



Effects of feedback using Objective Structured Assessments of Technical Skills (OSATS) versus Observational Clinical Human Reliability Analysis (OCHRA) in teaching medical students open inguinal hernia repair.

Running head: EFFECTS OF FEEDBACK ON MEDICAL EDUCATION

Name: Natalie Jasmin Schneider

Student number: s1084704

Date: 03-08-2019

Supervisor: Dr. G.P.H. Band

Second reader: Dr. F. Richter

Cognitive Psychology

Thesis MSc Applied Cognitive Psychology

Acknowledgements

I would like to start by thanking my thesis supervisor Guido Band for his supervision and support throughout this (often frustrated) journey. You kept believing in me and always encouraged me to bring out the best in myself, especially near the end, because you knew I could do it. Thank you for that.

Next, I would like to thank my boyfriend Haje Huffstadt for his programming skills, love and support. You spent hours on creating a program in Matlab for me which could perform calculations on the raw motion tracking data, even though you had minimal knowledge of the system and it was not even your job to begin with. Without you, this thesis could not have come to a successful completion.

At last, I want to thank my friends and entire family for their love and support throughout the whole master/thesis writing period. There were a lot of obstacles and frustrations during this period and you all helped me through it. Especially to my parents, thank you for everything!! I finally did it.

Abstract

Feedback always has been an important part in medical education as it can enhance the learning process by shortening the learning curve (Boyle et al., 2011). Especially for novice medical students, who have to learn and obtain medical knowledge and technical skills simultaneously, it is of great importance to investigate which type of feedback shows the greatest learning improvement. Two often used observational assessment tools in medical education are the OSATS and OCHRA. This research was conducted to investigate which assessment tool showed greater learning advances by means of motion tracking of the hand in teaching novice medical students an open inguinal hernia repair. It was **hypostasised** that the group assessed by the OCHRA would outperform students assessed with the OSATS on the second operation by showing smaller idle times and path lengths. 22 Dutch right handed medical students were randomly divided between the two feedback groups. First they were asked to fill in two spatial ability tests, the affect grid and a medical questionnaire. Secondly, they performed the first surgery followed by feedback. Lastly, they received another affect grid, followed by performing the second surgery. Results showed no significant **effects between** time and feedback group ($p > .05$). Furthermore, there were no significant effect of spatial ability and affective states on time. Follow-up research should combine the OSATS and OCHRA in order to investigate the learning performances. Furthermore, the effect of video-feedback compared to verbal feedback should be investigated.

Keywords: OSATS, OCHRA, feedback, medical education, technical skills, motion tracking, cognitive load theory, spatial ability, affect grid

Introduction

He then gives them kindly-meant advice in regard to their clothing, their behaviour, and the language they should use with patients: he recommends them cleanliness and a proper attention to their hair and forbids them to eat onions or garlic before visiting a patient, or to drink too much wine, lest they annoy the sufferer by the offensive odor from their mouths and stink like goats. (Puschmann, 1891/1966, p. 113)

Feedback, even centuries ago in the times of Plato and other ancient Greek philosophers, has been described as an important part of medical education (Puschmann, 1891/1966; Telio et al., 2015). Nowadays, feedback is still of great importance in medical education and can be described as a consequence of performance. Feedback, in order to be effective must be aimed at a learning context where the feedback can be directed to. The disparity between what a student understood and what is aimed to be understood can be decreased by feedback that provides information specifically to the task that has to be learned (Hattie & Timperley, 2007). On the other hand, when students lack certain knowledge that is required, feedback has little effect since they cannot link the new information with what they already do know. Prior researches in feedback reported different effect sizes between types of feedback, indicating that one type can be much more powerful than others (Hattie & Timperley, 2007). These general findings showed that information feedback about the task and how to do it more effectively showed the highest effect sizes. This finding also applies to medical education, where several researches showed that different types of feedback can have different effects depending on the proficiency level (novice, intermediate and expert) (Porte et al., 2007).

Medical education strives to teach the complex materials in an efficient and reliable way. The education is aimed at teaching not only anatomical knowledge, but also technical skills. For students, especially novices, who have very little experience, learning these complex tasks requires a lot of effort. Especially since the tasks that need to be learned demands simultaneous integration of knowledge, skills and behaviour.

With this writing, the effect of feedback by means of two different types of observational assessment tools often used in medical education are being outlined. We aim to investigate which type of feedback has a greater contribution in teaching novices technical

skills. Is there a difference in the effectivity between the two tools often used and if so, what is the difference?

Feedback and learning

In medical education, students do not start their medical internship until their fourth year of education. At this point, the students begin with the acquisition of specialised skills in diagnosis and make decisions about specific medical procedures (Ericsson et al., 2015). Traditionally, medical students first learn theoretical knowledge before they learn experiences in real-world situations, where performance could have potential consequences for patient's safety, during their medical internship (Ericsson et al., 2015). The focus on neglect of skill acquisition and knowledge by means of practice in medical education is defined by the expression about performing medical procedures and learning as "See one, do one, teach one", the so called learning according to the "master-apprentice" model (Ericsson et al., 2015). The study of Ericsson showed that receiving immediate feedback and performing the task at hand repeatedly, were important aspects in technical skill acquisition.

Feedback in medical education, is an important element of the learning process and can be defined as "the provision or return of performance-related information to the performer" (Boyle et al., 2011, p. 1). Furthermore, feedback has shown to be an essential aspect for surgeons to become an expert at a practiced skill (Porte et al., 2007). Especially for novice surgeons, the delivery of external feedback was thought to be crucial for technical surgical skill development (Porte et al., 2007). This assumption can be supported by Parmar (2011) who found that performance can be improved, when feedback is given quickly after the procedure (proximate feedback). Furthermore, feedback is likely to help speeding up the learning processes of the student for the procedure being taught (Parmar et al., 2011).

A study of Boyle (2011) showed that performance feedback can lower the amount of errors and reduce the learning curve, making it an important part of surgical skill training. Furthermore, receiving feedback conforming to a predetermined standard (standardised feedback) during medical training was associated with an enhanced learning curve and significantly fewer errors. Within this research, this occurred at the expense of instrument efficiency scores (Boyle et al., 2011). At last, Boyle stated that surgical training of medical interns improved by immediate performance feedback, which also enhanced patient safety.

In summary, feedback has shown to be a great contributor in surgical skill acquisition, which in turn can enhance patient safety. In medical education, different standardised

assessment tools on direct observation of technical skills are used to provide feedback. There are mainly two types of assessment tools being used; global and task-specific. It is of great importance to see which type of assessment has a greater contribution to teaching medical students technical skills. Is there a difference in the effectivity between the assessment tools and the given feedback and if so, what is the difference?

Global & specific assessment tools, feedback and learning

Providing feedback increases motivation, prevents incorrect actions, supplies reinforcement for correct actions and can provide information about errors as a basis for corrections (Xeroulis et al., 2007). The aims of giving good feedback are to motivate, to raise awareness of strengths, to highlight the blind spots, to encourage reflection and open discussion and to identify actions that enable improvements.

During surgical training, the delivery of external feedback has shown to be a crucial part in technical skill development of a novice surgeon (Porte et al., 2007; Xeroulis et al., 2007). Assessments can be based on direct observation of technical skills, which enhances the effectiveness of constructive feedback (Ahmed et al., 2011). Two observational assessment tools that can be used to provide constructive feedback are the Objective Structured Assessment of Technical Skill (OSATS) and Objective Clinical Human Reliability Analysis (OCHRA).

The OSATS is a reliable, valid and commonly used assessment tool that assesses the technical skills by rating seven technical- surgical competencies on a 5-point Likert scale (Martin et al., 1997). The trainer uses this multiple-item global rating scale to assess a surgical trainee on an entire surgical procedure, followed by providing constructive feedback. The OSATS is effective in assessment of surgical skills of trainees in the operating room and since it is not limited to a specific procedure, it can easily be applied for other technical skill assessments (Niitsu et al., 2013).

The OCHRA is a more specific method of technical skills assessment providing a step-specific evaluation (Tang et al., 2004). For the OCHRA, each procedure needs to be described per step, and for each step the potential errors and their consequences need to be identified. The OCHRA provides specific feedback on the steps during a particular surgery and has the advantage to show when a step has not been achieved and why. In laparoscopic surgery, the OCHRA is associated with a progressively decreasing number of errors (Angulo et al., 2014). In sleeve gastrectomy procedures, the OCHRA is considered a valuable tool for the

recognition of potential hazard zones and surgical performance (Van Rutte et al., 2017). Especially for assessing surgical performance at the specialist level, task specific feedback has shown to be a valid method (Miskovic et al., 2011)

In summary, feedback is associated with and contributes to greater learning performances. However, the process to successfully learn something depends on multiple interactions. So what happens when something is learned unsuccessfully? One theory that can explain this matter is the *Cognitive Load Theory* (Sweller, 1988).

Cognitive load theory

The process of successfully learning something, depends on multiple interactions such as those in the cognitive, social (experience of and interaction with others), environmental (setting or location), affective (emotions and motivation) and metacognitive (knowing about one's own knowing) areas (Young et al., 2014). One of the theories that can explain unaccomplished learning, is the *Cognitive Load Theory (CLT)* (Sweller, 1988).

The *CLT* should be considered an important key theory for medical education (Fraser et al., 2015). The theory implies that the *working memory (WM)* when receiving new information has a limited capacity and when the capacity is exceeded, performance and learning is impaired (Fraser et al., 2015). Three types of cognitive load that impacts the WM are identified by the CLT. The first one is intrinsic load which is load that is associated with the complexity of performing task at hand. The second one is extraneous load which is load that is not inherent to the task such as poor instructions or interruptions. The last one is germane load which is load that is associated with aspects of the task that demand skill acquisition (i.e., to learn) (Spruit et al., 2014; Young et al., 2014). In order to facilitate learning, the most desired outcome would be that extraneous load is decreased, intrinsic load is managed by for example simplifying tasks, so that unused WM capacity is committed to germane load (Spruit et al., 2014).

Especially in medical education, CLT is of great relevance, since the activities and tasks that need to be learned demand simultaneous integration of knowledge, behaviour and skills (Young et al., 2014). Furthermore, since tasks can be complex, they can demand a cognitive load that exceeds the learners WM capacity. As a result, the learner may be cognitively overloaded, at the expense of insight and memory consolidation (Young et al., 2014). CLT therefore can explain why and how medical students struggle with becoming experts by mastering the complex concepts of the tasks and skills that are required.

During procedural skills learning, the CLT can also be used to understand the influence of *feedback* (Hatala et al., 2014). Providing feedback during procedural skill learning could affect the cognitive load by either decreasing (structuring tasks for better understanding) or increasing (information-overload) it (Hatala et al., 2014). Young (2014) has found that in medical education when learning a new skill, providing the student with direct feedback about the task performance given by a clinical supervisor helps the student with the development of skill learning.

Hatala (2014) found that the students who received the most feedback during a laparoscopic task, showed the highest performance outcomes and rated their cognitive workload lowest (Hatala et al., 2014). Furthermore, they found that students performed better when receiving instructor feedback compared to simulator feedback. This was not only found for novice students learning simple tasks but also for medical residents learning more complex surgical skills (Hatala et al., 2014). At last, students reported a lower cognitive workload when presented with different types of feedback, including instructor feedback (Hatala et al., 2014).

In summary, the CLT implies that successful learning of technical skills is impaired when WM capacity is exceeded. Furthermore, feedback may facilitate in order to reduce the cognitive workload. Since successful learning also depends on affective states (emotion and motivation) it is of importance to see what the effects of affect is on learning. Do affective states contribute to learning performances of medical students and if so, how?

Affective states, arousal and learning

Over the years, affective states have received more attention in medical education. Studies have shown that learning and cognitive performance can be influenced by emotion and motivation. Positive emotions such as enjoyment have been linked to enhanced learning and performance and faster processing, whereas negative emotions such as anxiety have been linked to slower processing and decreased learning (Young et al., 2014).

These findings can be explained by Spachtholz (2014) who showed with use of the Affect Grid (Russell et al., 1989) that affective states have an influence on WM capacity where negative affect reduces WM capacity. It even showed that daily changes in negative affect are accompanied by daily changes in WM performance (Spachtholz et al., 2014). Two explanations of reduced WM capacity due to negative affect have been given. The first one implies that when someone experiences a negative affect, the emotion can shift the attention

from the task at hand onto itself. The second explanation states that experiencing a negative affect can decrease task motivation which affects the task at hand (Spachholz et al., 2014).

As described earlier, when a student is cognitive overloaded, WM capacity is surpassed and learning is impaired. Research has shown that affective states can have an influence on someone's cognitive load. Experiencing positive emotions can expand one's cognitive reserve, whereas negative emotions can reduce one's cognitive reserve (Kuhbandner et al., 2011; Young et al., 2014). In addition, as described above, negative affect reduces WM capacity which can cause to surpass a student's WM capacity faster.

Further research in affect and learning showed that affective states can also influence the breadth of information processing where negative affect can narrow and positive affect can broaden the informational access. Moreover, the amount of information processing at higher stages such as knowledge activation and/or attentional selection can be increased by positive affects (Kuhbandner et al., 2011). Additionally, a moderate degree of *arousal* can enhance different cognitive processes such as the working memory capacity, alertness and attentional focus (Hoogerheide et al., 2018). At last, it is reported that the relationship between task performance and arousal follows an inverted-U function (Hoogerheide et al., 2018).

In summary, positive affective states have a positive influence on learning, performance, cognitive load and working memory where negative affective states have a negative effect. Another cognitive ability that has an effect on technical skill learning is spatial ability. Therefore, it is of importance to investigate the following question: In what way does spatial ability contribute to learning performance of medical students?

Spatial ability and learning

Spatial ability in humans consists of orientation, visualisation and manipulation of structures in space (Langlois, Wells et al., 2015). In medical education, spatial abilities have been linked to three dimensional (3D) anatomy knowledge, 3D synthesis of two dimensional (2D) anatomical views, mental rotation of anatomical structures, topographical questions and cross-sections (Langlois, Wells et al., 2015). In health care, spatial abilities have been linked to performance in technical skills in several departments such as microscopic pathology and laparoscopic- and clinical surgery (Langlois, Wells et al., 2015).

Technical skills components are of great importance in health care professions (Langlois, Bellemare et al., 2015). They are involved in several medical aspect such as physical examination, clinical procedures, surgical procedures and image interpretation.

Spatial abilities have shown to be of importance in the cognitive phase of learning a new technical skill, since spatial abilities can be a determinant of individual differences during skill acquisition in this phase (Ackerman, 1992; Langlois, Bellemare et al., 2015). The cognitive phase is the first stage of skill acquisition where the novice attempts to understand the task at hand and what it requires and where the learner tries to develop strategies for task accomplishment. It is characterized by a high cognitive load on the learner (Ackerman, 1992).

Langlois (2015) showed that for novices, spatial abilities are an important aspect in training programmes. With this, if students with lower levels of spatial ability are given more time for learning new skills, they might achieve competency later (Henn et al., 2018; Langlois, Bellemare et al., 2015).

Henn et al. showed that objectively assessed spatial ability is also associated with performance. Furthermore, it is assumed that students with lower spatial ability might have a more limited capacity to respond to unexpected events such as intra-operative complications and recovery from error (Henn et al., 2018).

In summary, spatial ability is an important factor in successful acquisition of technical skills. Furthermore, feedback, affective states and motivation are all associated and contribute to greater learning performances. Therefore, with this study we aim to investigate which type of observational assessment tool with its feedback, shows a greater contribution in teaching medical students (novices) technical skills for performing an open inguinal hernia repair. The open inguinal hernia repair is a basal, nonspecialised operation which comprises a lot of steps and the motorial challenge is easy. Anatomical knowledge and following the step-by-step procedure is crucial with this operation, since errors can have a severe impact on patient safety. Furthermore, we want to investigate whether participants show different arousal and pleasure scores after receiving feedback. The spatial ability scores will be used as a sanity check, since many researches showed that students with a higher spatial ability outperformed students with a lower spatial ability in performance.

Hypotheses

Both the OSATS and OCHRA assessment tools have shown to be valid for providing constructive feedback (Ahmed et al., 2011; Niitsu et al., 2013). This study aims to compare the two methods and to investigate whether a more specific stepwise feedback (OCHRA) shows a greater and faster contribution in teaching skills to novice medical students in

performing an open inguinal hernia repair on a model compared to the global rating scale method (OSATS).

Spatial ability will be used as a covariate to correct the dispersion in learning for spatial ability as a main source of variance. Since all participants are university educated, the dispersion in learning would have been too small if intelligence would have been used as a covariate. Spatial ability is an important factor in successful acquisition of technical skills (Langlois, Bellemare et al., 2015). Here, it is reported that for beginners and intermediates, the quality of technical skill performance is positively correlated with spatial ability (Langlois, Bellemare et al., 2015).

Furthermore, since cognitive states can be subject to the effect of feedback and the ability to learn, it is of interest to see whether students have a change in arousal and pleasure scores after receiving feedback. Is there a difference between the two types? In order to answer this question, the Affect Grid (Russell et al., 1989), which is a single-item scale that can repeatedly assess people's subjective affective states is included in the study. Additionally, self-reported study time and year of medical school are included in the study to investigate any possible effects. The following two hypotheses have been formulated.

The **first** hypothesis in this study is that task-specific tools may provide more concise feedback to the student. With this in mind, expected is that the group assessed and given feedback using the OCHRA will receive specific feedback which results in learning more in between surgeries. This would lead to a relatively better performance of the second open inguinal hernia repair by showing a shorter idle time and path length compared to the medical students who were provided feedback using the OSATS.

Since many researches showed that students with high spatial ability outperform students with low spatial ability, spatial ability will be used as a sanity check.

Method

This study is part of a larger study with a total of 50 participants. Here, we analyse and report the results of the first 25 participants.

Participants

Medical students from year one to six at Leiden University Medical Center (40% male, 60% female, Mage = 20.64, age range = 18-25 years) were approached for participation by a

member of the research group. The invited students were given a week to consider participation. A total of N=25 medical students volunteered to participate of which 52% were first year medical students (Mage= 18.92), 4% second year (Mage= 19), 16% third year (Mage= 21.75) and 28% fourth year (Mage= 23.43). A power analysis could not be performed based on previous literature. Only right handed students were considered eligible for participation so the camera position as well as the motion tracking device could be standardised. The students were randomised in two groups OSATS or OCHRA, stratified for year of medical training.

Written consent was provided by all participants and the participants completed all the tasks. They were informed that all data were coded before processing. Furthermore, there was no disadvantage to the medical study progress when students decided not to participate or withdrew their consent. All data of participants who withdrew their consent for participation were destroyed. The students were informed that the aim of this study was to investigate the effects of feedback in learning an open inguinal hernia repair, but no further details were provided. Since this was a voluntary research study which was all additional to the students' curriculum, they were not compensated. Ethical approval was granted (dossier number 1013) by the Nederlandse Vereniging voor Medisch Onderwijs (NVMO) Ethical Review Board (NMVO-ERB). The general guidelines that the NMVO-ERB maintains in its decision making are based on ethical principles from existing frameworks and codes of conduct such as the Declaration of Helsinki (World Medical Association, 1964).

Materials, apparatus and measurements

Felt model

On the test day, the students performed the surgery on an inguinal hernia simulation felt model. The felt model mimics the abdominal wall layers of the human body as each felt layer corresponds with an abdominal wall layer. Positioned within the correct layers, the fragile nerves, e.g. ilioinguinal, iliohypogastric and genital branch of the genitofemoral nerve, can be found. This model is identical to the one used for the video demonstration in the online course, see Figure 1.

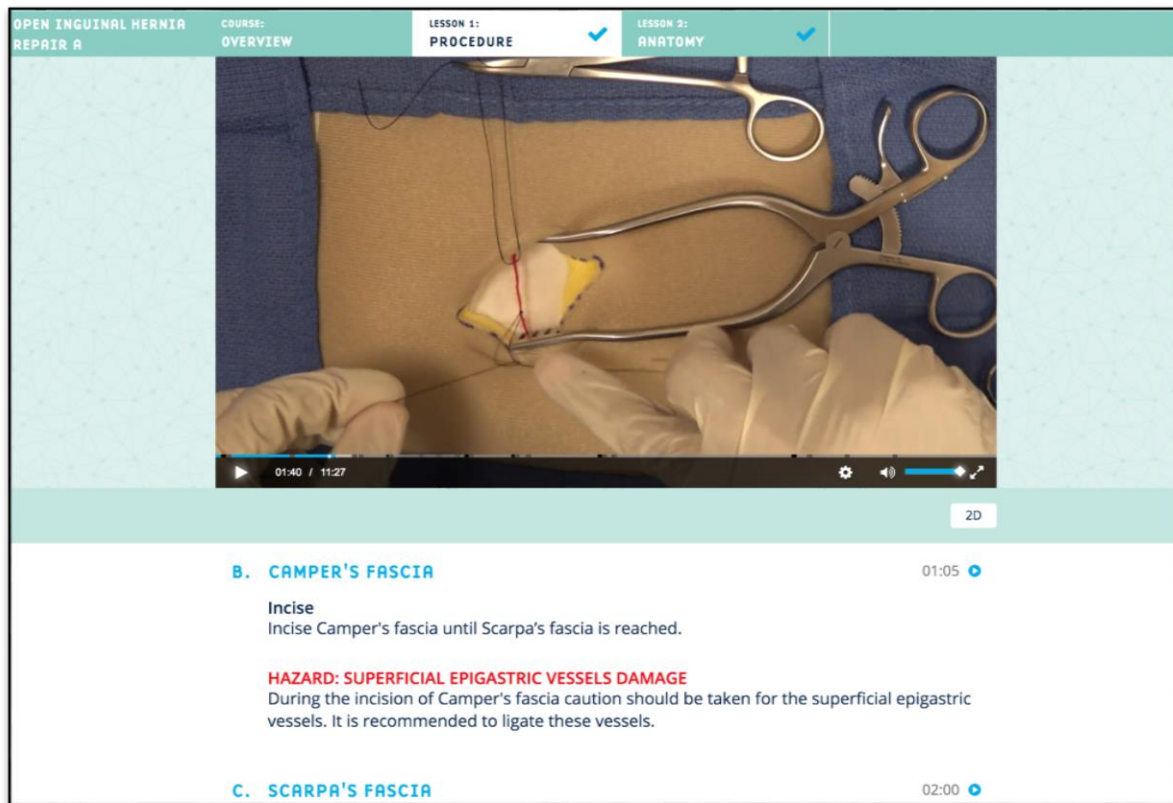


Figure 1: Felt model as demonstrated in the video of the training website (surgicalsteps.com).

Questionnaires - baseline measurements

Prior to performing the surgery, the medical students were asked to fill out a small questionnaire regarding influences on performance and demographics (see Appendix E). Within this writing, due to the small sample size only year of medical school and self-reported study time were included in data analysis as covariates. Additionally, prior to each surgery they were asked to rate their arousal and pleasure using the Affect Grid (see Appendix D). The Affect Grid is a single-item scale to repeatedly assess people's subjective affective states in a fast manner. The scale consists of a 9 x 9 grid, where the affective valence (ranging from unpleasantness to pleasantness) is presented on the horizontal axis and the arousal state (ranging from high arousal to sleepiness) is presented on the vertical axis (Russell et al., 1989).

Spatial ability test

The student performed the spatial relation ability tests directly after the baseline measurement. Spatial relation ability is the ability to mentally transform 2D representations into 3D images and mentally rotate them. Mental transformation and rotation domains were

tested respectively with the use of a validated 24-item Mental Rotation Test and 12-item Paper Folding test.

The Mental Rotation Test consists of 24 items where two-dimensional drawings of three-dimensional figures have to be compared with other similar figures (Langlois et al., 2017). The goal of the task is to determine which two of the four figures on the right side are correct rotations on the target figure on the left. The maximal outcome is 24, where one point per item is given only when the subject successfully identifies both correct rotations (Hegarty, 2010). Here, the mental rotation ability is tested. The Paper Folding Test consists of 12 items where mentally a square piece of paper is being folded and a hole has been perforated somewhere in the paper. The goal of the test is to determine which of the five figures on the right shows correctly how the paper will look like when it is unfolded again. One point is given per correct answer, with a maximum of 12 points (Hegarty, 2010). With this test, the mental folding ability is tested. Both abilities, mental rotation and mental folding are types of spatial ability.

Video recording and motion tracking

Per student both surgeries were recorded on video without sound by an assistant to obtain the model, hands and the actions performed. The camera was mounted above the felt model for optimal visualisation. Furthermore, the students' efficiency was tracked using motion tracking to determine path length, idle time, speed and acceleration. These are outcomes of performance efficiency during a surgical task. Since the total operation time was held constant, speed and acceleration were excluded in this writing since it did not have an additional value to the data. The motion tracker used in this study is the PST Base-55/100.

The PST Base (PS-Tech B.V., Amsterdam, The Netherlands) is an optical tracker that, in conjunction with the bundled software, can calculate a 6 degrees of freedom (6DoF) position and orientation of an object by means of retro-reflective markers. The system works by illuminating the measurement space with near-infrared (IR) light. The retro-reflective markers placed on the object to be tracked reflect this light back to a pair of stereoscopic cameras, see Figure 2. A glove was equipped with 27 retro-reflective markers and the system was trained to recognize this as a new object. The system exported a data file containing comma separated values of the objects position and orientation in X,Y and Z axis at a rate of 120 samples per second. These values were imported in Matlab and Microsoft Excel for further analysis. Specific informed consent is obtained from the students for recording and motion tracking their hands and actions.

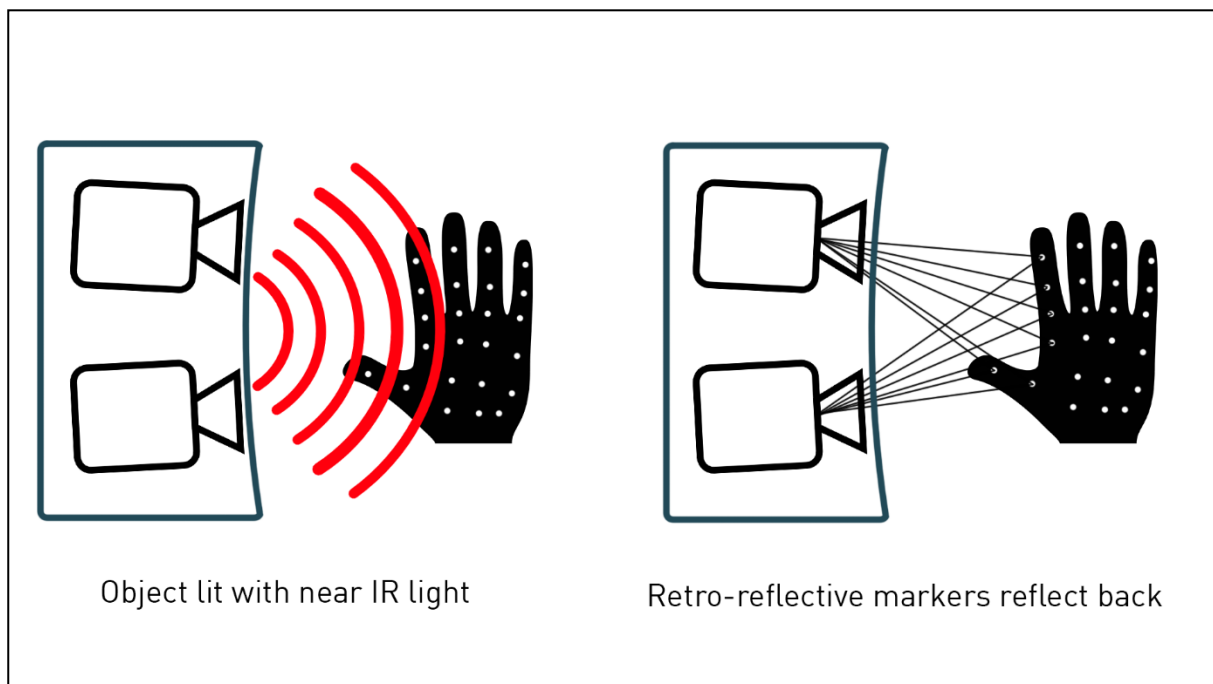


Figure 2: Measurement space is illuminated with near- infrared (IR) light. The retro-reflective markers attached to the glove reflect this light back to a pair of stereoscopic cameras.

Feedback forms

For giving feedback, two assessment forms were used, the OSATS (Appendix C) and the OCHRA (Appendix B). The OCHRA is a procedure specific step-by-step skills assessment checklist, where each step during surgery is evaluated. It is characterised by a breakdown of a procedure into tasks (Ahmed et al., 2011). Here, assessed is whether the steps are being performed correctly and if errors are made. Errors are classified into procedural and executional errors. Each error can be consequential or not (Tang et al., 2004).

The OSATS on the other hand is a seven-item global rating scale that assesses the overall competencies during surgery such as time, motion, knowledge of instruments and knowledge of specific procedure. Here, a 5-point Likert scale is used.

The difference between these two types of feedback is that the OCHRA is a specific checklist, where the OSATS is a global rating scale. Another difference is that the OCHRA is an assessment tool using error analysis adapted from human reliability, where the OSATS does not look at errors (Ahmed et al., 2011). The OSATS assesses operative skills in a less detailed way than the task-specific checklist (OCHRA), providing a ‘structured gestalt’ of performance (Martin et al., 1997). At last, the OSATS is not limited to any specific procedures, so it can be applied to any other skill assessment (Niitsu et al., 2013).

Evaluation instruments

In the larger study, the videos of the surgeries were assessed using the OSATS and the OCHRA by authors to determine accuracy of the performed surgery. In the current study, the OSATS and OCHRA were only used as the intervention instead of both intervention and measurement method.

Data collection

All the data collected (video's, motion tracking data, questionnaires) were stored anonymously on two separate external hard discs. Data will be kept for 10 years. In case a participant withdrew his or her consent to participation, the data of this participant have been destroyed.

Design

The study took place in the SkillsLab of the Leiden University Medical Center. The study was executed as a prospective randomised cohort and had a between subject design. This resulted in two groups (see Figure 3 for a complete overview of the study design):

- Group A received feedback using the OSATS
- Group B received feedback using the OCHRA

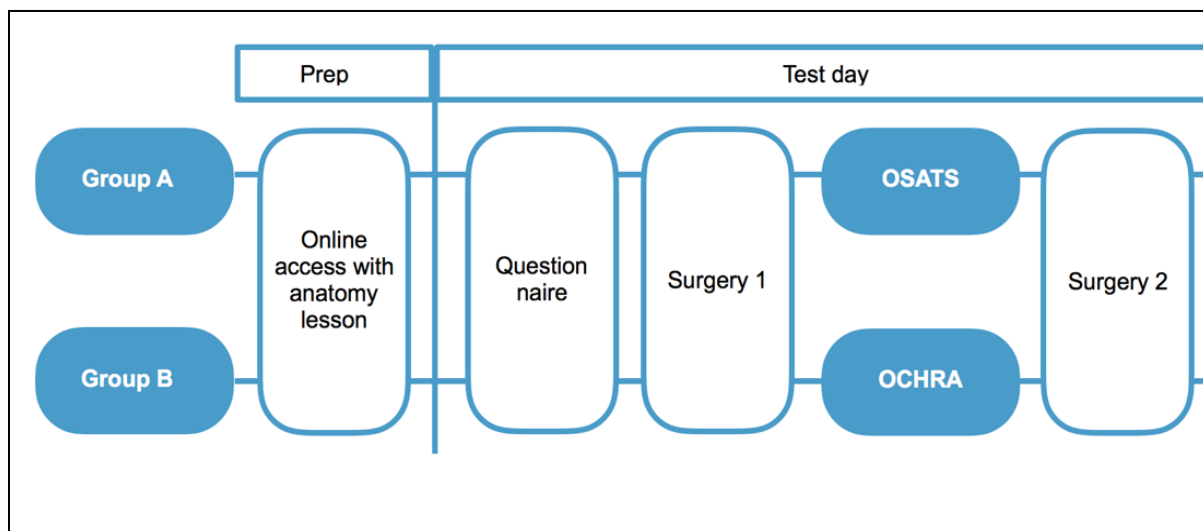


Figure 3: Study design, with Group A & B.

The dependent variables were path length and idle time. The motion tracking data were based on the position of the glove in a three dimensional field. The path length is the total three dimensional distance the hand travelled from the starting position, expressed in meters. The amount of idle time of each participant during both operations was expressed in seconds. Based on previous literature on motion tracking data, a predefined threshold for idle time was established at a velocity of 20 mm/seconds to identify idle periods. Furthermore, in order to avoid capturing very small pauses of movement, a minimum duration of 0.5 seconds was chosen to determine an unintentional idle period (D'Angelo et al., 2015).

The independent variable was the type of feedback (OCHRA/OSATS). Ordinal data (self-reported study time and year of medical school) and interval data (Affect Grid, Mental Rotation Test (MRT) and Paper Folding test (PFT)) were presented as numbers and continuous data (path length and idle time) were described by means and standard deviations or medians and ranges, according to the distribution. The primary outcomes were the motion tracking data before and after providing feedback using the OSATS and the OCHRA. Covariates included in the analyses were year of medical training, self-reported study time, the scores of the MRT and PFT.

Procedure

Online course

One week before the test day, the students were given access to the online course, which they could access at any place of preference. The course consisted of an anatomy lesson and the description and demonstration of the inguinal hernia repair. The surgery was described and demonstrated in a video on a felt model in a step by step manner. The preparation time was approximately one hour. The step by step description was prepared using the step by step framework and can be found in Appendix A. The felt model used for the video on the website is identical to the model used on the test day.

Test day

All medical students received access to an online course to study the inguinal region anatomy and the open inguinal hernia surgery in a step by step manner. On the test day, which took place at the Skillslab of the Leiden University Medical Center, the participants first received the information letter, where they were able to ask additional questions about the

research before signing the informed consent. Once the participant had agreed to participate, the research could begin.

First they received a short explanation on how to use the Affect Grid (Appendix D), Mental Rotation Test (Appendix F) and Paper Folding Task (Appendix G), after which they filled in the Affect Grid, a small questionnaire regarding influences on performance and demographics (Appendix E), the Mental Rotation Test and the Paper Folding Test. Each student was provided 10 minutes to complete the Mental Rotation Test and 3 minutes for completion of the Paper Folding Test. After completing the questionnaires, the participants performed the open inguinal hernia repair on a felt model (Pretest). The steps of the surgery are described in Appendix A (Ahmid, 2004; Lichtenstein et al., 1989). Each student was given 30 minutes time for the open inguinal hernia repair surgery. The students were notified every 10 minutes.

The medical students were given feedback according to the randomised group they were assigned to. Then, the students completed the Affect Grid, again followed by performing the second open inguinal hernia repair to test the effects of feedback on their skills and consolidation of memory on short term (Posttest). Again, they had 30 minutes time for the surgery and were notified every 10 minutes. The test day lasted approximately 90 minutes. In total, this research lasted about two and a half hours.

To perform the surgeries, each participant was given a scalpel, forceps, scissors, retractor, needle driver, sutures and a penrose drain. Both surgeries were video-recorded and motion tracked. Motion tracking was performed to measure performance efficiency through path length (distance travelled by medical students' hands) and idle time (amount of time medical students' hands were not moving).

Providing feedback

Feedback was provided by two postgraduate doctoral degree (PHD) in medicine students. Before the research started, they prepared themselves by watching the surgical video of the open inguinal hernia repair together, several times. Furthermore, they practiced scoring the video with the OSATS & OCHRA multiple times and compared their outcomes. Since the research is not a double-blind study and the authors were also the ones giving the final score, the surgical videos of the participants were coded in order to prevent bias in results. The inter-rater reliability will be calculated at the end of the larger research. Authors provided feedback in a structured manner using both the OCHRA (Appendix B) and the OSATS (Appendix C), based on the randomised group. The feedback needed to be given within 10 minutes. The total

feedback time was held constant to ensure that one type of feedback did not lead to better outcomes due to more time.

Analyses

During the process of data analysis, several steps were taken before the raw data were useful for analysis. First, a Matlab script was written and used to trim the raw localisation data to the allotted time of 30 minutes (1800 seconds). Subsequently, an effort was made to minimise the effect of missing data points by utilising interpolation. It was decided to only utilise interpolation of blocks of missing data smaller than .5 seconds (15 frames), as longer interruptions were deemed to be unreliable. Interpolation was performed using the distance between the start and end of the missing data block and distributing this movement over the number of timestamps. Next, the trimmed and (partially) complemented localisation data were used to calculate the participants movements during each operation. This resulted in the determination of the path length and idle time.

Furthermore, sensitivity analyses were performed within the Matlab script for three variables to investigate the effects of various threshold parameters. First, the length of the interpolation blocks were changed to 1 second (30 frames) and 1.5 seconds (45 frames). Second, since a predefined threshold for idle time was established, the idle velocity threshold used for the determination of idle time was changed to 1,5,35 and 50 mm/s. Finally, the idle time minimal duration threshold was investigated for times of .25 and .75 seconds with interpolation block lengths of .5, 1 and 1.5 seconds and an idle velocity of 20 mm/s. The idle times were represented as a relative number calculated by multiplying the absolute idle time with the percentage of registered data (100% - missing data). Results of the sensitivity analyses can be found in the *Exploratory tests* section within the *Results* section.

Before analyses of the MRT and PFT scores could be performed, the participants were divided for both MRT as PFT into low- and high scoring groups using a median split on the MRT scores and a mean split on the PFT scores. The median MRT score for all participants (N=24) was (Mdn= 18.00, range: 12.25-21.00) and the mean PFT score for all participants (N=24) was (M= 6.96, SD=1.85).

Next, the transformed data were inspected for any abnormalities and assumptions for analyses were checked. This included normality checks, homogeneity of variance, missing data and outliers. All data, ordinal (year of medical school and self-reported study time), interval (Affect Grid, Mental Rotation Test and Paper Folding test) and continuous data (path

length and idle time) were tested for normality using the Shapiro-Wilk test. Furthermore, Independent-Samples T-Tests were performed to check for homogeneity of variance. In order to identify outliers, possible univariate outliers have been detected via boxplots. No outliers were removed for further analysis since this would have made the sample size even smaller. Further details are discussed in the *Results* section.

To investigate the effects of feedback intervention on surgery performance, a Repeated Measures ANOVA was used with feedback (OSATS & OCHRA) as a between subjects factor and time (pre- and post-test) as a within subjects factor. Covariates included in the analyses were year of medical training, self-reported study time, MRT and PFT.

To investigate whether participants with a high spatial ability outperformed students with a low spatial ability on the first and second surgery, the following tests were performed. First, to see whether students with high mental rotations ability (MRT scores) outperformed students with low mental rotations ability on both the first and second operation, the idle times and path lengths were compared between both groups using the Mann-Whitney U Test. Path length scores on the first operation had no equal variances between MRT groups, which made it not possible to perform a Repeated Measures ANOVA. Second, to see whether students with high mental folding ability (PFT scores) outperformed students with low mental folding ability on the first and second operation, a Repeated Measures ANOVA with (high vs low) PFT scores as a between subjects factor and time (pre- and post-test) as a within subjects factor.

Furthermore, a joint spatial ability score was created where z-scores of MRT and PFT have been calculated and added together. Based on this, students were divided into high and low spatial ability groups. To investigate whether students with high spatial ability outperformed students with low spatial ability on the first and second operation, a Repeated Measures ANOVA with (high vs low) spatial ability scores as a between subjects factor and time (pre- and post-test) as a within subjects factor.

To investigate whether affect measures differed between feedback type, the difference in affect measures was compared using the Mann-Whitney U Test. Since the affect measures were not normally distributed, a Repeated Measures ANOVA could not be done.

In the *Results* section, the covariates year of medical school and the low vs high MRT and PFT groups are presented as M_{year} , $MRT_{\text{lowvshigh}}$ and $PFT_{\text{lowvshigh}}$. The idle time and path length outcomes as well as the arousal and pleasure scores for the first operation are presented as idle time 1/path length 1/arousal 1/pleasure 1 and for the second operation as idle time 2/path length 2/arousal 2 and pleasure 2.

P-values of less than .05 were considered statistically significant. All statistical analyses have been carried out by using IBM SPSS Advanced Statistics 21.

Ethics

Medical students were invited by the authors for participation. Students from Leiden University Medical Center (LUMC) in their first to sixth year were approached for participation. The aim was to include 50 students in six months. Since this aim was not reached, the inclusion time was extended until 50 students participated.

The invited students were given a week to consider participation. The advantage of participation was that the students were taught an open inguinal hernia repair surgery which they could perform twice on a felt simulation model. Another advantage was that this study may have had potential effect on the future of medical education. A disadvantage was the time required to invest in the experiment; approximately one hour preparation and approximately 90 minutes on the experiment day. Possible risks were privacy violation and study progress limitation. All measures were taken to prevent these risks. All data were coded to handle anonymously. Only the main researchers were able to access the file with the codes. The main researchers T. Nazari and K. Bogomolova were not in any form concerned with the education of the students.

There was no disadvantage to the medical study progress when students decide not to participate or to withdraw their consent. In case students withdraw their consent for participation, data of this participant were destroyed. The entire research team (all authors) may have had knowledge of a student's participation and possible withdrawal of participation. The entire research team was not directly involved in any **shape or** form with the education or the grading of the students.

As the students acquired new skills and as they were provided the possibility to perform surgery twice and receive feedback on their performance, they did not receive any form of compensation.

Results

Three participants were excluded before analysis of the data. One participant was left handed, while only right handed students were included in this research. For two participants motion tracking data for the first operation were lost, which meant they could not be included in data analysis. After excluding these participants, the sample size consisted of 22 participants

(N=22), with 12 participants in the OSATS group (Male=6, Female=6, $M_{age}=20.58$ and $M_{year}=2.17$) and 10 participants in the OCHRA group (Male=3, Female=7, $M_{age}=20.70$, $M_{year}=2.30$).

Missing data

Before analysis of motion tracking data was performed, the Excel files were checked for missing data points. Various causes led to dropouts in localisation data. The percentage of missing data was calculated for each operation for both feedback groups. This was calculated by dividing the amount of missing data points by the amount of data points after truncating at 1800 seconds, where one second contained 30 timeframes. In order to correct for the missing data points, several steps were taken.

First, blocks of missing data points smaller than 15 frames (0.5 second) were filled by interpolation. This showed a minimum missing data improvement of .61% and a maximum of 3.01%. Furthermore, an interpolation sensitivity analysis was performed for missing data blocks of 30 (1 second) and 45 (1.5 second) frames. These showed a minimum missing data improvement of 1.26% for 30 frames, 1.40% for 45 frames and a maximum improvement of 5.04% for 30 frames and 7.05% for 45 frames. Statistical analysis of motion tracking data on feedback groups with the interpolation sensitivity analysis can be found in the *Exploratory tests* section within the *Results* section. Mean missing data percentages were calculated per operation for both feedback groups before and after the interpolation of missing data blocks and can be found in Table 1.

Second, in order to reduce the impact of missing data on idle time, the absolute idle time period was corrected for the percentage of missing data. In other words, for each operation the percentage of missing data was removed from the total idle time which yields an idle time without missing data points. Statistical analysis of idle times without missing data on feedback groups can be found in the *Exploratory tests* section within the *Results* section.

Table 1. *Mean Percentages and standard deviations of missing data for both operations per feedback group before and after interpolation of missing data blocks.*

Operation	Missing data in %			
	No interpolation	Interpolation 15 frames (0.5 second)	Interpolation 30 frames (1 second)	Interpolation 45 frames (1.5 seconds)
1st	OSATS: 12.10 (SD= 4.94)	OSATS: 10.55 (SD= 4.68)	OSATS: 9.22 (SD= 4.44)	OSATS: 8.22 (SD= 3.98)
	OCHRA: 12.99 (SD= 7.35)	OCHRA: 11.68 (SD= 6.97)	OCHRA: 10.43 (SD= 6.45)	OCHRA: 9.50 (SD= 5.99)
2nd	OSATS: 11.36 (SD= 6.00)	OSATS: 9.97 (SD= 5.90)	OSATS: 8.71 (SD= 5.86)	OSATS: 7.82 (SD= 5.67)
	OCHRA: 11.07 (SD= 5.12)	OCHRA: 9.52 (SD= 4.71)	OCHRA: 8.27 (SD= 4.29)	OCHRA: 7.19 (SD= 3.88)

First Hypothesis: Feedback

The first hypothesis stated that students in the OCHRA group outperform students in the OSATS group on the second operation, showing smaller idle times and path lengths. To test this hypothesis, a Repeated Measures ANOVA was performed with time as the dependent variable and feedback group as the independent variable. Furthermore, year of medical school, self-reported study time, $MRT_{lowvshigh}$ and $PFT_{lowvshigh}$ were added as covariates.

Outliers: The total idle time and path length scores were visually inspected for outliers with boxplots for both feedback groups on both the first and second operation. Two outliers in idle time on the first operation in the OCHRA group were detected. The first outlier was observed in participant 15, with a score of 600 and the second outlier was observed in participant 24, with a score of 580. These scores were within the range of two standard deviations of the overall mean score for the OCHRA group ($N=10$, $M=414$, $SD=177$) and therefore not excluded from the results.

Normality tests: A Shapiro-Wilk's test ($p > .05$) and visual inspection of normal Q-Q plots, box plots and histograms showed that the idle time and path length scores on the first and second operation were approximately normally distributed for both feedback groups. All standardised z-scores of skewness and kurtosis were within the ± 1.96 limits, suggesting that the data is normally distributed.

Homogeneity of variance: There was homogeneity of variances ($p > 0.5$) for both feedback groups (OSATS & OCHRA), for idle time and path length scores on the first and second operation as assessed by Levene's test of homogeneity of variances. Furthermore, there

was a homogeneity of covariances, as assessed by Box's test of equality of covariance matrices ($p = .698$). Homogeneity of variance was assumed and a Repeated Measures ANOVA could be performed.

Repeated Measures ANOVA: The main effect of time showed a statistically significant difference in idle times between the first and second operation, $F(1,20) = 70.880$, $p < .001$, $partial \eta^2 = .780$. However, there was no significant interaction between time and feedback group on idle times $F(1,20) = .082$, $p = .777$, $partial \eta^2 = .004$. The main effect of time also showed a statistically significant difference in path length between the first and second operation, $F(1,20) = 7.214$, $p = .014$, $partial \eta^2 = .265$. Again however, no statistically significant interaction between time and feedback group on path lengths was found $F(1,20) = .562$, $p = .462$, $partial \eta^2 = .027$. Furthermore, there was no significant effect of covariates Year, Study time, $MRT_{lowvshigh}$ and $PFT_{lowvshigh}$; all F 's < 1 . For a complete overview of idle times and path length between the first and second operation, see Figure 4.

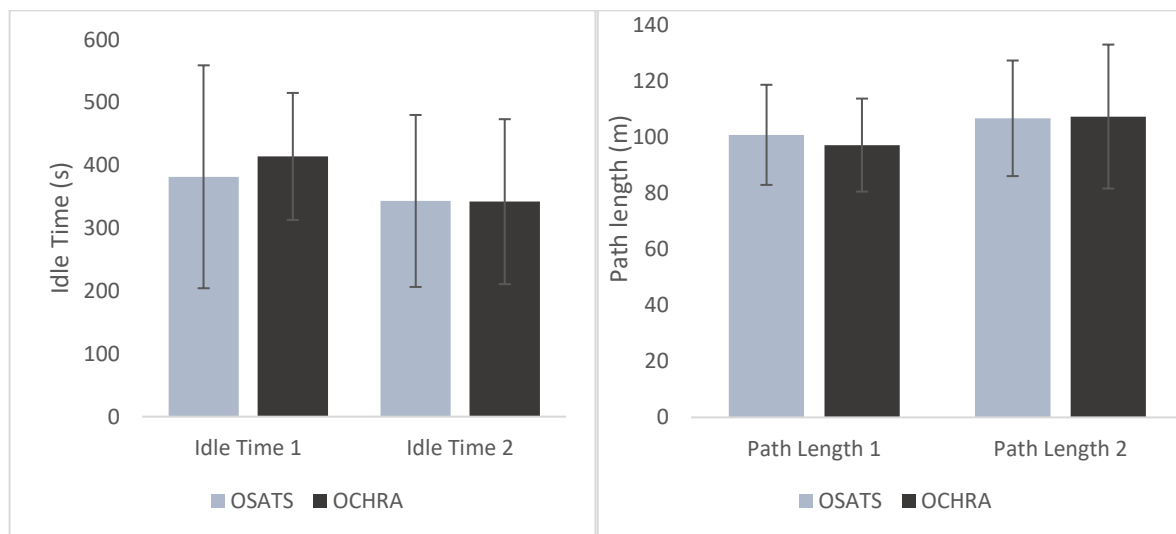


Figure 4: Mean scores and standard deviation of idle times and path lengths for the first and second operation for both feedback groups. Idle time in s and path length in m.

Second Hypothesis: Spatial Ability

The second hypothesis stated that students with high spatial ability outperform students with low spatial ability on both the first and the second operation.

Mental Rotation Test

Outliers: The idle time and path length scores were visually inspected for outliers with boxplots for both groups (low vs high) on both the first and second operation. Two outliers were detected. The first outlier was observed in the low MRT group on the second operation in path length score. This participant, number 5, had a score of 62.29. This score was within the range of two standard deviations of the overall mean score for the low MRT group (N=5, M= 104, SD= 26) and therefore not excluded from the results. The second outlier was detected in the low MRT group on the first operation in idle time. This participant, number 2, had a score of 110. This score was as well within the range of two standard deviations of the overall mean score for the low MRT group (N=5, M= 312, SD= 125) and therefore not excluded from the results.

Normality Test: A Shapiro-Wilk's test ($p > .05$) and visual inspection of normal Q-Q plots, box plots and histograms showed that the idle time and path length scores on the first and second operation were approximately normally distributed for both MRT (low vs high) groups. All standardised z-scores of skewness and kurtosis were within the ± 1.96 limits, suggesting that the data is normally distributed.

Homogeneity of variance: There was homogeneity of variances ($p > 0.5$) for both MRT (low vs high) groups, for idle time scores on the first and second operation as assessed by Levene's test of homogeneity of variances. The path length scores on the second operation showed equal variances for both MRT groups ($p > 0.5$) while path length scores on the first operation did not ($p < 0.01$). Homogeneity of variance could not be assumed.

Mann-Whitney U Test: Because of the violation of the assumption of homogeneity of variance for path length scores on the second operation, a non-parametric test was chosen. A Mann-Whitney U Test was run to determine if there were differences in idle times and path lengths between participants with low and high MRT scores on the first and second operation.

Median idle time scores were not statistically significantly different between the MRT groups on the first operation for the low group (Mdn=348) and for the high group (Mdn=403), $U=82$, $z=1.775$, $p= .082$. On the second operation, there was also no statistically significant difference between low and high MRT groups for idle time, with low (Mdn=243) and high (Mdn=393), $U=99$, $z=1.698$, $p= .096$. Although there is no significant difference between MRT groups and idle time, there is a trend towards longer idle times in the participants with higher mental rotation ability.

For path lengths, there was no significant difference between MRT groups within the first and second operation. On the first operation for the low group there was a median score

of (Mdn=103) compared with high (Mdn=92), $U=48$, $z= -.546$, $p= .616$. On the second operation, there were median path length scores of (Mdn=104) for the low group and (Mdn=108) for the high group with $U=88$, $z= 1.054$, $p= .312$. See Figure 5 for an overview of median idle time and path length scores on both operations for both MRT groups.

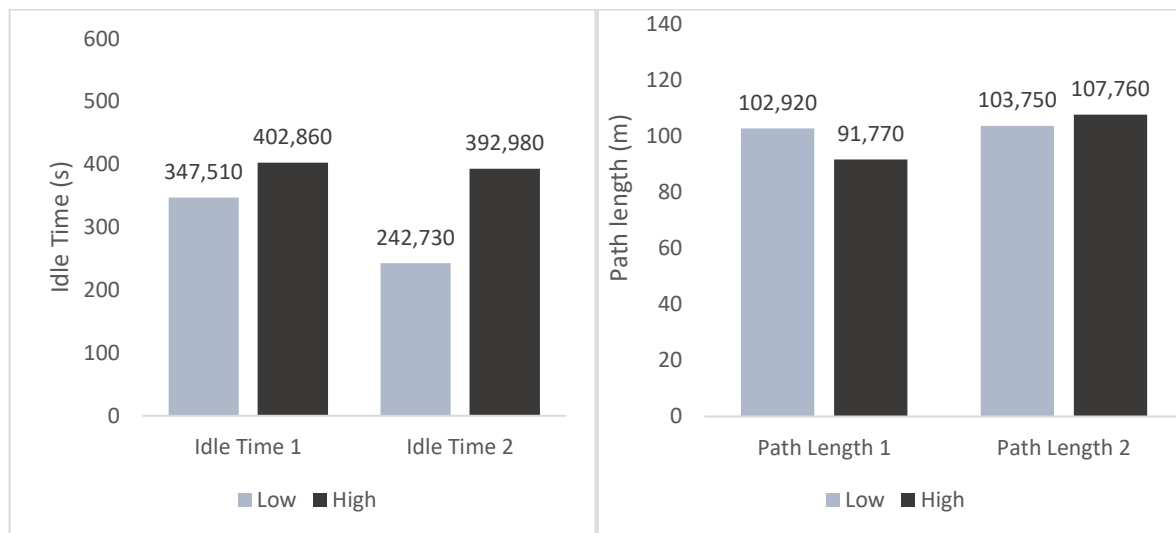


Figure 5: Median scores of idle times and path lengths for the first and second operation for both MRT groups (low vs high). Idle time in s and path length in m.

Paper Folding Test

Outliers: The total idle time and path length scores were visually inspected for outliers with boxplots for both groups (low vs high) on both the first and second operation. Two outliers were detected in the idle time scores on the first operation in the high PFT group. The first outlier was inspected in participant 6, with a score of 629 and the second outlier was observed in participant 7, with a score of 119. The first outlier was within the range of two standard deviations of the overall mean score for the high PFT group ($N=12$, $M=390$, $SD=132$) while the second outlier was not. Even though because of the small sample size, it was decided not to exclude these scores from the results.

Normality Test: A Shapiro-Wilk's test ($p > .05$) and visual inspection of normal Q-Q plots, box plots and histograms showed that the idle time and path length scores on the first and second operation were approximately normally distributed for both PFT (low vs high) groups. All standardised z-scores of skewness and kurtosis were within the ± 1.96 limits, suggesting that the data is normally distributed.

Homogeneity of variance: There was homogeneity of variances ($p > 0.5$) for both PFT (low vs high) groups, for idle time and path length scores on the first and second operation

as assessed by Levene's test of homogeneity of variances. Homogeneity of variance was assumed and a Repeated Measures ANOVA could be performed.

Repeated Measures ANOVA: A Repeated Measures ANOVA was performed to investigate whether idle times and path lengths differed significantly between the two PFT groups. No significant interaction effect between time and PFT scores on idle times were found $F(1,20) = .062, p = .805, \text{partial } \eta^2 = .003$. Furthermore, there was no significant interaction effect between time and PFT scores on path lengths, $F(1,20) = .006, p = .940, \text{partial } \eta^2 = < .001$. See Figure 6 for an overview of mean idle time and path length scores on both operations for both PFT groups.

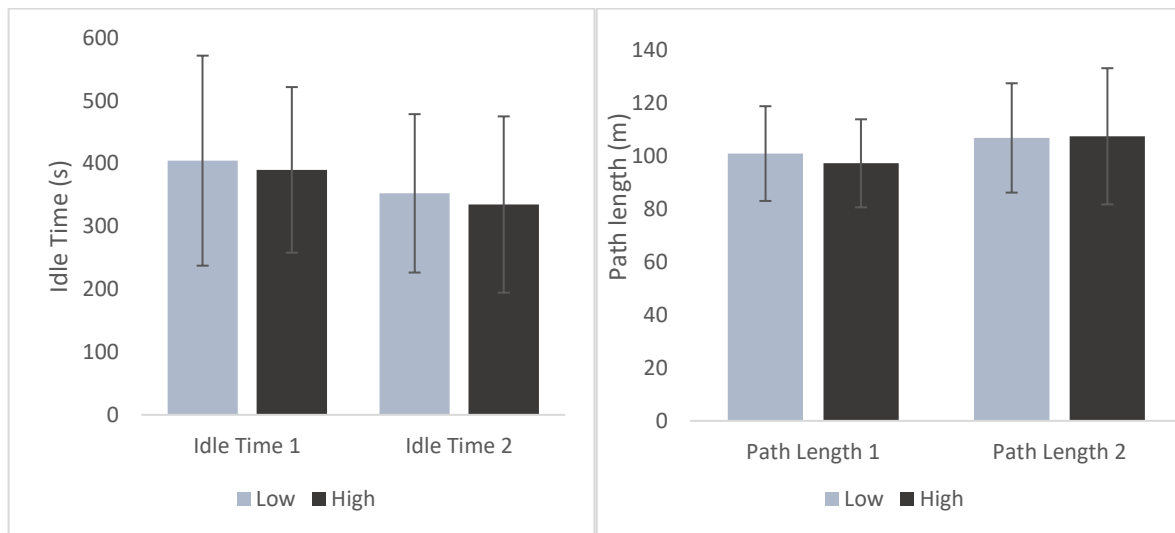


Figure 6: Mean scores and standard deviations of idle times and path lengths for the first and second operation for both PFT groups (low vs high). Idle time in s and path length in m.

Exploratory Tests

Arousal and pleasure scores

Since cognitive states can be subjected to the effect of feedback and the ability to learn, it was of interest to see whether students have a change in arousal and pleasure scores after receiving one of the feedback types. After looking at affect grid data, the following results were found.

Outliers: The pleasure and arousal scores were visually inspected for outliers with boxplots for both feedback groups on both the first and second affect grid scores. One outlier was detected in the second affect score for arousal in the OCHRA group. This outlier was

inspected in participant 4, with a score of 4. This outlier was within the range of two standard deviations of the overall mean score for the OCHRA group ($N=9$, $M=6.5$, $SD= 1.51$) and therefore not excluded from the results.

Normality test: A Shapiro-Wilk's test ($p < .05$) and visual inspection of normal Q-Q plots, box plots and histograms showed that the pleasure and arousal scores on both the first and second affect grid scores were not normally distributed for both feedback groups. Although all standardised z-scores of skewness and kurtosis were within the ± 1.96 limits, normality distribution of affect grid data could not be accepted.

Homogeneity of variance: There was homogeneity of variances ($p > .05$) for both arousal scores on both affect grids for both feedback groups as assessed by Levene's test of homogeneity of variances. The pleasure scores on the first affect grid showed equal variances for both feedback groups ($p > .05$) while pleasure scores on the second affect grid did not ($p < .05$). Homogeneity of variances could not be assumed.

Mann-Whitney U Test: Because of the violation of the assumption of a normal distribution and homogeneity of variance, a non-parametric test was chosen. A Mann-Whitney U Test was run to determine if there were differences in pleasure and arousal scores between participants in the two feedback groups on both the first as the second affect grid scores.

Median arousal scores were not statistically significantly different between the two feedback groups on the first affect grid with ($Mdn=7$) for the OSATS groups and ($Mdn=6$) for OCHRA, $U=45$, $z= -.656$, $p= .512$. On the second affect grid, there was also no significant difference between the two feedback groups for arousal scores, with ($Mdn=7$) for the OSATS group and ($Mdn=7$) for OCHRA, $U=47.5$, $z= -.476$, $p= .634$.

For pleasure scores, there were also no significant differences between feedback groups on both affect grids. On the first affect grid, there was a median score of ($Mdn= 7$) for the OSATS group and ($Mdn=6$) for OCHRA, $U=35$, $z= -1.380$, $p= .168$. On the second affect grid, there were median pleasure scores of ($Mdn=6.5$) for the OSATS group and ($Mdn=7$) for the OCHRA group, $U=52.5$, $z= -.109$, $p= .914$. See Figure 7 for an overview of arousal and pleasure scores on both Affect grids for both feedback groups.

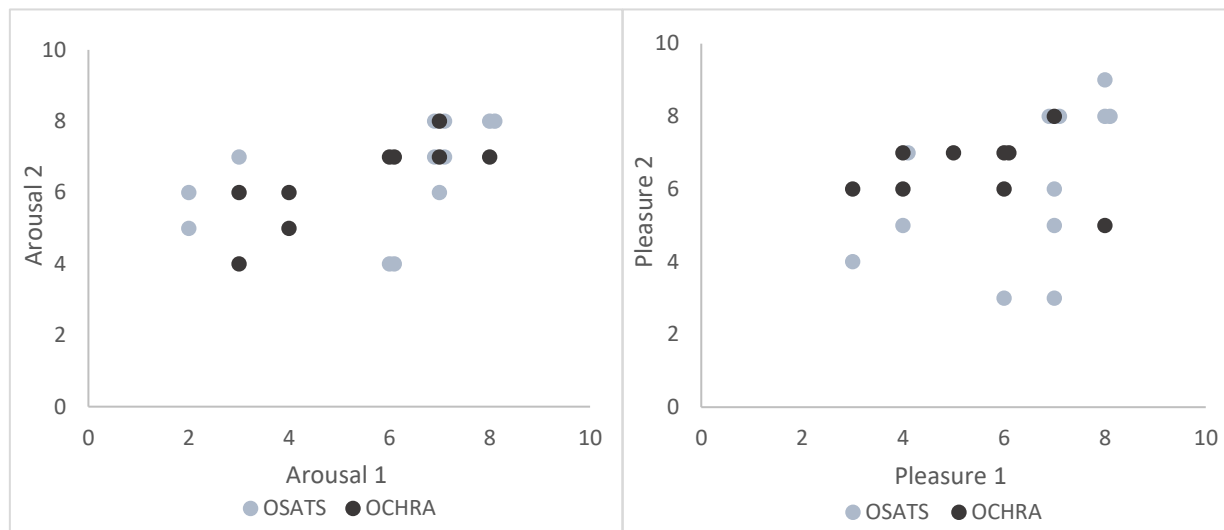


Figure 7: Scatterplots of median arousal and pleasure scores on the first and second Affect grid for both feedback groups.

Spatial ability

A joint spatial ability score was created by adding up MRT and PFT z-scores. Since the MRT and PFT tests on time (pre- and post-test) did not show any statistically significant differences between the two groups (high and low) on idle times and path lengths, it is of interest to investigate whether a joint spatial ability score will show significant differences.

Outliers: The total idle time and path length scores were visually inspected for outliers with boxplots for both groups (low vs high) on both the first and second operation. One outlier was detected in the idle time scores on the second operation in the high spatial ability group. The outlier was inspected in participant 5, with a score of 82.72. The outlier was outside the range of two standard deviations of the overall mean score for the high spatial ability group ($N=12$, $M=367$, $SD=127$). However, due to the small sample size, it was decided not to exclude this score from the results.

Normality Test: A Shapiro-Wilk's test ($p > .05$) and visual inspection of normal Q-Q plots, box plots and histograms showed that the idle time and path length scores on the first and second operation were approximately normally distributed for both spatial ability (low vs high) groups. All standardised z-scores of skewness and kurtosis were within the ± 1.96 limits, suggesting that the data is normally distributed.

Homogeneity of variance: There was homogeneity of variances ($p > 0.5$) for both spatial ability (low vs high) groups, for idle time and path length scores on the first and second

operation as assessed by Levene's test of homogeneity of variances. Homogeneity of variance was assumed and a Repeated Measures ANOVA could be performed.

Repeated Measures ANOVA: A Repeated Measures ANOVA was performed to investigate whether idle times and path lengths differed significantly between the two spatial ability groups. No significant interaction effect between time and spatial ability groups on idle times were found $F(1,20) = .736, p = .401, \text{partial } \eta^2 = .036$. Furthermore, there was no significant interaction effect between time and spatial ability groups on path lengths, $F(1,20) = .014, p = .907, \text{partial } \eta^2 = <.001$. See Figure 8 for an overview of mean idle time and path length scores on both operations for both spatial ability groups.

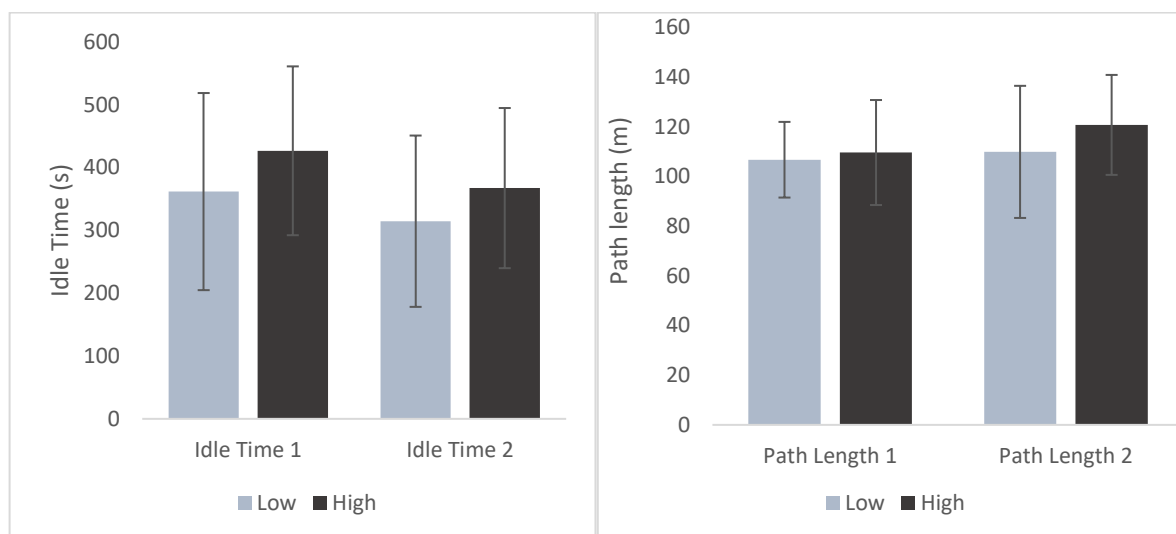


Figure 8: Mean scores and standard deviations of idle times and path lengths for the first and second operation for both spatial ability groups (low vs high). Idle time in s and path length in m.

Sensitivity analyses

Sensitivity analysis were performed for three different variables, resulting in a total of 21 repeated measures ANOVA's. The different investigated variables were the length of the interpolation blocks (.5, 1 and 1.5 seconds), idle time velocity (1,5,20,35,50 mm/s) and idle time minimal duration (.25, .5, .75 seconds). Furthermore, relative idle times were added in each test to investigate whether there is a difference on absolute and relative idle times between feedback groups.

No statistically significant interaction effects between time and feedback group on idle times and path lengths were found ($p > .05$). Moreover, results showed that the greater the idle

velocity the higher the p-value for idle times on feedback group. This was found within all different lengths of interpolation blocks.

Furthermore, the interaction effect between time and feedback group showed greater p-values on relative idle times compared to absolute idle times. This was found within all different lengths of interpolation blocks. At last, the threshold yielding the greatest differences in proportion among feedback group for both idle time and path length was an interpolation length of one second, an idle velocity of 20 mm/s with a minimal idle duration of .25 seconds.

Discussion

The aim of this study was to investigate which type of observational assessment tool with its feedback (OSATS & OCHRA), shows a greater contribution in teaching novice medical students technical skills for performing an open inguinal hernia repair. This study expected to find improved operation efficiency for the participants in the OCHRA group. This was foreseen to be characterised by shorter idle times and path lengths.

Feedback

However, the data suggested no statistically significant effect on the motion tracking data and feedback method, indicating that within this study, there is no difference in observational assessment tools when teaching novice medical students technical skills. Additionally, no interaction effect between time and feedback group on idle times and path lengths was found.

Novice medical students have less anatomical knowledge and knowledge of instrument use compared to specialists and residents. Therefore it was expected was that a more specific step-by-step assessment tool would provide novices with more detailed information of how to perform an operation compared to a global assessment tool. Our findings however go against this expectation. The OCHRA group did not show decreased idle times and path lengths compared to the OSATS group. These findings goes against prior research where the delivery of feedback is linked to an improved performance curve (Boyle et al., 2011; Parmar et al., 2011).

Various explanations were considered. Firstly, prior research showed that especially at the resident and specialist level, the OCHRA is a valid method for assessing surgical performance (Hiemstra & Jansen., 2010; Miskovic et al., 2011). Additionally, the OSATS was developed for assessing technical skills in surgical trainees/residents (Martin et al., 1997). This could indicate that when teaching novice students technical skills, who lack the technical competencies, the OCHRA nor the OSATA is a valid assessment tool.

Secondly, this non-significant interaction between time and feedback group could be explained by the amount of missing data. An interpolation sensitivity analysis showed no improvements on the outcomes for the feedback groups. Even an interpolation length of 45 frames (1.5 seconds) showed a mean missing data percentage of 8.86% on the first operation and 7.5% on the second. This is an indication of the relatively large sequences of missing data points present in the remainder of the localisation data. Further interpolation beyond 1.5 seconds was deemed meaningless. This occurrence of large chunks of missing data could have a detrimental effect on the analysis. Since the difference of missing data percentages between the two operations are minimal, it is debatable how much this would affect the localisation data. However, as it is unclear what occurs during the missing data, comparing missing data percentages between operations can give a distorted image of what is occurring and therefore as well on the path lengths and idle times.

Lastly, since the idle time calculations were predetermined using thresholds found in literature (20mm/s during at least .5 seconds) (D'Angelo et al., 2015) a sensitivity analysis of this threshold value was performed. The threshold analyses however, showed no improvements on the outcomes for the feedback groups.

Prior research showed that novices students showed greater amount of idle times compared with experts (D'Angelo et al., 2015). This study was conducted solely on novice students who lack technical competencies required for an open inguinal hernia surgery. It could be the case that the feedback methods provide insufficient improvement to these skills to observe a difference in idle times. Including a greater variation in experience levels within the research might show other findings.

Spatial ability

In this study, spatial ability was used as a sanity check since many researches already proved that students with higher spatial ability outperform students with low spatial abilities (Langlois, Wells et al., 2015; Ackerman, 1992; Langlois, Bellemare et al., 2015; Henn et al., 2018). However, our research was not in line with these results. No statistically significantly

interaction effects between time and MRT & PFT score were found. Moreover, there was a trend toward longer idle times in the participants with higher mental rotation ability which is in contrary with prior findings.

Additionally, there was no significant interaction effect between time and the joint spatial ability groups. In this study, the low spatial ability group contained (N=10) participants and the high spatial ability group (N=12). Because of these small groups, underestimates could have been made for individuals spatial ability due to the low sample size.

Significance

Even though no significant effects are found within this study, it is the first time that OSATS and OCHRA assessment tools are being compared by means of motion tracking data in novice medical students. Since medical education and thereby novices are the basis for later resident and specialist level, is it of great importance to investigate how medical students can be trained and educated in a matter that is most effective for all the knowledge and skills they have to learn and obtain. This research was the first step in understanding the differences in assessment tools and their effect on performance efficiency during a surgical task.

Limitations

As within any research, this study had several limitations. The first and one of the most important ones is the small sample size. Every feedback group contained not more than 12 subjects due to the small sample size. Because of this, it is hard to make any hard statements about the outcomes of the data. Furthermore, not all covariates of the medical students questionnaire could be included in the data analysis since it would not give any representative outcomes. For example, a total of 9 participants in this study were men compared to 13 women. In order to investigate whether there is a difference between men and women, a much larger sample size is needed to draw any conclusion.

Furthermore, a big issue in this study was the great amount of missing data. More than half of the localisation data contained more than 10% of missing data. This occurred even after interpolation of blocks of missing data smaller than .5 seconds. Within this study it was not realistic to remove all participants with high missing data percentages due to the small sample size. An explanation for the missing data is that the optical tracker only measures the hand when at least four of the markers are captured. When the hand for example is flat on its back, outside of the optical tracker's field, or when hands are crossed, no measurements are done.

Moreover, due to malfunctioning of the computer system, several motion tracking data were lost, leading to data files with different degrees of missing data, up to complete loss of motion tracking data. Improving the reliability of the motion tracking acquisition could have drastic effects on the outcome of this study. During the motion tracking data camera images were made to further evaluate the different feedback methods. These were not used during this study due to time constraints. The video analyses could potentially be used to fill in the empty spaces in the data files by cross-referencing the motion tracking data.

In this study the total operation time was held constant at 30 minutes. This was decided because of time restraints, since the total study time already lasted 90 minutes. Because of this however, none of the participants finished the open inguinal hernia repair surgery. Therefore, speed and acceleration data were not of any added value. Furthermore, in order to really investigate the effects of feedback on the student's performances, letting the students finish the operation and then compare the total operation time with the different feedback groups would give a greater insight in feedback and surgical performances.

Another limitation was that the felt models used in the operation were not equal. Some of the models missed anatomical parts which led to students being confused. This on its part could lead to different path lengths and idle times compared to the students who did not face this problem.

Future Research

For future research, several improvements can be made. Firstly, it is of great importance to have a clear overview of all important aspects in medical education. What are the greatest difficulties students face during their education and which assessment tools can provide the best possible feedback to enhance learning? The OSATS for example assesses surgical competencies instead of surgical steps only, whereas the OCHRA assesses whether the steps are being performed correctly and if errors are made. Future research should investigate whether for novices, a combination of global and task-specific assessment tools is the most effective approach in teaching technical skills to novice medical students.

Secondly, other research showed that surgical novices who received video-feedback outperformed novices who received verbal feedback in laparoscopic tasks (Farooq et al., 2018). An explanation of this finding is that students are able to observe and examine each procedure that they performed by means of video playback (Boyle et al., 2011). It would be interesting to investigate whether idle times and path lengths would decrease after receiving video-feedback compared to verbal feedback.

Thirdly, since research already showed that novice students show greater idle times than experts, further research could include a greater variation in experience levels within the research.

Lastly, improving the reliability of the motion tracking acquisition should be a priority for future research. Utilising the captured camera images could provide an extra insight into missing data points and the different feedback methods. At last, the repeatability of operations could be improved by more standardised felt models.

Conclusion

This study aimed to investigate whether there is a difference in assessment tools and its feedback, and other related aspects, in teaching novice medical students technical skills. The results of the motion tracking data showed no significant differences between the feedback groups. Furthermore, students with high spatial ability did not outperform students with low spatial ability on the first and second operation. Affect differences also did not show a significant effect between type of feedback. Due to missing data and a small sample size, it is difficult to draw accurate conclusions. Future research should combine the OSATS and OCHRA assessment tools in order to see whether that contributes in learning performances of novice medical students. Furthermore, the effect of video-feedback compared to verbal feedback in novices should be investigated further.

Reference list

- Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: dynamics of ability determinants. *Journal of applied psychology*, 77(5), 598.
- Ahmed, K., Miskovic, D., Darzi, A., Athanasiou, T., & Hanna, G. B. (2011). Observational tools for assessment of procedural skills: a systematic review. *The American Journal of Surgery*, 202(4), 469-480.
- Amid, P. K. (2004). Lichtenstein tension-free hernioplasty: its inception, evolution, and principles. *Hernia*, 8(1), 1-7.
- Angulo, V. M. P., Aguilar, C. Y. D., Álvarez, J. J. G., & Sáenz, G. G. B. (2014). The routine application system Observational Human Clinical Assessment Reliability (OCHRA) gradually decreases errors during laparoscopic cholecystectomy. *Journal of the American College of Surgeons*, 219(4), e161.
- Boyle, E., Al-Akash, M., Gallagher, A. G., Traynor, O., Hill, A. D., & Neary, P. C. (2011). Optimising surgical training: use of feedback to reduce errors during a simulated surgical procedure. *Postgraduate medical journal*, 87(1030), 524-528.
- D'Angelo, A. L. D., Rutherford, D. N., Ray, R. D., Laufer, S., Kwan, C., Cohen, E. R., ... & Pugh, C. M. (2015). Idle time: an underdeveloped performance metric for assessing surgical skill. *The American Journal of Surgery*, 209(4), 645-651.
- Ericsson, K. A. (2015). Acquisition and maintenance of medical expertise: a perspective from the expert-performance approach with deliberate practice. *Academic Medicine*, 90(11), 1471-1486.
- Farooq, A., Reso, A., & Harvey, A. (2018). Instant replay: Evaluation of instant video feedback in surgical novices for a laparoscopic gallbladder dissection. *The American Journal of Surgery*, 215(5), 943-947.
- Fraser, K. L., Ayres, P., & Sweller, J. (2015). Cognitive load theory for the design of medical simulations. *Simulation in Healthcare*, 10(5), 295-307.
- Hatala, R., Cook, D. A., Zendejas, B., Hamstra, S. J., & Brydges, R. (2014). Feedback for simulation-based procedural skills training: a meta-analysis and critical narrative synthesis. *Advances in Health Sciences Education*, 19(2), 251-272.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112.
- Hegarty, M. (2010). Components of spatial intelligence. *Psychology of learning and motivation*, 52, 265-297

- Henn, P., Gallagher, A. G., Nugent, E., Seymour, N. E., Haluck, R. S., Hseino, H., ... & Neary, P. C. (2018). Visual spatial ability for surgical trainees: implications for learning endoscopic, laparoscopic surgery and other image-guided procedures. *Surgical endoscopy*, 32(8), 3634-3639.
- Hiemstra, E., & Jansen, F. W. (2010). Richtlijn OSATS. *Tijdschrift voor medisch onderwijs*, 29(5), 135-141.
- Hoogerheide, V., Renkl, A., Fiorella, L., Paas, F., & van Gog, T. (2018). Enhancing example-based learning: Teaching on video increases arousal and improves problem-solving performance. *Journal of Educational Psychology*, 111(1), 45-56.
- Kuhbandner, C., Lichtenfeld, S., & Pekrun, R. (2011). Always look on the broad side of life: Happiness increases the breadth of sensory memory. *Emotion*, 11(4), 958.
- Langlois, J., Bellemare, C., Toulouse, J., & Wells, G. A. (2017). Spatial abilities and anatomy knowledge assessment: A systematic review. *Anatomical sciences education*, 10(3), 235-241.
- Langlois, J., Bellemare, C., Toulouse, J., & Wells, G. A. (2015). Spatial abilities and technical skills performance in health care: A systematic review. *Medical education*, 49(11), 1065-1085.
- Langlois, J., Wells, G. A., Lecourtois, M., Bergeron, G., Yetisir, E., & Martin, M. (2015). Spatial abilities of medical graduates and choice of residency programs. *Anatomical sciences education*, 8(2), 111-119.
- Lichtenstein, I. L., Shulman, A. G., Amid, P. K., & Montllor, M. M. (1989). The tension-free hernioplasty. *The American Journal of Surgery*, 157(2), 188-193.
- Martin, J. A., Regehr, G., Reznick, R., Macrae, H., Murnaghan, J., Hutchison, C., & Brown, M. (1997). Objective structured assessment of technical skill (OSATS) for surgical residents. *British journal of surgery*, 84(2), 273-278.
- Miskovic, D., Ni, M., Wyles, S. M., Parvaiz, A., & Hanna, G. B. (2012). Observational clinical human reliability analysis (OCHRA) for competency assessment in laparoscopic colorectal surgery at the specialist level. *Surgical endoscopy*, 26(3), 796-803.
- Niitsu, H., Hirabayashi, N., Yoshimitsu, M., Mimura, T., Taomoto, J., Sugiyama, Y., ... & Takiyama, W. (2013). Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. *Surgery today*, 43(3), 271-275.
- Parmar, S., & Delaney, C. P. (2011). The role of proximate feedback in skills training. *the surgeon*, 9, S26-S27.

- Porte, M. C., Xeroulis, G., Reznick, R. K., & Dubrowski, A. (2007). Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills. *The American journal of surgery*, *193*(1), 105-110.
- Puschmann T. A. (1966). History of Medical Education. New York, NY: Harner. (Original work published 1891).
- Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: a single-item scale of pleasure and arousal. *Journal of personality and social psychology*, *57*(3), 493.
- Schmitt, F., Mariani, A., Eyssartier, E., Granry, J.C., & Podevin, G. (2018). Skills improvement after observation or direct practice of a stimulated laparoscopic intervention. *Journal of Gynecology Obstetrics*, *47*, 101-106.
- Spachtholz, P., Kuhbandner, C., & Pekrun, R. (2014). Negative affect improves the quality of memories: Trading capacity for precision in sensory and working memory. *Journal of Experimental Psychology: General*, *143*(4), 1450-1456.
- Spruit, E. N., Band, G. P., Hamming, J. F., & Ridderinkhof, K. R. (2014). Optimal training design for procedural motor skills: a review and application to laparoscopic surgery. *Psychological research*, *78*(6), 878-891.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, *12*(2), 257-285.
- Tang, B., Hanna, G. B., Joice, P., & Cuschieri, A. (2004). Identification and categorization of technical errors by Observational Clinical Human Reliability Assessment (OCHRA) during laparoscopic cholecystectomy. *Archives of surgery*, *139*(11), 1215-1220.
- Telio, S., Ajjawi, R., & Regehr, G. (2015). The “educational alliance” as a framework for reconceptualizing feedback in medical education. *Academic Medicine*, *90*(5), 609-614.
- Van Rutte, P. W. J., Nienhuijs, S. W., Jakimowicz, J. J., & van Montfort, G. (2017). Identification of technical errors and hazard zones in sleeve gastrectomy using OCHRA. *Surgical endoscopy*, *31*(2), 561-566.
- Xeroulis, G. J., Park, J., Moulton, C. A., Reznick, R. K., LeBlanc, V., & Dubrowski, A. (2007). Teaching suturing and knot-tying skills to medical students: a randomized controlled study comparing computer-based video instruction and (concurrent and summary) expert feedback. *Surgery*, *141*(4), 442-449.
- Young, J. Q., Van Merriënboer, J., Durning, S., & Ten Cate, O. (2014). Cognitive load theory: Implications for medical education: AMEE guide no. 86. *Medical teacher*, *36*(5), 371-384.

Appendixes

Appendix A - Step by step description open inguinal hernia repair

Step	Substep (structure)	Action	Specification
1. External oblique aponeurosis exposure	Skin	Incise	Incise the skin for a length of approximately 5 cm in the line between anterior superior iliac spine to the pubic tubercle.
	Subcutaneous tissue	Incise	Incise the subcutaneous tissue until Scarpa's fascia is reached.
			HAZARD Superficial epigastric vessels damage During the incision of the subcutaneous tissue caution should be taken for the superficial epigastric vessels.
	Superficial epigastric vessels	Transect	Transect the superficial epigastric vessels in order to prevent postoperative hemorrhage.
	Scarpa's fascia	Incise	Incise Scarpa's fascia to expose the fat tissue overlying the external oblique aponeurosis.
	Subcutaneous tissue	Incise	Incise the subcutaneous tissue overlying the external oblique aponeurosis to expose it.
2. Inguinal canal exposure	External oblique aponeurosis	Identify	Identify the external oblique aponeurosis as a white plane with oblique fiber lines.
		Incise	Incise the aponeurosis of the external oblique muscle in the direction of the fibers. Extend the incision toward the external inguinal ring medially.
			Hazard - The ilioinguinal nerve on top of the spermatic cord should be identified in order to prevent injuring it.
		Dissect	Dissect the external oblique aponeurosis by developing the plane between the aponeurosis of the external oblique muscle and the internal oblique muscle caudally, while avoiding the ilioinguinal nerve.
3. Spermatic cord mobilization	Spermatic cord	Isolate	Isolate the spermatic cord completely from the floor of the inguinal canal, the transverse fascia and the pubic bone.
		Encircle	Encircle the spermatic cord with a penrose drain.
4. Hernia sac removal	Hernia sac	Identify	Identify the hernia sac.
		Remove	Remove the hernia sac by first rotating the hernia sac around its own axis. Then the hernia sac is transected after clamping and tying the rotated hernia sac.
5. Mesh placement	Inguinal ligament	Expose	Expose the inguinal ligament which is the lower edge of the inguinal canal.
	Mesh	Trim	Trim the mesh to fit the inguinal canal.
		Fixate - medial	Fixate the mesh medially to the distal anterior rectus sheath with 2 cm overlap over the pubic bone.

			HAZARD - Pubic periosteum damage During medial fixation of the mesh to the anterior rectus sheath, care should be taken not to include the pubic periosteum in this suture as this may result in chronic pain.
		Fixate - caudal	Fixate caudally by continuing with a running suture on the lower edge of the inguinal ligament until the level of the internal inguinal ring. Do not continue this suture beyond the lateral border of the internal inguinal ring.
			HAZARD - Femoral vessels and nerve damage During the caudal fixation, avoid damaging the femoral vessels and nerve which run just underneath the inguinal ligament. Damage is avoided by only taking small bites of the lower edge of the inguinal ligament (1-2 mm).
		Trim	Trim the mesh from lateral to medial until the medial border of the internal inguinal ring, creating two tails. The superior tail is approximately 2/3 of the width of the mesh and the inferior tail is approximately 1/3 of the width of the mesh.
		Position	Position the mesh under the spermatic cord, and subsequently maneuver the spermatic cord between the two tails. Then, pass the superior tail over the inferior tail to create a prosthetic internal inguinal ring.
		Fixate	Fixate both tails with a single non-absorbable suture to the inguinal ligament.
		Position	Position the tails of the mesh under the external oblique aponeurosis laterally and cranially.
		Fixate - cranial	Fixate the superior margin of the mesh with one or two interrupted absorbable sutures to the aponeurosis of the internal oblique muscle, while avoiding the iliohypogastric nerve.
			HAZARD - Iliohypogastric nerve damage Avoid damage or entrapment of the iliohypogastric nerve, preferably by identification. Make a transverse suture rather than a longitudinal suture, as this minimizes the risk of nerve entrapment. If necessary a separate cut can be made into the mesh to free the iliohypogastric nerve. If the nerve cannot be freed from the mesh, transect this nerve and bury the iliohypogastric nerve in the oblique internal muscle.
6. Wound closure	External oblique aponeurosis	Close	Close the aponeurosis of the external oblique muscle, creating a new external ring.
	Scarpa's fascia	Close	Close Scarpa's fascia.
	Skin	Close	Close the skin.

Appendix B - OCHRA checklist for open inguinal hernia repair

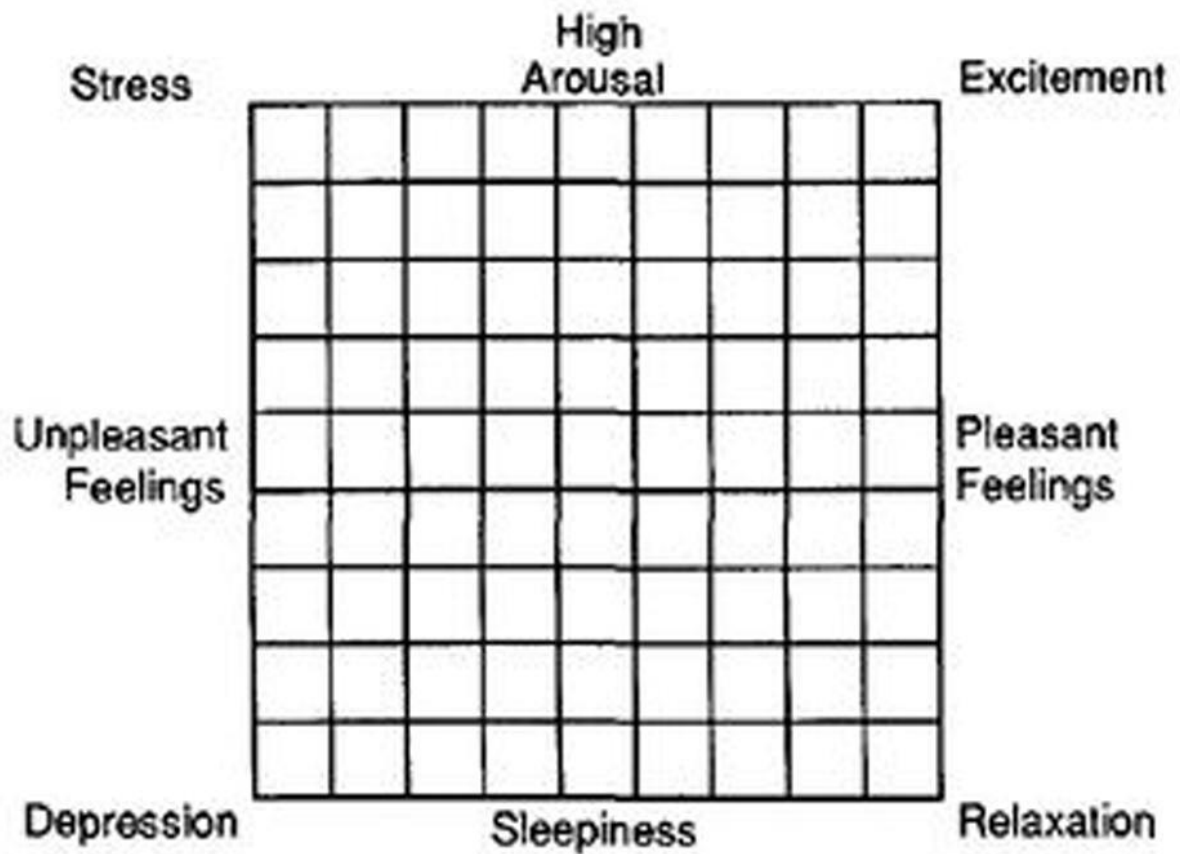
Surgical steps			Performed correctly?	Procedural error	Executorial error	Consequential?
Step	Substep	Action				
<i>1. External oblique aponeurosis exposure</i>	A. Skin	1. Incise				HAZARD – Iliohypogastric nerve damage
	B. Subcutaneous tissue	1. Incise				HAZARD - Superficial epigastric vessels damage
	C. Superficial epigastric vein	1. Transect				
	D. Scarpa's fascia	1. Incise				
	E. Subcutaneous tissue					
<i>2. Inguinal canal exposure</i>	A. External oblique aponeurosis	1. Identify				
		2. Incise				HAZARD - Ilioinguinal nerve damage
		3. Dissect				
<i>3. Spermatic cord mobilization</i>	A. Spermatic cord	1. Isolate				HAZARD - Genital branch of genitofemoral nerve
		2. Encircle				
<i>4. Hernia sac resection</i>	A. Hernia sac	1. Identify				
		2. Remove				
<i>5. Mesh placement</i>	A. Inguinal canal	1. Expose				
	B. Mesh	1. Trim mesh				
		2. Position mesh parallel to inguinal ligament				
		3. Fixate - medial to rectus sheath				HAZARD - Pubic periosteum damage
		4. Fixate - caudal to inguinal ligament				HAZARD - Femoral vessels and nerve damage
		5. Split - superior 2/3 inferior 1/3				
		6. Position - folding tails correctly				
		7. Fixate - lateral				HAZARD - prosthetic inguinal ring too wide or too small
		8. Trim mesh laterally				
9. Position mesh under aponeurosis						

		10. Fixate - cranial to internal oblique muscle				HAZARD - Iliohypogastric nerve damage
6. <i>Wound closure</i>	A. External oblique aponeurosis	1. Close				
	B. Scarpa's fascia	1. Close				
	C. Skin	1. Close				

Appendix C - OSATS checklist

Respect for tissue	1 Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments	2	3 Careful handling of tissue but occasionally caused inadvertent damage	4	5 Consistently handled tissues appropriately with minimal damage
Time and motion	1 Many unnecessary moves	2	3 Efficient time/motion but some unnecessary moves	4	5 Economy of movement and maximum efficiency
Instrument handling	1 Repeatedly makes tentative or awkward moves with instruments	2	3 Competent use of instruments although occasionally appeared stiff or awkward	4	5 Fluid moves with instruments and no awkwardness
Knowledge of instruments	1 Frequently asked for the wrong instrument or used an inappropriate instrument	2	3 Knew the names of most instruments and used appropriate instrument for the task	4	5 Obviously familiar with the instruments required and their names
Use of assistance	1 Consistently placed assistants poorly or failed to use assistants	2	3 Good use of assistants most of the time	4	5 Strategically used assistant to the best advantage at all times
Flow of operation and forward planning	1 Frequently stopped operating or needed to discuss next move	2	3 Demonstrated ability for forward planning with steady progression of operative procedure	4	5 Obviously planned course of operation with effortless flow from one move to the next
Knowledge of specific procedure	1 Deficient knowledge. Needed specific instruction at most operative steps	2	3 Knew all important aspects of the operation	4	5 Demonstrated familiarity with all aspects of the operation

Appendix D – The Affect Grid



[8] How many times have you seen the open inguinal hernia repair surgery in real life?

[9] To prepare yourself for the test day, did you use other sources or materials than the open inguinal hernia website?

Yes or No

In case Yes:

[10] Which other sources or materials did you use? books, articles, other websites, other videos, practice materials, others..

[11] How much time did you spend with these other sources or materials?

0 - 1 hour 1 – 2 hours 2 – 3 hours 3 – 4 hours More than 4
hours

[12] How many hours of sleep did you have last night?

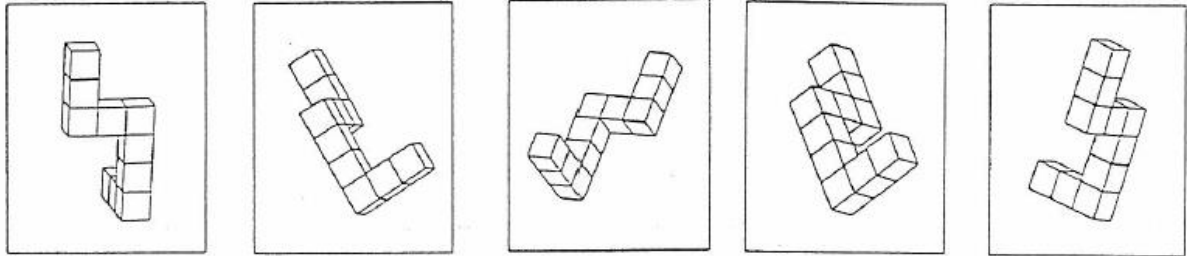
[13] Did you drink alcohol yesterday? If so, how many glasses?

[14] Did you drink any coffee or black tea today? If so how many cups?

[15] Are there any other circumstances (personal, environmental etc.) that could have effected your performance today?

Appendix F – Mental Rotation Test

Look at these five figures



Note that these are all pictures of the same object which is shown from different angles. Try to imagine moving the object, as you look from one drawing to the next.

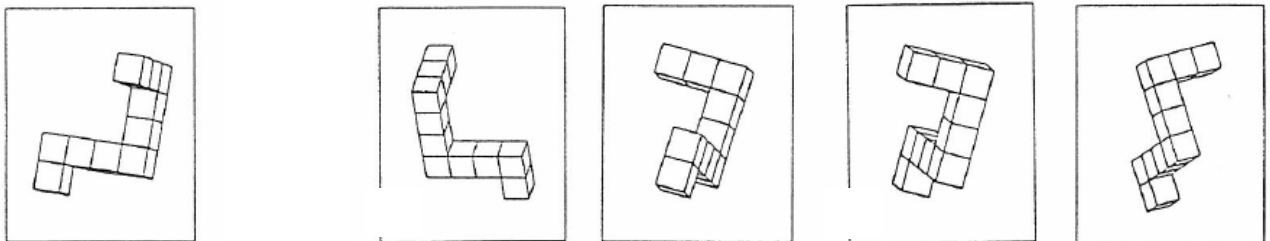
In the following 24 questions you will be asked to identify the two drawings which show the same object.

EXAMPLE:

Now look at this subject:

Two of these four drawings show the same object.

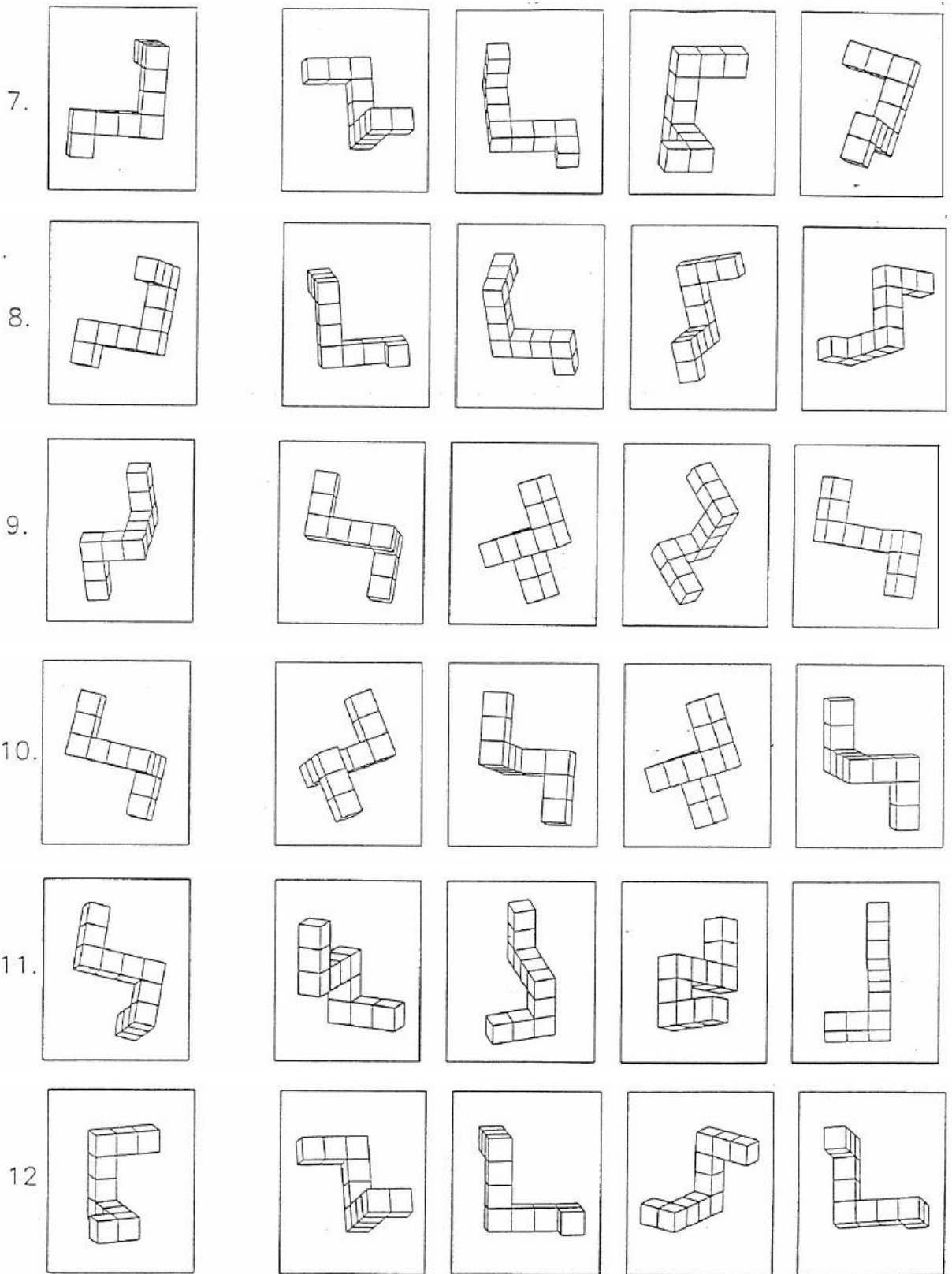
They are marked with X:

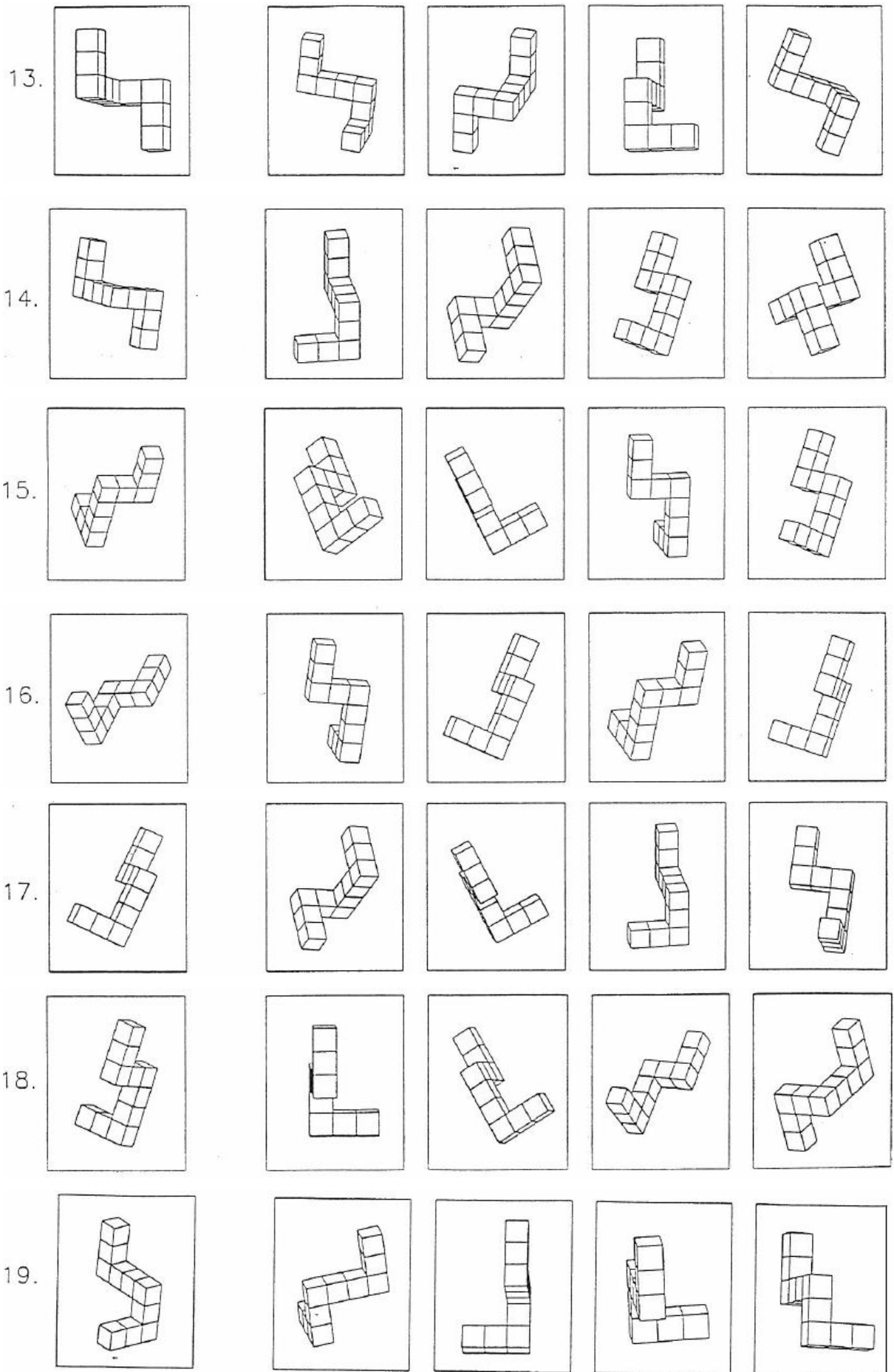


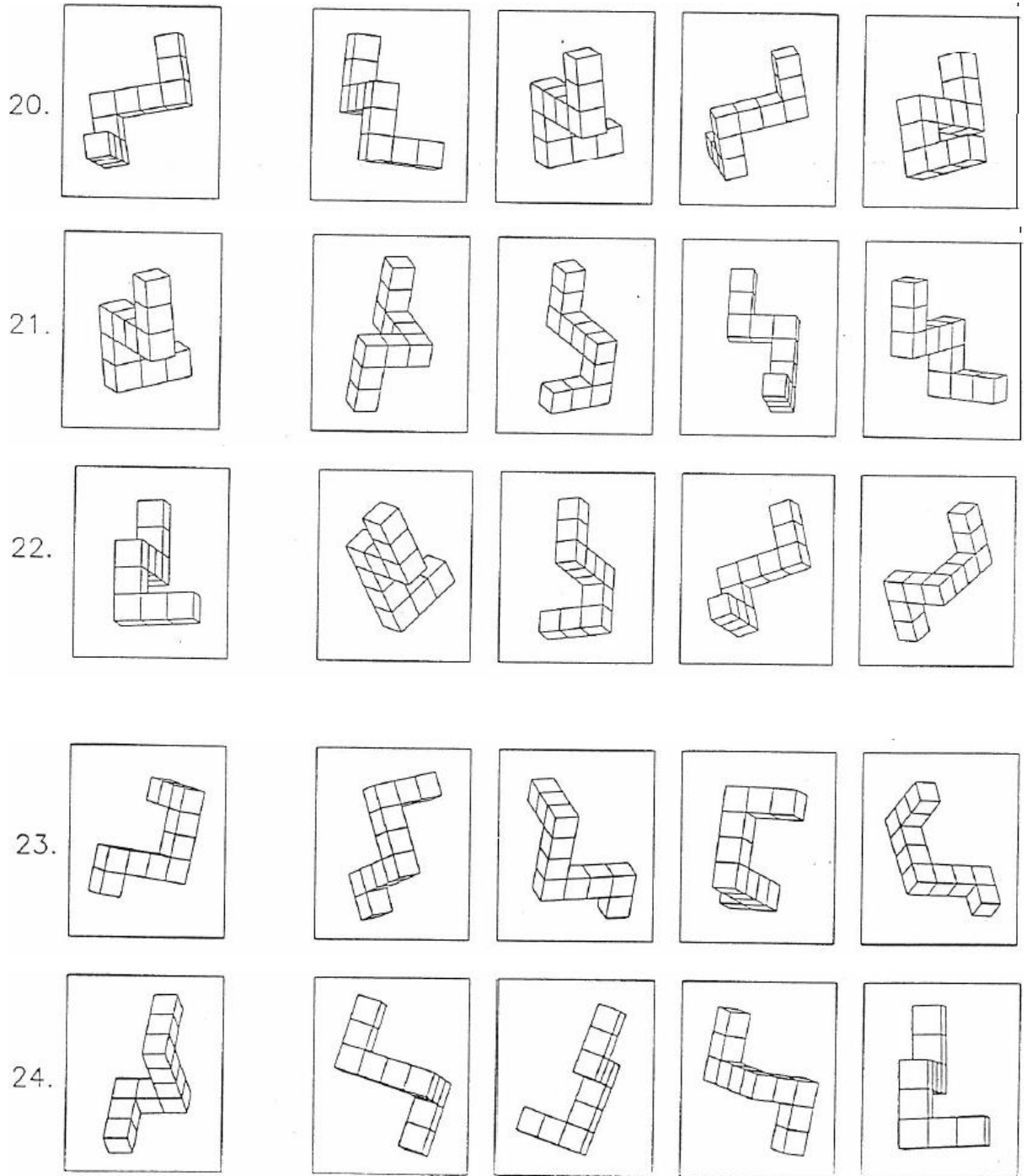
PLEASE DO NOT TURN PAGE UNTIL YOU ARE TOLD TO DO SO

Please identify the two drawings (on the right) which show the same object (on the left).
 Remember, for each problem set there are **only two** figures that match with target figure.

1.					
2.					
3.					
4.					
5.					
6.					







Appendix G – Paper Folding Test

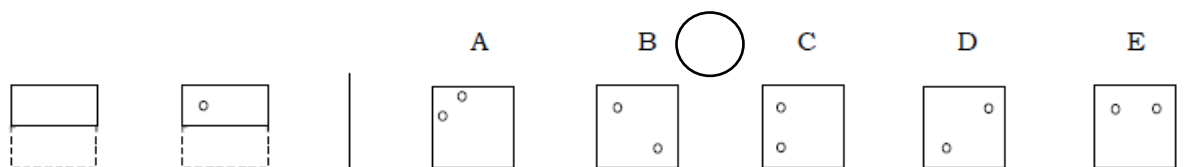
In this test you are to imagine the folding and unfolding of pieces of paper. In each problem in the test there are some figures drawn at the left of a vertical line and there are others drawn at the right of the line.

The figures at the left represent a square piece of paper being folded, and the last of these figures has one or two small circles drawn on it to show where the paper has been punched. Each hole is punched through all the thicknesses of paper at that point.

One of the five figures on the right of the vertical line shows where the holes will be when the paper is completely unfolded.

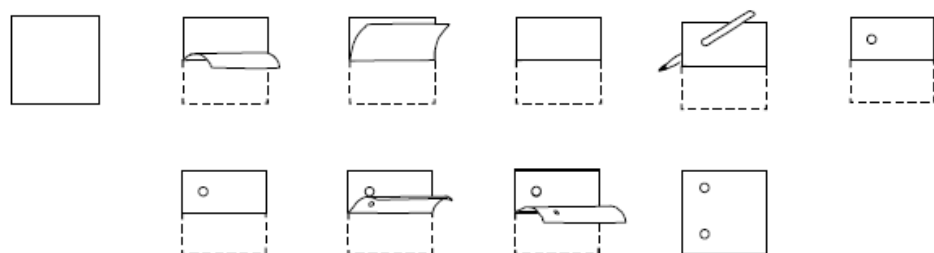
You are to decide which one of these figures is correct and encircle the correct option.

EXAMPLE: (In this problem only one hole was punched in the folded paper).

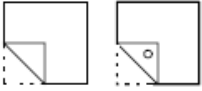





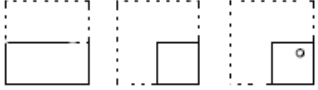





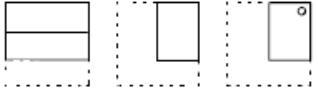











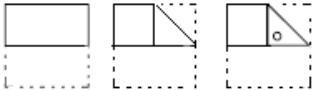





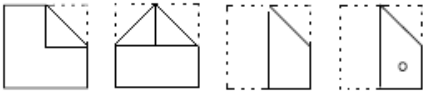





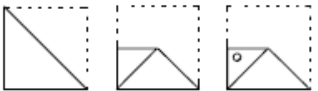





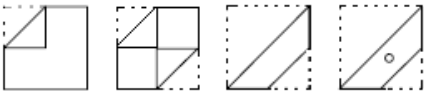





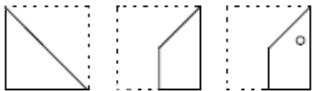







The correct answer to the sample problem above is **C**.

The figures below show how the paper was folded and why **C** is the correct answer.



PLEASE DO NOT TURN PAGE UNTIL YOU ARE TOLD TO DO SO

		A	B	C	D	E
1						
2						
3						
4						
5						
6						
7						
8						
9						
10	