rapidly, as in compliance with the model described in Figure 2. As for the error component, a Gamma distribution has been chosen instead of the Gaussian distribution. This decision was made for two reasons. Firstly, time errors are typically left skewed; secondly, Gamma features stronger error variance with larger time-on-tasks, for instance in early trials, which is more realistic given our case.

As described in section 1.5, a total of three parameters play a role in the learning curve model. The maximum performance and rate were transformed through link-functions that establish linearity. This is needed to be able to build proper random effects in the analysis. The applied link-functions scale maximum performance on the log-scale so that it runs from $-\infty$ to $+\infty$, and the rate parameter on the logit-scale so it runs from 0 to 1. After this transformation the now linear parameters for max performance and rate are referred to as eta-max performance and eta-rate respectively. This is also visible in the R syntax in Appendix F.

A prior distribution, or an assumed distribution before any evidence is collected, has been established by building a basic non-linear model. This has been done with the data of one participant and one task, in our case this was participant 5 and the Origami task 'time-on-task' were used. It has been used as the prior distribution in all subsequent analyses. The formula that was built was also used in every subsequent model. The final model estimates random effects for the three learning curve parameters, based on the time-on-task by all participants for each of the six tasks.

Un-motivated participants were excluded from the data. For any at random missing values, the Bayes estimation replaces these by the previously calculated prior. Then fixed- and group effects were drawn from the posterior distribution of the model. Finally, the correlations between each task's posterior parameters have been calculated. The correlations should be seen as an inter-item reliability measure. The more the tasks are alike, the higher the correlation is expected to be. For both effects and correlations, 95% credibility limits were established. A credibility limit denotes a range in-between which the true value can be found with a, in this case, 95% certainty.

The thesis of Arendt et al. (2017) presents the data concerning the innate ability test, including how the multi-level model is adapted for that data.

3 Results

3.1 Data exploration

Before doing any analysis, we first removed participant one. Another participant was later removed from the data as this participant wasn't motivated. For more in depth explanation, see Appendix G.

To make sure our expected prior model fits the gathered data, a plot was made in which all data points were plotted per participant, separated by task. This graph can be seen in Figure 8. This figure fits the learning curve model we assumed in Figure 2, meaning that our data fits the assumed prior.

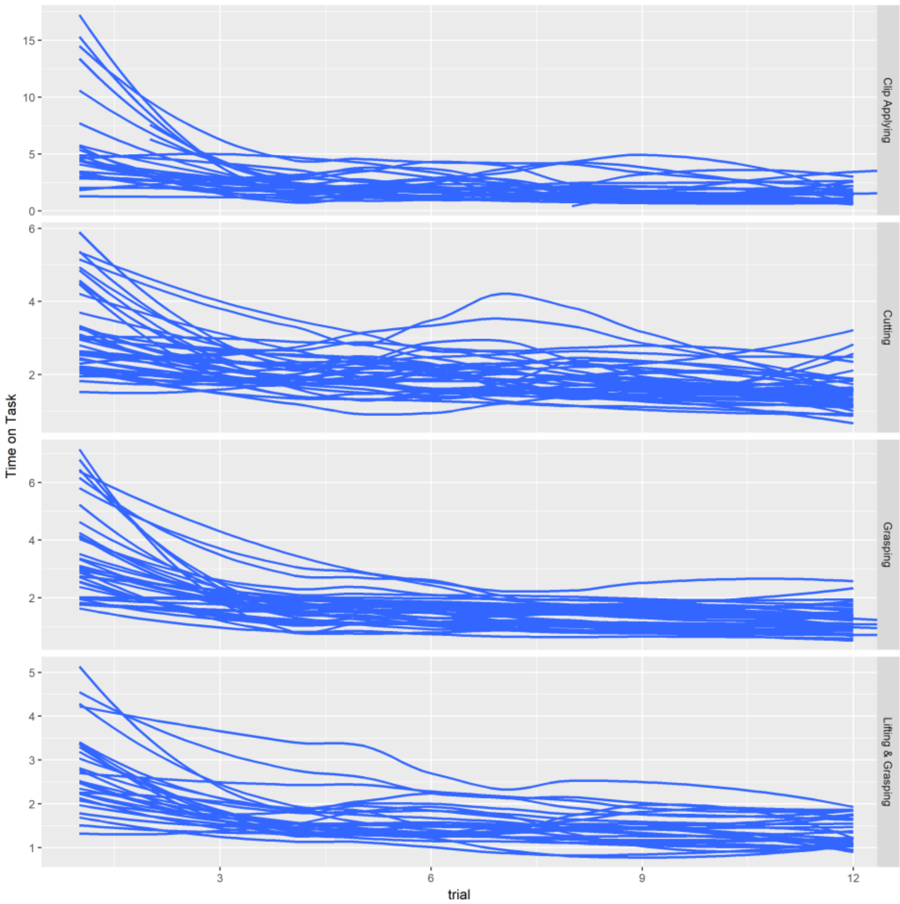


Figure 8 - Data curves per participant per task based on Time on Task

After the first exploration of the data, it was decided to use the Time-on-Task as this was the only variable available from all tasks that were conducted. As can be seen in Figure 9, almost everybody seemed to be able to go through the steepest part of their learning curve in a few trials. It should be noted however, that most people did not actually reach a passing grade from the software itself. Often this was due to high damage, or due the fastest performance still being seen as too slow. Participant 8 was the subject of a procedural error, this can still be seen in Figure 9 at the bottom, second to last graph. This is the only participant of all participants who has data points up and beyond trial 20.

When looking further at the learning curves per person in Figure 9, we can see that there is a lot of variation between people and in some cases we can see that this is also the case within one participant. A lot of variation within a participant is well visible in participant 31. However, a majority of the participants also show consistent results between tasks and within task.

3.2 Correlation between LapSim tasks

In the method section we discussed some tasks that were also performed for other theses that are being written concurrently with this one. Not all of the gathered data will be discussed in this thesis, for the other parts discussed in the Methods section, see the theses from Schmettow, Bennink, & Groenier (2017) or Arendt et al. (2017). The main model used here is derived from the data focused around all of the ability-learning tasks and can be found in the R syntax as M_6.

A Pearson Pairwise correlation coefficient was calculated to measure the association between all four simulator tasks and their results can be found in Table 1. The table shows the calculated correlations between estimated maximum performance for Time on Task of the four simulator tasks. The values between the brackets signify the 95% credibility limit.

According to Bolboaca & Jäntschi (2006), we can consider a correlation strong when its 0.8 or higher, and weak when it is lower than 0.5. According to this definition we have three moderate positive correlations, which can be seen between Cutting and Lifting & Grasping, Cutting and Grasping, Cutting and Clip Applying. A weak positive correlation has been found between Clip Applying and Grasping, Lifting & Grasping and Grasping, Clip Applying and Lifting & Grasping. Scatterplots of the correlations can be seen in Figure 10. These show a more visual representation of the found correlations.

Table 1 - Correlations of the Time on Task (etamaxp) parameter, CL-95%

	Correlation Coefficients			
	Grasping	Cutting	Clip Applying	Lifting & Grasping
Grasping	1	-	-	-
Cutting	.53 [.17, .78]	1	-	-
Clip Applying	.26 [-013, .60]	.54 [.15, .80]	1	-
Lifting & Grasping	.46 [.06, 0.75]	.68 [.33, .88]	.16 [-.25, .53]	1

It should be noted that for all correlations the credibility limit can be considered fairly broad. Still, 4 out of 6 correlations are positive with the exception of Clip applying with Grasping and Clip Applying with Lifting & Grasping. This means that we can say with 95%

certainty that the four tasks do all show a positive correlation, even on the lowest end of their respective credibility limits. For the other two, Clip Applying with Grasping and Clip Applying with Lifting and Grasping, we cannot say with certainty that they do truly correlate.

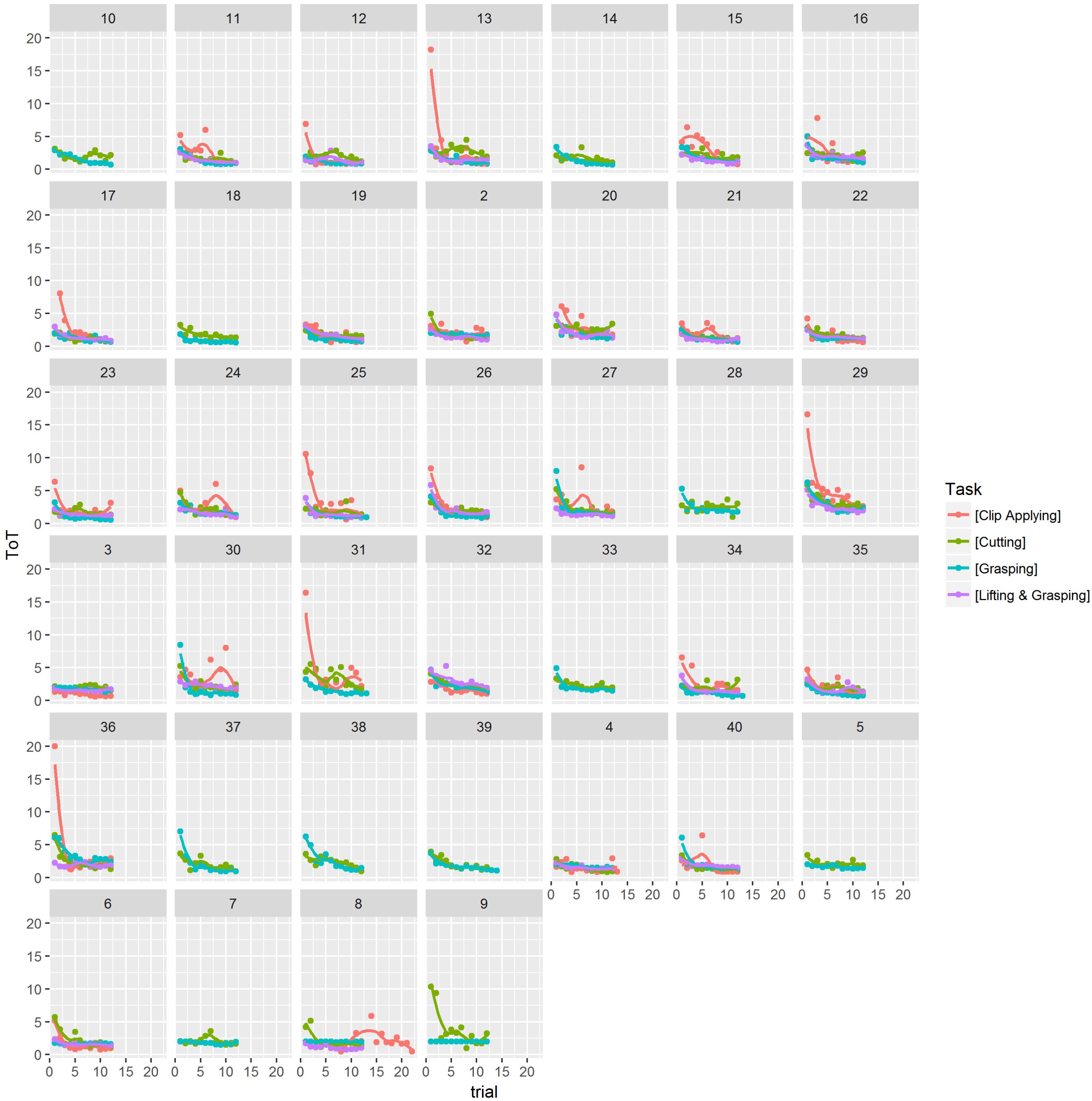


Figure 9 - Learning curves per participant

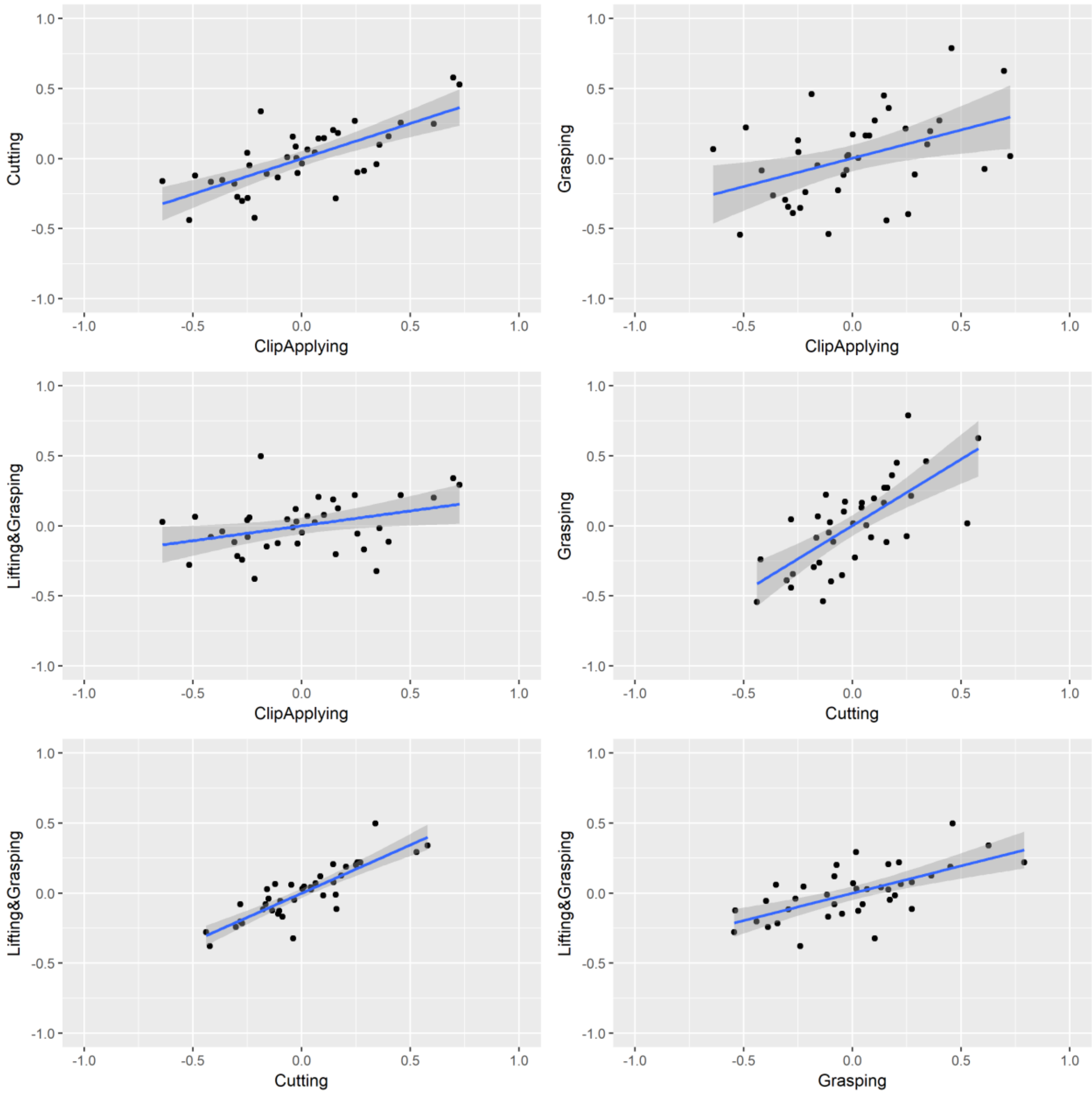


Figure 10 - Scatterplots of the etamaxp parameters between all Lapsim Tasks

4 Discussion

The aim of this study was to investigate if it is possible to create a reliable assessment suite for future laparoscopic task performance. To assess this we used the Time on Task max performance values after they were transformed to through the link functions into eta-max performance. The correlations of these eta-max performance values can be used in two different ways. Firstly, they can show how reliable the performance on one task predicts the performance on another task. Secondly, we can use the correlations to select items. The correlations we found show an overall moderate to weak correlation with each other. Three from the six correlations were found to have a moderate correlation. Of the other three, two have a weak correlation and one has an almost negligible correlation.

In normal psychometric terms the correlations found in this study would not justify building a simulator based assessment suite, as the correlations suggest that such a test would not be reliable enough. Even though the correlation might not directly suggest it, the results of this study seem to promise that it is highly likely to build a simulator-only based assessment suite. Most likely, the correlations are quite low due to the fact that laparoscopic tasks do not consist of merely one specific ability. As laparoscopic tasks such as the ones performed in this described as natural tasks, which means that they do not consist of one factor but are instead multidimensional and are expected to consist of multiple abilities. The multidimensionality is also something we see in the correlations found, which could be considered fairly high between a set of multidimensional items such as the tasks used. Most likely, the tasks all load on different factors with different weights, making them seem less correlated when analysed.

When looking at the correlations, it seems that Clip Applying has the lowest correlation overall to the other tasks. Clip Applying itself shows correlations with two other tasks which include 0 in their credibility limit, meaning that we can’t say with certainty that they are actual positive correlations. However, Clip Applying also includes a moderate correlation with Cutting which could be used as an argument to keep Clip Applying in a possible test set. Nonetheless, based on these results we should probably not include Clip Applying if we would build a simulator based assessment suite.

A possibility is that procedural errors created broader credibility limits and skewed correlations. To exclude this possibility as much as possible, a researcher manual was written to make everything as much standardized as possible. Furthermore, we kept a file to note down any procedural discrepancy. From this file only one major procedural error can be identified on task 8 where the participant accidentally was given the more difficult version of the Clip Applying task. The only other error to be identified from this document is the outlier produced by participant 11, trial six. This trial error was produced due to not reading the instructions given by the LapSim thoroughly enough, causing a delay until they asked the researcher for help.

It might also be that people are not able to reach their maximum performance within the 12 repetitions given. However, when looking at Figure 9 it seems to show that this assumption is not true. For almost everybody the time on task stays consistent throughout their trials, with sometimes a spike which would only indicate fatigue or a bad trial as the one discussed earlier. The spikes visible in the data might be due to the fact that none of the participants were explicitly instructed to pass the task, adding this into the verbatim instructions for every task might have given a different result. However, barely any of the participants showed any lack of motivation in participating and overall seemed eager to try and pass the simulation.

4.1 Alternative explanations

After considering the previous arguments we went looking for a possible alternative explanation what might cause these uncertainties in our data. An argument that could be given based on the correlations found is that the different usages of hands might play an important role in this. Considering the contents of the task, we can say that both Cutting and Lifting & Grasping are heavily based on using both instruments at the same time.

According to this reasoning participants would also have to be able to perform similarly well on Clip Applying. This however cannot be seen in the data. After some discussion among the researchers a possible option was found that the grasper instrument is used differently between these tasks. In Clip Applying the vessel will tear if you move too much during the applying or cutting. This makes the Clip Applying task appear more difficult. The correlations seem to show this mainly with the Grasping and Lifting & Grasping correlations towards Clip Applying. However, looking at the correlation between Cutting and Clip Applying, we see a moderately positive correlation which could be caused by the identical use of the instrument in both of these tasks.

A second explanation could be that the participant had a different approach on tasks. Grasping for might feel easy and people want to try and do this task fast. However, more complex tasks force people to make a decision in a speed-accuracy trade-off. A participant can either choose to do a task fast, or try and do it accurately. These two options are usually mutually exclusive for the participants in our sample. If a participant would choose for accuracy, this would increase the time. As we use Time-on-Task as our measurement, this causes our data to become slightly warped and show an unrealistic representation compared to the participants actual performance, of which the latter could be superior compared to a participant using a speed-focused approach to the task.

4.2 Generalisability

An important note to make about this entire study is that the participants overall have no direct interest in becoming a surgeon or learning the complete procedure. Furthermore, all participants can be described as novices in this area. This can cause motivation to waver during the experiment or between sessions. To have a better representation of the actual goal population, one would have to reproduce the research with a professionally interested group of participants. This is supported by the study from van Dongen, Tournoij, van der Zee, Schijven, & Broeders (2007), which showed that the Lapsim is also able to make a distinction between experts and novices. Both of these points could be of influence on our results.

A point can be made however that the results found can be used to start with building a simulator based assessment suite for laparoscopic skills. This study has shown that we can at least get reliable results from the simulators to assess later performance.

Looking at the Resemblance spectrum and its applicability, the thesis from Schmettow et al. (2017) shows that dexterity tasks do not have any predictive value towards laparoscopic performance. The thesis from Arendt et al. (2017) shows that innate abilities do not have any extra predictive power on laparoscopic skills either. This means that in the current version of the Resemblance Spectrum we cannot proof any of the validity that was expected left of the Simulator Tasks group. However, this study has found a possible predictive power in simulator tasks for future laparoscopic skill.

4.3 Future studies

As just has been stated, based on the conducted set of studies we cannot confirm a number of assumptions made in the Resemblance Spectrum. However, the correlations between LapSim tasks seems promising so far, making a case for further research on the right side of the Resemblance Spectrum. This would mean that we still need to show the validity between simulator tasks and a full on procedure within a similar setting. The latter part is important to make the results compatible and comparable with each other. This could be something future research could look into. We cannot prove the last part of the Resemblance spectrum at the current institution as this institution is not licensed for genuine surgery, so we need to rely on literature for data on this step of the model. It has been shown on different occasions that learning and improving skills on a simulator benefits operating room performance (Palter & Grantcharov, 2014; Seymour et al., 2002), thus giving support the last link of the Resemblance Spectrum between simulator and operating room performance.

Additionally, the amount of tasks could be expanded so that a more reliable test can build, or that this study's findings can be refuted. It would be of great interest for this future study to not deviate too far from this studies procedure so the results can be more easily compared. A broader set of simulator tasks might also give more insight into possible factors at play in laparoscopic tasks and procedures, which in turn could create a more personally tailored teaching program per person. A possible extra application in a broader set of tasks could be the use of simulators that also include haptic feedback, as it has been found in a systematic literature review by van der Meijden & Schijven (2009) that haptic feedback is important during the early phases of psychomotor skill acquisition.

Furthermore, a one hand vs two hand focus of tasks might be of influence on the eventual performance a person can achieve. Another issue that could be addressed concurrently would be looking at how the instruments are used differently and if this has any effect on performance. The influence this might have on laparoscopic proficiency is something that future research could focus on. From the results found this seems the best explanation. However, this is nothing more than mere speculation as no previous indications seem to have been found in the literature.

Additionally, the tasks have been set to an easy setting and most of the tasks were fairly repetitive. It could be that the correlations found in this research might be even lower due to the participants learning a sequence, instead of actually learning a skill that could be used during real surgery. A future study could try to avoid this by randomising the content of the task slightly, instead of using one and only one standard scenario from the software.

A worthwhile point of discussion is that, due to this study being a thesis, it is fairly difficult to gather a large set of participants willing to participate in an approximately three hour during study. Due to this, we automatically get a wide credibility limit which causes our study to reduce in certainty of the statements we can make. To reduce this uncertainty it would be advised to try and gather a larger amount of participants to further proof any research question one could derive from the Resemblance Spectrum.

5 Conclusion

In this study we have tried to approach the problem of predicting laparoscopic skill through a new model. This model, called the Resemblance Spectrum, shows how we expect different constructs to correlate with each other. This model tries to find predictive value of different tasks towards laparoscopic surgery. We expected that at least some of the constructs used in the Resemblance Continuum have some predictive value towards laparoscopic performance. This study has looked at the possibility of building a reliable assessment suite based on laparoscopic simulator tasks. Although only moderate and weak correlations were found, we argue for the case that laparoscopy is a natural task which consists of a multitude of factors. This multidimensionality causes our correlations to become weaker than those of a unidimensional test. Based on this reasoning we argue that it could be possible to build a reliable simulator based assessment suite, although more research is needed to prove how well simulator tasks would predict simulator procedure performance.

References

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117*(3), 288–318. https://doi.org/10.1037/0096-3445.117.3.288

Ackerman, P. L., & Cianciolo, A. T. (2000). Cognitive, perceptual-speed, and psychomotor determinants of individual differences during skill acquisition. *Journal of Experimental Psychology: Applied*, *6*(4), 259–290. https://doi.org/10.1037//1076-898X.6.4.259

Ackerman, P. L., & Cianciolo, A. T. (2002). Ability and task constraint determinants of complex task performance. *Journal of Experimental Psychology: Applied*, *8*(3), 194–208. https://doi.org/10.1037/1076-898X.8.3.194

Arendt, A., Schmettow, M., & Groenier, M. (2017). *Exploring the internal consistency and validity of basic laparoscopic tasks in the LapSim simulator*. University of Twente.

Berguer, R. (1999). Surgery and Ergonomics. *Archives of Surgery*, *134*(9), 1011. https://doi.org/10.1001/archsurg.134.9.1011

Bolboaca, S.-D., & Jäntschi, L. (2006). Pearson versus Spearman, Kendall’s tau correlation analysis on structure-activity relationships of biologic active compounds. *Leonardo Journal of Sciences*, *5*(9), 179–200.

Buckley, C. E., Kavanagh, D. O., Nugent, E., Ryan, D., Traynor, O. J., & Neary, P. C. (2014). The impact of aptitude on the learning curve for laparoscopic suturing. *American Journal of Surgery*, *207*(2), 263–270. https://doi.org/10.1016/j.amjsurg.2013.08.037

Cao, C. G. L., MacKenzie, C. L., & Payandeh, S. (1996). Task and motion analyses in endoscopic surgery. In *American Society of Mechanical Engineers, Dynamic Systems and Control Division (Publication) DSC* (Vol. 58, pp. 583–590). ASME.

Darzi, A., Smith, S., & Taffinder, N. (1999). Assessing operative skill. Needs to become more objective. *BMJ (Clinical Research Ed.)*, *318*(7188), 887–888. https://doi.org/10.1136/bmj.318.7188.887

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology.* (H. A. Ruger & C. E. Bussenius, Trans.). New York, NY, US: Teachers College Press. https://doi.org/10.1037/10011-000

Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). Kit of Factor-Referenced Cognitive Tests.

Fitts, P. M., & Posner, M. I. (1967). *Human Performance*.

Gallagher, A. G., Cowie, R., Crothers, I., Jordan-Black, J., & Satava, R. M. (2003). PicSOr: an objective test of perceptual skill that predicts laparoscopic technical skill in three initial studies of laparoscopic performance. *Surgical Endoscopy*, *17*(9), 1468–1471. https://doi.org/10.1007/s00464-002-8569-4

Gallagher, A. G., McClure, N., McGuigan, J., Ritchie, K., & Sheehy, N. P. (1998). An Ergonomic Analysis of the Fulcrum Effect in the Acquisition of Endoscopic Skills. *Endoscopy*, *30*(07), 617–620. https://doi.org/10.1055/s-2007-1001366

Gallagher, A. G., & Smith, C. D. (2003). Human-Factors Lessons Learned from the Minimally Invasive Surgery Revolution. *Surgical Innovation*, *10*(3), 127–139. https://doi.org/10.1177/107155170301000306

Gough, M., & Bell, J. (1989). Introducing aptitude testing into medicine. *BMJ*, *298*(6679), 975–976.

Grantcharov, T. P., & Funch-Jensen, P. (2009). Can everyone achieve proficiency with the laparoscopic technique? Learning curve patterns in technical skills acquisition. *American Journal of Surgery*, *197*(4), 447–449. https://doi.org/10.1016/j.amjsurg.2008.01.024

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: the case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*(2), 185–207.

Hegarty, M., Keehner, M., Cohen, C., Montello, D. R., & Lippa, Y. (2007). The Role of Spatial Cognition in Medicine: Applications for Selecting and Training Professionals. In G. L. Allen (Ed.), *Applied spatial cognition: From research to cognitive technology*. (pp. 285–315). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Henn, P., Gallagher, A. G., Nugent, E., Cowie, R., Seymour, N. E., Haluck, R. S., … Neary, P. C. (2017). A computerised test of perceptual ability for learning endoscopic and laparoscopic surgery and other image guided procedures: Score norms for PicSOr. *The American Journal of Surgery*, 1–5. https://doi.org/10.1016/j.amjsurg.2017.01.025

Henn, P., Gallagher, A. G., Nugent, E., Seymour, N. E., Haluck, R. S., Hseino, H., … Neary,

P. C. (2018). Visual spatial ability for surgical trainees: implications for learning endoscopic, laparoscopic surgery and other image-guided procedures. *Surgical Endoscopy*, *32*(8), 3634–3639. https://doi.org/10.1007/s00464-018-6094-3

Keehner, M. (2011). Spatial cognition through the keyhole: how studying a real-world domain can inform basic science--and vice versa. *Topics in Cognitive Science*, *3*(4), 632–647. https://doi.org/10.1111/j.1756-8765.2011.01154.x

Kramp, K. H., van Det, M. J., Hoff, C., Veeger, N. J. G. M., ten Cate Hoedemaker, H. O., & Pierie, J. P. E. N. (2016). The predictive value of aptitude assessment in laparoscopic surgery: A meta-analysis. *Medical Education*, *50*(4), 409–427. https://doi.org/10.1111/medu.12945

Kundhal, P. S., & Grantcharov, T. P. (2009). Psychomotor performance measured in a virtual environment correlates with technical skills in the operating room. *Surgical Endoscopy*, *23*(3), 645–649. https://doi.org/10.1007/s00464-008-0043-5

Louridas, M., Szasz, P., de Montbrun, S., Harris, K. a, & Grantcharov, T. P. (2016). Can We Predict Technical Aptitude? A Systematic Review. *Annals of Surgery*, *263*(4), 673–691. https://doi.org/10.1097/SLA.0000000000001283

Louridas, M., Szasz, P., Fecso, A. B., Zywiel, M. G., Lak, P., Bener, A. B., … Grantcharov, T. P. (2017). Practice does not always make perfect: need for selection curricula in modern surgical training. *Surgical Endoscopy*. https://doi.org/10.1007/s00464-017-5572-3

Luursema, J.-M., Rovers, M. M., Groenier, M., & van Goor, H. (2014). Performance Variables and Professional Experience in Simulated Laparoscopy: A Two-Group Learning Curve Study. *Journal of Surgical Education*, 1–6. https://doi.org/10.1016/j.jsurg.2013.12.005

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, *7*(2), 107–127. https://doi.org/10.1016/0160-2896(83)90023-5

Palter, V. N., & Grantcharov, T. P. (2014). Individualized Deliberate Practice on a Virtual Reality Simulator Improves Technical Performance of Surgical Novices in the Operating Room. *Annals of Surgery*, *259*(3), 443–448.

https://doi.org/10.1097/SLA.0000000000000254

Schmettow, M., Bennink, L., & Groenier, M. (2017). *Learning Laparscopy In A Simulator*. University of Twente.

Schmettow, M., Kaschub, V. L., & Groenier, M. (2016). *Learning complex motor procedures : can the ability to learn dexterity games predict a person's ability to learn a complex task?* University of Twente.

Seymour, N. E., Gallagher, A. G., Roman, S. a, O'Brien, M. K., Bansal, V. K., Andersen, D. K., & Satava, R. M. (2002). Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of Surgery*, *236*(4), 458-63; discussion 463-4. https://doi.org/10.1097/01.SLA.0000028969.51489.B4

Stefanidis, D., Korndorffer, J. R., Black, F. W., Dunne, J. B., Sierra, R., Touchard, C. L., … Scott, D. J. (2006). Psychomotor testing predicts rate of skill acquisition for proficiency-based laparoscopic skills training. *Surgery*, *140*(2), 252–262. https://doi.org/10.1016/j.surg.2006.04.002

The Southern Surgeons Club, Moore, M. J., & Bennett, C. L. (1995). The learning curve for laparoscopic cholecystectomy. *The American Journal of Surgery*, *170*(1), 55–59. https://doi.org/10.1016/S0002-9610(99)80252-9

Tinelli, A., Malvasi, A., Gustapane, S., De Nunzio, G., DeMitri, I., Bochicchio, M., … Tsin, D. A. (2012). The utilization of novel technology in risk reducing laparoscopic gynecological complications. In *Laparoscopy: New Developments, Procedures and Risks* (pp. 71–90). Nova Science Publishers, Inc.

van der Meijden, O. a J., & Schijven, M. P. (2009). The value of haptic feedback in conventional and robot-assisted minimal invasive surgery and virtual reality training: a current review. *Surgical Endoscopy*, *23*(6), 1180–1190. https://doi.org/10.1007/s00464-008-0298-x

van Dongen, K. W., Tournoij, E., van der Zee, D. C., Schijven, M. P., & Broeders, I. a M. J. (2007). Construct validity of the LapSim: can the LapSim virtual reality simulator distinguish between novices and experts? *Surgical Endoscopy*, *21*(8), 1413–1417. https://doi.org/10.1007/s00464-006-9188-2

Voitk, A. J., Tsao, S. G. ., & Ignatius, S. (2001). The tail of the learning curve for laparoscopic cholecystectomy. *The American Journal of Surgery*, *182*(3), 250–253. https://doi.org/10.1016/S0002-9610(01)00699-7

Wanzel, K. R., Ward, M., & Reznick, R. K. (2002). Teaching the surgical craft: From selection to certification. *Current Problems in Surgery*, *39*(6), 583–659. https://doi.org/10.1067/mog.2002.123481

Appendices

Appendix A – Informed Consent

CONSENT BY SUBJECT FOR PARTICIPATION IN RESEARCH PROTOCOL

Section A

Protocol Number: _______________ Participant Name: _______________

Participant Number: _____________

Title of Protocol: Learning laparoscopy in the LapSim

Doctor(s) Directing Research: Dr. Martin Schmettow, Dr. Marleen Groenier

Undergraduate Students Conducting Experiments: Alexander Arendt, BSc, Leslie Bennink, and Luuk Lenders

Phone: +49-1578-1580623 (Alexander Arendt)

You are being asked to participate in a research study. The researchers at the University of Twente want to explore the correlations between cognitive attributes and basic tasks and procedures in the LapSim simulator for laparoscopy. In order to decide whether or not you want to be a part of this research study, you should understand enough about its risks and benefits to make an informed judgment. This process is known as informed consent. This consent form gives detailed information about the research study, which will be discussed with you. Once you understand the study, you will be asked to sign this form if you wish to participate.

Section B

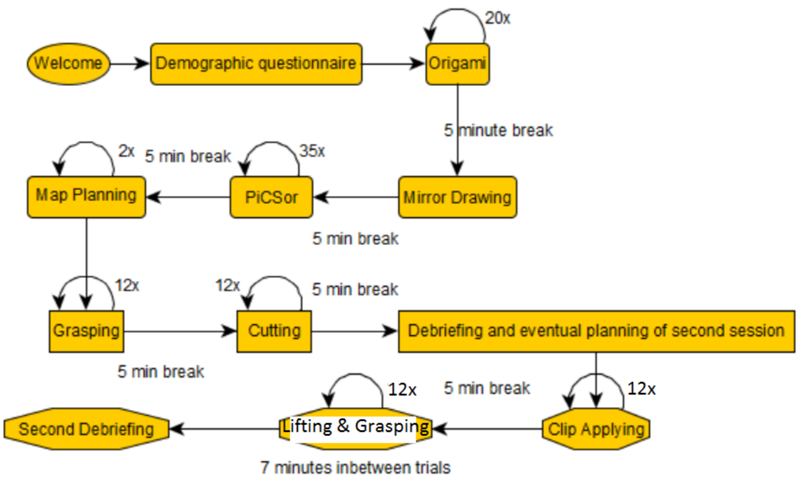
1. NATURE AND DURATION OF PROCEDURE(S):

The purpose of the current study is to examine the relationship between cognitive attributes and performance of basic laparoscopic tasks on a simulator. You can participate if you are: 1) not physically disabled, and 2) over 18 years of age, and 3) unexperienced with laparoscopy and/or the LapSim laparoscopy simulator.

The study consists of two parts across two non-consecutive days:

1. A set of cognitive attributes tests (Origami, Mirror Drawing, PicSOr, Map Planning) and two LapSim tasks (Grasping & Cutting), which will take no more than three hours.

2. An option to conduct twelve clip applying tasks and twelve lifting & grasping tasks on the LapSim. This, too, will not take longer than three hours.



Timeline of the study

On the first day you will complete the four cognitive attributes tests and the two basic simulator tasks. This will take no more than three hours. You will also be asked to fill in a standard demographics information form which will take no more than two minutes.

Timeline of the study.

Cognitive attributes test

The cognitive attributes test consists of several validated, psychological tests measuring visuo-spatial ability, rapid route-planning, mental rotation and manual dexterity. You will have to fold twenty simple Origamis, complete twenty Mirror Drawing patterns, do the PicSOr task and do the Map Planning Task.

Laparoscopic tasks

Two basic laparoscopic tasks are practiced on the simulator. You will repeat each exercise twelve times and then move on to the next. There is a time limit on the objectives for each task. Between each different exercise you will receive a break of five minutes. Should you also decide to take part in the second part you will get five minutes of break between and repeat the clip applying task and lifting & grasping task twelve times.

II. POTENTIAL RISKS AND BENEFITS:

The results from this study will help us to better design our training programs for surgical education on laparoscopy and will be used for one Master's thesis and two Bachelor's theses. A possible risk is fatigue during the study. We have planned many breaks between sessions to avoid fatigue. Also, you may quit the study at any moment without providing any reasons for your withdrawal from the study.

III. POSSIBLE ALTERNATIVES:

Your participation is voluntary. You may choose not to participate.

Section C AGREEMENT TO CONSENT

The research project and procedures associated with it have been fully explained to me. All experimental procedures have been identified and no guarantee has been given about the possible results. I have had the opportunity to ask questions concerning any, and all aspects of the project and any procedures involved. I am aware that participation is voluntary and that I may withdraw my consent at any time. I am aware that my decision not to participate or to withdraw will not restrict my access to health care services normally available to me. Confidentiality of records concerning my involvement in this project will be maintained in an appropriate manner. When required by law, government agencies and sponsors of the research may review the records of this research.

I understand that the sponsors and investigators have such insurance as is required by law in the event of injury resulting from this research.

I, the undersigned, hereby consent to participate as a subject in the above described project conducted at the ECTM on the campus of the University of Twente. I have received a copy of this consent form for my records. I understand that if I have any questions concerning this research, I can contact the doctor(s) and researcher(s) listed above. If I have further queries concerning my rights in connection with this research, I can contact the University of Twente Ethical Committee for Behavioural, Management and Social Sciences (BMS). Responsible for Cognitive Psychology and Ergonomics Research Dr. Rob H. J. van der Lubbe (r.h.j.vanderlubbe@utwente.nl).

After reading the entire consent form, if you have no further questions about giving consent, please sign where indicated.

Researcher:______________________ ______________________________

Signature of participant

Date: _____________ Time:__________AM PM

Appendix B – Norm scores for Map Planning

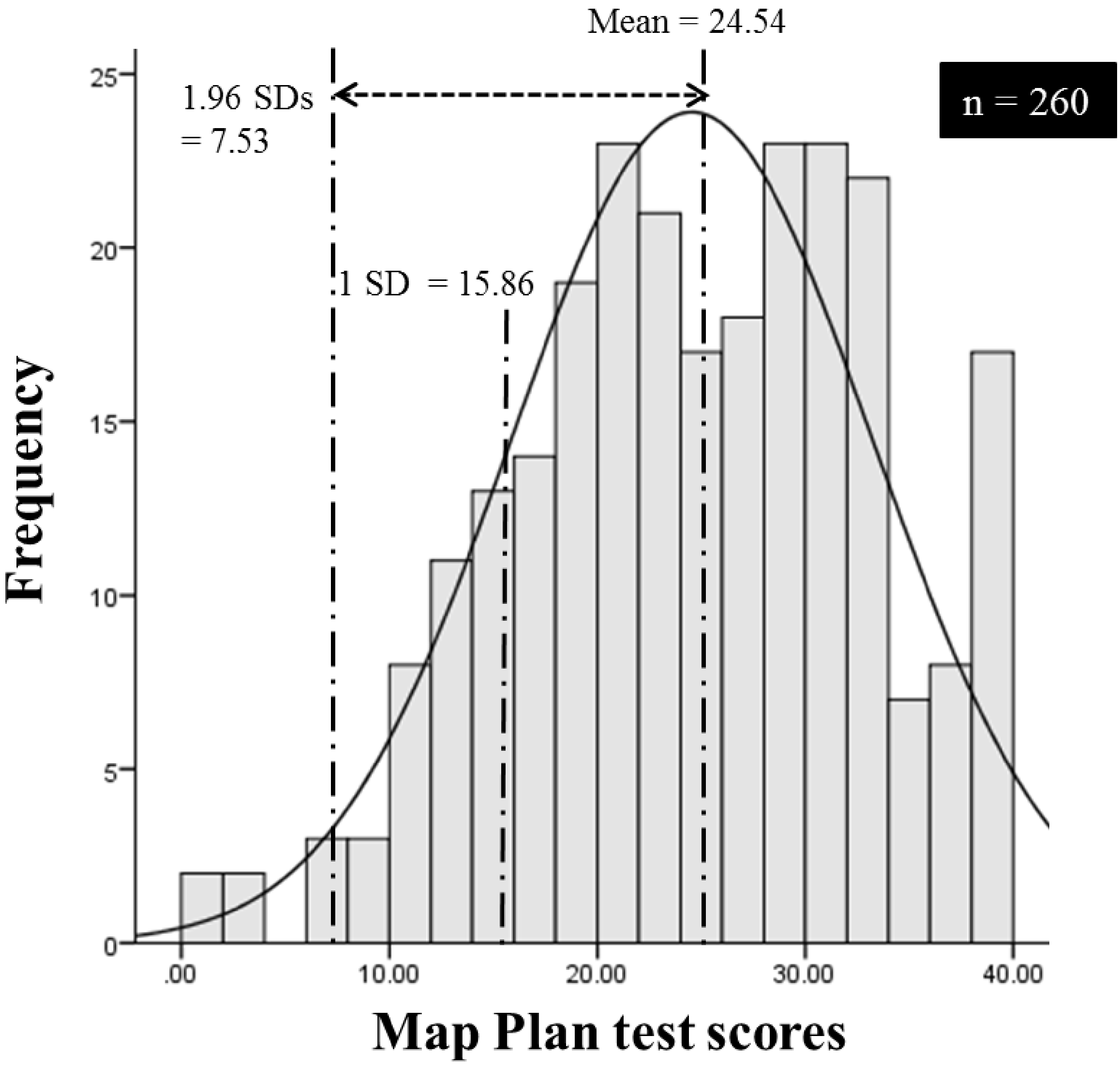


Figure 11 - Norm scores of the Map Planning test, taken from Henn et al., 2018

Appendix C – Questionnaire questions

- Please enter your participant number.
- What is your gender? Male / Female.
- Please enter your date of birth.
- Please enter your nationality.
- Are you left- or right-handed? Left-handed / Right-handed
- Do you have impaired vision? Yes, I wear glasses / Yes, I wear contacts / Yes, but I do not wear glasses nor contacts / No / Other (please explain)
- Are you colour-blind? Yes / No
- Are you dyslectic? Dyslexia is a condition which impairs the ability to read and understand written text fluently. Yes / No / I don’t know
- Are you experienced with playing video- or computer-games? Yes / No
    - If yes, how much time in hours do you spend in a week on average?
- Did you ever partake in a cognitive ability test? Yes / No / I don’t know

Appendix D – Verbatim Instructions

This study consists of four dexterity and two simulator tasks as I have just explained to you. You will receive instructions on the specific tasks before you start the exercise. First, you will have to fill out an informed consent form. [*Give informed consent form and make sure that participant signs it. Write down participant number on your form and add a new one to the Excel file.]* Next is a demographics questionnaire asking about some personal information, such as your age and handedness. This is online-based and takes about two minutes to complete. After that, you will have to do 20 Origamis, fill out a pattern in a mirror 20 times, complete a PiCSor and a Map Planning Task and then do two exercises on the simulator. You can ask questions any time. Do you have any questions thus far? [*Participant starts with the demographics questionnaire.]*

At the beginning of the study the following instructions were given:

The Mirror Drawing task:

Fill in the pattern with the pen while only observing your hand and the sheet in the mirror.

The PicSOr Test:

This is the PicSOr test. The PicSOr test measures perceptual ability. This test measures your ability to assess depth in a 3-dimensional picture. On each exercise, you will see a cube tilted at a certain angle. The point of a spinning arrowhead is touching the surface of the cube. You adjust the arrow until its shaft is perpendicular to the cube's surface at the point where they touch. The actual angle of the tilted cube is compared to your estimated angle as indicated by the positioning of the arrowhead's shaft.

You can use the 'up' and 'down' arrow keys on your keyboard to adjust the shaft of the spinning arrow. You can press 'Enter' as soon as you have positioned the arrowhead. You can click on 'Next' to continue to the next exercise. It is important that you work as fast and accurate as possible.

First, there are four practice exercises with feedback.

Then, at the beginning of the real experiment the following instructions were given:

The actual test consists of 35 exercises. Work as fast and as accurate as possible! There is no time limit for this test, but speed is important.

During the LapSim procedure the following instructions were given

You are about to start practicing two basic laparoscopic tasks which are part of a procedure called cholecystectomy, grasping and cutting. You can read the instructions for each exercise and view videos of performance of these tasks during an actual procedure as well as in the virtual environment. You cannot alternate between the two exercises of grasping and cutting. You must repeat each task twelve times. You will start with the grasping task and then continue to the cutting. There will be a time limit in which you have to complete the objectives of the grasping task. Please try to be as accurate and quick at the same time as possible. Do not falter just because you are getting low scores – this is a very difficult task which professionals train years for, and your actual performance does not matter as much as the progress, or the absence thereof, that we can observe. Only make sure that you do not hurt your patient, which is indicated by the screen flashing in red.

Per task specific instructions were given as follows:

Grasping

10 objects have to be picked up. 5 with the left and 5 with the right instrument. There is a time limit on the time-on-task, if the participant is too slow, they will automatically continue to the next object.

Cutting

3 segments of a vessel have to be cut off by carefully grasping the loose end of the vessel and then cutting off the grasped segment by moving the ultrasonic scissors over the vessel, pressing and then pushing the pedal to apply heat. There is no time limit on this task, but too severe stretch damage leads to a cancellation of the exercise.

Clip-applying

A vessel must be grasped and then the clip applier should be moved over it. By pressing the handle, a clip is applied. After both clips have been applied at the assigned positions, the participant has to cut through the vessel with the ultrasonic scissors.

Lifting & Grasping

A vessel must be lifted and an object grasped and removed from under it. For the lifting, a probe tool is employed and for grasping the standard grasper. The hand in which the probe, or respectively the grasper, are, alternates between objects.

Appendix E – Observation Form Origami/Mirror Drawing

Participant	Task	Trial	Time	Errors	Comments
x	Origami/Mirror Drawing				
		1 up to 20			
			Time in Min:Sec:MiliSec		
				Did participant make mistake?	
					Any comments from Researcher can be put in this column
					Used for writing observation so researchers can compare their assessment

Appendix F – R Syntax

```
---
title: "Learning MIS"
author: "Martin Schmettow"
date: "May 23, 2017"
---

library(devtools)
install_github("schmettow/mascutils")
install_github("schmettow/bayr")

library(tidyverse)
library(pryr)
library(brms)
library(mascutils)
library(bayr)
library(knitr)

purp.gather = F
purp.mcmc = F

## default parameters for MCMC
formals(brm)["chains"] <- 3
formals(brm)["iter"]   <- 3000
formals(brm)["warmup"] <- 1000

rstan::rstan_options(auto_write = T)
options(mc.cores = 3)

if (!purp.gather) load("Learning_MIS.Rda")
if (!purp.mcmc) {
  load("M_1.Rda")
  load("M_2.Rda")
  load("M_3.Rda")
  load("M_4.Rda")
  load("M_5.Rda")
  load("M_6.Rda")
  load("M_7.Rda")
}

opts_knit$set(cache = TRUE)
```

```r
## Plotting function

## MCMC chain plot
ggcaterp <-
  function(model, ...){
    filter_crit <- list(...)
    if("tbl_post" %in% class(model) == F) model = posterior(model)
    model %>%
      filter_(.dots = filter_crit)
    ggplot(aes(x = iter, y = value, col = chain)) +
      facet_wrap("parameter", scale = "free", ncol = 1) +
      geom_line()
  }

## posterior prediction plot of learning curve
gg_post_pred <-
  function(model,  ..., data = model$data, se = F) {
    filter_crit <- list(...)

    out <-
      data %>%
      bind_cols(predict(post_pred(model))) %>%
      filter_(.dots = filter_crit) %>%
      ggplot(aes(x = trial, y = ToT, group = Part, col = Task)) +
      geom_point() +
      geom_line(aes(y = center, group = Task))
    if(se){
      out <- out +
        geom_ribbon(aes(ymin = lower, ymax = upper), alpha = .1, col = 0) +
        scale_color_discrete(name=NULL)
    }
    out
  }

## extracts the posterior of correlatiuon parameters
## as this is not yet implemented in bayr
post_cor <-
  function(model){
    out <-
      brms::posterior_samples(model) %>%
```

```
      as.data.frame() %>%
      as_data_frame() %>%
      select(starts_with("cor")) %>%
      mutate(iter = row_number()) %>%
      gather(parameter, value, -iter) %>%
      separate(parameter, into = c("type", "re_factor",
                                   "nonlin_1", "level_1", "nonlin_2",
"level_2"), sep = "_+")
    class(out) <- append("tbl_post_cor", class(out))
    out
  }

cor.tbl_post_cor <-
  function(tbl_post_cor){
    out <-
      tbl_post_cor %>%
      group_by(nonlin_1, level_1, nonlin_2, level_2) %>%
      summarise(center = median(value),
                lower = quantile(value, .025),
                upper = quantile(value, .975)) %>%
      ungroup()
    out
  }

cor_cross <-
  function(tbl_post_cor, nonlin = "etamaxp"){
    out <-
      cor.tbl_post_cor(tbl_post_cor) %>%
      filter_(~nonlin_1 == nonlin) %>%
      select(level_1, level_2, center) %>%
      spread(level_2, center)
    out
  }

# Data preparation

## As advised by Groenier, Groenier, Miedema, & Broeders (2015), z-scores
have been calculated for damage and motion efficiency. Damage consists of
added z-scores of number of damages and maximum depth of damage divided by
two. Both left and right efficiency consist of added z-scores of instrument
angular path and instrument path length divided by two. ##
```

```r
collect_lapsim <-
  function(file = "Lapsim/20170324 lapsim data Alexander Arendt.xlsx",
           sheet = "Grasping"){
    readxl::read_excel(file,
                       sheet = sheet) %>%
      rename(Part = Login,
             Task = `Task Name`) %>%
      select(-Firstname:-`Course ID`, -`Task ID`:-`Start Time`) %>%
      group_by(Part, Task) %>%
      mutate(trial = row_number()) %>%
      ungroup() %>%
      gather(key = "Parameter",
             value = "value", -Part, -Task, -trial)
  }

collect_picsor <-
  function(file = "PicSOr/picsor.csv"){
    read.csv(file, header = T, fileEncoding="UTF-8-BOM", row.names = NULL)
%>%
      group_by(Part) %>%
      summarize(Corr=cor(RSSLA, OSSLA))
  }

collect_mapplanning <-
  function(file = "Map Planning/mapplanning_scores.xlsx"){
    readxl::read_excel(file) %>%
      group_by(Part) %>%
      summarize(RRP=mean(percentage_correct))
  }

collect_demographics <-
  function(file = "Demographics/demographics.xlsx"){
    readxl::read_excel(file)}

standardize <-
  function(var1, var2){
    ((((var1 - mean(var1, na.rm=T)) / sd(var1, na.rm=T)) + ((var2 -
mean(var2, na.rm=T)) / sd(var2, na.rm=T))) / 2)
  }
```

```
## There is no max_damage in "Clip Applying" and also no other value on the
same scale. Only damage-indices are bloodloss (l) and stretch damage (0 -
100%, whereby 100% is task essentially failed). Perhaps I missed an option
to include it while extracting the data from the LapSim because it is
possible to inflict the damages that are measured in the other tasks. ##

## "Total Time (s)" has to be calculated almost manually in Excel before
reading the file into R. Simply go to sheet "Grasping" and select column
AB, row 2. Type in "=N2+R2" (without the "") and drag the frame on its
lower-right corner to the lowest row of column AB. When you release the
mouse, the function "=Nx+Rx" (where x is row number) will be applied to
every cell you selected with corresponding values for x. Subsequently, name
the new column "Total Time (s)" ##

## To be able to catch instrument paths of "Cutting", rename "Cutter
Angular Path" and "Cutter Path Length" to "Left Instrument Angular Path /
Path Length" and "Grasper Angular Path" as well as "Grasper Path Length" to
"Right Instrument Angular Path / Path Length".  ##

D_Lapsim <-
  bind_rows(
    collect_lapsim("Lapsim/20170419 lapsim data Alexander Arendt.xlsx",
"Grasping"),
    collect_lapsim("Lapsim/20170419 lapsim data Alexander Arendt.xlsx",
"Cutting"),
    collect_lapsim("Lapsim/20170419 lapsim data Alexander Arendt.xlsx",
"Clip Applying"),
    collect_lapsim("Lapsim/20170419 lapsim data Alexander Arendt.xlsx",
"Lifting & Grasping")
  ) %>%
  mascutils::go_first(~Task, ~Part, ~trial) %>%
  filter(Parameter %in% c("Score", "Status", "Total Time (s)",
                          "Tissue Damage (#)", "Maximum Damage (mm)",
                          "Left Instrument Angular Path (degrees)",
                          "Right Instrument Angular Path (degrees)",
                          "Left Instrument Path Length (m)",
                          "Right Instrument Path Length (m)")) %>%
  spread(key = Parameter, value = value) %>%
  select(Task, Part, trial,
         score = Score, status = Status,
         ToT = `Total Time (s)`,
```

```
        damage = `Tissue Damage (#)`,
        max_damage = `Maximum Damage (mm)`,
        left_ang = `Left Instrument Angular Path (degrees)`,
        right_ang = `Right Instrument Angular Path (degrees)`,
        left_len = `Left Instrument Path Length (m)`,
        right_len = `Right Instrument Path Length (m)`) %>%
  mutate(damage = as.numeric(damage),
        ToT = as.numeric(ToT),
        max_damage = as.numeric(max_damage),
        left_ang = as.numeric(left_ang),
        right_ang = as.numeric(right_ang),
        left_len = as.numeric(left_len),
        right_len = as.numeric(right_len),
        damage = standardize(damage, max_damage),
        left_effic = standardize(left_len, left_ang),
        right_effic = standardize(right_len, right_ang)) %>%
  select(-max_damage, -left_len, -left_ang, -right_len, -right_ang) %>%
  filter(!grepl("Alexander93", Part),
        !grepl("cogpp1$", Part),
        !grepl("coggpp9$", Part)) %>%
  mutate(Part = stringr::str_replace(Part, "cogpp", ""),
        Task = stringr::str_replace(Task, "\\[", ""),
        Task = stringr::str_replace(Task, "\\]", ""),
        ToT = ToT/60) %>%
  mutate(ToT = ifelse(ToT == 0, NA, ToT)) %>%
  filter(!is.na(Task),
        !as.integer(Part) == 1,
        !as.integer(Part) == 9,
        !(as.integer(Part) == 8 & Task == "ClipApplying"))

D_Dext <-
  readxl::read_excel("Origami & Mirror Drawing/observations_total.xlsx")
%>%
  select(Task, Part = Participant, trial = Trial, ToT = Time) %>%
  mutate(Part = as.character(Part)) %>%
  mutate(ToT = (as.numeric(ToT) + 2209161600)/60) %>%
  mutate(ToT = ifelse(as.integer(Part) %in% c(14:17, 19, 25, 35:37), ##
wrong input in Excel
                      ToT/60, ToT)) %>%
  filter(!is.na(Task),
        !as.integer(Part) == 1,
```

```
         !as.integer(Part) == 9)

# D_Dext %>%
#   mutate(ToT = ToT/60) %>%
#   group_by(Part) %>%
#   summarize(mean(ToT))

D_Picsor <-
  collect_picsor(file="PicSOr/picsor.csv") %>%
  filter(!as.integer(Part) == 1,
         !as.integer(Part) == 9) %>%
  rename(picsor = Corr) %>%
  mutate(Part = as.character(Part))

D_Planning <-
  collect_mapplanning(file="Map Planning/mapplanning_scores.xlsx") %>%
  filter(!as.integer(Part) == 1,
         !as.integer(Part) == 9) %>%
  rename(planning = RRP) %>%
  mutate(Part = as.character(Part))

D_IAT <-
  full_join(D_Planning, D_Picsor) %>%
  mascutils::z_score(planning:picsor)


D_Demog <-
  collect_demographics(file="Demographics/demographics.xlsx")

save(D_Lapsim, D_Dext, D_Picsor, D_Planning, D_IAT, D_Demog, file =
"Learning_MIS.Rda")

check_n <-
  function(data) data %>%
  group_by(Part, Task) %>%
  summarize(n_Obs = n()) %>%
  ungroup() %>%
  spread(Task, n_Obs)

D_Lapsim %>% check_n()
D_Dext %>% check_n()
```

```
D_Picsor
D_Planning
D_Demog
```

```
# Data exploration

## Dexterity

```{r}
D_Dext %>%
  ggplot(aes(x = trial, y = ToT, group = Part)) +
  facet_grid(Task~., scales = "free_y") +
  geom_smooth(se = F)

D_Dext %>%
  filter(trial == 20) %>%
  spread(Task, ToT) %>%
  ggplot(aes(x = `Mirror Drawing`, y = Origami)) +
 geom_point() +
  geom_smooth(method = "lm", se = F)

## approximating the maxp cor between Origami and Mirror
## on the log scale to mimick the link function

T_Dext_cor <- D_Dext %>%
  filter(trial > 15) %>%
  group_by(Part, Task) %>%
  summarize(ToT = log(mean(ToT))) %>%
  ungroup() %>%
  spread(Task, ToT) %>%
  select(Part, `Mirror Drawing`, Origami)

T_Dext_cor %>% select(2:3) %>% cor()


T_Dext_cor %>%
  #  filter(Part != "29") %>%
  ggplot(aes(x = `Mirror Drawing`,
             y = Origami)) +
  geom_point() +
```
```

```
  geom_smooth(method = "lm")


## Lapsim


D_Lapsim %>%
  ggplot(aes(x = trial, y = ToT, color = Task)) +
  geom_point() +
  geom_smooth(se = F) +
  facet_wrap(~Part)

ggsave("lapsim_tot_eda.png", width = 10, height = 10)

D_Lapsim %>%
  ggplot(aes(x = trial, y = left_effic, color = Task)) +
  geom_point() +
  geom_smooth(se = F) +
  facet_wrap(~Part)

ggsave("lapsim_lefteffic_eda.png", width = 10, height = 10)

D_Lapsim %>%
  ggplot(aes(x = trial, y = right_effic, color = Task)) +
  geom_point() +
  geom_smooth(se = F) +
  facet_wrap(~Part)

ggsave("lapsim_righteffic_eda.png", width = 10, height = 10)

D_Lapsim %>%
  ggplot(aes(x = trial, y = damage, color = Task)) +
  geom_point() +
  geom_smooth(se = F) +
  facet_wrap(~Part)

ggsave("lapsim_damage_eda.png", width = 10, height = 10)



## PicSOr
```

```
D_Picsor %>%
  ggplot(aes(x = Corr)) +
  geom_histogram()
ggsave("picosr_histo.png", width = 5, height = 5)

## Map Planning

D_Planning %>%
  ggplot(aes(x = RRP)) +
  geom_histogram()
ggsave("mapplanning_histo.png", width = 5, height = 5)

# Regression


# The only performance variable available for all tasks is time. We build
the learning curve model stepwise (for testing and instructional purposes):
# 1. one participant and one task
# 2. all participants and one task
#.3. all participants and all task, including correlations between random
effects. This model should suffice to answer all questions regarding
reliability and validity
# 4. adding the IAT scores as predictors

## Basic non-linear model

#We use the Origami task of part 1 to build the basic learning curve model.
Parameters maxp and rate are used through link functions (log, logit) that
establish linearity. That is needed to build proper random effects later
on.
#The following formulas specify the non-linear likelihood and vague priors
for learning parameters. For the random component the Gamma distribution is
favored over the Gaussian for (ie.e. Normal) two reasons:
# + time errors are typically left skewed
# + Normal errors would have the same variance across the whole process,
whereas Gamma has stronger error variance with large ToT, i.e. in early
trials. This is just way more realistic.

## formula
F_nl_1 <- formula(ToT ~ exp(etamaxp) + exp(-inv_logit(etarate) * (trial +
```

```
pexp)))

## fixed efefcts priors priors
F_pr_1 <-
  c(set_prior("normal (0, 3)", nlpar = "etamaxp"),   # vague
    set_prior("normal (0, 25)", nlpar = "pexp"),  # vague
    set_prior("normal (0, 5)", nlpar = "etarate"))   # vague


F_re_1 <- formula(etamaxp + etarate + pexp ~ 1)

D_1 <- D_Dext %>%
  filter(Part == 5, Task == "Origami")


M_1 <- brm(bf(F_nl_1, F_re_1, nl = T),
           prior = F_pr_1,
           family = Gamma(link = "identity"),
           data = D_1,
           control = list(adapt_delta = 0.999),
           iter = 2000, warmup = 1000,
           seed = 42)

save(M_1, file = "M_1.Rda")

fixef(M_1)

# Posterior checks:

posterior(M_1) %>%
  ggcaterp()

# Estimates

fixef(M_1)

# Posterior prediction

gg_post_pred(M_1, data = D_1)

## multiple participants model
```

```
F_re_2 <- formula(etamaxp  + etarate + pexp ~ (1|corr1|Part))

D_2 <- D_Dext %>%
  filter(Task == "Origami") %>%
  filter(!is.na(ToT))


M_2 <- brm(bf(F_nl_1, F_re_2, nl = T),
           prior = F_pr_1,
           family = Gamma(link = "identity"),
           data = D_2,
           control = list(adapt_delta = 0.999),
           iter = 6000, warmup = 3000,
           seed = 42)

M_2
save(M_2, file = "M_2.Rda")
# Posterior prediction

ggcaterp(M_2)

gg_post_pred(M_2, data = D_2) +
  facet_wrap(~Part)

## multiple participants model, both dexterity tasks

F_re_3 <- formula(etamaxp  + etarate + pexp ~ (0 + Task|corr1|Part))

D_3 <- D_Dext %>% filter(!is.na(ToT))

M_3 <- brm(bf(F_nl_1, F_re_3, nl = T),
           prior = F_pr_1,
           family = Gamma(link = "identity"),
           data = D_3,
           control = list(adapt_delta = 0.999),
           iter = 4000, warmup = 2000,
           seed = 42)

M_3
save(M_3, file = "M_3.Rda")
```

```
# Posterior prediction

ggcaterp(M_3)

gg_post_pred(M_3) +
  facet_wrap(~Part)

P_3 <- posterior(M_3)
P_3

## multiple participants model, all tasks

# ** this model is flawed, do not interpret results**

 # + tasks are modelled as random effects, we might change that to get more
useful fixed effects
#+ cross-task correlations are included for all three learning parameters

D_4 <-
  select(D_Dext, Part, Task, trial, ToT) %>%
  bind_rows(select(D_Lapsim, Part, Task, trial, ToT)) %>%
  filter(!is.na(ToT))

F_re_4 <- list(formula(etamaxp ~ (0 + Task|corr2|Part)),
               formula(etarate ~ (0 + Task|corr3|Part)),
               formula(pexp ~ (0 + Task|corr4|Part)))

M_4 <- brm(bf(F_nl_1, flist = F_re_4, nl = T),
           prior = F_pr_1,
           family = Gamma(link = "identity"),
           data = D_4,
           control = list(adapt_delta = 0.999),
           iter = 3000, warmup = 1000,
           seed = 42)

M_4
save(M_4, file = "M_4.Rda")

# Posterior
```

```
P_4 <- posterior(M_4)
P_4

fixef(P_4)
grpef(P_4)

gg_post_pred(M_4, ~Task == "Origami") + facet_wrap(~Part, ncol = 8)
gg_post_pred(M_4, ~Task == "Mirror Drawing") + facet_wrap(~Part, ncol = 8)
gg_post_pred(M_4, ~Task == "[Clip Applying]") + facet_wrap(~Part, ncol = 8)
gg_post_pred(M_4, ~Task == "[Cutting]") + facet_wrap(~Part, ncol = 8)
gg_post_pred(M_4, ~Task == "[Grasping]") + facet_wrap(~Part, ncol = 8)
gg_post_pred(M_4, ~Task == "[Lifting & Grasping]") + facet_wrap(~Part, ncol
= 8)

P_4_cor <- post_cor(M_4) %>%
  mutate(level_1 = stringr::str_replace(level_1, "Task", ""),
         level_2 = stringr::str_replace(level_2, "Task", ""))

(T_4_cor <- cor.tbl_post_cor(P_4_cor))

T_4_cor %>%
  #filter(nonlin_1 == "etamaxp") %>%
  discard_redundant()

cor_cross(P_4_cor) %>% knitr::kable(digits = 2)

## Corrected all-participants, all-task model

# **M_4 is flawed**. It omits the fixed effect Task, which seems to produce
spurious corelations between random effects.  By introducing the Task
factor M_6, the correlation between Origami & Mirror collapses and other
correlations change, too. The spurious negative correlations almost
completely disappear.

D_6 <-
  select(D_Dext, Part, Task, trial, ToT) %>%
  bind_rows(select(D_Lapsim, Part, Task, trial, ToT)) %>%
  filter(!is.na(ToT))

F_re_6 <- list(formula(etamaxp ~ 0 + Task + (0 + Task|corr2|Part)),
               formula(etarate ~ 0 + Task + (0 + Task|corr3|Part)),
```

```
              formula(pexp    ~ 0 + Task + (0 + Task|corr4|Part)))

M_6 <- brm(bf(F_nl_1, flist = F_re_6, nl = T),
           prior = F_pr_1,
           family = Gamma(link = "identity"),
           data = D_6,
           control = list(adapt_delta = 0.999),
           iter = 3000, warmup = 1000,
           seed = 42)

M_6
save(M_6, file = "M_6.Rda")

# Posterior

P_6 <- posterior(M_6)
  mutate(fixef = stringr::str_replace(fixef, "Task", "")) %>%
  posterior()
P_6


fixef(P_6) %>%
  mascutils::discard_redundant() %>% kable(digits = 2)
grpef(P_6) %>%
  mascutils::discard_redundant() %>% kable(digits = 2)


P_6_cor <- post_cor(M_6) %>%
  mutate(level_1 = stringr::str_replace(level_1, "Task", ""),
         level_2 = stringr::str_replace(level_2, "Task", ""))

(T_6_cor <- cor.tbl_post_cor(P_6_cor))

T_6_cor %>%
  #filter(nonlin_1 == "etamaxp") %>%
  discard_redundant()

cor_cross(P_6_cor) %>% knitr::kable(digits = 2)

## All tasks and innate ability predictors on maxp
```

```
D_5 <-
  select(D_Dext, Part, Task, trial, ToT) %>%
  bind_rows(select(D_Lapsim, Part, Task, trial, ToT)) %>%
  left_join(D_IAT) %>%
  #filter(stringr::str_detect(Part, "2")) %>%
  filter(!is.na(ToT)) %>%
  filter(!is.na(zpicsor)) %>%
  filter(!is.na(zplanning))

#F_nl_1 <- formula(ToT ~ exp(etamaxp) + exp(-inv_logit(etarate) * (trial +
pexp)))

F_re_5 <- list(formula(etamaxp ~ 0 + Task + zplanning:Task + zpicsor:Task +
(0 + Task|corr2|Part)),
               formula(etarate ~ 0 + Task + (0 + Task|corr3|Part)),
               formula(pexp    ~ 0 + Task + (0 + Task|corr4|Part)))

M_5 <- brm(bf(F_nl_1, flist = F_re_5, nl = T),
           prior = F_pr_1,
           family = Gamma(link = "identity"),
           data = D_5,
           control = list(adapt_delta = 0.99),
           iter = 20, warmup = 10,
           seed = 42)

M_5 <- brm(fit = M_5,
           data = D_5,
           control = list(adapt_delta = 0.999),
           iter = 3000, warmup = 1000,
           chains = 3,
           seed = 42)



M_5
# save(M_5, file = "M_5.Rda")

# Posterior
```{r}
P_5 <-
  posterior(M_5) %>%
  mutate(fixef = stringr::str_replace(fixef, "Task", "")) %>%
```
```

```
  posterior()

fixef(P_5)
grpef(P_5)

fixef(P_5) %>%
  #filter(nonlin == "etamaxp") %>%
  mascutils::discard_redundant() %>%
  separate(fixef, c("Task", "IAT"), sep = ":") %>%
  filter(!is.na(IAT)) %>%
  arrange(Task) %>%
  ggplot(aes(x = center, xmin = lower, xmax = upper, y = Task)) +
  geom_errorbarh(height = 0) +
  geom_vline(xintercept = 0, col = "red", linetype = "dashed") +
  geom_point() +
  facet_grid(~IAT)

P_5_cor <- post_cor(M_5) %>%
  mutate(level_1 = stringr::str_replace(level_1, "Task", ""),
         level_2 = stringr::str_replace(level_2, "Task", ""))

T_5_cor <-
  cor.tbl_post_cor(P_5_cor) %>%
  discard_redundant()

T_5_cor %>% knitr::kable(digits = 2)

cor_cross(P_5_cor) %>% knitr::kable(digits = 2)

## Re-analysis of Kaschub

MS_15_2 <- new.env()
load("Data_analysis_motor_skills_15_2.Rda", envir = MS_15_2)

D_7 <-
  MS_15_2$D_$LCM %>%
  select(Part, Task, trial, ToT = time) %>%
  mutate(ToT = ToT/60) %>%
  mutate(Task = stringr::str_replace(Task, "_", "")) %>%
  filter(!is.na(ToT))
```

```
unique(D_7$Task)
#F_nl_1 <- formula(ToT ~ exp(etamaxp) + exp(-inv_logit(etarate) * (trial +
pexp)))

F_re_7 <- list(formula(etamaxp ~ 0 + Task + (0 + Task|corr2|Part)),
               formula(etarate ~ 0 + Task + (0 + Task|corr3|Part)),
               formula(pexp    ~ 0 + Task + (0 + Task|corr4|Part)))

M_7 <- brm(bf(F_nl_1, flist = F_re_7, nl = T),
           prior = F_pr_1,
           family = Gamma(link = "identity"),
           data = D_7,
           control = list(adapt_delta = 0.9),
           iter = 2000, warmup = 1000,
           chains = 3,
           seed = 42)

M_7
save(M_7, file = "M_7.Rda")

# Posterior

P_7 <- posterior(M_7)

P_7_cor <-
  post_cor(M_7) %>%
  mutate(level_1 = stringr::str_replace(level_1, "Task", ""),
         level_2 = stringr::str_replace(level_2, "Task", ""))

cor_cross(P_7_cor) %>% knitr::kable(digits = 2)

##Create graphs from the saved maxp_values


#Plot the six correlation graphs as one figure

z <- T_6_etamaxp_wide
a <- z %>%
  ggplot(aes(x = ClipApplying, y = Cutting)) +
  geom_point() +
  geom_smooth(method="lm") +
```

```
  coord_cartesian(xlim=c(-1,1),ylim=c(-1,1))
b <- z %>%
  ggplot(aes(x = ClipApplying, y = Grasping)) +
  geom_point() +
  geom_smooth(method ="lm") +
  coord_cartesian(xlim=c(-1,1),ylim=c(-1,1))
c <- z %>%
  ggplot(aes(x = ClipApplying, y = `Lifting&Grasping`)) +
  geom_point() +
  geom_smooth(method="lm") +
  coord_cartesian(xlim=c(-1,1),ylim=c(-1,1))
d<- z %>%
  ggplot(aes(x = Cutting, y = Grasping)) +
  geom_point() + geom_smooth(method="lm") +
  coord_cartesian(xlim=c(-1,1),ylim=c(-1,1))
e<-z %>%
  ggplot(aes(x = Cutting, y = `Lifting&Grasping`)) +
  geom_point() +
  geom_smooth(method="lm") +
  coord_cartesian(xlim=c(-1,1),ylim=c(-1,1))
f <-z %>%
  ggplot(aes(x = Grasping, y = `Lifting&Grasping`)) +
  geom_point() +
  geom_smooth(method="lm") +
  coord_cartesian(xlim=c(-1,1),ylim=c(-1,1))

ggsave("correlations_lapsim_graph.png", arrangeGrob(a,b,c,d,e,f,ncol = 2),
width = 10, height = 10)


#8 correlations for the dext to lapsim
g <-z %>%
  ggplot(aes(x = MirrorDrawing, y = `Lifting&Grasping`)) +
  geom_point() +
  geom_smooth(method="lm") +
  coord_cartesian(xlim=c(-1,1),ylim=c(-1,1))
h <-z %>%
  ggplot(aes(x = MirrorDrawing, y = Grasping)) +
  geom_point() +
  geom_smooth(method="lm") +
  coord_cartesian(xlim=c(-1,1),ylim=c(-1,1))
```

```
i <-z %>%
  ggplot(aes(x = MirrorDrawing, y = Cutting)) +
  geom_point() +
  geom_smooth(method="lm") +
  coord_cartesian(xlim=c(-1,1),ylim=c(-1,1))
j <-z %>%
  ggplot(aes(x = MirrorDrawing, y = ClipApplying)) +
  geom_point() +
  geom_smooth(method="lm") +
  coord_cartesian(xlim=c(-1,1),ylim=c(-1,1))
k <-z %>%
  ggplot(aes(x = Origami, y = `Lifting&Grasping`)) +
  geom_point() +
  geom_smooth(method="lm") +
  coord_cartesian(xlim=c(-1,1),ylim=c(-1,1))
l <-z %>%
  ggplot(aes(x = Origami, y = Grasping)) +
  geom_point() +
  geom_smooth(method="lm") +
  coord_cartesian(xlim=c(-1,1),ylim=c(-1,1))
m <-z %>%
  ggplot(aes(x = Origami, y = Cutting)) +
  geom_point() +
  geom_smooth(method="lm") +
  coord_cartesian(xlim=c(-1,1),ylim=c(-1,1))
n <-z %>%
  ggplot(aes(x = Origami, y = ClipApplying)) +
  geom_point() +
  geom_smooth(method="lm") +
  coord_cartesian(xlim=c(-1,1),ylim=c(-1,1))

ggsave("correlations_lapsim_dext_graph.png",
arrangeGrob(g,h,i,j,k,l,m,n,ncol = 2), width = 10, height = 10)


#Plot all pp per task as points, with average line in middle.
D_Lapsim %>%
  ggplot(aes(x = trial, y = ToT )) +
  geom_point() +
  geom_smooth(se = F) +
  facet_wrap(~Task)
```

```
ggsave("lapsim_tot_eda_task.png", width = 10, height = 10)

#Plot the graph of all curves per task, show that data fits the prior model
D_Lapsim %>%
  ggplot(aes(x = trial, y = ToT, group = Part)) +
  facet_grid(Task~., scales = "free_y") +
  coord_cartesian(xlim=c(1,11.8)) +
  geom_smooth(se = F) +
  labs(y="Time on Task")

ggsave("lapsim_tot_eda_allpp.png", width = 10, height = 10)
```

Appendix G – Participant exploration

Participant 1 did not finish the entire procedure and was therefore removed from the data. After this the model was ran and plots were made. During the data gathering notes were made about the performance of participants on all tasks. Only once did a participant show lack of motivation. This was confirmed after running the model the first time and visualizing this participant's graph. For this reason participant 9 was removed from further inclusion in the data analysis. The results of participant 9 are shown in Figure 12.

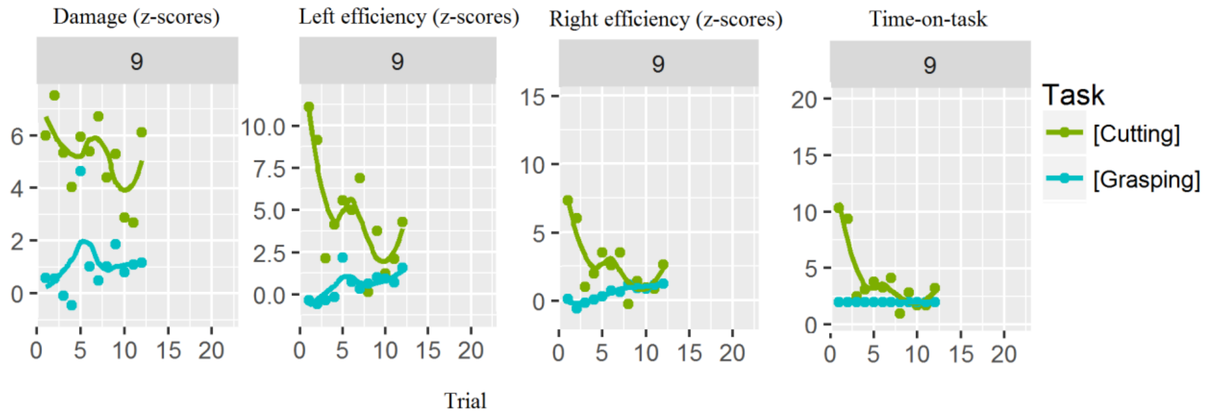


Figure 12 - Scores of partipant 9