Thore Schröder

28-10-2019

Master Thesis

The influence of time pressure on MIS simulation tasks

1st supervisor: Dr. Martin Schmettow

2nd supervisor: Dr. Marleen Groenier

Introduction

Minimal invasive surgery (MIS) has become more and more established over the last decades (Darzi & Munz, 2004). Advantages of MIS compared to traditional surgery are reduced blood loss and pain and a faster recovery rate which also leads to shorter stays in hospitals and less need for pain medication (Müller-Stich & Büchler, 2015; Ponsky, 1991). However, the skills that are needed to perform MIS differ from traditional surgery and require extensive training for surgeons. In recent years, virtual reality (VR) simulators have been established as a training method for MIS (Patel & Patel, 2012). However, systematic training and assessment methods for VR simulators have not been established, yet (Gardner et al., 2016). This study focusses on how to improve MIS training for surgeons on VR simulators.

During MIS procedures endoscopic tools and a camera are inserted into the body. Examples for MIS are *laparoscopy* in the abdomen or pelvis or *bronchoscopy*, where the endoscopic tools and camera are inserted through the nose or mouth. The camera displays the images on a 2d-screen while the surgeon performs the procedure. Compared to traditional surgery, this leads to some challenges for the surgeon: the absence of haptic feedback, a more difficult depth perception because the images are displayed in 2d instead of 3d and a more demanding hand-eye coordination since the surgeon does not see his hands while operating (Müller-Stich & Büchler, 2015, Perez-Cruet, Fessler & Perin,2002). MIS therefore asks for specific visual-spatial abilities and psychomotor skills of the surgeon (Kramp et al., 2016).

Extensive training is necessary to develop these skills. A study by Quellette (2006) showed higher complication rates in bronchoscopy among inexperienced surgeons compared to experienced surgeons. In the past, trainees would observe surgeons during their MIS and eventually perform supervised procedures themselves, followed by a subjective evaluation from the respective supervisors (Fielding, Maldonado & Murgu, 2014). This method however is suboptimal because of patient safety concerns. Training and assessment methods in safe environments that provide surgical ‘trainees with the skill set necessary without putting patients at risk are preferable (Fielding et al., 2014).

VR simulators fulfill this purpose. Simulator training provides a safe environment that does not put the life of patients at risk and trains the skills necessary for MIS procedures possible (Wanzel, Hamstra, Anastakis, Matsumoto, & Cusimano, 2002). Performance metrics like time necessary for the task and mistakes/wall contacts provide tailored feedback about the level of expertise of the individual and provide continuous feedback and motivation for learning and evaluating the training program (Epstein & Hundert, 2002; Patel & Patel, 2012). However, simulators are expensive and not always available for all trainees. The training durations vary due to individual differences in talent and training, which can cause scheduling problems (Sadideen, Hamaoui, Saadeddin, & Kneebone, 2012). Because of this, this study will focus on improving the efficiency of simulator training.

This study will use learning curves to assess the performance of participants on VR simulators over time and the effectiveness of a training intervention. Learning curves have been shown to be effective as prediction and assessment tools for surgical performance (Wanzel, Ward and Reznick, 2002, Pusic, M. V., Boutis, K., Pecaric, M. R., Savenkov, O., Beckstead, J. W., & Jaber, M. Y. 2017). Factors that influence the shape of the learning are the nature of the task, experience, manual dexterity and anatomical knowledge (White, Rodger & Tang, 2016). The longer learning curve for minimally invasive procedures compared to open surgical procedures is an example of this (White et al., 2016).

Other than one-time performance measurements, learning curves enable quantative measurements of an individual's learning process by monitoring the progress over repeated trials. This allows an estimation of when an individual will be proficient instead of determing if the individual is proficient or not after one assessment. Among novices, improvement rates like increased speed and fewer errors will be higher during the first trials but will slow down with more practice, which forms a learning curve (Heathcote, Brown & Mewhort, 2000). Eventually the individual will reach its maximum performance where it is not possible to be significantly faster or to do fewer mistakes due to physical limits and boundaries in performance. This is also known as the saturation effect: the more a task is practiced, the closer the individual gets to the natural boundaries. Improvement will be minimal at this point.

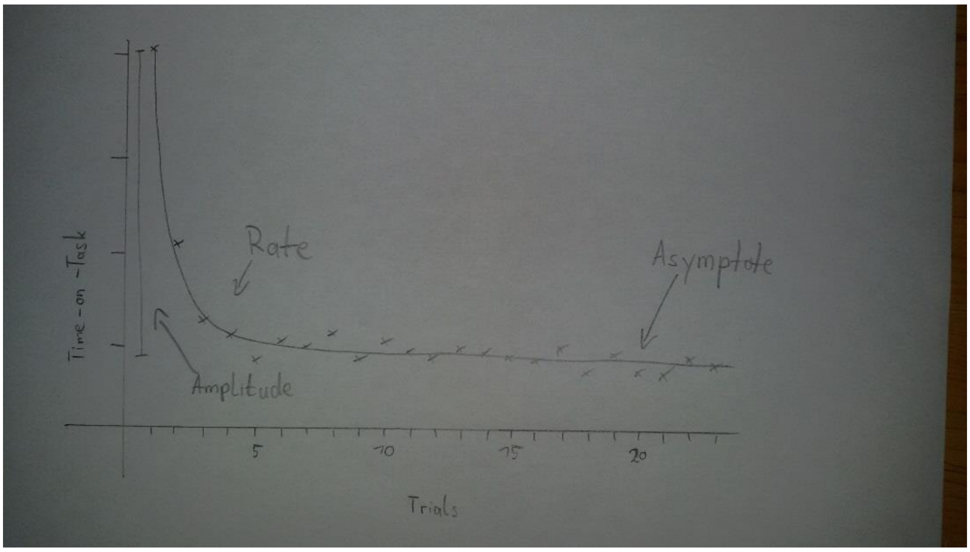

Figure 1 - learning curve with the three parameters rate, amplitude and asymptote. Trials on the x axis, time on task on the y axis.

A learning curve consists of three parameters that determine its shape: amplitude (δ, amount of learning), rate (ρ, speed of learning) and asymptote (ω, maximum performance). The amplitude shows the amount of improvement of the individual (i.e. the difference between initial performance and asymptote). The rate represents the speed of learning. The higher the rate, the faster the individual is learning and reaching its maximum performance. This maximum performance is referred to as asymptote which will be reached after some amount of practice. An example of a learning curve can be found in figure 1, where the progress on time-on-task is exemplified.

To analyze the learning curves, this study will focus on the asymptotes of participants as a measurement for performance since it projects how well somebody can perform laparoscopic tasks given continued training. It is not based on a single value, but on the continued observation of the learning process. It functions a predictor for the maximum performance, which also could be described as talent. This approach has also been used in earlier studies (Arendt, Schmettow & Groenier, 2017; Schmettow, Kaschub, Groenier, 2016).

If individuals reach their maximum performance regarding the time needed for a task but still try to be faster and push their limits, their performance will become prone to errors. In medicine, especially in the surgical field, practitioners are under constant pressure to optimize efficiency to make exhausting procedures as short as possible for patients and to act quickly in emergency situations. At the same time, surgeons have to be accurate and minimize the amount of errors to guarantee the safety of the patient (Gas, Buckarma, Cook, Farley & Pusic, 2018). The dilemma between being fast and accurate at the same time forms the speed-accuracy trade-off

(SAT) (Standage, Wang, Heitz& Simen, 2015). Speed-accuracy trade-offs describe the conflicting demands of being fast on the one hand and precise on the other. As people speed up, they make more mistakes, whereas when they focus on being accurate, they become slower. (MacKay, 1982; Soukoreff & MacKenzie, 2009). The present study will therefore not only analyze the variable time-on-task, but also the accuracy. The use of accuracy as a performance variable has been recommended earlier (van Dongen, 2007).

Most research on SAT has been done in non-medical domains. A comprehensive overview by Heitz (2014) shows the addition of time limits in tasks helps pinpointing the locus of the deficit among practitioners. However, according to some studies tasks using SAT methodologies seem to work better for practitioners with some level of experience compared to practitioners who are novices. A study by Beilock and colleagues (2004) showed that during a putting task, skilled golfers were even more accurate when they were prompted to putt faster than when they had no time pressure. Novices on the other hand showed the opposite pattern: their accuracy was better when they had more execution time compared to less execution time. It has been argued that this difference is rooted in the use of different cognitive processes in various stages of skill acquisition. Novices use their working memory whereas advanced practitioners also use automated processes and routines that do not require a lot of attention to the task (Beilock and colleagues, 2009). That would also explain why experts in other movement tasks like soccer dribbling or baseball batting can handle distracting conditions (e.g. secondary tasks that demand attention) better than novices (Beilock, S. L., Carr, T. H., MacMahon, C., & Starkes, J. L., 2002; Castaneda & Gray, 2007, Jackson, Ashford, & Norsworthy, 2006)

A study by Gas et al (2018) also found that adding time pressure during a simulated, traditional surgical task increased the challenge level leading to more errors which could help learners identify deficiencies in their skill development. In their study, they put participants from different skill levels under chronometric pressure: participants were asked to perform 20% faster than during their 5th repetition of the task. Novices (in this case: medical students) did not show a speed-accuracy tradeoff since they did not reach the maximum performance before. As their speed increased, their accuracy was maintained or improved. Intermediates (first-year residents) and experts (senior residents) showed the speed-accuracy tradeoff – increased speed was traded for decreased accuracy – since they had less potential for improvement than novices had.

The research on SAT shows potential for improving training programs with VR simulators. Other than in the traditional apprentice approach, training with VR simulators provides surgeons with the possibility to take risks and push them out of their comfort zones without risking the life of the patient. Assuming the SAT appears when trainees are put under time pressure, it is possible that trainees will learn from their mistakes and improve their skill.

Research question:

Due to the unclear influence of time pressure on learning in novices in MIS procedures, this study will investigate on this topic. Its goal is to find out if novices who experience time pressure during practice tasks will benefit from this experience afterwards. The research question is:

How does the implementation of time pressure trials influence the performance of novices in MIS practice tasks?

Answering this research question will bring clarity in the role of time pressure inducement into MIS simulator training. This could potentially increase the efficiency of MIS training and offer new instructional strategies to optimize the skill of trainees.

Method

2.1. Participants

40 participants (22 women and 18 men) who are students or former students from the University of Twente were recruited. The participants' nationalities were Dutch (42,5%), German (37,5%) or from other countries (20%). Some participants were recruited via SONA Systems, an online platform for undergraduate Psychology students, the rest were directly recruited on the Campus of the University of Twente or via Messengers (Facebook/ Whatsapp). All participants filled in an informed consent form.

2.2. Design

A within-subject design was used. Every participant received the same instructions. All participants performed a total of 70 trials on the VR simulator *LapSim*. The trials were split into the two basic laparoscopic tasks "cutting" and "lifting & grasping". The 35 trials per task were split into 3 phases: the initial phase (15 trials), the time pressure phase (10 trials) and the removed time pressure phase (10 trials). During the initial phase, the participants were asked to complete the tasks as fast and accurate as possible. Originally this phase was planned to consist of 10 trials as well, but after pilot testing of the design it became clear that 15 trials were necessary for participants to get used the LapSim mechanics and the tasks and also to compute a more accurate learning curve for the initial phase. With the 15th trial finished, the required time limit for the time pressure phase was calculated. Since the study by Gas et al (2018) was able to produce the desired SAT effect by instructing the participants to be 20% faster, we decided to do the same. During the time pressure phase, the participants were therefore asked to be 20% faster than during their 15th trial. For the third phase, the participants were instructed to be as fast and accurate as possible (same as in the initial phase). Additionally, all participants answered a one-item question on mental demand after every trial to introduce a cover story (see 2.4. for elaboration). Detailed instructions for the trials/phases can be found under Appendix 1.

2.3. Materials

Baseline questionnaire. The baseline questionnaire was created and filled in by all participants. It included questions for age, gender, occupation and nationality. It also asked for physical disabilities, especially for visual impairments since visual abilities are specifically relevant for this study. The baseline questionnaire can be found under Appendix 2

LapSim. The LapSim is a virtual reality simulator. It is used as a training and assessment tool for laparoscopic procedures. Three SimBall-modules enable the insertion of endoscopic tools, a desktop computer runs the simulation, which is displayed on a LC-display. The endoscopic tools resemble the ones used in the OR, except that their ends do not have actual forcipes. The modules register how far the tools are inserted, to what angle they are turned from their starting point and to what degree the user pushes the handles. With this simulation, it is possible to practice basic skills like grasping, cutting and clip applying or more complex procedure modules that simulate operations. A study by van Dongen et al. (2007) showed that its performance measures showed significant differences between experts and novices in MIS procedures, making it a suitable simulator for the assessment of technical MIS skills. See figure 1 for photos of the setup.

Figure 2 - Setup of the Lapsim: Three SimBall-modules, as well as the desktop computer and LC-display. The table can be adjusted in height to allow for a comfortable posture during task performance.

Arendt et al. (2018) tested the internal consistency of 4 LapSim tasks, namely grasping, cutting, clip applying and lifting & grasping. However, the internal consistency was mediocre at best. The highest internal consistency was found between lifting & grasping and cutting, which was also the only sufficiently certain one. We therefore used only these two tasks since they seemed to be the most promising test suites for laparoscopic skills.

Cutting

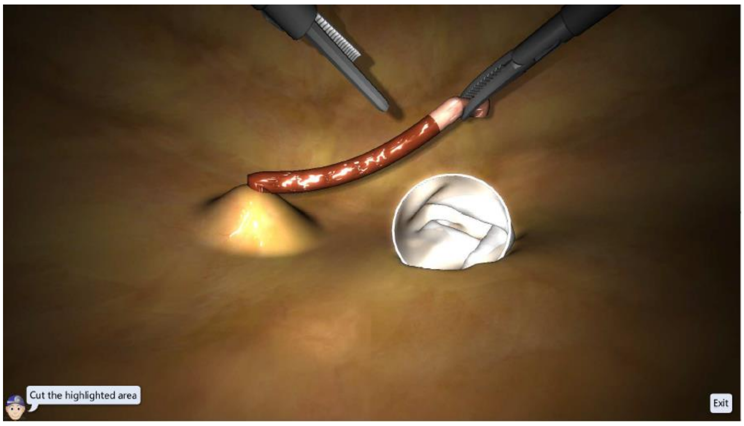

Figure 3 - The Cutting Task. The left instrument are ultrasonic scissors, the right one is a grasper. An endoscopic bag is provided for deposit of the excised tissue. The elongated vessel features differently colored areas to mark cutting spots.

Cutting is a two-handed technique which consists of grasping tissue with one hand and approaching it with the opened ultrasonic-scissors forceps with the other hand. By closing the handle, the scissors close on the tissue and by a press of the pedal, heat is applied to cut and cauterize simultaneously.

To accomplish the task in the LapSim, three pieces of one vessel must be cut and removed. The vessel is fixed on two positions of the abdominal tissue. The participant has to take care that no stretching damage is inflicted during the procedure. The removed pieces have to be put in the endoscopic bag. Figure 2 shows a screenshot of the task.

<u>Lifting & Grasping</u>

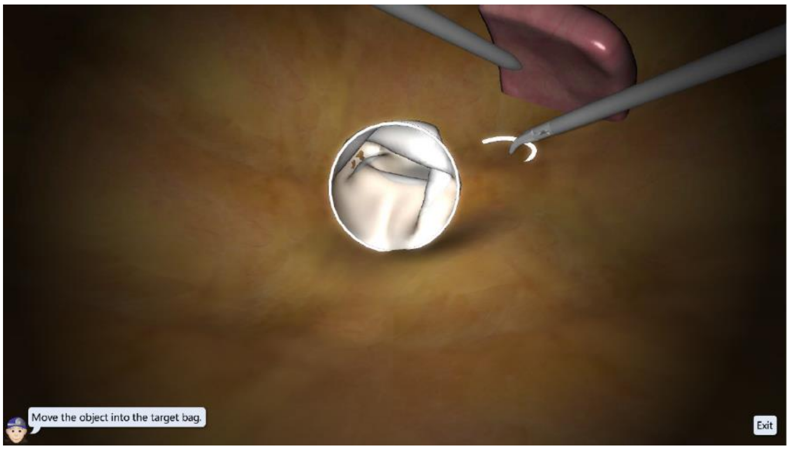

Figure 4 - The Lifting & Grasping task. The screenshot shows a probe instrument on the right and a grasper on the left. the probe is currently lifting the tissue box to allow for removal of the needle, which is the white object

Lifting & Grasping is also a two-handed technique where the surgeon uses a probe instrument. This is used to gently push under certain tissues and lift them up so the surgeon can operate underneath them. The other instrument in this task is the grasper. The instruments alter between hands after every object. There are six objects (suturing needles) which have to be picked up from under the tissue box in total.

One item question on mental demand

On basis of the NASA-TLX, a one item question was formed to measure the mental demand experienced by the participants after every trial. The NASA-TLX itself consists of 6 scales. For our research however, we simply used the scale “mental demand” to ask the question “how mentally demanding was the task?”. Furthermore, instead of the 21 graduations used in the

NASA-TLX, we changed the number of graduations to 10 so we could ask a more common “on a scale of 1-10….?” question.

2.4. Procedure

At first, participants were orally instructed to the study, its goal, duration and content. An informed consent paper was then signed, after which they were introduced to the LapSim and the task they were asked to do. The participants could then ask questions and otherwise start with the trials. The trials were presented as described in the section 2.2. After every trial the participants took a 30-second break. During that break the participants could drink some water and were asked to rate the mental demand the participants felt during that trial. The time limit for the time pressure phase was computed based on the time needed for the 15th trial. The participants were not made aware of this. Instead, we used a cover story where an anesthesiologist asks them to be faster and stay under the time limit we computed. The priority of the time pressure goal at the expense of accuracy was emphasized in the instruction of the time pressure phase and reinforced after every trial in this phase. After the 25th trial, we told the participants that we were concerned over the “high level of mental demand” they reported during the time pressure phase (regardless whether those levels were high or not) and instructed them to be “as fast and accurate as possible” for the rest of the task like in the initial phase. “Mental demand” was hereby used as an experimental manipulation to support the cover story and make the participants abandon the time pressure goal. All participants completed all 35 trials of the task “cutting”. After a break of at least one hour, they continued with the second task “lifting & grasping”. By doing so we tried to prevent the participants from fatigue.

2.5. Measurements

This study used two parameters provided by the LapSim, namely damage rate and completion time. These performance variables from the simulator have been shown to correlate with the performance in a naturalistic setting (Kundhal & Grantcharov, 2009). Damage rate consists of the values of a count of how often damage was done. Time-on-task was measured in seconds. With these values we were able to compute learning curves for all phases and tasks.
Mental demand was measured by verbally asking “how mentally demanding was the task” after each trial.

2.6 Data analysis

The raw data was explored by using scatterplots to check for general patterns in the data. Scatterplots were created per task, parameter (time-on-task, damage rate and mental demand) and participant. It was expected to see strong improvements on time on task during the initial phase and the SAT during the time-pressure phase on time-on-task (faster during the time pressure phase) and damage rate (more damage during the time pressure phase).

A non-linear mixed effects model with a learning curve as a likelihood function was used to run the regression analysis and estimate three individual learning curves on Time-on-Task and damage rate respectively. The exponential model of learning, which we called ARY, formed three learning curves that models the three experimental phases of the tasks: *general learning, learning under time pressure* and *learning after time pressure*. The first part of the learning curve models the Time-on-Task (or damage rate respectively) of the first 15 trials without time pressure, followed by the 10 trials under time pressure and 10 trials after time pressure.

2.6.1 Statistical Model and analysis

Learning curves were modelled with three learning factors: The amount of learning within the study (amplitude, δ), learning speed (rate, ρ) and the maximum learning capacity (asymptote, ω). These factors form up an exponential function for learning curves that has been frequently used in the past to power functions. Learning curves have been computed for each task, phase and measurement. The number of trial repetition per phase t_i, t_t and t_r is also modeled. The performance over trials is represented in the following formulas:

Learning curve for <u>initial</u> phase (*i*): Performance = $\omega + \delta_i\, e^{-\rho_i t_i}$

Learning curve for <u>time pressure</u> phase (*t*): Performance = $\omega + \delta_t\, e^{-\rho_t t_t}$

Learning curve for <u>removed time pressure</u> phase (*r*): Performance = $\omega + \delta_r\, e^{-\rho_r t_r}$

To perform regression analysis, the package brsm 2.1 **(**Bruckner, 2017) was used. The non-linear functions were built using the dedicated library from the package asymptote (Schmettow, 2017). Population-level effects were estimated for analysis.

The analysis focused on the three asymptotes within each task, assuming that differences among these asymptotes would reflect the effect of adding time pressure and removed time pressure. We created three ARY model whose parameters were linearized, running on a log-scale ranging from -∞ to +∞. With this it was possible to capture random effects which showed the variance caused by individual differences. The bayr 0.8.10 package (Schmettow, 2018) enabled the investigation on a participant-level.

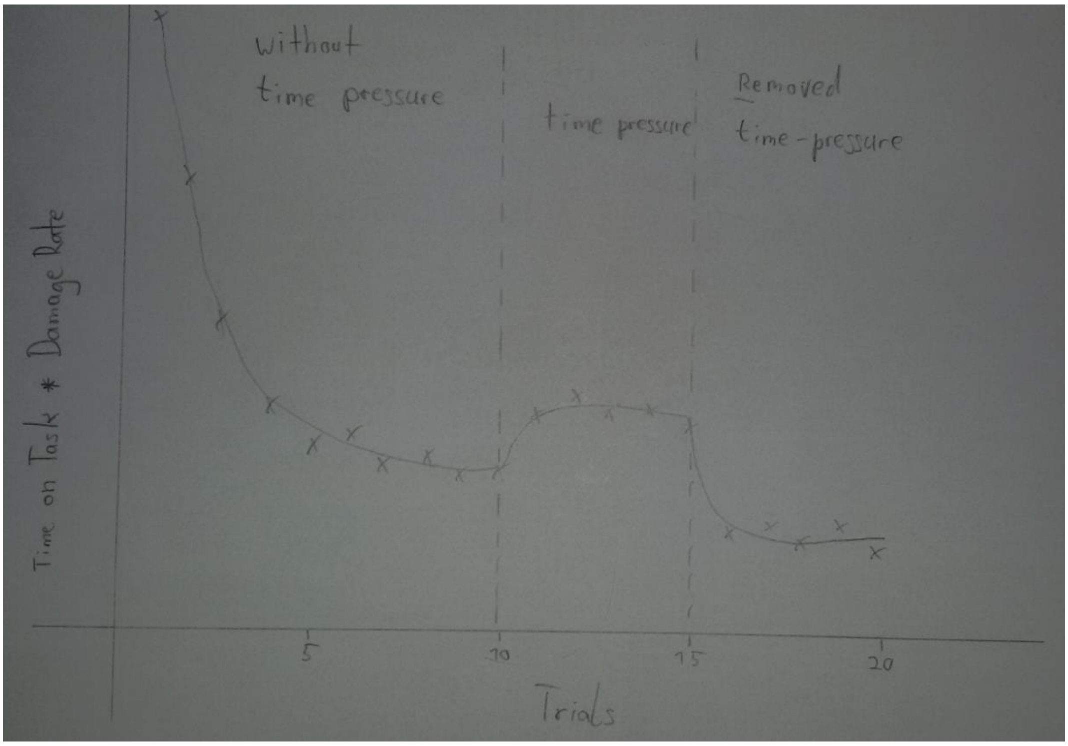

Figure 5 - learning curve of Speed Accuracy.

The population level analysis focused on the average asymptotes of the models per phase and task. Its goal was to find patterns and trends for the different tasks and phases to determine the effect of time pressure on a population level. This was followed by an individual level analysis to determine if these trends apply for every participant, as well or if there are participants who deviate from the norm.

It was not possible to compute learning curves for mental demand since the data did not converge with the ARY-model. To compare the self-reported mental demand over the three phases and two tasks, a multi-level comparison of groups model (CGM) has been computed, instead. The last five trials of each phase were compared between the three phases for both tasks on both population and individual level.

Results

Data exploration

This study investigated the influence of time pressure trials on novices performing basic MIS tasks. In this section we take a look at the data to observe if the manipulations we implemented had effect on the performance of participants. Some data for representative individuals is presented to give an impression over the performance of the participants.

The effects we expect to see in the data are the speed-accuracy tradeoff due to the implementation of time pressure during the second phase and an improvement of the performance in the third phase compared to the initial phase. Based on the visual exploration of the raw data (Figure 6), a general pattern is higher damage during the Lifting & Grasping task than during the Cutting task and more damage during the time pressure phase than in the other two phases, creating a “bump” in the middle. The time on task (ToT) seems higher during the initial phase than during the second and third phases. The decreased ToT during the second phase combined with the increased damage is in line with our expectations for the speed-accuracy tradeoff. More specifically, the ToT is highest during the first trials of the initial phase but declines rapidly. Mental demand often shows a similar pattern during the initial phase, but often goes also high in the beginning of the time pressure phase. Figure 6 shows an example for these observations.

Figure 6 – representative participant

However, there are individual differences among some participants, mainly among damage and mental demand which deviate from the general pattern. The observations on ToT however seem to be stable among most participants. The next two participants are examples of individual performances that deviate from the general pattern: Participant 13 (figure 7) did more damage on the supposedly easier Cutting task and reported more mental demand at the end of the time pressure phase (instead of the beginning of the time pressure phase).

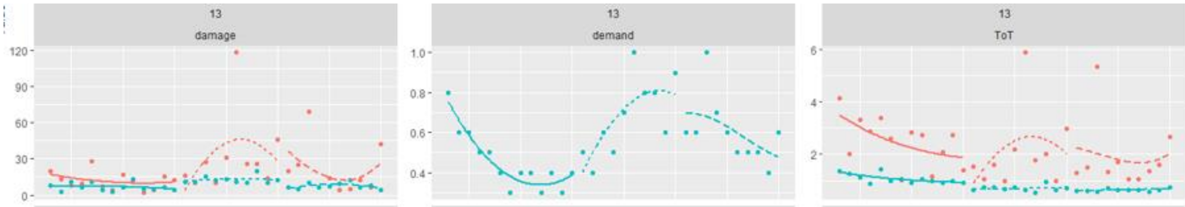

Figure 7 - participant with atypical mental demand scores, also with higher damage rates on Cutting than Lifting and Grasping

Another deviation from the general pattern that was observed among some participants was and absent decrease in damage when time pressure was introduced, even if ToT became lower. Figure 8 shows an example of that deviation.

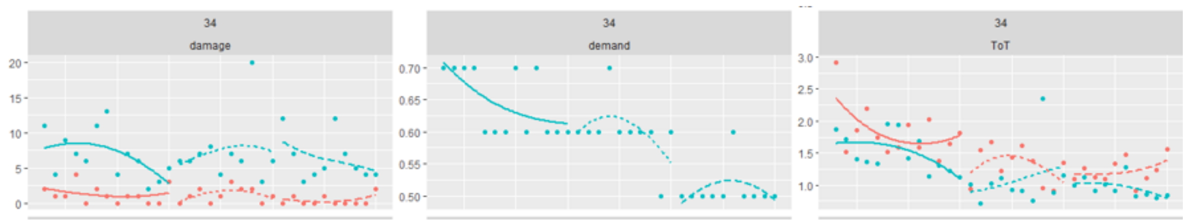

Figure 8 - participant without damage increase during time pressure

Population level analysis

This study analyzed the influence of time pressure trials on the performance of novices in MIS practice tasks and examined the speed-accuracy tradeoff in the second phase and an improvement of the performance in the third phase. The raw data was analyzed by using a multi-level nonlinear regression model to plot individual learning curves for all conditions (three phases and two tasks), resulting in 6 learning curves per participant with the parameters rate, amplitude and asymptote.

Since we were interested in the maximum performances, the focus of the analysis were the asymptotes. A graphic presentation of the asymptotes in all conditions is presented (Figure 9). The exact values are listed in table 1. On average, the asymptote of ToT during the initial phase is clearly higher during the second, which was expected due to the introduction of time pressure in the second phase. The average ToT of the third phase was similar to the second phase, which shows that participants did not become slower even after time pressure was removed.

The asymptotes of damage in both tasks showed an increase from the first task towards the second task. This pattern was expected by the introduction of time pressure and the resulting speed-accuracy trade-off during the second phase. The asymptotes for damage in the third phase for both tasks were similar to the damage in the first phase. This shows that on average the participants inflicted a similar amount of damage in the 1st and 3rd phase but are faster in the 3rd phase. Therefore it can be concluded that on average participants increased their performance.

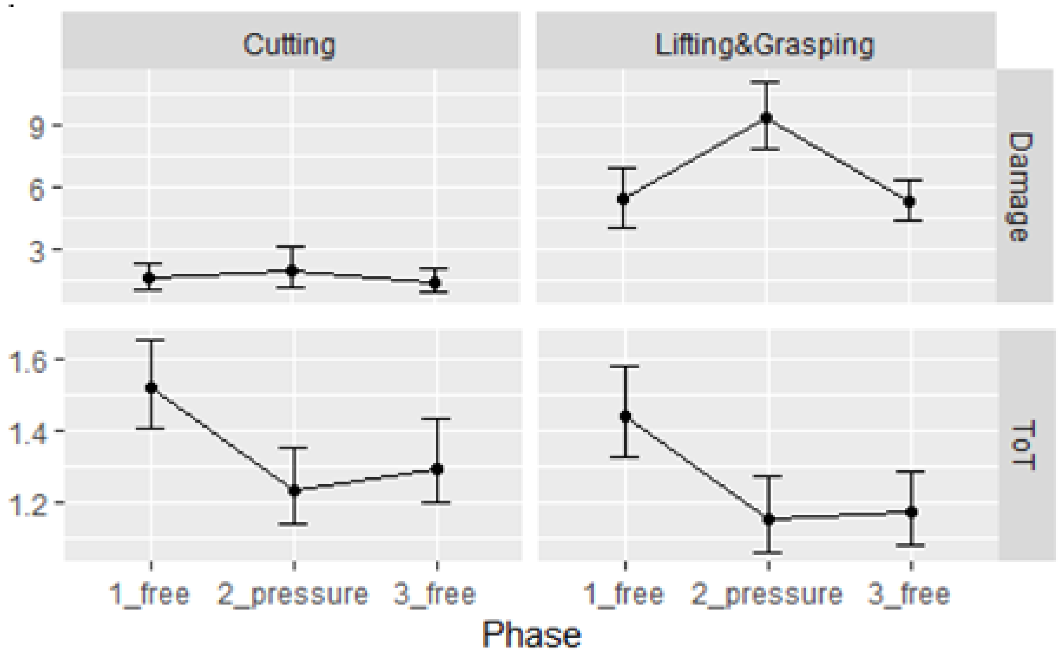

Figure 9: estimated asymptotes per learning curve on population level

Outcome	Phase	Task	center	lower	upper
Damage	1_free	Cutting	1.6901223	1.1373622	2.3165358
Damage	2_pressure	Cutting	2.0101292	1.1588933	3.1186773
Damage	3_free	Cutting	1.4553069	0.9341355	2.0998235
Damage	1_free	Lifting&Grasping	5.4604096	4.1446987	6.9178989
Damage	2_pressure	Lifting&Grasping	9.3537973	7.8566903	11.1370845
Damage	3_free	Lifting&Grasping	5.3795711	4.4249580	6.4491387
ToT	1_free	Cutting	1.5234300	1.4061874	1.6527037
ToT	2_pressure	Cutting	1.2377641	1.1430995	1.3566799
ToT	3_free	Cutting	1.2957013	1.1986496	1.4350136
ToT	1_free	Lifting&Grasping	1.4418816	1.3280576	1.5831874
ToT	2_pressure	Lifting&Grasping	1.1547266	1.0649640	1.2760797
ToT	3_free	Lifting&Grasping	1.1748865	1.0821815	1.2881611

Table 1 - asymptotes per task, phase and outcome

Individual level analysis

The learning curves per participant were estimated. As an example, the predicted learning curves of the 3 participants mentioned above are displayed (Figures 9-11). The observed “bump” at the damage scale becomes more visible. There is also a notable drop in ToT with beginning of the time pressure phase among all participants.

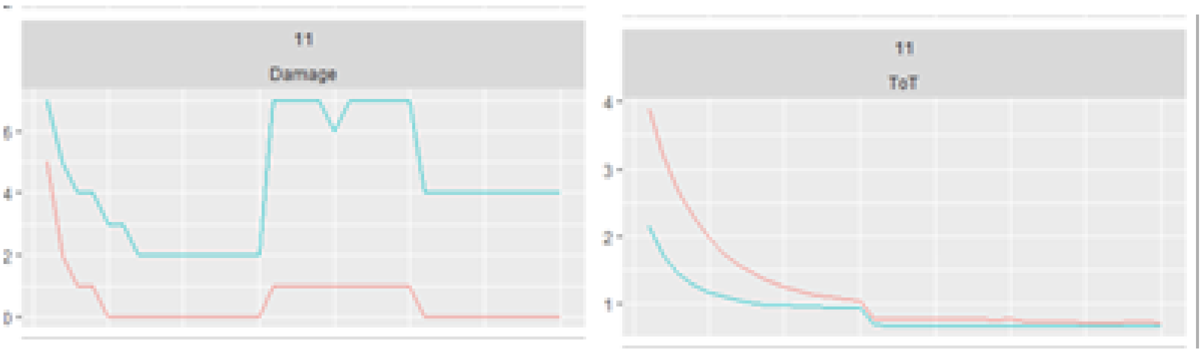

Figure 9 – Estimated learning curves

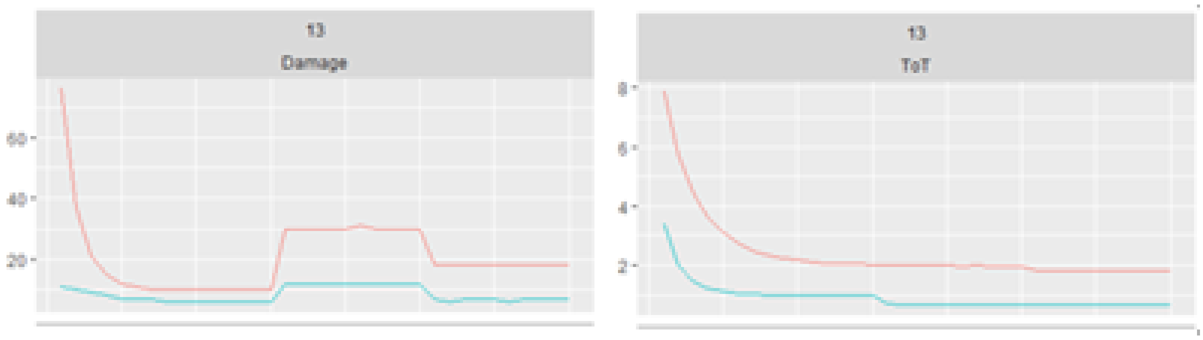

Figure 10 - Estimated learning curves

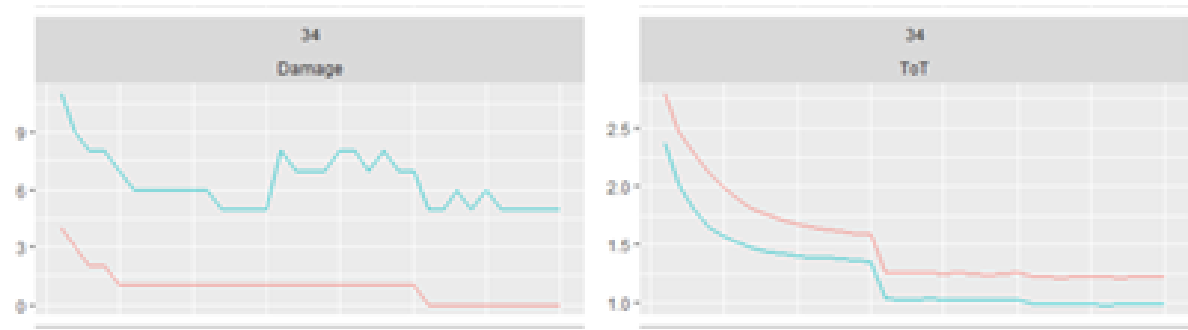

Figure 11 - Estimated learning curves

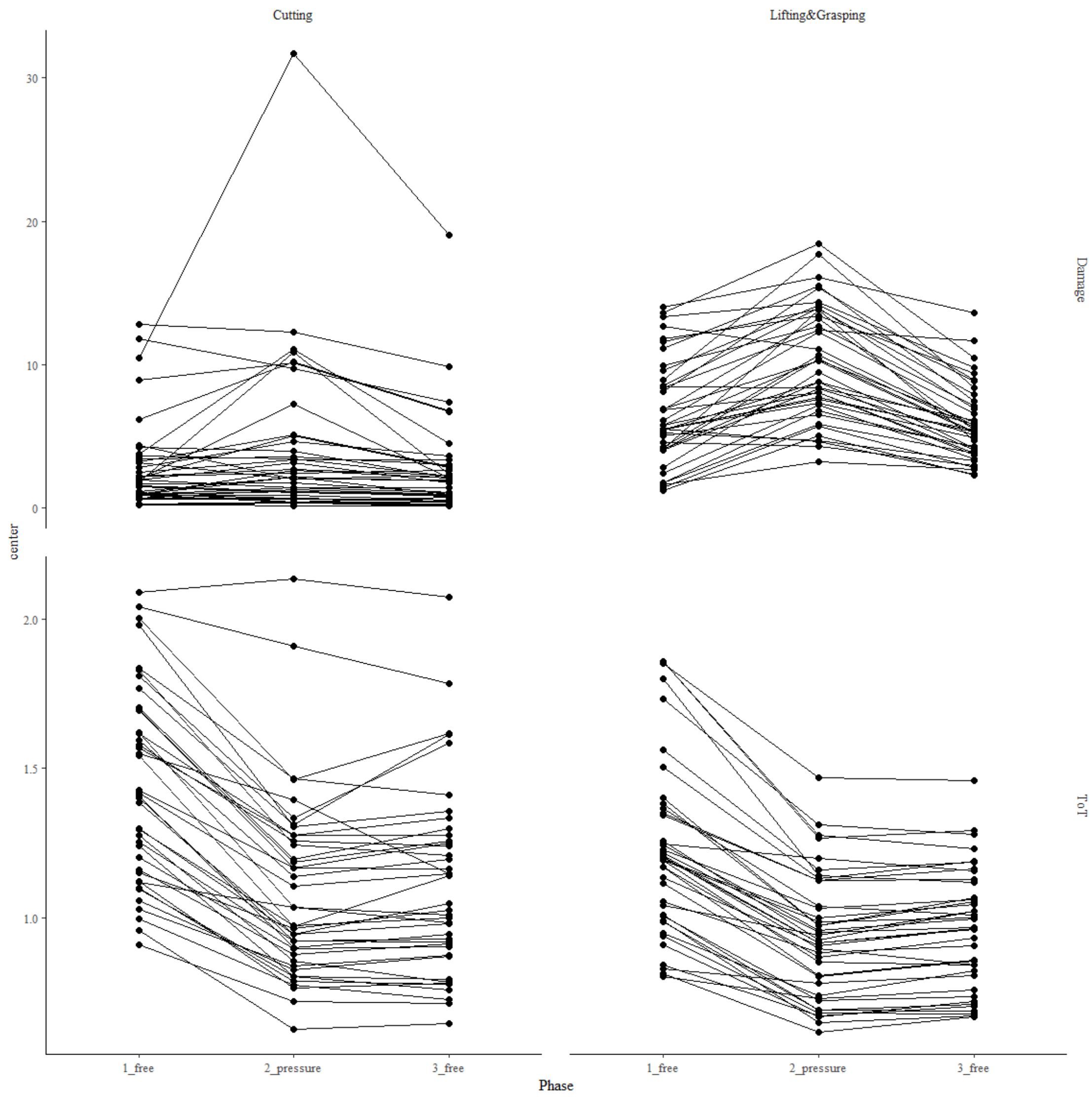

Figure 12 - individual asymptotes per learning curve

The individual learning curves were modeled to check for individual differences. This way it can be determined if the effects of time pressure hold for all participants. Figure 12 therefore presents the asymptotes on the individual level. Based on these asymptotes it can be concluded that the findings of the population level for ToT and damage also apply on the individual level.

Mental demand

On a population level, we found basically no difference on mental demand between the three phases (table 2).

fixef	center	lower	upper
Task[Cutting]	0.5470124	0.4990150	0.5973951

Task[Lifting & Grasping]	0.4938401	0.4473656	0.5418667
Task[Cutting]:Phase2_pressure	-0.0117513	-0.0501821	0.0268181
Task[Lifting & Grasping]:Phase2_pressure	0.0210039	-0.0198890	0.0608385
Task[Cutting]:Phase3_free	-0.0785467	-0.1132090	-0.0442667
Task[Lifting & Grasping]:Phase3_free	-0.0461134	-0.0782648	-0.0144721

Table 2 CGM values on population level for mental demand

If the effect of the three phases on mental demand on the population level would also apply to the individual level, we would expect many flat lines in the spaghetti plots. However, the opposite it the case (figure 13).

When taking a closer look on the individual data, there seem to be two types of participants. One type of participants reports higher mental demand during the time pressure phase and a similar level during the initial and the removed time pressure phase, whereas the other group reports a steady decline in mental demand. The ratio between participants reporting higher mental demand during time pressure and participants reporting steadily declining time pressure was 70:30. On population level, those effects are invisible. A complete overview of the individual scores on mental demand can be found in appendix 3.

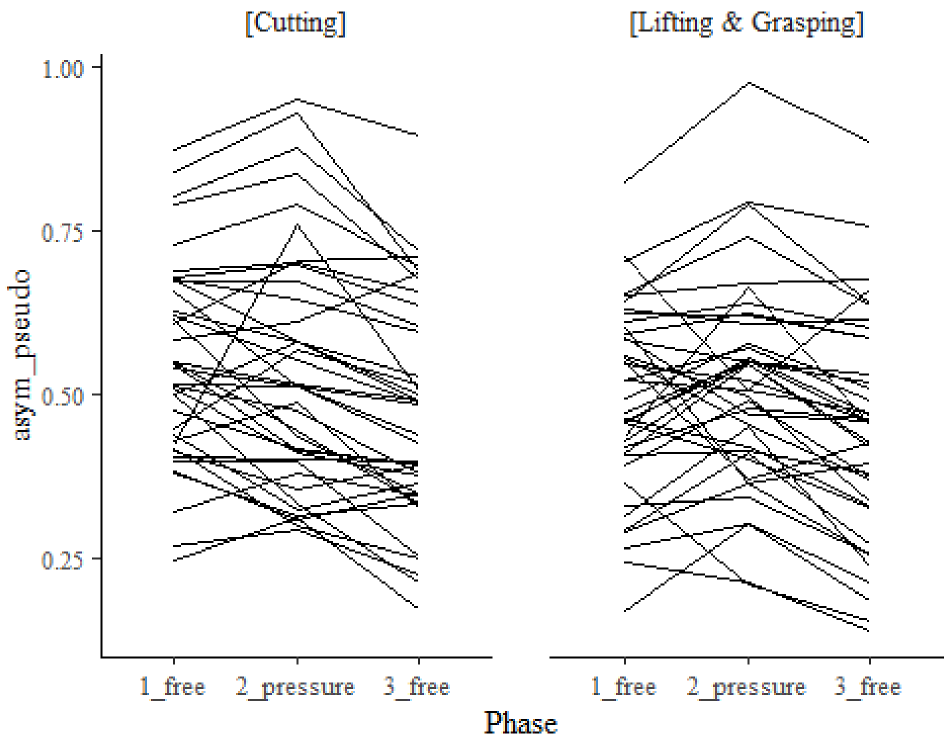

Figure 13 Mental demand on individual level

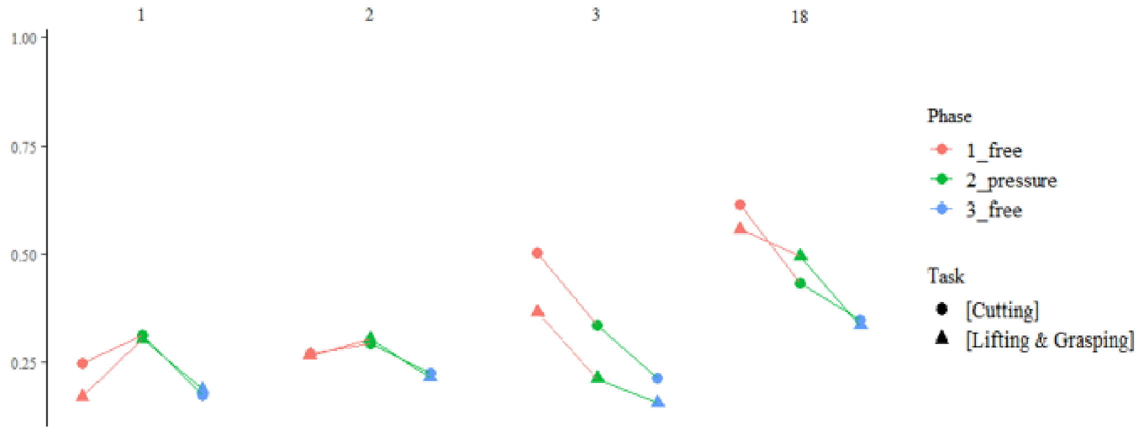

Figure 14 Different types of responses on mental demand on the individual level: participants 1 and 2 report a higher mental demand during the time pressure phase, whereas participants 3 and 18 report a steady decline in mental demand.

Discussion

This study's purpose was to contribute to the improvement of simulator-based training for MIS. In this case the implementation of time pressure in MIS was studied. This was done by analyzing the influence of time pressure trials on the performance of two basic laparoscopy tasks among novices. Learning curves for time on task and damage during the three phases were computed. Regarding our research question, we saw that during the time pressure trials (second phase) the participants were faster but had a lower accuracy compared to the initial phase, which is in line with the speed-accuracy tradeoff. However, after removing the time pressure (third phase), the participants showed an increased performance regarding time on task while showing a similar accuracy to the performance in the initial phase. On average, participants performed both tasks approximately 20% faster after the time pressure phase while maintaining their accuracy.

It is therefore possible to improve VR simulator training by introducing time pressure episodes in the training schedule to enhance the performance of trainees. This "performance boost" however does come with a price – at least for some learners. Mental demand during time pressure phases was rated higher for a considerable number of participants whereas a smaller number (ratio 70:30) reported steadily declining mental demand.

The findings in this study have practical implications for VR training and assessment. We were able to demonstrate that during a training session it is possible to enhance the performance of trainees by adding time pressure phases. This can be specifically useful in assessment situations where the potential of possible trainees is tested. Time pressure phases in assessment can push trainees over their supposed boundaries and reveal potential otherwise unnoticed as Gas et al. (2018) have stated.

With that in mind, it would be unwise to put all trainees under constant time pressure. Instead, tailored training schedules can be the answer. Trainees who are not bothered by time pressure training can implement them more in their own training, while those who are affected by time pressure should practice without or less time pressure to prevent them from negative consequences due to constant high mental demand.

Fundamentally, these findings raise questions about learning curves, specifically the standing of the asymptote. Heathcote and Brown (2000) described the asymptote as a boundary because of physical constraints. Asymptotes were viewed as an indicator for maximum performance in several studies (Schmettow et al, 2016; Pusic, M.V., Boutis, K., Pecaric, M.R., Savenkov, O., Beckstead, J.W., Jaber, M.Y., Martín-Láez, R, Martínez-Agüeros, J.Á, Suàrez-Fernandez, D., Montiaga-Núnez, F., Fázquez-Barquero, A.). Our findings suggest that either individuals can overcome their supposed limits or our model for computing the asymptote is insufficient.

Most effects on time pressure have looked into immediate effects of time pressure on performance. They naturally found speed-accuracy effects (e.g. Cook, Aljamal, Pankraty, Sedlack, Farley and Brydges, 2019, Gas et al., 2018). However, the performance was not measured after the time pressure instructions were removed. It has been argued before that by changing the difficulty of a task learners could be pushed outside their comfort zones and latent errors or inefficient behavior could be exposed (Korndorffer, Scott, Sierra, Brunner, Dunne et al, 2005, Stefanidis, Korndorffer, Scott, 2007). It is plausible that in our experiment this happened after changing its conditions by adding time pressure. The LapSim environment provides conditions that help learners to explore possibilities and limits of the tasks. On inflicting damage to the patient, the screen flashes red immediately and every trial is followed by useful performance statistics that help understand what went well and where further improvement is possible. It is speculative, but an explanation for the vast improvement during the third phase would be that participants broadened their skillset during the time pressure phase where they pushed their limits and learned from their errors.

The focus in this study was to analyze the role of time pressure, which is a speed emphasis. Since speed instructions are not advised for learners who do a task for the first time, we implemented an initial phase where participants were instructed to be as "fast and accurate as possible". However, Magill (2011, page 338) recommends to start with a focus on accuracy when practicing for a speed-accuracy skill. They refer to a study by Blais (1991) and motor program and dynamical systems theories to argue that practice should focus on accuracy at first and on speed on a later stage. It is possible that trainees for MIS surgery could also benefit from this approach.

To help explain the differences in mental demand during the third phase compared to the first two phases one can argue that the speed-accuracy trade-off as a theory is inconclusive. Stefanidis et al (2007) argued that speed and accuracy as measurements of performance do not provide enough information about the skill level of the performer. Differences in experience and mental capacities are not necessarily reflected in a performance measured by speed and/or accuracy. Mental demand certainly plays a role, as well, since participants reported higher ratings on that scale during time pressure trials. But even a speed-accuracy-mental-effort trade-off would not tell the whole story, since not all participants reported higher mental demand during time pressure phases. Learning types/styles could offer an explanation. Kolb (2005) argued that different learning types benefit from different learning experiences. It is possible that some participants experienced positive motivation from the challenge provided by the time pressure phase while others were put under negative stress.

Regarding alternative explanations for the results in this study, the more explicit goal during the time pressure phase could give an explanation. During the time pressure phase, participants had a precise goal (to be 20% than during their 15th trial). In the other phases, the instruction was to be "as fast and accurate as possible" without specific goals in terms of time limit or damage inflicted. Goals direct and energize effort, promote effective learning strategies and increase persistence (Locke and Latham, 2002). It is therefore possible that specific goals contributed to the improvement of participants. However, a study by Cook at al. (2019) found that goals focused on quality did not seem to influence performance.

Strengths and limitations

One of this study's strength was its within subjects design, which has more power compared to a between subjects design. This was also necessary since data collection was time consuming and the time available on the LapSim simulators at the University of Twente was limited. The within subjects design ensured us the efficiency needed. The use of short breaks after each trial and a longer break between the two tasks also prevented the participants from fatigue. Another strength was the use of learning curves since it allowed us to predict the talent of the participants after constant training instead of just the skill at this moment. The innovative implementation of a post-timepressure phase in this experiment is an advantage of this study compared to earlier studies examining the SAT since it allowed us to analyze the performance after induced time pressure.

This study also had some limitations. The original purpose of mental demand in this study was to use it as fake reason to remove the time pressure element after the 25th trial due to "too high levels" of mental demand. We therefore decided that one item would be sufficient to fulfill this purpose. No specific definition of mental demand was given beforehand. It is therefore unclear what participants interpreted when it comes to this term. Frustration about mistakes or time pressure, perceived energy levels, difficulty of the task could have played into the answers given by the participants, we cannot say how participants interpreted the concept of mental demand in this experiment.

We cannot make assumptions over long-term effects of time pressure training since our experiment consisted of two basic laparoscopic tasks which did not take longer than 3 hours in total. This applies both for the improvement of performance and the effects of higher experience of mental demand for a longer time.

To induce time pressure in the experiment, we asked the participants to be 20% faster than during the 15th trial. 20% faster was also used in the study by Gas et al (2018), but is rather arbitrary in general. To make the time pressure dependent on the 15th trial makes it also vulnerable to outliers. It is easier to be 20% faster than a slow 15th trial compared to a fast one. For instance, a more stable reference would be a value that refers to the average of the last five trials. There may be also an "ideal" percentage that provides a more effective time pressure stimulus. However, the 20% goal proved to be productive in general and led consistently to the SAT.

Future research should further investigate the benefits of time pressure for MIS training. This study analyzed time pressure effects on two basic laparoscopic tasks. We cannot say if those effects also exist for longer and/or more complex tasks. Studies should also focus on the role of mental demand. A more precise concept and research methods are necessary to formulate how mental demand relates do the speed-accuracy trade-off.

Conclusion

This study showed that time pressure can be an effective tool to improve the performance of learners on basic laparoscopic tasks. On average, participants were estimated to be 20% faster with a time pressure phase than without a time pressure phase. Accuracy was similar in both conditions. A change in mental demand during the last phase suggests that the theory of the speed-accuracy trade-off can be expanded. These findings could improve training of MIS surgeons and should be further investigated. Furthermore, the concept of learning curves needs to be reviewed since this study has shown that the asymptote does not necessarily symbolize the maximum performance that performers can reach due to natural limits.

References

Arendt, A., Schmettow, M, Groenier, M. (2017). *Towards reliable and valid prediction of MIS-performance with basic laparoscopic tasks in the LapSim and low-fi dexterity tasks.* University of Twente, Enschede, Netherlands.

Beilock, S., Bertenthal, B., Hoerger, M., & Carr, T. (2009). When Does Haste Make Waste? Speed-Accuracy Tradeoff, Skill Level, and the Tools of the Trade (Vol. 14). *Journal of Experimental Psychology*: *Vol. 14, No. 4, 340–352*

Beilock, S. L., Bertenthal, B. I., McCoy, A. M., & Carr, T. H. (2004). Haste does not always make waste: Expertise, direction of attention, and speed versus accuracy in performing sensorimotor skills. *Psychonomic Bulletin & Review, 11*(2), 373-379. doi:10.3758/BF03196585

Beilock, S. L., Carr, T. H., MacMahon, C., & Starkes, J. L. (2002). When paying attention becomes counterproductive: Impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. *Journal of Experimental Psychology: Applied, 8,* 6–16.

Castaneda, B., & Gray, R. (2007). Effects of focus of attention on baseball batting performance in players of differing skill levels. *Journal of Sport & Exercise Psychology, 29,* 59–76.

Cook, D. A., Aljamal, Y., Pankratz, V. S., Sedlack, R. E., Farley, D. R., & Brydges, R. (2019). Supporting self-regulation in simulation-based education: a randomized experiment of practice schedules and goals. *Advances in Health Sciences Education, 24*(2), 199-213. doi:10.1007/s10459-018-9860-z

Darzi, S. A., & Munz, Y. (2004). The Impact of Minimally Invasive Surgical Techniques. *Annual Review of Medicine, 55*(1), 223-237. doi:10.1146/annurev.med.55.091902.105248

Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *JAMA*, *287*(2), 226–235. doi: 10.1001/jama.287.2.226

Fielding, D. I., Maldonado, F., & Murgu, S. (2014). Achieving competency in bronchoscopy: Challenges and opportunities. *Respirology, 19*(4), 472-482. doi:10.1111/resp.12279

Gardner, A. K., Ritter, E. M., Paige, J. T., Ahmed, R. A., Fernandez, G., & Dunkin, B. J. (2016).
Simulation-based selection of surgical trainees: considerations, challenges, and opportunities. *Journal of the American College of Surgeons*, *223*(3), 530–536. Doi: 10.1016/j.jamcollsurg.2016.05.021

Gas, B. L., Buckarma, E. H., Cook, D. A., Farley, D. R., & Pusic, M. V. (2018). Is Speed a Desirable Difficulty for Learning Procedures? An Initial Exploration of the Effects of Chronometric Pressure. *Academic Medicine, 93*(6), 920-928. doi:10.1097/ACM.0000000000002167

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. doi: 10.3758/BF03212979

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*(8 JUN). doi:10.3389/fnins.2014.00150

Jackson, R. C., Ashford, K. J., & Norsworthy, G. (2006). Attentional focus, dispositional reinvestment and skilled motor performance under pressure. *Journal of Sport & Exercise Psychology, 28,* 49–68.

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.

Korndorffer, J., J Scott, D., Sierra, R., C Brunner, W., Bruce Dunne, J., Slakey, D., . . . L Hewitt, R. (2005). Developing and Testing Competency Levels for Laparoscopic Skills Training. *Archives of surgery (Chicago, Ill. : 1960), 140*, 80-84. doi:10.1001/archsurg.140.1.80

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*(9), 705-717. doi:10.1037/0003-066X.57.9.705

Kolb, A. Y., & Kolb, D. A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning and Education, 4*(2), 193-212. doi:10.5465/AMLE.2005.17268566

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review, 89*(5), 483-506. doi:10.1037/0033-295X.89.5.483

Magill, R. A. (2011). *Motor learning and control: Concepts and applications*. New York: McGraw-Hill.

Martín-Láez, R., Martínez-Agüeros, J. Á., Suárez-Fernández, D., Montiaga-Núñez, F., & Vázquez-Barquero, A. (2012). Complications of endoscopic microdiscectomy using the EASYGO! system: Is there any difference with conventional discectomy during the learning-curve period? *Acta Neurochirurgica, 154*(6), 1023-1032. doi:10.1007/s00701-012-1321-5

Mührmann, L. (2018). *Towards the prediction of bronchoscopic skill acquisition on a low-fidelity endoscopic prototype.* University of Twente. Retrieved from http://purl.utwente.nl/essays/75187

Müller-Stich, B. P., & Büchler, M. W. (2015). Komplikationsmanagement in der minimal-invasiven Chirurgie. *Der Chirurg, 86*(12), 1095-1096. doi:10.1007/s00104-015-0109-6

Patel, H. R. H., & Patel, B. P. (2012). Virtual reality surgical simulation in training. *Expert Review of Anticancer Therapy, 12*(4), 417-420. doi:10.1586/era.12.23

Perez-Cruet, M. J., Fessler, R. G., & Perin, N. I. (2002). Review: Complications of minimally invasive spinal surgery. *Neurosurgery, 51*(5 SUPPL.), 26-36.

Ponsky, J. L. (1991). Complications of laparoscopic cholecystectomy. *The American Journal of Surgery, 161*(3), 393-395. doi:10.1016/0002-9610(91)90605-D

Pusic, M. V., Boutis, K., Pecaric, M. R., Savenkov, O., Beckstead, J. W., & Jaber, M. Y. (2017). A primer on the statistical modelling of learning curves in health professions education. *Advances in Health Sciences Education, 22*(3), 741-759. doi:10.1007/s10459-016-9709-2

Pusic, M. V., Brydges, R., Kessler, D., Szyld, D., Nachbar, M., & Kalet, A. (2014). What's your best time? Chronometry in the learning of medical procedures. *Medical Education, 48*(5), 479-488. doi:10.1111/medu.12395

Ouellette, D. R. (2006). The safety of bronchoscopy in a pulmonary fellowship program. *Chest, 130*(4), 1185-1190. doi:10.1378/chest.130.4.1185

Reis, J., Schambra, H. M., Cohen, L. G., Buch, E. R., Fritsch, B., Zarahn, E., & Krakauer, J. W. (2009). Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. Proceedings of the National Academy of Sciences, 106(5), 1590–1595. doi:10.1073/ pnas.0805413106

Sadideen, H., Hamaoui, K., Saadeddin, M., & Kneebone, R. (2012). Simulators and the simulation environment: Getting the balance right in simulation-based surgical education. *International Journal of Surgery*, *10*(9), 458–462. doi: https://doi.org/10.1016/j.ijsu.2012.08.010

Schmettow, M (2018) r package for unified reporting of Bayesian regression results Retrieved from: https://github.com/schmettow/bayr

Schmettow, M (2018) provides functions and formulas for exponential learning curves Retrieved from: https://github.com/schmettow/asymptote

Schmettow, M (2018) *New Statistics for the Design Researcher*. Retrieved from https://schmettow.github.io/New_Stats/index.html

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. Retrieved from http://essay.utwente.nl/70010/1/Kaschub_BA_psychology.pdf.

Shmuelof, L., Krakauer, J. W., & Mazzoni, P. (2012). How is a motor skill learned? Change and invariance at the levels of task success and trajectory control. Journal of Neurophysiology, 108 (2), 578–594. doi:10.1152/jn.00856.2011

Soukoreff, R. W., & MacKenzie, I. S. (2009). An Informatic Rationale for the Speed-Accuracy Trade-Off. *Medical Education 2014: 48: 479–488 doi: 10.1111/medu.12395*

Stefanidis, D., Scerbo, M., Korndorffer, J., & J Scott, D. (2007). Redefining simulator proficiency using automaticity theory. *American journal of surgery, 193,* 502-506. doi:10.1016/j.amjsurg.2006.11.010

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), 1413–7. Doi: 10.1007/s00464-006-9188-2

Wanzel, K. R., Hamstra, S. J., Anastakis, D. J., Matsumoto, E. D., & Cusimano, M. D. (2002).
Effect of visual-spatial ability on learning of spatially-complex surgical skills. *The Lancet*, *359*, 230–231.

Westerhof, Marlise (2018) *Towards Predicting Bronchoscopic Skill Acquisition Using Basic Bronchoscopic Simulator Tasks.* Retrieved from: http://purl.utwente.nl/essays/76502

White, C., Rodger, M., & Tang, T. (2016). Current understanding of learning psychomotor skills and the impact on teaching laparoscopic surgical skills. *The Obstetrician & Gynaecologist*, *18*(1), 53–63. doi: 10.1111/tog.12255

Zhang,J. and Rowe,J.B.(2014). Dissociable mechanisms of speed-accuracy trade-off during visual perceptual learning are revealed by a hierarchical drift-diffusion model. *Front.Neurosci.* 8:69.doi:10.3389/fnins.2014. 00069

Soukoreff, R. W., & MacKenzie, I. S. (2009). An Informatic Rationale for the Speed-Accuracy Trade-Off. *Medical Education 2014: 48: 479–488 doi: 10.1111/medu.12395*

Stefanidis, D., Scerbo, M., Korndorffer, J., & J Scott, D. (2007). Redefining simulator proficiency using automaticity theory. *American journal of surgery, 193,* 502-506. doi:10.1016/j.amjsurg.2006.11.010

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), 1413–7. Doi: 10.1007/s00464-006-9188-2

Wanzel, K. R., Hamstra, S. J., Anastakis, D. J., Matsumoto, E. D., & Cusimano, M. D. (2002).
Effect of visual-spatial ability on learning of spatially-complex surgical skills. *The Lancet*, *359*, 230–231.

Westerhof, Marlise (2018) *Towards Predicting Bronchoscopic Skill Acquisition Using Basic Bronchoscopic Simulator Tasks.* Retrieved from: http://purl.utwente.nl/essays/76502

Zhang,J. and Rowe,J.B.(2014). Dissociable mechanisms of speed-accuracy trade-off during visual perceptual learning are revealed by a hierarchical drift-diffusion model. *Front.Neurosci.* 8:69.doi:10.3389/fnins.2014. 00069

Appendix 1 - Instructions

This study consists of 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 two exercises on the simulator. You can ask questions any time. Do you have any questions thus far?
[*Participant starts with the demographics questionnaire.]*

Appendix 2 – Demographic questionnaire

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 colorblind? 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

LapSim

You are about to start practicing two basic laparoscopic tasks which are part of a procedure called cholecystectomy, 1) lifting and grasping and 2) 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 lifting and grasping and cutting. You must perform each task 35 times. You will start with the lifting and grasping task and then continue to the cutting. 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.

LapSim + time pressure

The conditions of the task are going to change now. Until now, you focused on being as fast and accurate at the same time. However, we will now focus more on speed than accuracy and introduce time pressure to the task: Your goal is to be 20% faster than during the last (15^{th}) trial. The last trial took you x seconds. Try to do the task within 8/10x seconds now.

[After each trial under time pressure, the participant simply gets informed if he succeeded in beating the time limit. No matter the outcome, the goal to stay under the time limit is reinforced in the instructions.]

Try again to stay under the time limit.

Lapsim without time pressure

The time pressure element will now be removed from the task. Please try to be as accurate and quick at the same time as possible. Only make sure that you do not hurt your patient, which is indicated by the screen flashing in red.

Appendix 3 – individual CGM scores on mental demand