BACHELOR THESIS PSYCHOLOGY

Towards the prediction of bronchoscopic skill acquisition on a low-fidelity endoscopic prototype

Author: Lisa Mührmann (l.muhrmann@student.utwente.nl)
Std. number: s1605135

Primary Supervisor: Dr. Martin Schmettow (CPE)

Secondary Supervisor: Dr. Marleen Groenier (TG)

Department of Cognitive Psychology and Ergonomics (CPE)

UNIVERSITY OF TWENTE.

Abstract

Introduction A great development in the field of medicine is the admittance of minimally invasive surgery, as it allows operations with reduced blood loss, pain, hospitalization time, and improved cosmetic. One well-known method is the flexible bronchoscopy, in which the lungs could be examined for abnormalities. However, the risk for patients is increased, as the intervention takes place on a vitally necessary organ. In order to minimize the risk, professional surgeons are needed, who are selected for their suitability during different training programs, such as the virtual reality (VR) simulator training. Nevertheless, adequate methods have to be further developed. The original goal of this study was to test, if training on a low-fidelity endoscopic prototype (boxtrainer), which simplify the real bronchoscopic procedure, can improve the VR-simulator task performance. Due to occurring technical problems with the VR-simulator, we focused now only on the boxtrainer task-performance by approaching the performance variables time on task, wall contacts and task success. Another unexpected problem arose, as the estimation of learning curves failed. However, this allowed us to concentrate on different aspects of performance, such as the speed-accuracy tradeoff, without the difficulty to appreciate learning curves. The resulting goal of this research was then to explore the association between the performance variables time on task and wall contacts.

Method Twenty four students of the University of Twente participated. A one-hour training on a low-fidelity boxtrainer is administerd from an allocentric and an egocentric perspective. All participants did the same tasks. Stopping rule was time. A time series design was applied. The original goal of estimating learning curves with a non-linear mixed effect model based on the performance variable time on task, wall contact and task success, failed. Insteed we used the linear multi-level model in order to obtain the association between the performance variables time on task and wall contact.

Results The estimation of learning curves failed. Performance did not improved after prolonged training on the box tainer. Therefore, the predictor trial could be disrepected. The problem simplifies to a multi-level linear model where trials become exchangeable repeated measures. Through employing a generalized linear model (GML) with a poisson distribution a linear association between the performance varbiables time on task and wall contact could be noticed on a population level, as well as on a participant level. Participants made more mistakes, the more time they needed for completing a task.

Conclusion The low-fidelity boxtrainer is not an adequate substitute for the high-fidelity VR-simulators, as the estimation of learning curves is not possible. However, instead, training on

the box can measure certain other important aspects of performance, such as the speed accuracy tradeoff during the execution of bronchoscopic tasks. Further research should consider both performance variables time on task and wall contacts during MIS-training in order to obtain a realistic assessment of the potential of a person.

Keywords: Bronchoscopy; skill acquisition; virtual-reality simulator; box trainer; learning curve; non-linear multilevel mixed effect model

Table of Content

General Introduction

Minimally invasive surgery (MIS) is one of the greatest developments in the field of medicine and has become the gold standard approach in many surgical conditions (Fuchs, 2002). MIS differs from conventional open surgery in the use of endoscopic, minimally invasive access, special instruments and techniques, which results in a reduced blood loss, pain, hospitalization time, and improved cosmetic (Dorr, Maheshwari, Long, Wan & Sirianni, 2007). However, in addition to the benefits of minimally invasive surgery over conventional surgery, it also offers some disadvantages. Two major drawbacks have arose with the admittance of this new technique. On the one hand, there is a great increase in costs due to the investment in the special training programs for surgeons, the equipment required, as well as longer operating times. On the other hand, surgeons have a prolonged learning curve, in comparison with the learning process in open surgery, because of the necessary acquisition of complex technical skills needed for performing MIS (Fuchs, 2002).

One of the well-established minimally invasive procedures is the bronchoscopy. During the insertion of a bronchoscope into the lung of a patient, the surgeon is now able to discover any abnormalities, which is a big step forward in the medicine (Rogalla et al., 2001). To ensure the humans safety during this complex surgical intervention on a vitally necessary organ, surgeons have to be selected for their suitability and proficiency within their training (Hassan et al., 2007).

However, there are currently no adequate ways to decide, whether an aspiring surgeon will have the potential to become a suitable and professional surgeon in the field of minimally invasive surgery. Traditionally, surgeons will be judged only on letters of recommendations or interviews, assessment of scientific knowledge and the achievements on medical school, which refers to the apprenticeship model (Basdogan, Sedef, Harders, & Wesarg, 2007). The subjective opinion of an expert, who will oversee the learning process of the surgeon and decide about their potential, is responsible for the selection process of surgeons. A solution for the missing objective assessment of surgical competence came through the development of virtual reality simulators, as they provide computer-based modules of realistic surgical procedures, which objectively determine the surgeons potential (Schell & Flynn, 2004). Nevertheless, adequate methods have to be developed.

The original goal of this research was to explore whether previous training on a low-fidelity inanimate box trainer, as a simplified version of the VR-simulator for bronchoscopy, can improve the simulator task performance. However, there were unexpected technical problems with the VR-simulator, which causes us to concentrate only on the task performance of the participants on the box trainer. A set of surgical tasks on the box trainer were provided. For estimating individual's performance after continued training, we chose for the maximum performance parameter of a non-linear multilevel mixed effect model, based on the three performance variables, time on task, task success and wall contact. With this, we want to find out whether the performance parameters are correlated in such a way, that combining the measures yields a more precise estimate of performance. Unfortunately, the estimation of individual learning curves failed, as there were no learning effects. Therefore, we can say in advance that the boxtrainer is not an adequate substitute for the VR-simulator.

However, the renewed problem simplifies to a multi-level linear model where trials become exchangeable repeated measures, which allows us to concentrate on different aspects of performance, such as the speed accuracy tradeoff, without the difficulty to appreciate learning curves. Therefore, our current goal of this research is to explore the association between the performance variables time on task and wall contact.

Traditional Training Method for MIS

Current studies suggest a great deficit of adequate surgical training, which is important for the patient's safety (Rosser, Murayama & Gabriel, 2000). The traditional training of minimally invasive surgery is the apprenticeship model, by which aspiring surgeons acquire skills by observing a senior surgeon performing surgical procedures and vice versa (Basdogan, Sedef, Harders, & Wesarg, 2007). In addition, there is no uniform curricula or standardized objective metrics for education in bronchoscopy to ensure acquisition and measurement of skills needed to achieve competence. The subjective opinion of the experts about the performance of the trainee surgeons is crucial for the selection process. Thereby the decision, whether a surgical applicant will become a successful bronchoscopic surgeon is not only depended on the examiner who gives the grade, but it can even differ on a day-to-day basis with the same examiner. One of the most likely cause of human error within the means of an objective assessment is the fatigue assessors experience after taking several examinations on a single day (Gardner et al., 2016). In addition, experts assert that training is the most important factor to become professional and competent and trainees should at least perform 100 flexible bronchoscopies (Konge, Arendrup, Von Buchwald & Ringsted, 2011). With regard to the cost containment and increasing oversight of professional competency, alternatives to the conventional apprenticeship model are necessary (Rosen et al., 2002).

VR-Simulator Training

In the recent years, the virtual reality (VR) training method for the development and refinement of surgical skills come to the forefront, as it allows a standardized, metric-based training and have a high level of resemblance towards the real surgical procedures (Gallagher et al., 2006). VR - simulators can be used to establish a benchmark (i.e. the level of proficiency), by providing a more homogeneous skill-set in the assessment of trainees. In addition, they can be applied to any level of training. The computer-based modules allow a targeted training of different techniques in freely selectable scenarios (Schell & Flynn, 2004). A great advantage of VR – training is the possibility of working in a safe and controlled area, away from the patient, as well as repeating the exercises with no limits and having only low operation costs (Gaba, 2004). Therefore, VR-simulators lend themselves to assessment, too, but methods need to be developed.

Low-Fidelity Assessment

A good alternative to the high-fidelity simulator, like the VR-simulator for learning bronchoscopic skills, could be a low-fidelity simulator. Using low-fidelity simulators could be a much cost-effective variant, as these tangible simulators are made of common and simple materials, like households items. In contrast to the high-fidelity simulators, it is possible to create a low-fidelity simulator in just a few minutes to a few hours. With regard to the acquisition of surgical skills, low-fidelity simulators could contain real aspects of surgical procedures in a simplified way, which make them easy to understand, whilst being complex enough to learn the complex cognitive and psychomotor skills. When the skills level of the surgeons could be systematic estimated through using a low-fidelity simulator, hospitals and society will also benefit from these innovations, because only suitable surgeons are allowed to engage in bronchoscopic procedures on an alive human. Because of the possibility of continuous training on the low-fidelity simulator, surgeons probably already have a much greater wealth of experience in this starting phase, which could result in a high degree of surgical manual self-esteem. Therefore, using low-fidelity simulators can lay a foundation for the future use of simulator-based training in the medical field, as it allows a cost-effective, simple and low-risk way of assessment of the required surgical skills.

Learning Curves

As a method to determine individual differences based on the surgical skill acquisition, learning curves were established, to define at which point(s) practice is most efficient and how much practice is required to achieve a defined level of mastery. According to Heathcote, Brown & Mewhort (2000) the exponential law of practice is composed out of three parameters. The first parameter asymptote refers to the level of maximum performance a person can achieve after prolonged training. In addition, the amount of improvement will be represented through the parameter amplitude, which show exactly the performance difference between the first trial and the asymptote. The last parameter of learning curves is the rate parameter, which indicates the overall speed of learning.

However, the focus here is more on the individual's process than on a one-time measurement. In general, learning curves display the relationship between the performance variable (i.e. time on task) on the vertical axis and experience (related to the number of trials) on the horizontal axis. Mostly the learning curves rise quickly, approach asymptotically a limit and then stabilizes (Fuchs, 2002). However, it depends on the amount of experiences a person already has, where learning curves will have their starting point. In addition, the time at which people reach their maximum performance (asymptote) and the progress of the learning curve (rate parameter) can vary between persons, which makes it possible to compare the learning processes of individual people.

Using maximum performance as a performance measure allows different advantages. On the one hand, maximum performance remains stable with every new trial in contrast to the two other learning curve parameter, rate and previous experience (Schmettow, Kaschub, & Groenier, 2016). On the other hand, the parameter acts as a predictor for the maximum possible performance a person can achieve, which allows an adequate selection process of talented persons (Arendt, Schmettow & Groenier 2017). This requires individual learning curves, rather than averaged.

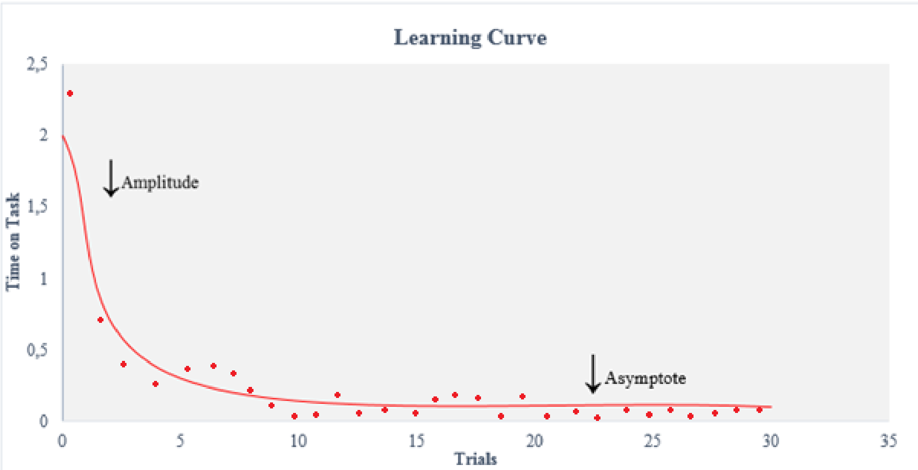


Figure 1. Exponential learning curve. The x-axis represents the experience (number of trials) of the participants, whereas the y-axis represents the learning variable (time on task).

Flexible Bronchoscopy

Flexible bronchoscopy represents one of the clinically well-established invasive diagnostic tools, as it allows the visualization of the inner respiratory tract for diagnostic and therapeutic purposes (Rogalla et al., 2001). During bronchoscopy, surgeons insert an endoscopic instrument (bronchoscope) through the nose or mouth of the patient in the trachea to the large and middle bronchi (see figure 2). Usually, the bronchoscope is made of a flexible fiber-optic material and has a light source and a camera on the end for transmitting an image from the tip of the instrument to an eyepiece or video camera at the opposite end. Using Bowden cables, which are connected to a lever at the hand piece, the tip of the instrument can be oriented (Radosevich, 2013). This allows the surgeon to navigate the instrument into individual lobe or segment bronchi and examine the patient's airways for abnormalities such as foreign bodies, bleeding, tumors, or inflammation (Nakhosteen et al., 2009). Since flexible bronchoscopy only requires local anesthesia, mild tranquilizers and sides effects usually occur in a mild form, patients will have less discomfort compared to open surgery (Ni, Lo, Lin, Fang & Kuo, 2010).

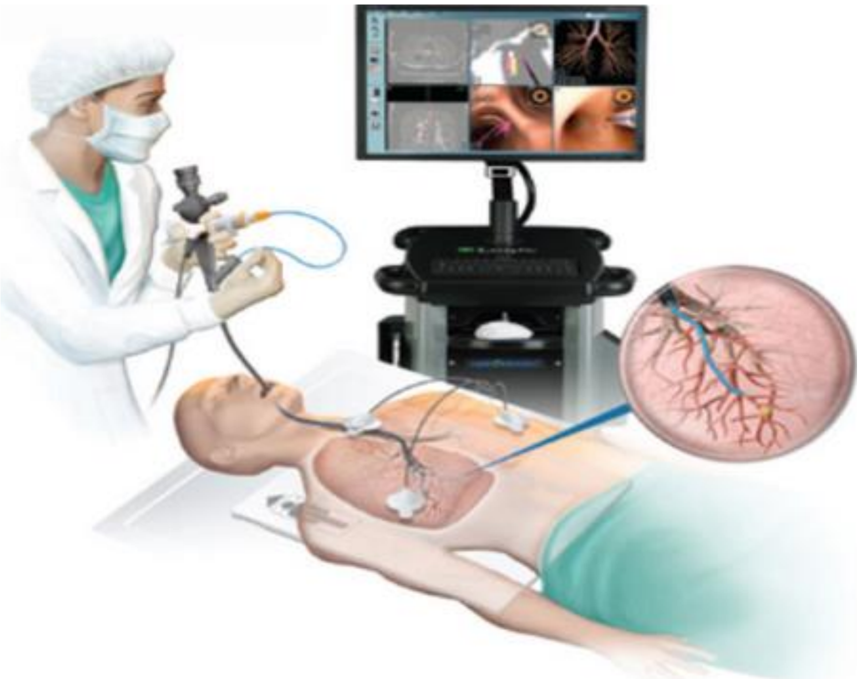


Figure 2. Simulation of a flexible bronchoscopy. Retrieved from https://www.sydneyrespiratoryspecialist.com.au/flexible-bronchoscopy.html on 22-06-2018

However, with regard to the patient’s risks bronchoscopy cannot be compared with other endoscopic (especially gastroenterological) procedures. The risk of complications is increased because of the surgical intervention on a vitally necessary organ, which requires a high safety standard (Geraci et al., 2007). The lung is one of the vital organs in the human body as it is a central component of breathing and therefore indispensable for life. If the respiratory tract or pulmonary alveolus are injured during bronchoscopy, it could lead to pneumothorax, the collapse of the lung, which is life threatening. Unlike the kidneys, the stomach or the intestine, for example, no machine can permanently replace its function. If the lungs fail, only a transplant will help (Pereira, Kovnat & Snider, 1978). Therefore, surgeons have to dispose a wide spectrum of skills, such as spatial ability, perceptual motor skill and complex surgical motor skills, which are needed for performing bronchoscopy. The integration of muscle function, strength, speed, precision, dexterity, balance and spatial perception, makes bronchoscopy a highly complicated technical skill for surgeons (Silvennoinen, Mecklin, Saariluoma & Antikainen, 2009).

Perspective Shifting

A result of computer technology guiding the surgical gesture is the dramatically reduced depth perception in minimally invasive surgery (Norton & Ischy, 2017). During bronchoscopy, a surgeon observes the endoscopic camera picture of the bronchoscope on a monitor and is guided by this. Therefore, the perspective of surgeons is different during bronchoscopic procedures than during traditional open surgery. Because of the perspective change from an allocentric (object-to-object) perspective like in traditional operations, to an egocentric (self-to-object) perspective, as during a bronchoscopy, the cognitive ability of processing visual information about spatial relations between objects and performing mental spatial transformations and manipulations, is an additional skill surgeons have to possess. An allocentric coordinate system represents locations as coordinates in a system centered on entities other than a navigator, such as an object array and the surrounding room. A higher spatial ability refers to the egocentric coordinate system locations, which are represented relative to the body-orientation of a navigator (Klatzky, 1998). During the execution of bronchoscopic procedures from the egocentric perspective, immersion plays an important role, as it provides adequate information for building a spatial reference frame crucial for high-order motor planning and egocentric encoding (Slater & Wilbur, 1997).

Hand-Eye Coordination

With regard to the perspective change in bronchoscopic procedures compared to traditional open surgery, the impaired hand-eye coordination plays a crucial role in performing bronchoscopy. The hand-eye coordination allows the hands to be guided by the visual feedback the eyes receive. It is the coordinated control of eye movement with hand movement (Wentink, 2001). Spatial information help the motor cortex to determine which hand movements are needed to perform the task, and to produce stimulation signals for the muscles in the upper-arms and forearms of the surgeon. Through the process of visual motor transformation, which is responsible for the decision, which muscles need to be stimulated to produce a desired hand movement that fits to the retinal image, the stimulation signals are produced (Dankelman, Grimbergen & Stassen, 2004).

However, during bronchoscopy a main difficulty is the impaired hand-eye coordination for the surgeons. In traditional open surgery, the coordination of hand movements is based on a direct view on the hands and the resulting mapping between action and perception is well known to the brain. During a bronchoscopy, however, the direct view on the hands is replaced by an indirect view via a camera picture of the bronchoscope on a monitor. Usually, the bronchoscope has a different point of view than the natural point of view of the surgeon’s eyes (Wentink, 2003). In addition, the hands are replaced by instruments, which is responsible for a reduced haptic feedback. As a result, the mapping between action and perception is significantly changed and a relatively long learning curve is required for the brain to adapt to the changing mapping during bronchoscopy, as the union of visual and motor skills during bronchoscopy represents a complex cognitive ability (Arsenault & Ware, 2000).

Speed Accuracy Tradeoff

While performing perceptual-motor tasks, there is a tradeoff between how fast a task can be performed and how many mistakes are made in performing the task. When asking people to perform the perceptual-motor task as well as possible, they have to negotiate between the competing demands of response speed and response accuracy and will probably apply various strategies which may optimize speed or accuracy, or which may optimize speed and accuracy together (Bogacz, Wagenmakers, Forstmann & Nieuwenhuis, 2010). According to the speed-accuracy tradeoff, people who finished a task very fast will probably make many errors. On the other side, people who perform a task very slow will have only a few errors (Fairbrother, 2010).

If the speed accuracy tradeoff exists during the execution of bronchoscopic task and if it is an interindividual factor, both performance variables time on task and wall contact have to be considered. Otherwise, if only the performance variable time on task is considered, a person who tries to make few mistakes will be disadvantaged. Especially previous researches, which study the practice of cognitive skills, concentrate on improvements in response times (RT) and therefore only concentrate on the performance variable time on task (Liu &

Watanabe, 2012). Regarding the surgeon's selection process, this would mean that only surgeons who will be very fast would be preferred, regardless of the variety of mistakes they make. Exactly this decision would be life threatening for patients during bronchoscopy, as a simple injury to the lung tissue can cause the lung to fail (see flexible bronchoscopy). Therefore, the disregard of a possible speed accuracy tradeoff during the execution of bronchoscopic task can have fatal consequences for future research in estimating surgeon's potential. Comparing the performance during the execution of bronchoscopic task of different people cannot be done based on speed or accuracy alone, but both values need to be known.

However, there is an additional possibility regarding the relationship between the performance variables time on task and wall contacts. If people make many mistakes during the execution of a perceptual-motor task and yet are slow, what would be expressed in a linear relationship, it would imply that there are two reasonably independent measurements for the same latent ability. That is, the manifest variables such as time on task and wall contact would be indicators of a (postulated) latent dimension (Harvey & Hammer, 1999).

Previous Studies on the Prediction of MIS-Performance

A current study of Arendt, Schmettow and Gronier (2017) approached the question whether a reliable and valid prediction of MIS-performance with basic laparoscopic tasks in the LapSim and low-fi dexterity tasks is possible, which would allow systematic and controlled ways of selection and assessment for surgeons. Two dexterity tasks and four basic laparoscopic tasks in the LapSim were provided for the participants, which they had to repeat a predefined number of times. Exponential learning curves were estimated per participant and task. The primary measurement of talent for technical laparoscopic skills was the population-average maximum performance parameter, based on time-on-task. For assessing the internal consistency reliability inside a test suite and validity between test suites a pairwise correlations have been calculated. The participant-level maximum performance parameters were extracted to make statements about the feasibility of psychometrics for prediction of technical laparoscopic skills.

However, Arendt, Schmettow and Gronier (2017) found that the correlation between the two dexterity tasks was small and the correlations between the four basic laparoscopic tasks in the LapSim were small to medium. Moreover, correlations between the two sets of tasks were small to non-existent. However, individual differences in maximum performance have been found.

Research question

As previously models for estimating individual's surgical potential are mostly based only on the performance variable time on task, this research especially focus on the speed accuracy tradeoff during the execution of bronchoscopic tasks on a low-fidelity boxtrainer. We thus assessed the performance variable time on task and wall contact. If a speed-accuracy tradeoff can be obtained within training on the low-fidelity boxtrainer, this must be necessarily taken into account in further research, in order to ensure an adequate selection process of suitable surgeons. Therefore, our research question is whether there is an association between the two performance variables time on task and wall contact.

Method

Participants

A convenience sample consisting of 24 students from the University of Twente was taken. The students were recruited via the SONA system of the University. After participation, they have received two credit points. The Ethics Committee at the Faculty of Behavioral Science (BMS) of the University of Twente assessed the research as being ethically. Overall, 21 persons of the sample were students of psychology and three persons of the sample were students of Communication Science. All of them were students from the University of Twente. In total, 17 women and seven men took part in this research. There were 20 German participants, two participants were Dutch, one was Bulgarian and one was Iranian. The average age was 22.58 years (*min.* = 19, *max.* = 28, M = 22.58, SD = 2.08). None of the participants had previous experiences in the field of endoscopy. In addition, 23 students were right-handed and one student used both hands equally. A total of 19 students have no impairments. Only one student has color blindness and four students wore glasses.

Design

All participants performed trials on a low-fidelity endoscopic prototype (boxtrainer). Stopping rule is time. All performed task from an egocentric and allocentric perspective.

Materials

Demographic questionnaire. A questionnaire was designed via Survey Monkey asking about demographics, like gender, age, nationality as well as about prior knowledge and experiences in the field of endoscopy, handedness, color blindness, motion sickness, disruption of sensory integration, and limitation of visual strength.

Endoscopic prototype. A low-fi endoscopic prototype (see Figure 3) was designed for testing skills necessary for performing bronchoscopy. A rectangular polymer box, which was measured 17.5 x 30 cm, simulated the human bronchial system in a simplified way. The four outside walls of the box were covered with homogenous holes of 6.5 mm diameter. Every broadside of the box contained 20 holes and every longitudinal side contained 36 holes in total. The number ‘1’ at the broadside of the box indicated the starting point of the first session. A row under the first starting point a second starting was located, which was indicated by a ‘2’ (see Figure 3). The box was prepared with a dividing wall, which could be manually, and variable placed along the width of nine holes inside the box. Through this flexible dividing wall consisting of small openings next to each other, the inside of the human bronchial system were simulated as the bronchi or bronchioles continue to branch out with different distances to the smallest alveoli. This dividing wall represented exactly the broadsides of the box, by covering the same number of holes in the same position and with the same diameter. However, after a pilot test was done, the holes of the dividing wall were broadened to a 12.5 mm diameter for the second part of the box training, because of the high degree of severity. For reaching a better resemblance to the simulator tasks, researcher gave the participants fixed routes through fixed numbers at each hole of the dividing wall of the endoscopic prototype. A coloring pattern was applicate on the dividing walls for reasons of orientation, instruction simplicity and comprehension of participants (see Figure 4). In addition, the colors represented different levels of difficulty: yellow = very simple, green = simple, blue = moderate, red = little difficult, black = very difficult. Another additional aspect was to improve measurability. Through the coloring, a better visual contrast was reached in the black box, which helped observing the success of reaching a destination via the video recordings.

To simulate the flexible endoscope needed for performing bronchoscopy, a USB Android wire camera (see Figure 5) was used as it represents a provisional and simplified version of a real endoscope. This device was equipped with Led and offered the possibility of HD recordings. The provisional-endoscope had a diameter of 5.5 mm diameter and a length of 2 m. The device could be connected to Android and Windows XP/VISTA/7/8 and enabled to take snapshots and video recordings in sound and vision. In addition, filming was done with the software “ViewPlayCap”.

Figure 3. Low-fi endoscopic prototype

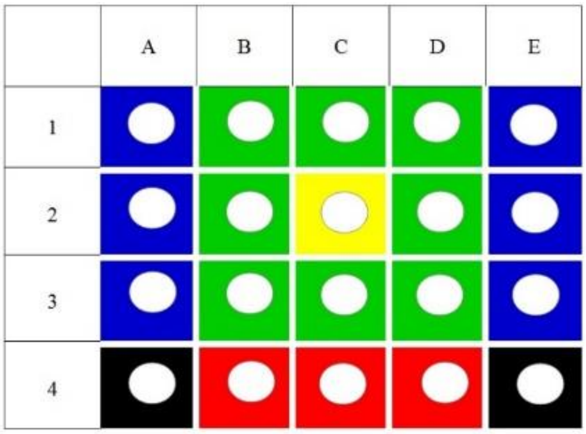


Figure 4. Dividing wall: Colors show levels of complexity

Figure 5. Provisional endoscope: USB Android wire camera

Figure 6. Endoscope introduced in the box

Procedure

Location. Experiments took place in the MIS-simulator room 2 in the Experimental Center for Technical Medicine (ECTM) at the University of Twente. The room had a good lightening and consisted of partitions, which could avoid possible distractors and allow silence during the execution of the boxtraining.

Greetings & instructions. Participants were greeted and thanked for their participation. They received information about the nature of the research, the different tasks to be performed and related rules via verbal instructions (see Appendix 3), as well as via the informed consent (see Appendix 1), which they had to carefully read and sign at first. Before starting the first session, participants were asked to fill in an online demographic questionnaire, to be sure, that they did not have experiences in the field of endoscopy or visual impairments, which were the exclusion criteria of the research. If there were any questions, they were answered carefully.

First session - dexterity task I -allocentric view. Participants of the first group were asked to perform simple dexterity tasks on the endoscopy-prototype, which rose in complexity with the amount of repetitions. At the beginning of the first session, the wall was placed at the third hole on the longitudinal side, approximately 8.0 cm from the starting point. Now participants were instructed to insert the endoscope through all openings on the wall one time following fixed routes, which were provided by the researchers and which switched between the sequence of complexity, presented by different colors (yellow = very simple, green = simple, blue = moderate, red = little difficult, black = very difficult). Expected time for this second sequence amounted around ten minutes. In a second sequence, the dividing wall of the endoscopic prototype was placed on the second row, counted on the longitudinal side. With reduced freedom of movement, the task becomes more difficult. The exact distance from the starting point ‘1’ to the dividing wall was 5.0 cm. Participants had to insert the provisional endoscope first one time through the yellow and then through all green holes, dependent on the fixed route, which were provided by the researcher. This task sequence was expected to take around ten minutes. Ultimately, in the third sequence, the diving wall was placed in the fourth row with an exact distance of 10.5 cm to the starting point ‘1’. This sequence was also expected to take around five minutes. The expected time for the first dexterity task of the first session was approximately 25 minutes.

First session - dexterity task II - egocentric view. In the second half of the first session, the perspective and position of the participants was changed. The box was turned upside down and participants viewed the movements of the endoscope through an integrated camera from an egocentric perspective on a laptop screen via the ‘ViewPlayCab’ software. Participants did dexterity task 2 in a standing position in order to reach a better resemblance of the simulator for bronchoscopy. The starting point ‘2’ was chosen for this procedure, which was one row under the starting point ‘1’. Because the box had now been turned over, starting point ‘2’ was equivalent to starting point ‘1’ from the first half of the first session (see Figure 7). Since the egocentric view represented a higher level of difficulty, because the location of objects in space are relative to the body axes of the self and not like the allocentric view relative to other objects, an adjusted wall with bigger openings (12.5 mm) was positioned on the third row. Participants again were asked to insert the endoscope into all twelve openings (yellow, green and red holes) one time, dependent on the fixed route they got from the researcher. The expected duration for this second dexterity task was expected to take around 30 minutes. Altogether, the expected time for the whole first session was 60 minutes.

Figure 7. Upturned box for the second half of the boxtraining.

Debriefing. Participants were asked about their experiences and if they liked the training on the low-fi simulator for bronchoscopy. In addition, they were asked if they had any questions and if they would like to receive their results via email. It was also stated, that questions, which would arise later, would be answered via email, as it is described in the informed consent. The data participants would receive via email would include the data of the performance variables, such as the time on task, wall contact and task success. In this regard, it was mentioned again that all data would be processed in a confidential and anonymized way by the researcher. Finally, participants were thanked for participation.

Measurements

The original three main variables relevant for the current research paper were task duration (time on task), failures of touching the walls of the box representing human tissue (wall contact) and success of reaching or passing a goal (task success). Time on task was measured with a stopwatch. Task success and wall contact were recorded via observation by the researcher. However, after the estimation of learning curves based on the three performance variables failed, we only focused afterwards on the two performance variables time on task and wall contact. Collected data was written down in a participant protocol. Data will be saved via SharePoint.

Statistical Analysis

The original research plan involved learning curves. It turned out that these were practically inestimable, as no learning seems to have taken place. For the sake of transparency, we describe the planned model, first, and then continue to describe the linear multi-level model, that was used, instead.

A non-linear multilevel mixed effect model with an exponential learning curve was approached in order to create regression models for the learning curves of our participants. According to Heathcote, Brown & Mewhort (2000) a learning curve is composed of the following formula:

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

The asymptote refers to the performance level a participant is expected to achieve after prolonged training. In addition, the amplitude parameter describes the actual rate of learning and improvement. The third parameter is the rate parameter, which represents the general speed of learning (Arendt, Schmettow & Groenier 2017). In this study, learning curves should display the relationship between the aspired performance variables time on task, task success and wall contact on the vertical axis and experience (number of trials) on the horizontal axis.

The nonlinear multilevel mixed effect model has been built with the package ‘brms’ for the statistical programming language R 3.4.4, which provide an alternative type of analysis for univariate or multivariate analysis of repeated measures. Multilevel models are able to estimates individual learning curves, which can hence differ in all three parameters. It is based on the within-subject design by estimating individual learning curves for all participants per task and then bundles them for an analysis on population-level, which allows the measuring of the variation in a population. The LARY model was created whose parameters (amplitude, rate and asymptote) were linearized through link functions and running on a log-scale ranging from $-\infty$ to $+\infty$. Through this process random effects can be obtained, which show the variance caused due to individual differences. Because the random pattern of response times is left-skewed and the variance residuals decreases by approaching the asymptote, we chose the Gamma distribution instead of the Gaussian distribution. Ultimately, the correlations between each task’s population-level maximum performance parameter have been calculated along with their estimated 95% credibility intervals. For analyzing performance during boxtraining, the most important parameter for our analysis was the maximum performance.

Because the estimation of learning curves failed, we used instead the linear multi-level model, where trials become exchangeable repeated measures, in order to explore the relationship between the performance variables time on task and wall contact. First of all a violin plot was made for demonstrating the continuous distribution of the routes (1-20). In a following step a scatter plot of slopes for route (fixed effect + random effects) and task (fixed effect + random effects) was obtained by fitting the multi-level linear model. In order to establish whether there is a statistically significant relationship between the two performance variables time on task and wall contacts, a simple linear regression plot was created. However, because linear models make assumptions that are never truly met by real data, we chose in a following step for a generalized linear model (GML) with a poisson distribution, which can re-established linearity through link functions, and allows the variance of each measurement to be a function of its predicted value. With this, we wanted to obtain learning effect on the population-level. Through the linear predictor scale, we compared the population-level effects to the standard deviation of the individual deviations of the participant level in order to obtain random effects, which show the variance caused due to individual differences. For calculating the linear association between time on task and wall contact and the interaction effect route by task, a further model was approached that accounted for overdispersion, by using the negative binomial distribution instead of Poisson. It showed fixed-effects on population-level, which was transformed back to the original scale through the link function. Comparison between population-level effects to the standard deviation of the individual deviations were done on a linear predictor scale, as standard deviations cannot be transformed to original scale.

Results

Explorative Analysis

Non-linear mixed-effect regression. The estimation of learning curves failed. The effects for the variable time on task on maximum performance showed an unexpected result, namely that performance did not improve after consecutive trials. The needed time for carrying out the four different dexterity tasks on the low-fi simulator from the allocentric and egocentric perspective varied among the trials but become neither less on a population level, nor on a participant level (see Figure 8, 9 and 10). Therefore, this model did not converge and the predictor trial as an independent variable of the non-linear mixed-effect model could be disrespected. Simultaneously this meant that the asymptote could not be pursued any further and therefore, in this study the individual’s bronchoscopic skill acquisition could not be predicted through the estimation of learning curves.

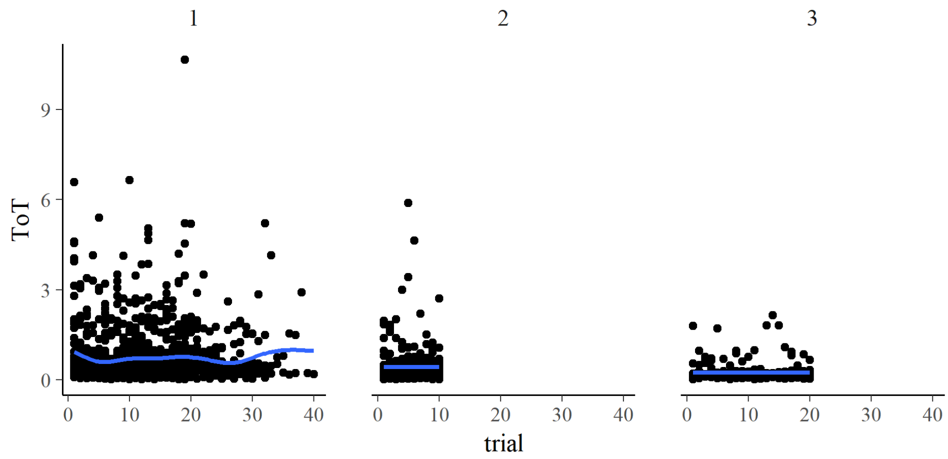


Figure 8. Raw learning curve, estimated on a population level, displays the relationship between the learning variable (time on task) on the vertical axis and experience (related to the number of trials) on the horizontal axis.

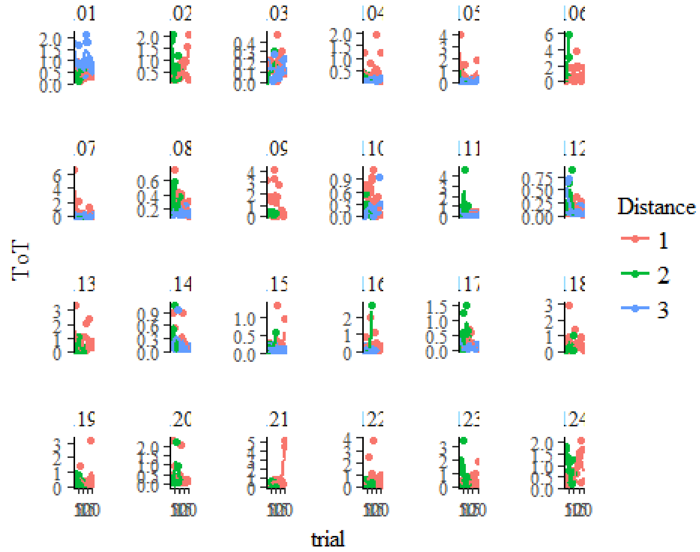


Figure 9. Raw learning curve, estimated on a participant level, displays the relationship between the learning variable (time on task) on the vertical axis and experience (related to the number of trials) on the horizontal axis for tasks carried out from the allocentric perspective.

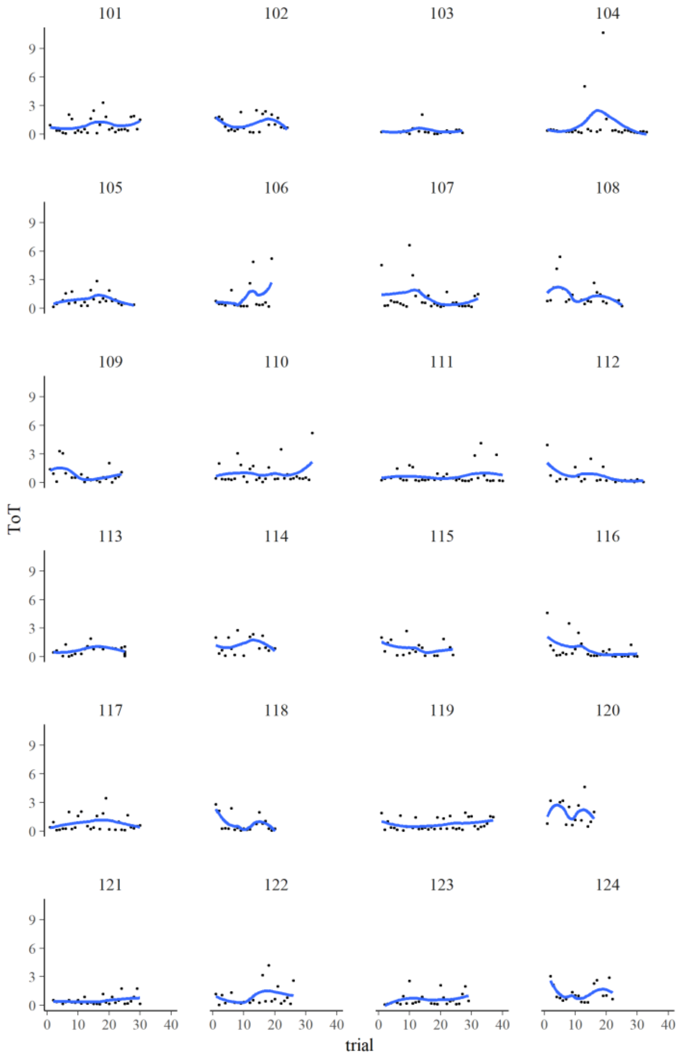


Figure 10. Raw learning curve, estimated on a participant level, displays the relationship between the learning variable (time on task) on the vertical axis and experience (related to the number of trials) on the horizontal axis for tasks carried out from the egocentric perspective.

However, the problem simplifies to a multi-level linear model where trials become exchangeable repeated measures, which leads to our main question: how are the two performance indicators, wall contact and time on task related. The model has shown that the population effects and random effects are almost complete flatlines (see Appendix 7).

Linear association between time on task and wall contact and interaction effect route by task. A further model was used in order to complete the previous by adding the possibility that the route effects differ by task. This model accounted for overdispersion by using the negative binomial distribution instead of Poisson. Table 3 shows the fixed-effects on population-level, which are transformed back to the original scale through the link function. This means that the value of the intercept represents the numbers of wall contact, which are multiplicative. A clear effect of time on task on wall contacts is dramatic, as with every minute longer, the number of wall contacts multiplies by 1.77. A comparison between the population-level effects to the standard deviation of the individual deviations is done on a linear predictor scale, as standard deviations cannot be transformed to original scale. By comparing the population mean of the linearized scale (0.55) with the standard deviation of the individual deviations (0.24) one can be relative certain that all participants have a positive slope (see Table 4).

Table 3. Showing fixed-effects on population-level which are transformed back to original scale

fixef	center	lower	upper
Intercept	1.2474457	0.8273745	1.8173290
cToT	1.7726131	1.5533064	2.0606988
Taskallo_2	1.2207341	0.6539765	2.1196508
Taskallo_3	0.5067341	0.3619497	0.6950428
Taskego_1	0.7398900	0.5305070	1.0052777

Estimates with 95% credibility limits

Table 4. Comparison between population-level effects to the standard deviation of the individual deviations on linear predictor scale, as standard deviations cannot be transformed to original scale

fixef	center	lower	upper	center_sd	lower_sd	upper_sd
<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1 Intercept	0.221	-0.189	0.597	0.553	0.394	0.795
2 cToT	0.552	0.440	0.723	0.241	0.105	0.416

Estimates with 95% credibility limits

Visualization of results. When looking at Figure 11 it is obvious that the number of wall contact is positively associated with ToT for all and every participant.

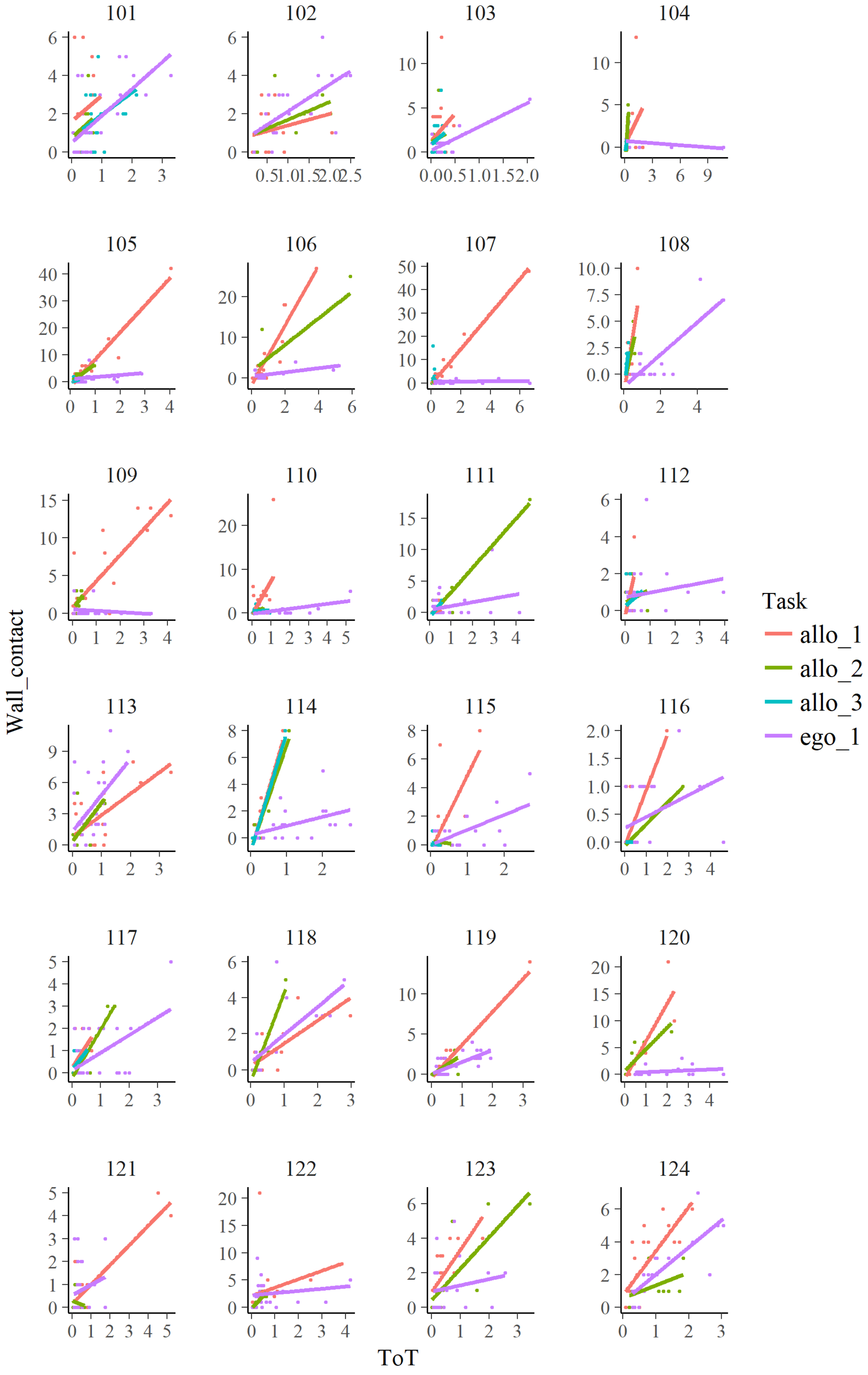


Figure 11. Linear regression model, showing the relationship between the performance variable time on task (depicted on the x-axis, and wall contact (depicted on the y-axis).

Discussion

Our original research goal was to find out whether previous training on a low-fidelity inanimate box trainer, as a simplified version of the VR-simulator for bronchoscopy, can improve the simulator task performance. Moreover, we wanted to know if the three performance variables time on task, task success and wall contact are correlated in such a way that combining the measures yields a more precise estimate of performance. Therefore, we sought to estimate individual learning curves, as they provide an adequate way to know exactly when someone is reaching their maximum performance level. In addition, learning curves make it possible to visualize individual's differences regarding the bronchoscopic skill acquisition (Heathcote, Brown & Mewhort, 2000). Certainly, the estimation of learning curves failed. It pointed out that the predictor trial had no influence on the performance variable, which automatically denies us the possibility for predicting individual's skill acquisition by approaching the asymptote.

However, these unexpected problems allowed us to examine different aspects of performance, such as the speed accuracy tradeoff, without the difficulty of the estimation of learning curves. Therefore, our current goal of this research was to explore the association between the performance variables time on task and wall contact.

Consequently, the discussion will emerge around two questions: why no learning took place and what do the results tell about the association between the performance measures.

Differences between low-fi endoscopic simulator training and VR- simulator training

To approach the question why there were no learning effects visible after training on the low-fidelity boxtrainer we first examine the differences between the low-fidelity boxtrainer and the VR-simulator for bronchoscopy.

The first difference refers to the coarseness of the box compared to the very sensitive VR-simulator. Smallest movements on the VR-simulator could lead to great tissue injuries, which are made visible. In contrast, if the participant slipped the endoscope during box training and touched the walls, there was only a slight noise. In addition, it was partly necessary to apply force to push the endoscope through the narrow holes during box training. Moreover, the holes of the box were all the same size, whereas the different openings of the bronchi got tighter as the endoscope gets deeper into the lungs. The difference of coarseness can also be transferred to the endoscope itself. During box training participants get a kind of a thicker cable as a simplified version of an endoscope, which they moved with their fingertips. In contrast, the VR-simulator allows a real bronchoscope, which participants held in their palms and moved the flexible tip of the end of the endoscope with a fingertip switch. These differences could lead to the assumption, that participants will learn fine motor skills during simulator training, which is not required during training on our low-fidelity box. According to Chung et al., 2017, training on a VR-simulator for learning laparoscopic skills will improve fine motor skills after prolonged training. Their goal was to determine the effect of fine motor activity and nondominant-hand training by medical students. Students have to perform three surgical simulator tasks: navigation, forceps, and bimanual. All showed statistically significant improvements in all three tasks at follow-up after a single baseline evaluation on the surgical simulator.

In addition, there is a difference with regard to the achievement of the holes, which simulate the different openings of the bronchi. During the box training, participants just have to pierce the endoscope straight through the hole by moving the endoscope with their fingertips in the right position, while participants have to rotate the endoscope of the VR-simulator partially up to 360 degrees with their hand and by moving the whole body in order to move it in the right direction. This leads to the assumption, that the box trainer is maybe too simple, as it demands less degree of freedom.

Analysis of the results

Because the LARY model for estimating proper learning curves for the participants based on the performance variable time on task, task success and wall contact did not converge, we chose in a following step for multi-level linear model to analyses the correlation between the performance variables. The results stated that there is a positive linear association between the performance variables time on task and wall contact on the population-level, as well as on the participant-level, which means, that the longer participants needed for the execution of the bronchoscopic tasks, the more mistakes they made. Therefore, the speed-accuracy tradeoff does not seem to be exist, since it claims that persons who need longer for performing a task, probably are more accurate (Fairbrother, 2010).

At the same time, the question arises, whether the relationship can be purely caused by time for error recovery. That would be the case, if a wall contact causes a severe delay in task completion. However, this can be completely ruled out, since slipping with the endoscope lasted no more than 2 seconds and the participant could continue directly with his task without having to start again.

Another cause for the linear association could result from the verbal instruction the participants got from the researcher: ‘Please try as fast as possible to achieve the trials and thereby make as few mistakes as possible.' If participants had noticed that they needed a longer time for the execution of one task, they were impatient and wanted to hurry and thus they did more errors. However, this could almost be completely ruled out, as there was a linear relationship between time on task and wall contact by each individual participant. In this case, they all must have become restless during box training, because they have needed too long for a task, and would have subsequently tried to hurry and therefore have to make more mistakes. This would be unlikely.

A possible explanation for the linearity of the two performance variables time on task and wall contacts could therefore be the existence of a third latent variable, like ability for example. This would imply that there are two independent measurements for the same latent ability. Therefore, the manifest variables time on task and wall contact would be both indicators of a (postulated) latent ability, which would be great finding.

Possible limitations of the experiment

Participants mentioned multiple utterance of vertigo during the endoscopic task. One third of the participants had difficulty looking at the screen for 30 minutes to get their bearings and move the endoscope to the correct hole. Two of them had to stop the execution of the endoscopic task for a few minutes. These phenomena occur more frequently in the context of virtual reality exercises (Schuemie, Van Der Straaten, Krijn & Van Der Mast, 2001).

Further research

As our results are based on ad hoc measures because of the unexpected problems, the speed accuracy tradeoff should be further examined by considering the two performance variables time on task and wall contacts during the execution of bronchoscopic tasks on the low-fidelity box trainer. Especially in the selection process of suitable surgeons the two performance variables should be further examined in order to get a realistic picture of the potential of the surgeons. Therefore, an adequate method for testing the speed accuracy tradeoff should be to give people the instruction, to perform the bronchoscopic task on the box trainer as fast as possible and then people should execute the bronchoscopic task as accurately as possible (Bogacz, Wagenmakers, Forstmann & Nieuwenhuis, 2010).

Conclusion

In conclusion, it can be stated that the boxtrainer is not an adequate substitute for the VR-simulator. Although this simulator has a certain resemblance to the virtual reality simulator, which is a current trainings method for learning bronchoscopic skills, it also differs from it in any aspects, which can be responsible for the absence of learning. However, these allows us to examine further aspects of performance on the low-fidelity box, such as the speed accuracy tradeoff, which can make a great profit with regard to an adequate selection process of surgeons.

References

Adamovich, S. V., Fluet, G. G., Tunik, E., & Merians, A. S. (2009). Sensorimotor training in virtual reality: a review. *NeuroRehabilitation*, *25*(1), 29-44.

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. Retrieved from http://essay.utwente.nl/73211/1/arendt_MA_bms.pdf

Arsenault, R., & Ware, C. (2000). Eye-hand co-ordination with force feedback. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 408-414). ACM

Basdogan, C., Sedef, M., Harders, M., & Wesarg, S. (2007). VR-based simulators for training in minimally invasive surgery. *IEEE Computer Graphics and Applications*, *27*(2).

Bogacz, R., Wagenmakers, E. J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed–accuracy tradeoff. *Trends in neurosciences*, *33*(1), 10-16.

Chung, A. T., Lenci, L. T., Wang, K., Collins, T. E., Griess, M. D., Oetting, T. A., & Shriver, E. M. (2017). Effect of fine-motor-skill activities on surgical simulator performance. *Journal of Cataract & Refractive Surgery*, *43*(7), 915-922.

Colt, H. G., Crawford, S. W., & Galbraith, O. (2001). Virtual reality bronchoscopy simulation: a revolution in procedural training. *Chest*, *120*(4), 1333-1339.

Dankelman, J., Grimbergen, C. A., & Stassen, H. G. (Eds.). (2004). *Engineering for patient safety: issues in minimally invasive procedures*. CRC Press.

Diesen, D. L., Erhunmwunsee, L., Bennett, K. M., Ben-David, K., Yurcisin, B., Ceppa, E. P., & Pryor, A. (2011). Effectiveness of laparoscopic computer simulator versus usage of box trainer for endoscopic surgery training of novices. *Journal of surgical education*, *68*(4), 282-289.

Dorr, L. D., Maheshwari, A. V., Long, W. T., Wan, Z., & Sirianni, L. E. (2007). Early pain relief and function after posterior minimally invasive and conventional total hip arthroplasty: a prospective, randomized, blinded study. *JBJS*, *89*(6), 1153-1160

Fairbrother, J. T. (2010). *Fundamentals of motor behavior*. Champaign, IL: Human Kinetics.

Fuchs, K. H. (2002). Minimally invasive surgery. *Endoscopy*, *34*(02), 154-159.

Gaba, D. M. (2004). The future vision of simulation in health care. *BMJ Quality & Safety*, *13*(suppl 1), i2-i10.

Gallagher, A. G., Ritter, E. M., Champion, H., Higgins, G., Fried, M. P., Moses, G., ... Satava, R. M. (2005). Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Annals of surgery*, *241*(2), 364.

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.

Geraci, G., Pisello, F., Sciumè, C., Li, F. V., Romeo, M., & Modica, G. (2007). Complication of flexible fiberoptic bronchoscopy. Literature review. *Annali italiani di chirurgia*, *78*(3), 183-192.

Hassan, I., Dayne, K. B., Kappus, C., Gerdes, B., Rothmund, M., & Hellwig, D. (2007). Comparison of minimally invasive surgical skills of neurosurgeons versus general surgeons: is there a difference in the first exposure to a virtual reality simulator?. *min-Minimally Invasive Neurosurgery*, *50*(02), 111-114.

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

Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In *Spatial cognition* (pp. 1-17). Springer, Berlin, Heidelberg.

Konge, L., Arendrup, H., Von Buchwald, C., & Ringsted, C. (2011). Using performance in multiple simulated scenarios to assess bronchoscopy skills. *Respiration*, *81*(6), 483-490.

Liu, C. C., & Watanabe, T. (2012). Accounting for speed–accuracy tradeoff in perceptual learning. *Vision research*, *61*, 107-114.

Munz, Y., Kumar, B. D., Moorthy, K., Bann, S., & Darzi, A. (2004). Laparoscopic virtual reality and box trainers: is one superior to the other?. *surgical endoscopy and other interventional techniques*, *18*(3), 485-494.

Nakhosteen, J. A., Khanavkar, B., Darwiche, K., Scherff, A., Hecker, E., & Ewig, S. (Eds.). (2009). *Atlas und Lehrbuch der thorakalen Endoskopie: Bronchoskopie, Thorakoskopie*. Springer-Verlag.

Ni, Y. L., Lo, Y. L., Lin, T. Y., Fang, Y. F., & Kuo, H. P. (2010). Conscious sedation reduces patient discomfort and improves satisfaction in flexible bronchoscopy. *Chang Gung Med J*, *33*(4), 443-52.

Norton, M. J., & Ischy, N. D. (2017). *U.S. Patent No. 9,820,771*. Washington, DC: U.S Patent and Trademark Office.

Pereira, W., Kovnat, D. M., & Snider, G. L. (1978). A prospective cooperative study of complications following flexible fiberoptic bronchoscopy. *Chest*, *73*(6), 813-816.

Pisano, G. P., Bohmer, R. M., & Edmondson, A. C. (2001). Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery. *Management Science*, *47*(6), 752-768.

Pusic, M. V., Boutis, K., Hatala, R., & Cook, D. A. (2015). Learning curves in health professions education. *Academic Medicine, 90*(8), 1034-1042.

Radosevich, J. A. (Ed.). (2013). *Head & neck cancer: Current perspectives, advances, and challenges*. Springer Science & Business Media.

Rogalla, P., Rückert, J. C., Schmidt, B., Witt, C. H., Meiri, N., & Hamm, B. (2001). Virtuelle Bronchoskopie. *Der Radiologe*, *41*(3), 261-268.

Rosen, J., Brown, J. D., Barreca, M., Chang, L., Hannaford, B., & Sinanan, M. (2002). The blue dragon-a system for monitoring the kinematics and the dynamics of endoscopic tools in minimally invasive surgery for objective laparoscopic skill assessment. *Studies in health technology and informatics*, 412-418.

Rosen, M. J., & Ponsky, J. L. (2006). Minimally invasive surgery, 2004-2005. *Endoscopy*, *38*(02), 137-143.

Rosser, J. C., Murayama, M., & Gabriel, N. H. (2000). Minimally invasive surgical training solutions for the twenty-first century. *Surgical Clinics*, *80*(5), 1607-1624.

Schell, S. R., & Flynn, T. C. (2004). Web-based minimally invasive surgery training: competency assessment in PGY 1-2 surgical residents. *Current surgery*, *61*(1), 120-124.

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.

Schuemie, M. J., Van Der Straaten, P., Krijn, M., & Van Der Mast, C. A. (2001). Research on presence in virtual reality: A survey. *CyberPsychology & Behavior*, *4*(2), 183-201.

Silvennoinen, M., Mecklin, J. P., Saariluoma, P., & Antikainen, T. (2009). Expertise and skill in minimally invasive surgery. *Scandinavian Journal of Surgery*, *98*(4), 209-213.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *6*(6), 603-616

Wentink, B. (2001). Eye-hand coordination in laparoscopy-an overview of experiments and supporting aids. *Minimally Invasive Therapy & Allied Technologies*, *10*(3), 155-162.

Wentink, M. (2003). Hand-eye coordination in minimally invasive surgery: Theory, surgical practice & training, 2-3.

Appendices

Appendix 1 - Informed Consent

<u>Section A:</u>

Protocol number: ________________ Participant number: __________________

Participant name: _____________________

Dear participant,

We are Ace Küpper and Lisa Mührmann and we are currently writing our master and bachelor thesis in "Human Factors and Engineering Psychology" at the University of Twente. Our topic is "Learning minimally invasive surgery" and we want to test whether a specific training of dexterity tasks on a low-fi endoscopic prototype can influence the simulator task performance of bronchoscopy. We are going to give you information and invite you to be part of this research. Please ask us to stop as we go through the information and we will take time to explain.

Purpose of the research

Minimally invasive surgery (MIS) is one of the preferred approach in surgical procedures (Rosen & Ponsky, 2006). In comparison to conventional open surgery, MIS offers many advantages like a reduced blood loss, pain, complications, hospitalization time, and improved cosmetic (Hu et al., 2009). However, these differences make performing minimally invasive surgery a great challenge for surgeons, who need a broad spectrum of cognitive and psychomotor skills. It is obvious that not all surgeons can perform minimally invasive surgery as adequate as necessary to reduce the risks for the patients. There are many inter- and intra-personal differences while MIS-training (Pisano, Bohmer & Edmondson, 2001). We want to explore whether a specific training of dexterity tasks on a low-fi endoscopic prototype can influence the simulator task performance of the surgeons.

Voluntary Participation

Your participation in this research is entirely voluntary. It is your choice whether to participate or not. You may also stop participating in the research at any time you choose.

<u>Section B:</u>

Description of the Process

In a first session, you will train different dexterity tasks on a endoscopic prototype.

In a second session, you will train on a professional simulator for surgeons.

Duration

The research consists of two session and each takes approximately one hour.

Confidentiality

The information that we collect from this research project will be kept confidential. Information about you that will be collected during the research will be put away and no one but the researchers will be able to see it. Any information about you will have a number on it instead of your name. Only the researchers will know what your number is and we will lock that information up with a lock and key. It will not be shared with or given to anyone except we both (Ace and Lisa).

Sharing the Results

The knowledge that we get from doing this research will be shared with you via email if you want that. We will publish the results in order that other interested people may learn from our research in an anonymous way.

Certificate of Consent

I have read the foregoing information carefully. I have had the opportunity to ask questions about it and any questions that I have asked have been answered to my satisfaction. I consent voluntarily to participate as a participant in this research.

Print Name of Participant______________________________

Signature of Participant ______________________________

Date ____

Day/month/year

I have witnessed the accurate reading of the consent form to the potential participant, and the individual has had the opportunity to ask questions. I confirm that the individual has given consent freely.

A copy of this ICF has been provided to the participant.

Name of Researcher _________________________________

Signature of Researcher ______________________________

Date ____

Day/month/year

Who to Contact

If you have any questions, you may ask them now or later, even after the study has started. If you wish to ask questions later, you may contact any of the following:

Ace Küpper: a.kupper@student.utwente.nl

Lisa Mührmann: l.muhrmann@student.utwente.nl

Appendix 2 - Demographic Questionnaire

What is your gender? Male / Female

Please enter your date of birth: ____

Please enter your nationality: ____

Please enter your study: ____

What is your dominant hand? Left-hand / Right-hand

Do you have impaired vision (i.e. debility of sight, color-blindness, eye cataract or glaucoma)?

Yes / No

If yes, please give a description: ____

Do you have already made experiences in the field of endoscopy?

Yes / No

If yes, please give a description of the amount of experience:

____

____

____

Name participant: ______________________ Participant number: ________________

Date: ______________________________ Protocol number: ____________________

Appendix 3 - Task Instructions

Session I - Dexterity task I. You may use one or even both hands for the task, but your hands have to stay outside the box all the time. The box may not be moved. You have to sit on a chair while performing the first task and start at the starting point ‘1’. After each trial, you have to go back to the starting point and reorientate and start again. If you have any questions please ask us after the first half of the first session, because we as researcher has to listen to the damages you make and are very concentrated while we fill in the test protocol.

Session I - Dexterity task II. Your view has to be fixed on the screen as you maneuver the endoscope. You will carry out this task while standing. If the endoscope stucks inside an opening, the researcher can help you to extract it or allow you to do it yourself. You start every trial at the starting position ‘2’. If you reached a goal, you turn back to the starting position to begin a new trial. If you completely lose your orientation, you may go back to the starting point to reorientate and start again.

Appendix 4 - Test Protocol Session 1+2

Test Protocol Session 1

Protocol No.: __________ Date: ____________________ Sona No.: ________________

Task 1 - Allocentric Perspective (sitting posture) → 25 Min Task 1 (1.1.-1.3.)

Sub-Task 1.1. : Plate in Line III, colors: all					
Trial No.	Route No.	Task success = wrong hole!	Damage	Time	Skipped
1					
2					
3					
4					
5					
6					
7					
8					
9					

10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
Total					

Notes & observations:

____

____

____

Task 1 - Allocentric Perspective (sitting posture) → 25 Min Task 1 (1.1.-1.3.)

Sub-Task 1.2. : Plate in Line II, colors: yellow & green (No.1-9)					
Trial No.	Route No.	Task success = Wrong hole!	Damage	Time	Skipped
1					
2					
3					
4					
5					
6					
7					
8					

9					
10					
Total					

Task 1 - Allocentric Perspective (sitting posture) → 25 Min Task 1 (1.1.-1.3.)

Sub-Task 1.3. : Plate in Line IV, colors: all					
Trial No.	Route No.	Task success = Wrong hole!	Damage	Time	Skipped
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

Total					

Test Protocol Session 2

Protocol No.: ____________ Date: ____________________ Sona No.: ____________

Task 2 - Egocentric Perspective (standing posture) → 30 Min!

Sub-Task 2.1. : Plate in Line III, colors: yellow, green & red					
Trial No.	Route No.	Task success = Wrong hole!	Damage	Time	Skipped
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
Total					

Appendix 5 – 1. Pilot Test Session 1

Protocol No.: __99________ Date: 09.03.2018_______ Participant No.: ___________

1. Allocentric Perspective

Task	Duration	Damage	Success	Quantity

1. Plate in Line II (only yellow/green)	1.14 Min	1	Yes	Total: 9 Done: 9
2. Plate in Line III	3.29 Min	6	Yes	Total: 20 Done: 20
3. Plate in Line IV	2.34 Min	5	Yes	Total: 20 Done: 20
4. Plate in Line V (as time permits)	2.00 Min	2	Yes	Total: 20 Done: 20
Total	9.17 Min	14	Yes	69

2. Egocentric Perspective (with PC – only green & yellow holes, No. 1-9)

Task	Duration	Success	Task success	Quantity
1. Plate in Line II	8.21 Min	Yes	100,00%	Total: 9 Done: 9
2. Plate in Line III	7.49 Min	Yes	66% (three times false hole = repeat)	Total: 9 Done: 12
3. Plate in Line IV	12.59 Min	Yes	33% (six times false hole = repeat)	Total: 9 Done: 15
Total	28.09 Min	Yes	18 from 27 at once corrrect = 66%	Total: 27 • failures = 9 • = 36

• tested to do all yellow and green openings

• If tested in a range 3 failures/unwanted other openings entered!

• Failures happen fast with fast movements nearly in front of the opening. Piercing may lead to entering another opening

2.1. egocentric view (standing posture) – just yellow and green

Task	Duration	Success	Task success	Quantity
1. Plate in Line II	3.58 Min	Yes	100%	Total: 9 Done: 9
2. Plate in Line III	4.05 Min	Yes	100%	Total: 9 Done:
3. Plate in Line IV	3.42 Min	Yes	100%	Total: 9 Done:
Total	11.45 Min	Yes	27	

• Makes the task much more simple maybe nearly as the allocentric view

• Advantage: Posture resembles the one used for the simulator

• Advantage: to control the task success is much more simple for the researcher

Plenty of additional video recording tests done. Nearly all failed. Camera is necessary!

Appendix 6 – 2. Pilot Test Session 1 + 2

Protocol No.: __100_____ Date: ___27.03.2018__ Participant:______________

Task 1 - Allocentric Perspective (sitting posture)

Sub-Task 1.1. : Plate in Line II, colors: yellow & green (No.1-9)					
Trial No.	Route No.	Task success = Wrong hole!	Damage	Time (in seconds)	Skipped
1	1	No	0	13	No
2	2	No	1	74	No
3	7	No	0	85	No
4	3	No	0	21	No
5	5	No	0	26	No
6	4	No	0	82	No

7	6	No	1	25	No
8	8	No	1	21	No
9	9	No	1	48	No
10	2	No	1	63	No
Total	10	No	5	403	No

Task 1 - Allocentric Perspective

Sub-Task 1.2. : Plate in Line III, colors: all					
Trial No.	Route No.	Task success = Wrong hole!	Damage	Time	Skipped
1	6	No	0	11	No
2	17	No	0	25	No
3	15	No	1	61	No
4	6	No	0	25	No
5	12	No	1	31	No
6	2	No	0	16	No
7	13	No	0	23	No
8	1	No	0	15	No
9	20	No	0	178	No
10	7	No	0	17	No
11	8	No	0	12	No
12	18	No	0	68	No
13	11	No	0	19	No
14	4	No	0	16	No
15	10	No	0	32	No
16	19	No	0	75	No
17	3	No	0	36	No
18	9	No	0	15	No

19	14	No	0	39	No
20	16	No	0	15	No
Total	20	No	2	744	No

Task 2 - Egocentric Perspective (standing posture)

Sub-Task 2.1. : Plate in Line III, colors: yellow, green & red					
Trial No.	Route No.	Task success = Wrong hole!	Damage	Time	Skipped
1	9	No	0	218	No
2	15	No	0	412	Yes
3	4	No	0	120	No
4	18	No	1	366	No
5	13	No	0	327	Yes
6	6	No	0	120	No
7	16	No	1	199	Yes
8	14	No	0	39	No
9	1				
10	8				
11	2				
12	17				
13	10				
14	5				
15	11				
16	12				
17	3				
18	7				
19	20				
20	19				

21	3				
22	2				
23	14				
24	17				
25	18				
26	15				
27	6				
28	9				
29	10				
30	14				
31	9				
32	2				
33	8				
34	7				
35	6				
36	5				
37	3				
38	11				
39	13				
40	12				
Total	40 (only 7 done)	No	2	1827	3x skipped

Notes & Observations:

- participant found first task harder than second- less freedom of movement- change order of tasks
- very slow and very anxious to make a mistake- participant complained of exhaustion
- he did not pay attention to the time- absolutely wanted to reach the target hole
- therefore, no possibility to do task 1.3 and 2.1 completely- maybe it is better to motivate participants after a certain time to skip the trial?

Appendix 7 – R Syntax

Author: Martin Schmettow, Date: 09 June, 2018

```
knitr::opts_knit$set(warning = F, message = F)

purp.data = F
purp.mcmc = T

library(tidyverse)

## -- Attaching packages ---------------------------------------------
------------------------------ tidyverse 1.2.1 --

## v ggplot2 2.2.1     v purrr   0.2.4
## v tibble  1.4.2     v dplyr   0.7.4
## v tidyr   0.8.0     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0

## -- Conflicts ------------------------------------------------------
------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(readxl)

library(brms)

## Loading required package: Rcpp

## Loading 'brms' package (version 2.3.0). Useful instructions
## can be found by typing help('brms'). A more detailed introduction
## to the package is available through vignette('brms_overview').
## Run theme_set(theme_default()) to use the default bayesplot theme.

library(rstanarm)

## rstanarm (Version 2.17.4, packaged: 2018-04-13 01:51:52 UTC)

## - Do not expect the default priors to remain the same in future rs
tanarm versions.

## Thus, R scripts should specify priors explicitly, even if they are
just the defaults.

## - For execution on a local, multicore CPU with excess RAM we recom
mend calling

## options(mc.cores = parallel::detectCores())

## - Plotting theme set to bayesplot::theme_default().

##
## Attaching package: 'rstanarm'

## The following objects are masked from 'package:brms':
##
##     exponential, kfold, lasso, ngrps
```

```
options(mc.cores = 6)
library(mascutils)
library(asymptote)

##
## Attaching package: 'asymptote'

## The following objects are masked from 'package:mascutils':
##
##     inv_logit, logit

library(bayr)

##
## Attaching package: 'bayr'

## The following objects are masked from 'package:rstanarm':
##
##     fixef, ranef

## The following objects are masked from 'package:brms':
##
##     fixef, ranef

## The following object is masked from 'package:stats':
##
##     predict

load("AK18.Rda")
```

Data preparation

Box study

```
box_files <-
  dir(path = "raw_data/AK/",
      pattern = "^Participant\\d{3}_Box\\.xls",
      full.names = T,
      recursive = T)

check_box <- function(x){
  colnames(read_xls(x))
}

read_box <- function(x){
  print(x)
  read_xls(x) %>%
    select(-Date) %>%
    #select(Participant, Task, Repetition, Route, Task success, Wrong_Route
, Skipped, Wall_contact, Time, TimeOnTask) %>%
    mutate(Route = as.character(Route),
           #Task success= as.character(Damage),
           Wall_contact = as.character(Wall_contact),
           Time = as.numeric(Time),
           TimeOnTask = as.numeric(TimeOnTask))
}
```

```
AK18 <-
  set_names(box_files) %>%
  map_df(read_box) %>%
  select(-TimeOnTask) %>%
  filter(!is.na(Time)) %>%
  tidyr::separate(Task,
                  into = c("View", "Distance", "Marvin"),
                  sep = "\\.") %>%
  mutate(Part = str_extract(Participant, "\\d+"),
         trial = as.integer(Repetition),
         Route = as.character(Route),
         Wrong_Route = as.numeric(Wrong_Route),
         Setup = "Box",
         Wall_contact = as.integer(Wall_contact),
         View = if_else(View == "1", "allo", "ego"),
         Task = str_c(View, Distance, sep = "_"),
         ToT = Time/60,
         Task_success = if_else(Task_success == "TRUE", T, F)) %>%
  filter(!is.na(Repetition), is.na(Wrong_Route)) %>%
  select(Setup, Part, View, Distance, Task,
         Route, trial, ToT, Wall_contact, Task_success) %>%
  print()


save(AK18, file = "AK18.Rda")
```

Data exploration

```
      load("AK18.Rda")
```

Descriptives

Number of observations

```
      AK18 %>%
  group_by(Setup, Part, Task) %>%
  summarize(N_trials = n()) %>%
  ungroup() %>%
  group_by(Setup, Task) %>%
  summarize(N_Part = n(),
          min(N_trials), median(N_trials), max(N_trials), sd(N_trials)) %
>%
  knitr::kable()

      AK18 %>%
  group_by(Route, View, Distance) %>%
```

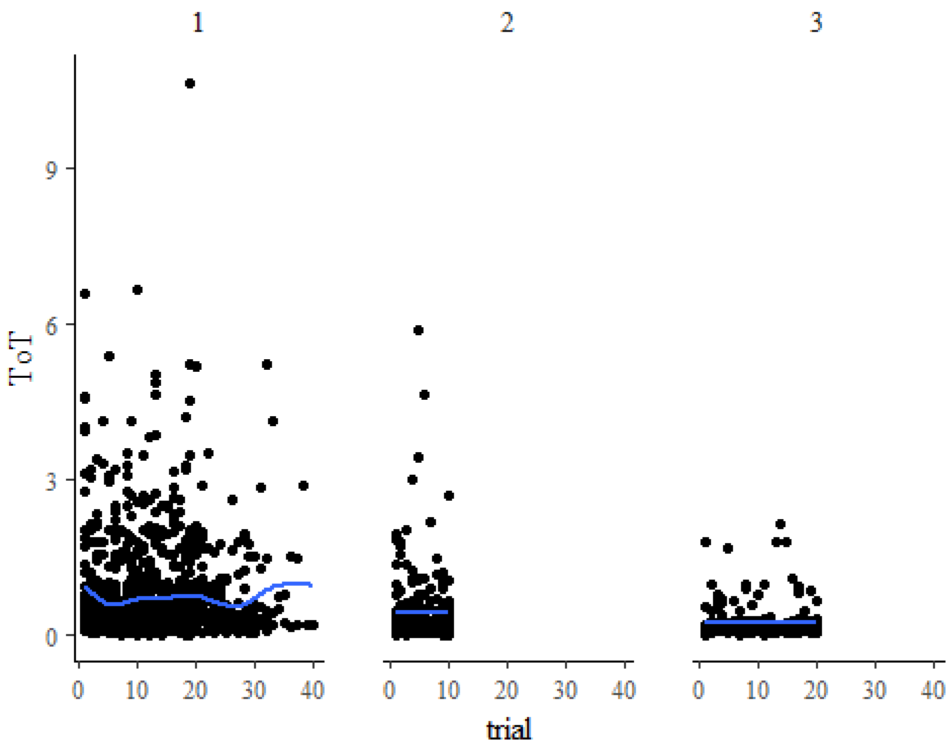
Setup	Task	N_Part	min(N_trials)	median(N_trials)	max(N_trials)	sd(N_trials)
Box	allo_1	24	20	20	20	0.000000
Box	allo_2	24	5	10	10	1.020621
Box	allo_3	13	14	20	20	1.664101
Box	ego_1	24	14	25	39	6.192416

```
summarize(n_obs = n()) %>%
spread(View, n_obs)

    ## # A tibble: 50 x 4
## # Groups:   Route [20]
##     Route Distance  allo   ego
##     <chr> <chr>    <int> <int>
##   1 1     1           24    53
##   2 1     2           26    NA
##   3 1     3           11    NA
##   4 10    1           25     2
##   5 10    2            2    NA
##   6 10    3           10    NA
##   7 11    1           27    NA
##   8 11    3           12    NA
##   9 12    1           22     3
##  10 12    3           11    NA
## # ... with 40 more rows
```

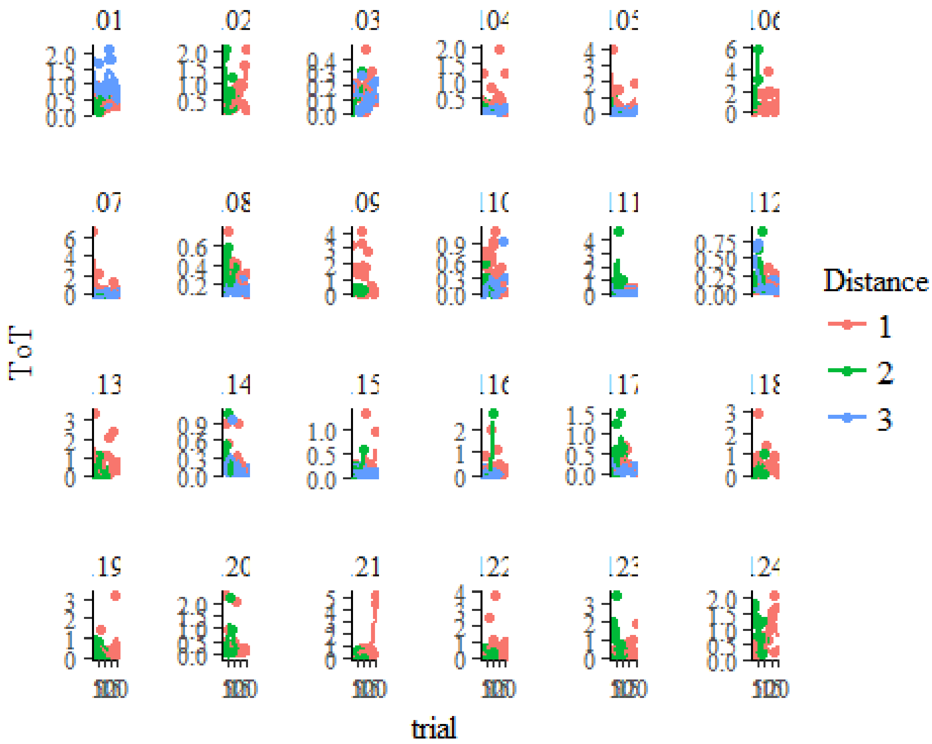
Explorative analaysis

Raw learning curves

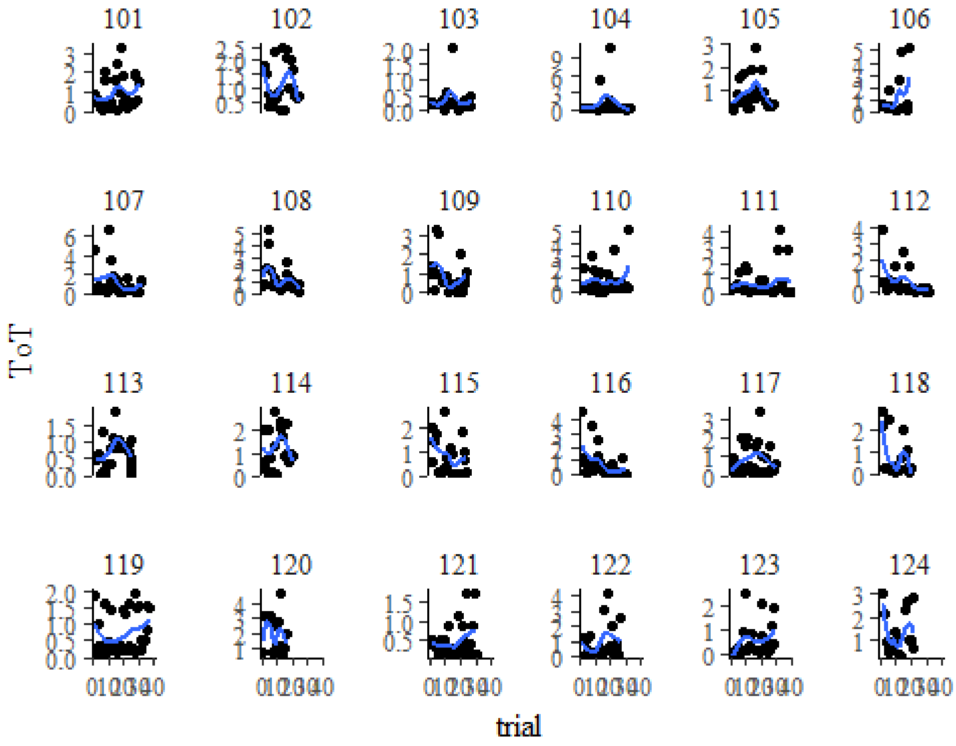
```
    AK18 %>%
ggplot(aes(x = trial, y = ToT)) +
facet_grid(~Distance, scale = "free_y") +
geom_point() +
geom_smooth(se = F)

    ## `geom_smooth()` using method = 'gam'
```



```
    AK18 %>%
filter(View == "allo") %>%
ggplot(aes(x = trial, color = Distance, y = ToT)) +
facet_wrap(~Part, ncol = 6, scale = "free_y") +
geom_point() +
geom_smooth(se = F)

    ## `geom_smooth()` using method = 'loess'
```



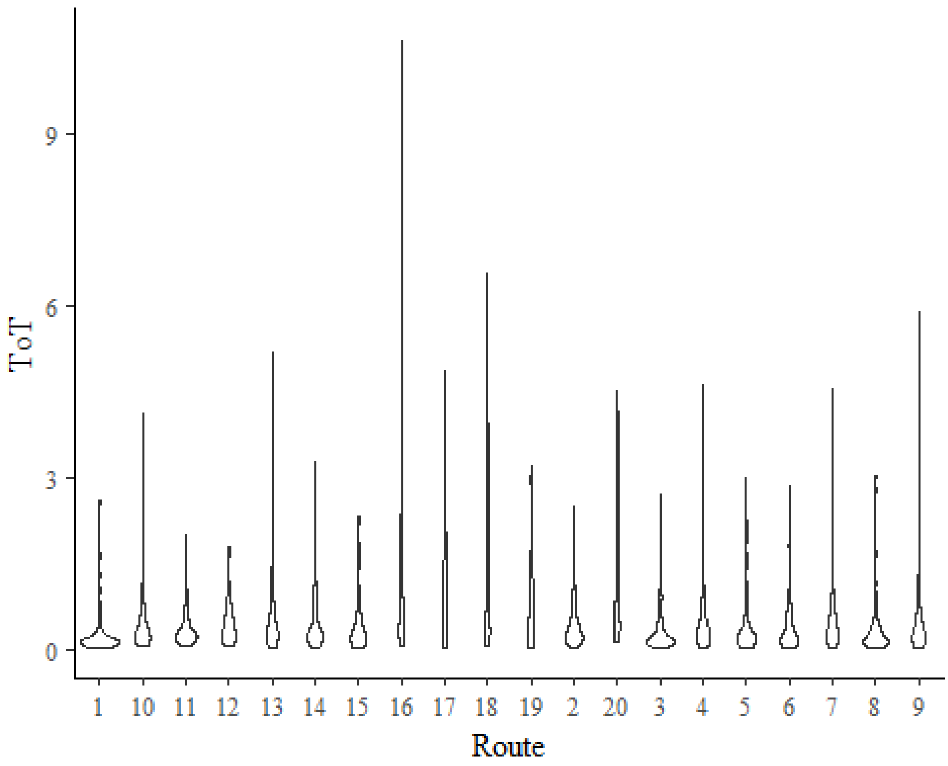

```
    AK18 %>%
filter(View == "ego") %>%
ggplot(aes(x = trial, y = ToT)) +
facet_wrap(~Part, ncol = 6, scale = "free_y") +
geom_point() +
geom_smooth(se = F)

    ## `geom_smooth()` using method = 'loess'
```



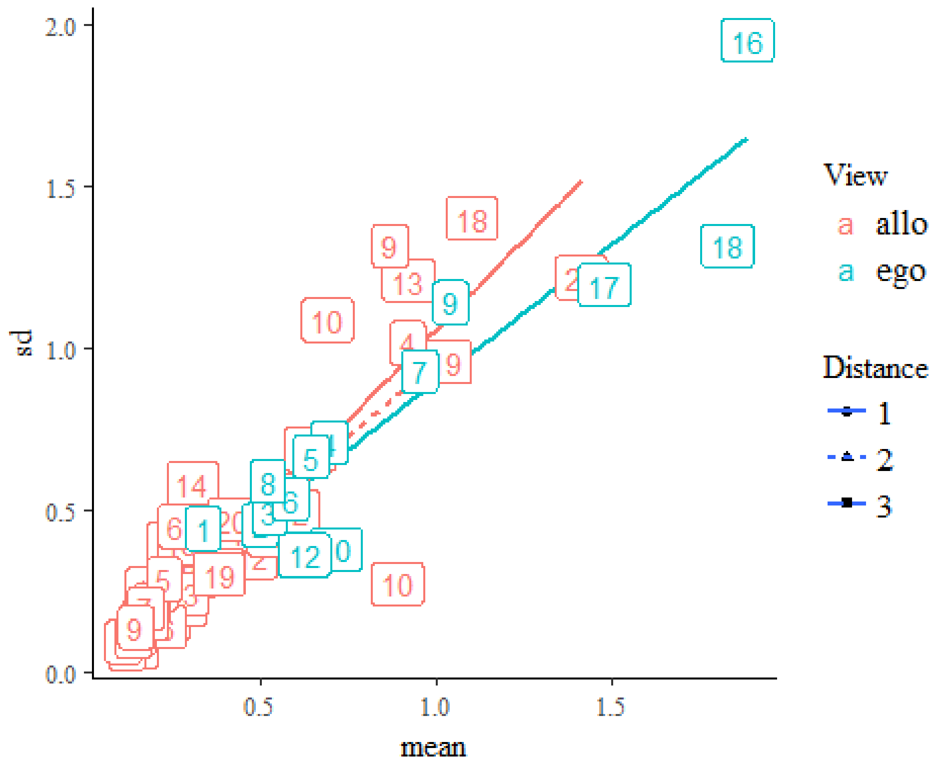
Examining Routes

```
AK18 %>%
ggplot(aes(x = Route, y = ToT)) +
geom_violin()
```

A violin plot was made for demonstrating the continuous distribution of the routes (1-20), which were depicted on the horizontal axis, whereas the performance variable time on task were depicted on the vertical axis. A great variance between the routes and time on task with regard to all participants and tasks was given. However, for route 16, 17 and 18 participants needed overall the most time. On the opposite, participants needed overall the least time for route 1, 11 and 12.



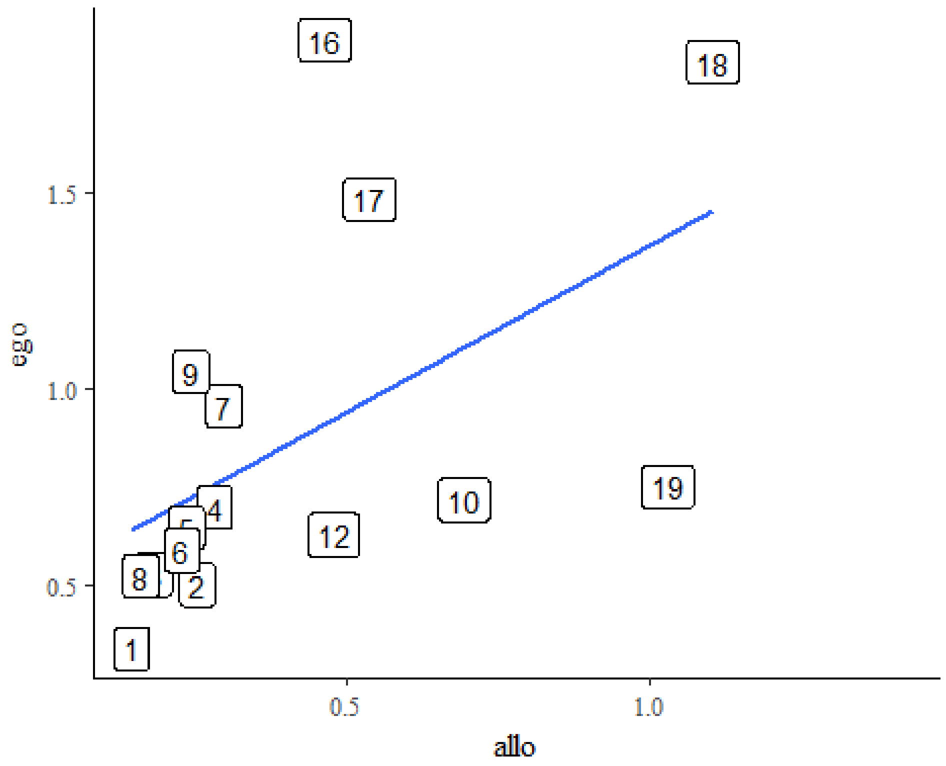

```
AK18 %>%
  group_by(View, Distance, Route) %>%
  summarize(mean = mean(ToT),
            sd = sd(ToT)) %>%
  ggplot(aes(x = mean, y = sd, color = View, linetype = Distance, shape = D
istance, label = Route)) +
  geom_point() +
  geom_smooth(se = F, method = "lm") +
  geom_label()

## Warning: Removed 1 rows containing non-finite values (stat_smooth)
.

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_label).
```



- linear relationship between mean and sd. That should be covered by the model.

```
    AK18 %>%
  filter(Distance == 1) %>%
  group_by(View, Route) %>%
  summarize(mean = mean(ToT)) %>%
  spread(View, mean) %>%
  ggplot(aes(x = allo, y = ego, label = Route)) +
  geom_point() +
  geom_smooth(se = F, method = "lm") +
  geom_label()

    ## Warning: Removed 5 rows containing non-finite values (stat_smooth)
.

    ## Warning: Removed 5 rows containing missing values (geom_point).

    ## Warning: Removed 5 rows containing missing values (geom_label).
```

A scatter plot of slopes for route (fixed effect + random effects) and task (fixed effect + random effects) was obtained by fitting the multi-level linear model. Only a poor correlation between the difficulty of routes across the allocentric and egocentric perspective could be noticed. Routes, which were performed from an allocentric perspective, seem to be easier than routes, which were performed from an egocentric view. Especially route 16, 17 and 18 was extremely difficult under the egocentric perspective and produced extreme variance.




```
AK18_allo <-
  AK18 %>%
  filter(View == "allo")
```

Learning curves

Any attempt to estimate proper learning curves failed. See below for an analysis. In brief: not a single participant showed signs of learning with the box.

Setting up the LARY model:

Estimated curves

Effects

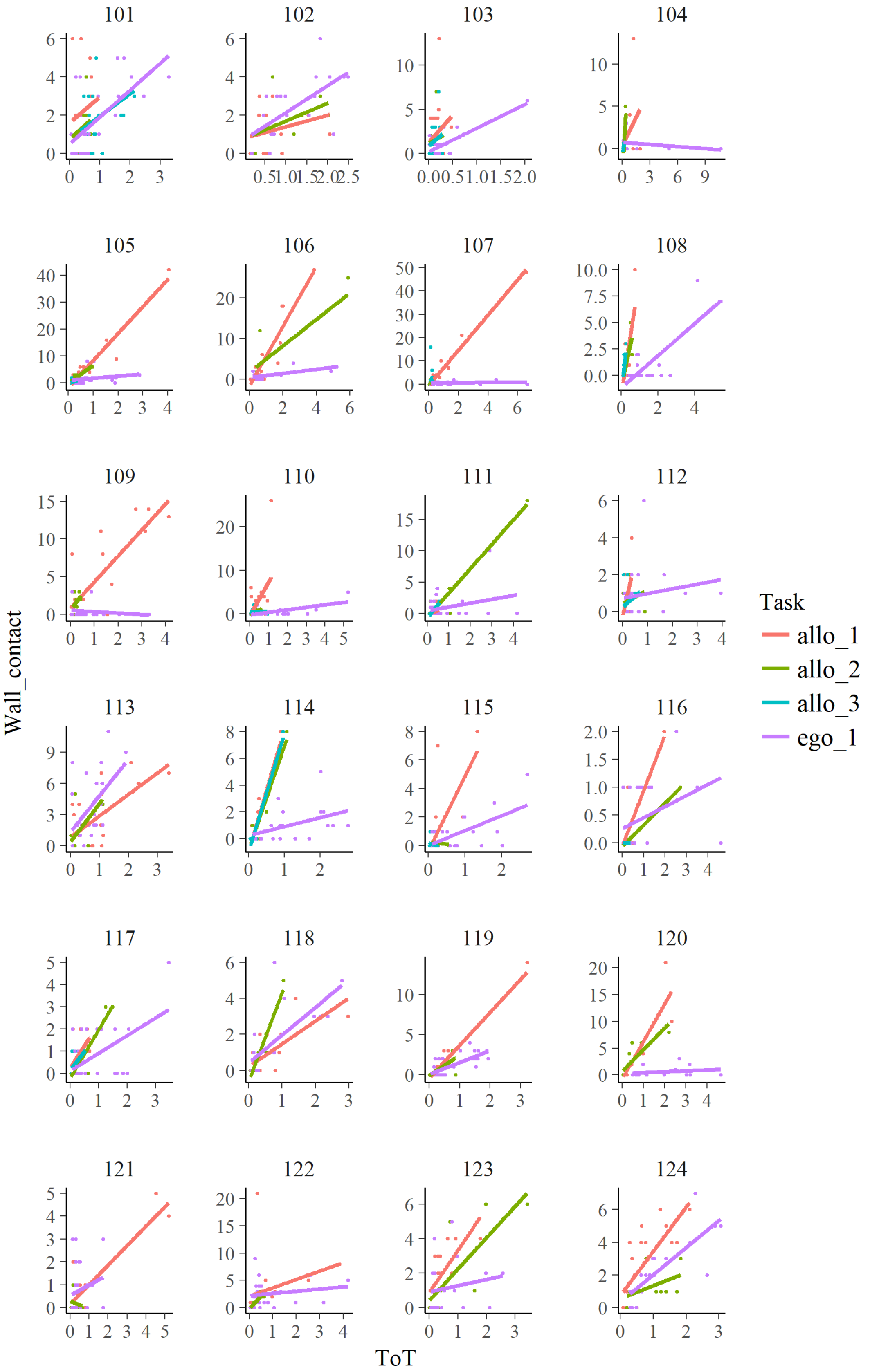
Individual differences as standard deviations by task and parameters:

Association between performance measures

Exploratory analysis

Is there any sign of learning when averaging over all participants?

```
AK18 %>%
  ggplot(aes(x = ToT, y = Wall_contact, color = Task)) +
  facet_wrap(~Part, scale = "free", ncol = 4) +
  geom_point(size = .4) +
  geom_smooth(se = F, method = "lm")
```



The relationship looked as being linear, because the equation represents a straights line in a bi-dimensional plot. However, because linear models make assumptions that are never truly met by real data, we chose in a following step for a generalized linear model (GML) with a poisson distribution, which can re-established linearity through link functions, and allows the variance of each measurement to be a function of its predicted value.

Regression

M_6: Checking for learning effects

The following regression model estimates the *individual* associations between number of wall contacts by ToT. We control for trial, Distance and View.

This analysis will, among others, produce intercept effects (participant-level and population-level). To make this parameter more meaningful, we center ToT at the population average. Similarly, the variable trial is shifted by one to make the intercept represent the first trial.

```
tot_pop_avg <- mean(AK18$ToT)

AK18 <-
  AK18 %>%
  mutate(cToT = ToT - tot_pop_avg,
         trial = trial - 1) %>%
  as_tbl_obs()
```

First, we examine if there is any learning effect at all, meaning on the population-level and the participant-level.

```
M_6 <- stan_glmer(Wall_contact ~ 1 + trial +
                  (1 + trial|Part) +
                  (1|Route) +
                  (1|Obs),
              family = poisson,
              data = AK18,
              init = "0")

P_6 <- posterior(M_6)

## Warning in sqrt(value): NaNs produced

save(M_6, P_6, file = "M_AK_6.Rda")

load("M_AK_6.Rda")

fixef(P_6, mean.func = exp)

##
##
## Table: Estimates with 95% credibility limits
##
## fixef           center       lower       upper
## ----------   ----------   ----------   ---------
## Intercept     0.8288882    0.5342585    1.268346
## trial         0.9919659    0.9760908    1.008340
```

A generalized linear model (GML) with a poission distribution was estimated. This model confirmed no learning effect on the population-level as the trial effect was 1% (*95% CI* [0.98;

1.01]). Therefore, one can be certain that on average, people do not learn to have fewer wall contacts.

```
      left_union(fixef(P_6) %>% discard_redundant(),
           grpef(P_6) %>%
             filter(re_factor == "Part") %>%
             rename_if(is.numeric, funs(str_c(., "_sd"))) %>%
             discard_redundant()
)

      ## # A tibble: 2 x 7
##   fixef       center   lower   upper  center_sd lower_sd upper_sd
##   <fct>        <dbl>   <dbl>   <dbl>      <dbl>    <dbl>    <dbl>
## 1 Intercept -0.188   -0.627  0.238      0.719   0.495     1.04
## 2 trial     -0.00807 -0.0242 0.00831    0.0277  0.00972   0.0475
```

In order to obtain random effects, which show the variance caused due to individual differences, we compare the population-level effects to the standard deviation of the individual deviations of the participant level. This was done on the linear predictor scale, which is the logarithm of the multiplicative scale. In this model, things were additive, where 0 means no change. Compared to the very small effect of trials 1% (*95% CI* [-0.02; 0.01) and 2% (*95% CI* [0.01; 0.04]), there was substantial variation with regard to the intercept effects -19% (*95% CI* [-0.62; 0.24]) and 72% (*95% CI* [0.49; 1.04]) (see Table 2). However, that by no means makes anyone in the sample a visible learner. Therefore, we could exclude learning on average and for practically all individuals.

M_5: ToT and wall contact

```
      M_5 <-
  stan_glmer(Wall_contact ~ 1 + cToT + Task +
                 (1 + cToT|Part) +
                 (1|Route) +
                 (1|Obs),
             family = poisson,
             data = AK18)

P_5 <- posterior(M_5)

      ## Warning in sqrt(value): NaNs produced

      save(M_5, P_5, file = "M_AK_5.Rda")

      load("M_AK_5.Rda")
```

The effect is dramatic: On average, with every minute longer, the number of wall contacts multiplies by 1.79, or increases by almost 80%, with good certainty.

Compared to the average effects size (0.58), the variation is existent, but not overwhelming. By far most participants have a positive relation between ToT and wall contact.

```
      left_union(fixef(P_5) %>%
             filter(fixef  %in% c("Intercept", "cToT")) %>%
             discard_redundant(),
           grpef(P_5) %>%
             filter(re_factor == "Part") %>%
             rename_if(is.numeric, funs(str_c(., "_sd"))) %>%
             discard_redundant()
)

      ## # A tibble: 2 x 7
##   fixef     center  lower upper center_sd lower_sd upper_sd
##   <fct>      <dbl>  <dbl> <dbl>     <dbl>    <dbl>    <dbl>
## 1 Intercept -0.122 -0.495 0.228     0.525    0.377    0.734
## 2 cToT       0.586  0.443 0.734     0.278    0.166    0.435
```

Compared to the average effects size (0.58), the variation is existent, but not overwhelming. By far most participants have a positive relation between ToT and wall contact.

M_7: ToT and wall contact, interaction effect Route by Task

The following model etnends the previous by adding the possibility that the route effects differ by task

```
      M_7 <-
  brm(Wall_contact ~ 1 + cToT + Task +
                (1 + cToT|Part) +
                (1 + Task|Route),
              family = negbinomial(link = log),
              iter = 4000,
              warmup = 3000,
              data = AK18)

      ## Compiling the C++ model

      ## Start sampling

      P_7 <- posterior(M_7)

save(M_7, P_7, file = "M_AK_7.Rda")

      load("M_AK_7.Rda")

      fixef(P_7, mean.func = exp)

      ##
##
## Table: Estimates with 95% credibility limits
##
## fixef             center       lower       upper
## -----------  ----------  ----------  ----------
## Intercept     1.2474457   0.8273745   1.8173290
## cToT          1.7726131   1.5533064   2.0606988
## Taskallo_2    1.2207438   0.6539765   2.1196508
```

```
## Taskallo_3    0.5067341   0.3619497   0.6950428
## Taskego_1     0.7398900   0.5305070   1.0052777
```

```r
grpef(P_7)
```

```
##
##
## Table: Estimates with 95% credibility limits
##
## fixef         re_factor       center       lower       upper
## -----------  ----------  ----------  ----------  ----------
## Intercept    Part         0.5525471   0.3941567   0.7952082
## cToT         Part         0.2411365   0.1045807   0.4163128
## Intercept    Route        0.6506511   0.4464556   0.9400092
## Taskallo_2   Route        0.5235004   0.0874271   1.2007623
## Taskallo_3   Route        0.4120467   0.0470928   0.8848689
## Taskego_1    Route        0.4610276   0.1991460   0.7875025
```

```r
left_union(fixef(P_7) %>%
             filter(fixef  %in% c("Intercept", "cToT")) %>%
             discard_redundant(),
           grpef(P_7) %>%
             filter(re_factor == "Part") %>%
             rename_if(is.numeric, funs(str_c(., "_sd"))) %>%
             discard_redundant()
)
```

```
## # A tibble: 2 x 7
##   fixef     center  lower upper center_sd lower_sd upper_sd
##   <fct>      <dbl>  <dbl> <dbl>     <dbl>    <dbl>    <dbl>
## 1 Intercept  0.221 -0.189 0.597     0.553    0.394    0.795
## 2 cToT       0.572  0.440 0.723     0.241    0.105    0.416
```

Model selection

Does M_7 fit the data better than M_5?

```r
waic(M_5)
```

```
## Warning: 300 (19.0%) p_waic estimates greater than 0.4. We recomme
nd trying
## loo instead.
```

```
##
## Computed from 4000 by 1582 log-likelihood matrix
##
##           Estimate   SE
## elpd_waic  -1999.4 38.1
## p_waic       416.4 11.2
## waic        3998.7 76.2
```

```
## Warning: 300 (19.0%) p_waic estimates greater than 0.4. We recomme
nd trying
## loo instead.
```

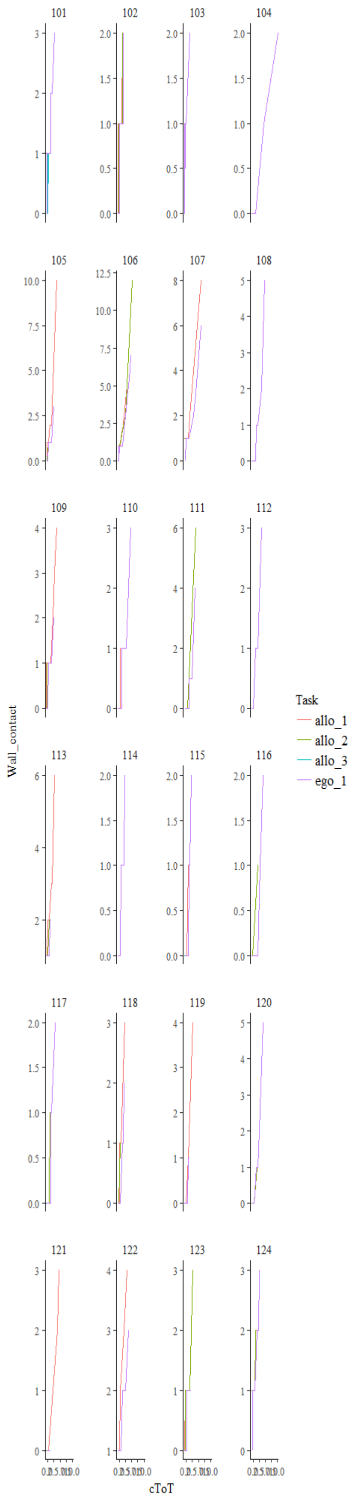
```r
waic(M_7)
```

```
##
## Computed from 4000 by 1582 log-likelihood matrix
##
##           Estimate   SE
## elpd_waic  -2180.5 47.4
## p_waic        76.8  6.7
## waic        4361.1 94.8

## Warning: 32 (2.0%) p_waic estimates greater than 0.4. We recommend
trying
## loo instead.
```

It does.

Visualization of results

```
load("M_AK_7.Rda")

AK18_pred <-
  AK18 %>%
  mutate(Route = 1)

T_predict_Route1 <-
  predict(M_7, newdata = AK18_pred) %>%
  mutate(pred = "Rt_1")

T_predict_Route1 %>%
  left_join(AK18) %>%
  ggplot(aes(x = cToT, y = Wall_contact, col = Task, col = pred)) +
  facet_wrap(~Part, ncol = 4, scale = "free_y") +
  geom_line(aes(y = center))

## Joining, by = "Obs"

## Warning: The plyr::rename operation has created duplicates for the
## following name(s): (`colour`)
```



This confirms that the number of wall contact is positively associated with ToT for all and every participant. However, one issue remains to discuss: could the relationship purely be caused by time for error recovery. That would be the case, if a wall contact causes a severe delay in task completion.