DEVELOPING A VIRTUAL REALITY ADAPTATION OF THE LAPAROSCOPIC SURGICAL TRAINING: A MULTIMODAL STUDY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

EGE YOSUNKAYA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF MULTIMEDIA INFORMATICS

SEPTEMBER 2023

Approval of the thesis:

DEVELOPING A VIRTUAL REALITY ADAPTATION OF THE LAPAROSCOPIC SURGICAL TRAINING: A MULTIMODAL STUDY

Submitted by **EGE YOSUNKAYA** in partial fulfillment of the requirements for the degree of **Master of Science in Modeling and Simulation Department, Middle East Technical University** by,

_____

_____

_____

_____

_____

_____

_____

Date: _____ *01.09.2023*

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Ege Yosunkaya

Signature : _____

ABSTRACT

DEVELOPING A VIRTUAL REALITY ADAPTATION OF THE LAPAROSCOPIC SURGICAL TRAINING: A MULTIMODAL STUDY

Yosunkaya, Ege

MSc., Department of Modeling and Simulation

Supervisor:

Co-Supervisor:

September 2023, 65 pages

Laparoscopic surgical training has established methods with physical training tools that aim to improve the psychomotor skills of surgeons. Virtual reality is a technology that can greatly improve the effectiveness of training while increasing accessibility, especially by removing the physical devices needed. In this study, we developed a virtual reality adaptation of an established method of laparoscopy training, the box trainer. We implemented the simulation in two phases. We used the feedback from the three surgeons and questionnaire responses from 28 participants to develop the second version of the simulation. With a second group of 30 participants, we validated the second version with the same set of questionnaires. The responses to the questionnaires showed that the version with tutorials, haptic feedback, and assisted grabbing physics is more usable and better accepted by the participants. Additionally, we recorded the motion of the graspers in the second version and concluded that the performance can be distinguished by the kinematic characteristics of the motion, similar to its physical counterpart. Lastly, we examined the physiological response with electrocardiogram (ECG) recordings with 15 people from the second group. We found that, during the tasks, heart rate increased, and heart rate variability (HRV) decreased compared to the rest state. We approached the development of the simulation from a multimodal standpoint, focusing on the user's self-reported experience, kinematic behavior of the hands, and physiological response. These results suggest that VR is a promising technology for laparoscopic surgical training. The VR adaptation of the box trainer provides a realistic and immersive experience that can help surgeons develop the psychomotor skills necessary for laparoscopic surgery.

Keywords: Virtual Reality, Laparoscopy, Usability, Motion Analysis, Physiological Response

ÖZ

LAPAROSKOPİK CERRAHİ EĞİTİMİNİN SANAL GERÇEKLİK UYARLAMASININ GELİŞTİRİLMESİ: ÇOK KİPLİ BİR ÇALIŞMA

Yosunkaya, Ege

Yüksek Lisans, Modelleme ve Simülasyon Anabilim Dalı

Tez Yöneticisi:

Eş-Danışman:

Eylül 2023, 65 sayfa

Laparoskopik cerrahi eğitiminde cerrahların psikomotor becerilerini geliştirmeyi amaçlayan fiziksel eğitim araçları kullanılmaktadır. Sanal gerçeklik, özellikle ihtiyaç duyulan fiziksel cihazları ortadan kaldırarak erişilebilirliği artırırken eğitimin etkinliğini büyük ölçüde artırabilecek bir teknoloji olarak önem arz etmektedir. Bu çalışmada, oturmuş bir laparoskopik cerrahi eğitim yöntemi olan eğitim kutularının sanal gerçeklik uyarlaması geliştirilmiştir. Üç cerrahtan alınan geri bildirimler ve 28 katılımcının anket yanıtları ile, geliştirilen ilk versiyonun kullanıcı deneyimi incelenmiş ve yapılan çıkarımlar ikinci versiyonun tasarımında kullanılmıştır. Alınan geri bildirimler ile geliştirilen ikinci versiyon 30 kişilik ikinci bir grupla ve aynı anket setiyle doğrulanmıştır. Anketlere verilen yanıtlar, öğretici başlangıç, dokunsal geri bildirim ve yardımlı tutma fiziği içeren versiyonun daha kullanışlı olduğunu ve katılımcılar tarafından daha iyi kabul edildiğini göstermiştir. Ek olarak, ikinci versiyonda kavrayıcı cihazların hareketi kaydedilmiş ve kullanıcı performansının, fiziksel cihazlar ile gerçekleştirilen eğitime benzer şekilde, hareketin kinematik özelliklerinin performans açısından ayırt edici olduğu belirlenmiştir. İkinci grupta yer alan 15 kişiden elektrokardiyogram (EKG) kayıtları toplanmış ve fizyolojik değişim incelenmiştir. Görevler sırasında dinlenme durumuna göre kalp atış hızının arttığını ve kalp atış hızı değişkenliğinin (HRV) azaldığı görülmüştür. Bu çalışmada simülasyonun geliştirilmesine, kullanıcının kendi deneyimine, ellerin kinematik davranışına ve fizyolojik yanıtına odaklanan çoklu modüllü bir bakış açısıyla yaklaşılmıştır. Sonuçlar, sanal gerçekliğin laparoskopik cerrahi eğitimi için gelecek vaat eden bir teknoloji olduğunu göstermektedir. Fiziksel eğitim kutularının sanal gerçeklik uyarlaması, laparoskopik cerrahi için gerekli psikomotor becerileri geliştirmeye yardımcı olabilecek gerçekçi ve çevreleyici bir deneyim sunduğu görülmüştür.

Anahtar Sözcükler: Sanal Gerçeklik, Laparoskopi, Kullanılabilirlik, Hareket Analizi, Fizyolojik Tepki

To My Grandfather

ACKNOWLEDGMENTS

I would like to express my gratitude to my thesis advisor, [redacted]
for their mentorship throughout my time at METU; their dedication, expertise, and support have been instrumental in shaping the direction and quality of this research. I would also like to thank my thesis co-advisor [redacted] for their collaborative efforts. I deem myself fortunate to have the opportunity to collaborate with them.

Besides my supervisors, I want to express my appreciation to all the participants of this study. Their willingness to sacrifice their time and valuable input enriched the validity of our findings. Thank you for making an indispensable contribution to the completion of this study.

I would like to thank Prof. Dr. Süleyman Serdar Kozat, for his encouragement in pursuing my education and for providing the conducive conditions that allowed me the necessary time and space to bring this work to completion.

I am thankful to my family and friends for their support throughout the journey. Their belief in me and understanding during the moments of intense focus have been the pillars that upheld me.

I would like to express my sincere gratitude to the members of my thesis examining committee for their time and insightful evaluation of my work.

To my dear love, your presence has been my constant source of inspiration, and soft landing when I feel overwhelmed. Your faith in my potential shaped my academic path. From the core of my heart, thank you for your immeasurable contributions to my success.

TABLE OF CONTENTS

ABSTRACT iv
ÖZ v
DEDICATION vi
ACKNOWLEDGMENTS vii
TABLE OF CONTENTS viii
LIST OF TABLES x
LIST OF FIGURES xi
LIST OF ABBREVIATIONS xiii
INTRODUCTION 1
1.1 Motivation 1
1.2 Purpose of the Study 3
1.3 Contributions 3
1.4 Thesis Outline 4
LITERATURE REVIEW 5
2.1 Medical Education Through Virtual Reality 5
2.2 Motor Skill Training 6
2.3 Approaches in Surgical Training 6
2.4 Available Tasks for Training 9
2.5 Utilization of Motion Data In Laparoscopy Training 10
2.6 Use of Electrocardiogram in Laparoscopy Training 11
DESIGN AND DEVELOPMENT OF THE SIMULATION 13
3.1 Implementation and Features of the First Version 14
3.2 Implementation and Features of the Second Version 18
3.2.1. User Interface 18
3.2.2. Physics and Interactions 19
3.2.3. Tutorials 19
3.2.4. Motion Data Collection 20
EVALUATION 23

4.1 Ethics 23
4.2 Evaluating the First Version 23
4.2.1. Participants 23
4.2.2. Experiment Setup 24
4.3 Evaluating the Second Version 25
4.3.1. Participants 25
4.3.2. Experiment Setup 25
4.4 Evaluation Methods 25
4.4.1. Questionnaires 25
4.4.2. Think-Aloud Playtest with Surgeons 26
4.4.3. Incorporating Motion Data 27
4.4.4. Features of physiological signals 28
RESULTS 29
5.1 Feedback for the First Version 29
5.1.1. Think Aloud Sessions with Surgeons 29
5.1.2. Questionnaires 30
5.2 Questionnaires on the Second Version 33
5.3 Kinematic Analysis 36
5.4 ECG Analysis 40
DISCUSSION 45
CONCLUSIONS AND FUTURE WORK 49
REFERENCES 51
APPENDICES 59
APPENDIX A 59
APPENDIX B 61

LIST OF TABLES

Table 1: Most frequent codes and their corresponding segments after analyzing the think-aloud session recordings of surgeons. 30
Table 2: TAM Questionnaire mean, and standard deviation of the questions answered for the first version under the topic “ease of use”. 32
Table 3: Comparison of versions and groups on GAMEFULQUEST for each topic according to a two-tailed t-test. 34
Table 4: Descriptive statistics of mean HR and mean distance between normal RR peaks (MeanNN) of the participants for each task between resting and playing states. 43

LIST OF FIGURES

Figure 1: Modelling of task objects for the Peg Transfer task. 15
Figure 2: Modelling of the task objects for the String Pass task. 15
Figure 3: Axes relative to the grasper and the pivot point in the box trainer setup from the editor screen. 16
Figure 4: Flowchart of grasper and object grab interaction. 16
Figure 5: Peg Transfer task from in-game HMD recording. 17
Figure 6: Peg Transfer task from in-game HMD recording top of the box removed to show inside with the naked eye. 17
Figure 7: Flowchart of deciding if a collision is a successful move or an error. 18
Figure 8: Time elapsed, error count and remaining task indicators added in the second version. 19
Figure 9: Object models of (A) rigid object for moving, (B) torus to move the object through, and (C) drop zone to put the object to complete the tutorial. 20
Figure 10: Sample scenes from (a) Peg Transfer tutorial (b) String Pass Tutorial. 20
Figure 11: Sample scene from the test conducted with Meta Oculus Quest 2. 24
Figure 12: Boxplot of the responses to the GAMEFULQUEST for the first version, grouped by subtopics. 31
Figure 13: Visualization of SUS scores of the first version per user with “acceptable” and “good” threshold markings. 32
Figure 14: Boxplot visualization of the GAMEFULQUEST answers in the “Guided” topic of both versions and groups. 34
Figure 15: Visualization of SUS scores of the second version per user with “acceptable” and “good” threshold markings. 35
Figure 16: Boxplot visualization of the SUS scores of both versions and groups. 35
Figure 17: Boxplot visualization of TAM scores of both versions and groups in Ease of Use and Perceived Usefulness topics. 36
Figure 18: Movement of the grasper tip during the Peg Transfer task for two sample users representing the top half (Left) and bottom half (Right). 37
Figure 19: Movement of the grasper tip during the String Transfer task for two sample users representing the top half (Left) and bottom half (Right). 37
Figure 20: Results of outlier detection for two sample users from each group on the Peg Transfer task. 39
Figure 21: Boxplot visualization of the outlier occurrences during the Peg Transfer and String Pass tasks of each hand. 39
Figure 22: Boxplot visualization of the mean path length and mean jerk values corresponding to the top and bottom half of users for the Peg Transfer and String Pass tasks. 40
Figure 23: Visualization of ECG recording of a sample user during the initial 12 seconds of resting state. 41
Figure 24: Visualization of ECG recording from channel 5 that has background noise of a sample user during the initial 12 seconds of the Peg Transfer task. 42

Figure 25: Visualization of ECG recording from channel 1 of a sample user during the initial 12 seconds of the Peg Transfer task. 42
Figure 26: Word cloud representation of free text responses of overall opinion on the second version. 47
Figure 27: Skip the tutorial button located on the left of the box. 61
Figure 28: Timer and remaining tasks text components located between the box and the screen. 61
Figure 29: First step of the Peg Transfer tutorial explained with the on-screen text located on the right side of the screen. 62
Figure 30: The second step of the Peg Transfer tutorial explained with the on-screen text located on the right side of the screen. 62
Figure 31: The third step of the Peg Transfer tutorial explained with the on-screen text located on the right side of the screen. 62
Figure 32: The last step of the tutorial is explained with the on-screen text located on the right side of the screen. 63
Figure 33: Finished tutorial screen for Peg Transfer tutorial with finish time and the button to continue. 63
Figure 34: First step of the String Pass tutorial explained with the on-screen text located on the right side of the screen. 64
Figure 35: Second step of the String Pass tutorial explained with the on-screen text located on the right side of the screen. 64
Figure 36: Finished tutorial screen for String Pass tutorial with finish time and the button to continue. 65

LIST OF ABBREVIATIONS

VR	Virtual Reality
XR	Extended Reality
AR	Augmented Reality
APK	Android Package Kit
HR	Heart Rate
HRV	Heart Rate Variability
BPM	Beats Per Minute
UI	User Interface
HMD	Head Mounted Display
DOSPERT	Domain Specific Risk Taking
EEG	Electro-Encephalography
ECG	Electro-Cardiogram
FLS	Fundamentals of Laparoscopic Surgery
MISTELS	McGill Inanimate System of Training and Evaluation of Laparoscopic Skills
OSATS	Objective Structured Assessment of Technical Skill
RCT	Randomized Controlled Trial
OVR	Oculus Virtual Reality
HCI	Human-Computer Interaction
NASA-TLX	National Aeronautics and Space Administration Task Load Index
LMIC	Low-Middle Income Countries
GAMEFULQUEST	Gameful Experience Questionnaire
PQ	Presence Questionnaire
TAM	Technology Acceptance Model
SUS	System Usability Scale
CSV	Comma Separated Values

CHAPTER 1

INTRODUCTION

1.1 Motivation

Virtual Reality (VR) has become a game-changing technology with a wide range of uses in a number of different fields, effectively overcoming the limitations of physical tools and the restrictions of physical presence. This innovative technology has paved the way for new approaches that challenge and reshape traditional methods, including the field of training. The capacity to generate immersive simulation settings that offer a genuine experience, obviating the necessity for physical presence, is a very advantageous attribute for training purposes. The capacity to generate immersive simulation settings that offer a genuine experience, obviating the necessity for physical presence, is a very advantageous attribute for training purposes. Furthermore, given that the experience is inherently computer-assisted, it is possible to leverage prior advancements in human-computer interaction (HCI) to enhance the educational effects.

Many surgeons have traditionally learned laparoscopic techniques through an informal "see one, do one" approach [1], which, while offering some amount of hands-on experience, frequently falls short of providing full and structured training. The utilization of this informal approach, while occasionally necessitated, has the potential to undermine the safety of patients and introduce variability in the outcomes achieved [2], [3]. Simulations are used at various phases of laparoscopic surgical training. Simulations in the medical field are a method, rather than a standalone technology, that seeks to substitute or enhance genuine experiences with guided experiences, typically immersive in nature, and these simulations attempt to represent or recreate significant components of the real world in an entirely interactive manner [4]. These methodologies exhibit variations in terms of their form and the degree to which they serve as substitutes for real-world analogues. The various approaches encompass multiple methods, including physical box trainers, computer-assisted simulations with physical graspers, simulation software featuring a screen interface, and head-mounted displays equipped with physical tools [5]. It should be noted, however, that all the aforementioned approaches need the usage of actual surgical instruments or expensive hardware to serve as a substitute for these instruments. This makes creating novel scenarios challenging and exacerbates the problem of limited accessibility due to the high price connected with obtaining the necessary technology and tools. The ability to recognize grasper motion and visually

perceive a three-dimensional world using a two-dimensional screen necessitates the development of hand-eye coordination, particularly in domains requiring extreme precision.

The advancement of VR simulations with readily available hardware represents a crucial milestone in tackling these concerns. The cheap and extensive accessibility of these simulations can greatly mitigate the difficulties encountered by lower-middle-income nations, where the expense of adequate laparoscopic training can be excessively burdensome [6], [7]. By providing an alternate solution that upholds both quality and safety standards, the implementation of accessible virtual reality (VR) simulations has the potential to overcome the accessibility hurdles that have hindered effective training and enhance the level of care in regions with limited resources. The democratization of surgical education guarantees that surgeons worldwide are provided with systematic, evidence-based training that improves their abilities and protects the welfare of patients, irrespective of their financial situation.

In order to tackle the obstacles presented by the dependence on physical equipment and the accompanying expenses in laparoscopic surgical training, the creation of virtual reality simulations utilizing readily available hardware emerges as a potentially viable resolution. Virtual reality simulations possess the capacity to significantly transform laparoscopic training through the provision of a platform that is both more cost-effective and accessible to a larger population. Trainees get the opportunity to engage in realistic surgical scenarios by utilizing virtual reality headgear and motion-tracking controllers, thereby eliminating the requirement for costly physical instruments. This innovation has the dual effect of reducing the obstacles faced by prospective surgeons and facilitating the development of a wider array of situations that can augment the educational process. The replication of complex hand-eye coordination tasks within a three-dimensional virtual environment facilitates the cultivation of fundamental abilities necessary for executing precise surgical techniques. The utilization of virtual reality simulations in surgical training facilitates a highly engaging and interactive experience, enabling trainees to repetitively engage in practice sessions and enhance their skills. This immersive aspect of simulation-based training is particularly advantageous in the context of laparoscopic operations, as it promotes the development of muscle memory and spatial comprehension, both of which are vital for achieving effective outcomes in such procedures [8]. With the continuous advancement and increasing affordability of technology, the incorporation of easily available virtual reality simulations into surgical training curricula has the potential to democratize knowledge and skills, enhance patient outcomes, and fundamentally transform the field of surgical education.

In conclusion, the use of virtual reality simulations in conjunction with readily available gear signifies a significant advancement in the enhancement of laparoscopic surgical training. Through the utilization of virtual reality technology, educators could construct training experiences that are immersive, flexible, and economically efficient. These experiences not only serve to bridge the knowledge

gap between inexperienced individuals and experts but also extend the accessibility of surgical education to a broader audience. With the continuous advancement and increasing affordability of technology, there is a growing potential for the incorporation of virtual reality simulations in surgical training. This integration holds the promise of improving patient safety and enhancing surgical proficiency.

1.2 Purpose of the Study

The purpose of this study is to cultivate a comprehensive comprehension of the adaptation of the virtual reality of the box trainer, a widely employed technique, especially in earlier stages of laparoscopic surgical training. The objectives of this study are as follows:

- Developing a simulation to conduct early stages of training in laparoscopic surgery, utilizing affordable and easily accessible hardware.
- Conduct a usability assessment of the application and delineate the key components that enhance user experience, thereby facilitating future research endeavors.
- Demonstrate the potential applications of motion data generated within the VR iteration of psychomotor skill training, ultimately leading to the formulation of informed conclusions.
- Examine the physiological response observed in VR environments when individuals engage in mentally demanding tasks, and further investigate the potential relationship between such responses and individuals' skill levels.

1.3 Contributions

The aim of this study is to improve the understanding of psychomotor training by employing VR technology. Our research primarily centers on adapting an existing method training of for laparoscopy, a challenging activity that has been widely studied through the utilization of physical equipment and computer-assisted simulations on readily available hardware platforms. However, the growing ubiquity of readily available VR devices holds the promise of improving the accessibility of training programs. Consequently, this facilitates the utilization of our preexisting study discoveries and deductions by content generators to construct more engaging and efficient instructional encounters for people participating in these programs. The final conclusions are derived from an analysis of quantitative and qualitative data obtained from participants with varying levels of competence, as well as subject matter experts in the field of usability and acceptance of the application. Furthermore, we employ motion and physiological data to demonstrate similarities as well as differences between our adaptation and the existing simulations.

Additionally, we provide methods for integrating this data to evaluate skill levels and pave the way for future research aimed at enhancing the learning experience.

1.4 Thesis Outline

The rest of the thesis is outlined as follows:

- Chapter 2 presents the previous studies on how computer-assisted simulations and physical tools are used in laparoscopic surgical training and the emerging standards for the validation of surgical skills in laparoscopy. The approaches in using kinematic and physiological data in other methods of training.

- Chapter 3 gives a detailed explanation of the design and development process of the first and second iterations of the simulation.

- Chapter 4 explains the evaluation methods used for usability analysis, kinematics, and physiological response.

- Chapter 5 provides an exposition of the outcomes pertaining to assessment methodologies.

- Chapter 6 presents a comprehensive analysis of the primary findings, exploring their correlation with the design decisions made and drawing comparisons with prior research in the field.

- Chapter 7 summarizes the study and discusses the potential directions for research and its limitations.

CHAPTER 2

LITERATURE REVIEW

2.1 Medical Education Through Virtual Reality

The increasing accessibility of VR has prompted substantial research into improving training methods utilizing this technology. The integration of technology in the education of health professionals is a possible avenue for enhancing or substituting traditional methods. According to Kyaw et al. [9], their comprehensive review of research spanning from 1990 to 2017 demonstrates substantial evidence supporting the notion that the utilization of VR enhances knowledge and skill acquisition among healthcare professionals when compared to conventional or alternative digital educational methods, regardless of whether they are delivered online or offline. While acknowledging the absence of a uniform approach for reporting data and the potential limitations in generalizing findings across various socioeconomic and geographic contexts due to the diversity of training topics, this study represents a significant initiative to examine the comparative effectiveness of VR and other educational modalities. In his study on current trends of change in medical education, Pottle [10], evaluates and compares several forms of VR-based teaching, analyzing their present and potential applications. The educational methods under consideration include screen-based instruction, 360-video experiences, and interactive VR. The screen-based instruction methods are rather confusing but have been referred to as virtual reality frequently in the medical domain. It has been argued that the utilization of VR simulation technology has resulted in, and will continue to yield, enhancements in the efficacy of training. This is akin to its shown impact in diverse industries, as a result of its advantageous attributes such as cost-effectiveness, replicability, and standardized procedures. The inclusion of screen-based instruction is rather confusing but is a common practice in the medical domain. In their research on VR training within the medical field. Mantovani et al. define VR as encompassing computer-generated representations of both real and imagined environments. [11]. However, this definition conflicts with the definition of others, since the aspects included in VR are listed as virtual world, immersion, sensory feedback, and interaction in other domains [12]. In the field of medicine, virtual reality training may not necessarily encompass all of these characteristics.

2.2 Motor Skill Training

Motor skill training is a specific area of inquiry within the broader domain of education. Luft and Buitrago [13], in their study about motor skill training and its inner workings, claim that the acquisition of motor skills occurs gradually across multiple training sessions, and once acquired, these skills tend to be retained over an extended period. Furthermore, with regard to the meaning of motor skill learning, it is postulated that motor skills encompass the mastery of complex movement sequences achieved by repetitive practice. Motor skill education is identified under the latter of the three domains of learning; cognitive, affective, and psychomotor formulated by the important works of three book series by Benjamin S. Bloom, Taxonomy of Educational Objectives [14]–[16]. According to the classification, the psychomotor domain, which was subsequently investigated and refined [17]–[19], encompasses the domains of learning related to physical movement, coordination, and the utilization of motor abilities. These skills are assessed by measures of speed, precision, and the application of specific processes or techniques during execution. As stated in the literature, motor skills are developed with the repetition of tasks, which can be provided with immersive environments that are specifically designed for the task. In Krathwohl's revision of the initial taxonomy [19], after remembering and understanding the topic, the next step is the ability to "carry out or use a procedure in a given situation". The primary emphasis of this work is on the incorporation of laparoscopic training, specifically on strengthening the implementation and execution components of the learning process.

2.3 Approaches in Surgical Training

In the domain of surgical education, the conventional method is commonly referred to as "see one, do one" or "see one, do one, teach one," which emphasizes the apprenticeship-based aspect of this discipline and comes from the method formalized by William Stewart Halsted, first Chief of Surgery at John Hopkins Hospital of USA in 1890 [20]. The concept of "see one, do one" refers to the process of initially witnessing a technique being performed, and subsequently acquiring the ability to independently execute the same activity. Additionally, the teach-one part of the concept entails the underlying assumption that an individual who is capable of instructing one trainee is also competent in instructing another learner [21]. Although the "see one, do one" approach is not only used in surgical training, our review of the surgical training literature discusses surgical training approaches especially ones concerned with laparoscopy. There are many takes on the conventional method in the literature, some studies argue that the method is still applicable and with points open to improvement [21]. Hermann-Werrner et al. [22] address shortcomings in assessing the training effectiveness of short-term studies by conducting 6 months randomized controlled trial (RCT) with 94 first-year medical students and comparing "see one, do one" with a method where the teacher shows the procedure two times and conduct a verbal examination on the material before the trainee performs the task. The authors concluded that modification of teaching methods could improve

learning outcomes. Conversely, numerous studies propose the incorporation of laboratory training using some form of computer-assisted simulation prior to engaging with actual patients, as a result of the notable prevalence of avoidable errors, as emphasized in the report published by the Institute of Medicine in 1999 [23]. The scope of this work is not in a position to speculate on why an error occurs in a medical procedure, however, A notable proportion of doctors, namely 42%, indicate their perception of insufficient training to independently perform a practical procedure with proper safety measures [24]. This statistic suggests that a significant proportion of surgeons express dissatisfaction with the level of training they have received prior to independently performing surgical procedures It is imperative to examine this technique within the context of the aforementioned cognitive process of learning, wherein learning occurs in a multidimensional manner and the capacity to execute and implement the procedure takes precedence over the ability to analyze, evaluate, and create [19]. Therefore, it is suggested that a thorough understanding of a subject can only be achieved by meticulous training in the application of the relevant procedures.

At the core, Halsted's approach suggests not only a way of teaching but also how mentorship should be incorporated into surgical training. The "learning by doing" approach, still holds its validity and many studies focus on the increasing quality and fidelity of "doing" with computer-assisted simulations. Given the multitude of studies conducted on various medical procedures, our research has mostly concentrated on influential surveys and studies within the laparoscopic training field that have successfully captured suggested methodologies and assessed their efficacy.

Simulation is used rather loosely in surgical training literature since it encompasses everything from animal and cadaver labs to physical training boxes and computer-assisted virtual reality. On top of that, virtual reality (VR) does not imply an immersive environment, especially in earlier medical investigations, but rather computer software presents the simulation to a screen and receives input through robotic graspers similar to those used in laparoscopy. Simulation-based training facilitates the establishment of a secure and regulated setting, enabling trainees to engage in the practice of surgical techniques, refine their psychomotor skills, and acquire valuable experience before performing procedures on actual patients [25]. With the rapid advancement of VR technology, immersive and simulated settings in training are now becoming available. The current landscape in laparoscopy training has various levels of virtualization and paradigms in terms of realism. In their review study, Hammoud et al. examine the existing simulations in two main categories in terms of realism, and high and low fidelity [26]. The high-fidelity simulators incorporate laparoscopic instruments and a screen, whereas low-fidelity instruments often consist of box trainers that contain a box with slits, a camera, and a screen. The authors also note that low-fidelity simulations are hindered by their reliance on direct observation. In contrast, computer-assisted simulations have the capacity to employ objective assessment methods. The primary purpose of utilizing low-fidelity trainers, specifically box trainers, is mainly to train inexperienced surgeons on motor skill

acquisition. These trainers allow trainees to become acquainted with the manipulations of graspers and develop their video-hand-eye coordination skills [27].

Numerous endeavors have been made to assess and contrast the efficacy of various simulation-based training methodologies across various types of laparoscopic methods. The existing literature on comparing the efficacy of various training methods is characterized by a wide range of techniques employed and the specific body areas targeted, hence posing challenges in conducting direct comparisons. Regardless, the aim is to present an overview that highlights the benefits and drawbacks of using various techniques in literature without delving extensively into medical intricacies. Box trainers have been employed in surgical training since the early 2000s due to their ease of production and lack of reliance on computers. It is shown by RCTs that, training with box trainers results in better skill acquisition both in drills and in the operating room [28], [29]. The one obvious topic of interest is, how the box trainer and the computer-assisted simulations are compared in the acquisition of psychomotor skills. The meta-analysis by Guades et al. [30] consisting of 14 publications comparing box trainers concluded that the box trainer and virtual reality simulation performed practically identically except for the time to complete the initial peg transfer task. Moreover, additional studies show this skill acquisition through computer-assisted simulations also translates in the operating room [31], [32]. However, both methods of training utilize physical experience on graspers, whereas a VR simulation with a generic device would lack the feeling of the force as a main differentiator of the simulation used in our study and the other widely used commercial computer-assisted counterparts, MIST-VR®, ProMIS™, LapSim®[1] and LapMentor™ [33]–[36] In their recent research conducted in 2019, Fritz et al. [37] examine the utilization of virtual reality (VR) in the context of surgical training. The authors assert that simulators equipped with pre-established training programs prove advantageous in terms of minimizing individual expenses and offering trainees intuitive advice. However, the study indicates that the availability of these technologies is limited, and while demonstrating their effectiveness, it is necessary to do further investigations on how they might be used in the operating room.

An additional aspect of the simulations involves the evaluation of trainees' proficiency. Chaudry et al. [38], in their study using MIST-VR to asses motos skills between experienced surgeons and inexperienced individuals found out it is possible to predict the level of laparoscopic skills performance and experienced surgeons perform better than the inexperienced group in all aspects of the training. Gallagher and Savata [39], conducted a study examining the relationship between skill level and simulator competency. Specifically, they explored how the experience of surgeons, categorized into three distinct groups based on their expertise, transfers into their proficiency in VR. The researchers reached the conclusion that, based on

[1] Surgical Science Ltd., Gothenburg, Sweden, https://surgicalscience.com/simulators/lapsim/, Last accessed: 2023-07-01.

their utilization of MIST-VR, there were significant discrepancies in the participants' performance with regard to speed, frequency of errors, and efficiency of movement [39]. Furthermore, Mason et al. [40] conducted a survey to evaluate the proficiency of laparoscopic abilities using motion analysis. Their findings indicated that duration, path length, and number of hand movements are reliable indicators for assessing the skill level of trainees. Hence, in the assessment of motor skill proficiency, it is crucial to take into account factors such as task completion time, error frequency, and efficiency of movement.

The presence of diverse surgical training methodologies, tools, and assessment methods poses additional challenges in establishing a standardized curriculum. Although the effort to standardize the training in laparoscopic surgery, many surgeons trained in an informal fashion, ultimately resulting in differences in skill levels. High-end training computer-assisted training hardware is expensive, especially in lower-middle-income countries (LMICs). This accessibility barrier is also reported as an important factor restraining wider adaptation of the procedure by surgeons in the United States [41]. Moreover, lack of case volume and insufficient exposure to the devices result in poor video-eye-hand coordination, and unexpected scenarios are also reported as limiting factors [6]. Another review by Wilkinson et al. underlines the lack of equipment as one of the biggest issues and concludes low-cost solutions with better implementation of laparoscopic training can improve access to surgical care in LMICs [7]. Their conclusion also supports our claim that training applications on affordable VR devices are an important direction for research. The study conducted by Kyaw et al. [9] acknowledges the absence of data on the usefulness of VR for health professionals' training in LMICs, thereby highlighting a gap in the existing literature.

The absence of standardized curriculum and training tools in the sector leads to a wide range of methodologies, as previously emphasized. Different approaches have been researched in the forming of a curriculum for simulators [42]–[45]. Aggarval et al. suggest competency-based curricula and focus on basic and procedural tasks [46]. Furthermore, their additional research tries to define an objectively measured training methodology that results in a more desirable learning curve [47]. However, different tools that have been used in these studies and the plenitude of variations, such as surgical methods, trainee skill level, and using different evaluation methods make it harder to draw conclusions on compiling a uniform curriculum.

2.4 Available Tasks for Training

In the certification of laparoscopic surgery skills, the most known program Fundamentals of Laparoscopic Surgery (FLS) uses a manual skills component based on the McGill Inanimate System of Training and Evaluation of Laparoscopic Skills (MISTELS) program [48]–[50]. Thus, our selection of tasks was based on this rigorously evaluated and verified program [50]. There are five tasks used in the FLS manual skills exam. The Peg Transfer task is one of the standard tasks in the FLS

examination. The task was also studied in other VR-based simulations, and shown to be valid and parallel to its physical counterparts [51].

The other task designed for the experiment was the string pass, also called rope pass in some studies. This task was not in the standard examination but used in different studies in the field. The challenging part of this task is, apart from the hand-eye coordination the hardship of understanding the distance between the rope and circle. Another reason for using both of these tasks was to avoid the usage of soft tissue physics that was involved in some other tasks. The realistic soft tissue physics in VR was out of the scope of this study and would create another aspect to delve into in terms of user experience and task validity. These tasks are selected because they emphasize video-hand-eye coordination skill level [52], [53].

2.5 Utilization of Motion Data In Laparoscopy Training

Objectively assessing the surgical skill is important to validate the proficiency of the trainee. With that need, before motion capture was widely available, the most used method was the Objective Structured Assessment of Technical Skill (OSATS) technique [54]. This technique uses a 5-point scale on seven topics that are filled out by the teacher. The OSATS method is shown to be correlated with motion analysis of hands using trackers in general surgery [55]. Two parts of the scale, time and motion, and instrument handling are especially connected to the motion. The five points of these topics correspond to “Economy of movement and maximum efficiency” and “Fluid moves with instruments and no awkwardness” respectively. However, these observations are not defined formally and are not trivial to translate into automated assessment.

The data-driven approach has also been studied both in physical and computer-assisted training methods. Ebina et al. use a box trainer with physical instruments and capture the motion data of surgical instruments [56]. As seen in their study, they set up six infrared cameras and used tracking software, which is costly and hard to use as opposed to computer-assisted simulations of our VR adaptation. Our proposed method can use motion tracking with no extra equipment and can produce more robust results since all participants use the same virtual environment. According to their results, for each task, there were distinguishing characteristics of higher-skilled participants such as shorter path length or faster overall velocity.

Other studies have shown similar results in the laparoscopy field, where expert and novice groups showed significant differences in terms of economy of movement (path length) [57]. Moreover, Uemura et al. [58] incorporated interesting methods of motion analysis corresponding flexibility and stability of the trainee using time series analysis methods from different domains. Their method also suggests hand motion can be a distinguishing factor between surgeons who have varying laparoscopy experience.

2.6 Use of Electrocardiogram in Laparoscopy Training

Psychomotor skill training, especially in the field of laparoscopy, is a task with high mental demand. Moreover, the increased workload is shown to be related to inferior task performance and a higher likelihood of errors [59]. The aforementioned studies use the National Aeronautics and Space Administration Task Load Index (NASA-TLX) scales [60] that participants use for self-assessment after the procedures. However, the advancements in wearable technologies made physiological data collection more accessible. This physiological data can be utilized to assess the mental workload of the trainee during the training session. ECG signals are studied deeply in the medical field for diagnosing cardiovascular diseases and abnormalities; however, it is also shown that they are useful for understating stress levels and mental workload [61]. Many features can be extracted from the ECG signals for different purposes. Some of those features are unique to ECG signals, utilizing the QRS complex, waveform that reflects the main electrical activity. Moreover, there are other features such as statistical features, spectral domain features, and time-frequency domain features used in various fields.

Main time features of ECG, heart rate (HR), and heart rate variability (HRV) are shown to be indirect markers of stress in surgeons [62]. Moreover, HRV is shown to be significantly different between experts and non-surgeons on box trainers using similar tasks to our selection [63]. Zakeri et al. explain the underlying physiological mechanisms for this phenomenon by the relation between cognitive load and respiratory sinus arrhythmia [64]. Moreover, it suggests that HRV might detect autonomic changes even when the HR is constant.

CHAPTER 3

DESIGN AND DEVELOPMENT OF THE SIMULATION

In preparation for the design and development phase, an examination was carried out at the Ankara University Department of Biomedical Engineering to get insight into the technique, objectives, and tactile experience of physical instruments. This investigation involved the utilization of the box trainer. The simulation underwent a two-part development process. In the initial phase, the first version was developed and refined based on input received from a group of 28 individuals. This feedback was collected through questionnaires and free-text questions, which sought suggestions for improvement. Additionally, the first version was subjected to testing sessions involving three surgeons.

The Meta Quest 2 (formerly known as Oculus) is used for the purposes of development and experimentation as a virtual reality device. The hardware was selected based on its accessibility and versatile design. It is considered one of the most cost-effective options available in the market, offering an immersive environment with support for hand-tracking controllers. The Unity 3D Engine[2] (2020.3.23f1) has been selected as the game engine for the implementation of the simulation. The integration of Oculus and Unity 3D entails the utilization of many libraries accessible through the online package manager, which has been produced by Unity Technologies. The "XR Interaction Toolkit" facilitates the seamless integration of VR and augmented reality (AR) applications with the Unity input system. The provided interaction encompasses several aspects. Firstly, there is a canvas where the user interface components are rendered and can be engaged with. Additionally, haptic feedback is sent through the controllers, enhancing the user's sensory experience. Moreover, the default select, grab, and interact capabilities are available for manipulating objects within the scene. In addition, the package serves as a versatile, overarching component that provides compatibility for Meta Quest 1 and Quest 2, Windows Mixed Reality, and a generic interface known as OpenXR for additional devices. Our application makes use of version 2.2.0 of the package. The XR Origin is the component responsible for replacing the default main camera within a scene. It offers built-in functionality to facilitate a room-scale VR experience. The inclusion of the ability for the user to freely move and navigate on foot during the task was important in our experiment design.

---

[2] Unity Technologies. (n.d.). Unity - Game Development Platform. Retrieved from https://unity.com/.

Secondly, the “XR Plugin Management” (4.3.0) package is utilized to achieve a seamless development experience. The package provides the VR environment to be loaded to Quest 2 when the development version is started through the Unity 3D Editor. This provides a faster iteration cycle where the effects of changes in the project code or models can be seen without explicitly building the application into a .apk file and loaded into the device. Lastly, the “Oculus Integration SDK” which is developed by Meta, contains the necessary libraries to utilize Quest device capabilities. The main component used in the application was the OVRCameraRig which substitutes the Main Camera component of the Unity 3D engine with this component that enables the development of the immersive experience through the head-mounted display. The plugin includes the controller models and the interface for the buttons on the VR controllers to be mapped to the game engine. Moreover, the OVRManager component allows the configuration of device-related properties. The tracking origin type property governs the camera’s initial position upon scene loading. The program utilizes eye-level tracking technology in order to offer a comparable user experience for individuals of varying heights. An alternative option floor level, involves drawing the objects always according to their heights in terms of their defined height in the editor; however, it is important to note that those with greater height may experience increased difficulty in reaching the graspers due to the stretched hand position required.

Prior to conducting the tests, the Unity project was built into an Android Package Kit (APK) format and subsequently installed onto the Quest 2 device, as it operates on the Android operating system. The Android 10 version was utilized for the development process, and the app was compiled accordingly.

3.1 Implementation and Features of the First Version

The first version is planned to be unornamented but functional to receive feedback and draw conclusions on how to improve the experience before settling for design decisions. The main menu and two tasks (1) Peg Transfer and (2) String Pass are implemented in this version. The user is presented with a main menu when the application opens, where both tasks are selectable, however when the task starts, no in-game explanation or guidance is provided. Furthermore, elapsed time and error count are also not present as a UI element. For ease of development, the operation table, box, the room environment, and the graspers are reused and prevent scene reloading only the task objects are changed between the tasks. The task objects are modeled with inspiration from their real-life versions. In Figure 1, the green and red cylinders are the immovable objects, and blue toruses are the pegs that should be transferred inside the red cylinders. Moreover, the green squares under the red cylinders are the colliders that detect the successful transfer. For the String Pass task, the modeling of the toruses that the rope will pass through was relatively easier. However, the rope mechanic should be not computation-heavy and still be realistic to mimic the real movement patterns. In Figure 2, the rope with the toruses can be seen from a screen capture from the engine editor. The rope is dynamically created when the scene loads. The string consists of small cylinders, attached to each other with configurable joints. The script creates each one by one and then crates the joints between them. Every small cylinder is a rigid body and has its own collision box.

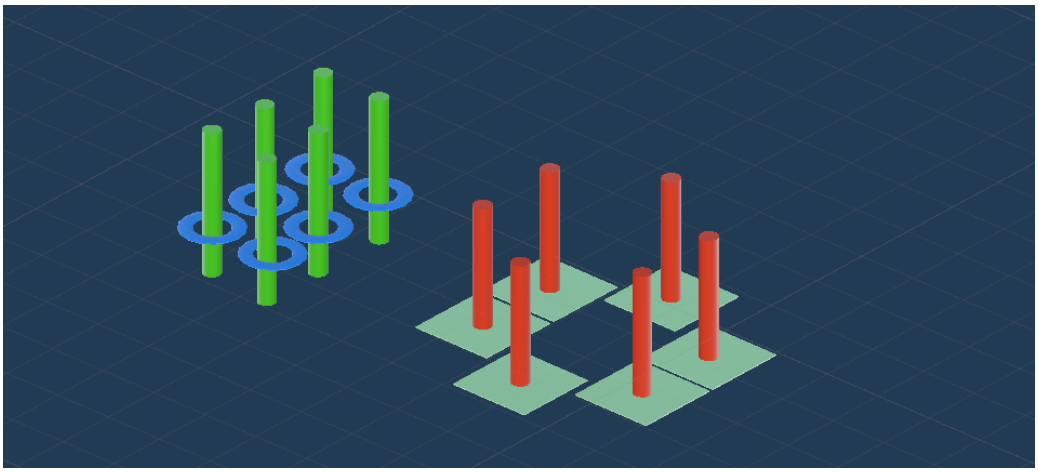


Figure 1: Modelling of task objects for the Peg Transfer task.

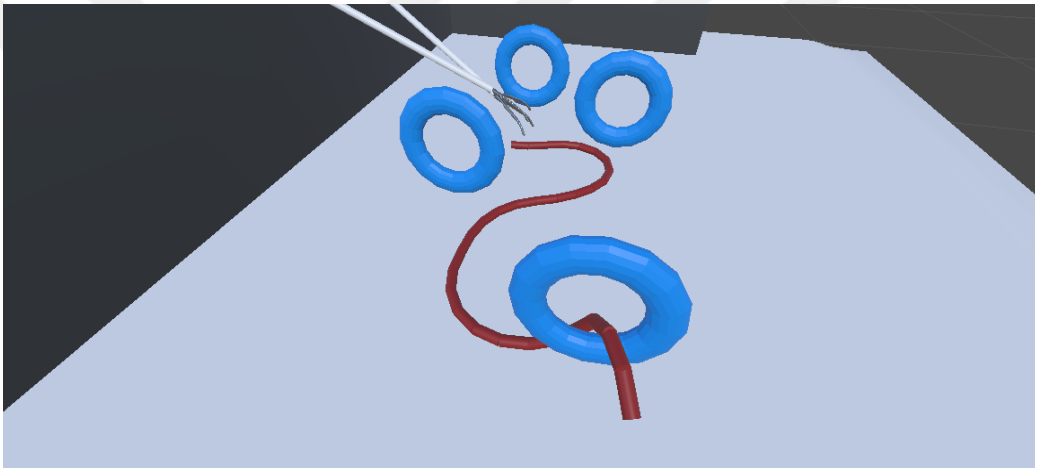


Figure 2: Modelling of the task objects for the String Pass task.

The primary instrument employed in laparoscopy is the grasper, which exhibits an inverted movement pattern as a result of the entry point serving as the pivot. The range of motion is constrained by the presence of a pivot point, limiting movement to angular motion along the X and Y axes with respect to the grasper object. The Z-axis movement pertains to the linear motion along the hole in the box, while rotation refers to the rolling motion that alters the angle of the grasper's tip (see Figure 3). The imposition of movement limitations is accomplished through the utilization of the 3D configurable joint component of the Unity engine. The configurable joint is situated at the point of penetration of the grasper into the box. Analogous to the movement along the X and Y axes, it also imposes limitations on rotational motion along those axes.

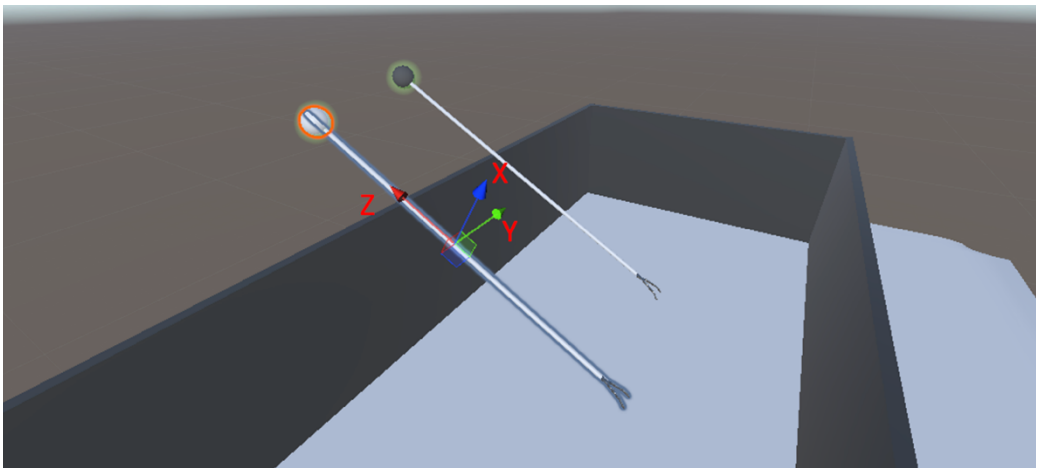


Figure 3: Axes relative to the grasper and the pivot point in the box trainer setup from the editor screen.

In the box trainer, graspers are held by hand and the index finger lays on the trigger that opens and closes the tip of the grasper. Moreover, the tip is rotatable by a disk, controlled by the thumb. In order to imitate the box trainer, the initial step involves holding the graspers. The detection of the proximity between the holding part of the grasper and the object is accomplished by the use of ray casting. Upon activation of the secondary trigger, placed on the side of the controller, the user applies pressure with their middle finger, resulting in a motion that closely resembles the action of genuine graspers. If the ray makes contact with the head of the grasper, it becomes attached to the user's hand. Additionally, the process of object pickup involves the utilization of ray casting to detect events of collision between the objects and the two rays projected from the lower extremities of the grasper. If both rays cast from the tip of the graspers collide with the object then the object is attached to the grasper with a joint and a force threshold is applied for the joint to be broken. The tip of the graspers listens to the trigger button action and activates when pressed (see Figure 4). The use of ray casting as a substitute for object-to-object collision and reliance on friction leads to the occurrence of jittering due to the diminutive size of the objects.

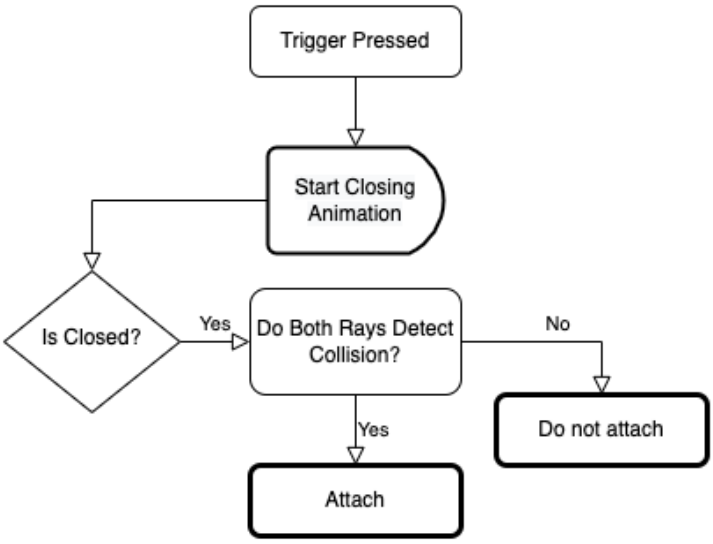


Figure 4: Flowchart of grasper and object grab interaction.

One of the most challenging aspects of laparoscopy involves the task of perceiving the three-dimensional (3D) environment through a two-dimensional (2D) screen. A camera was integrated within the box, enabling the rendering of its perspective onto a 2D texture. This texture was then applied to the screen model, thereby replicating the visual representation of the box trainer. As depicted in the in-game sample scene, illustrated in Figure 5, the screen displays the interior of the box from an angled perspective. The requirement for video-hand-eye coordination can be understood by the examination of the inverse motion and altered perspective from which items are observed. Additionally, the sample scene without the cover of the box trainer further demonstrates the difference between viewing the graspers with the camera and with the naked eye in Figure 6.

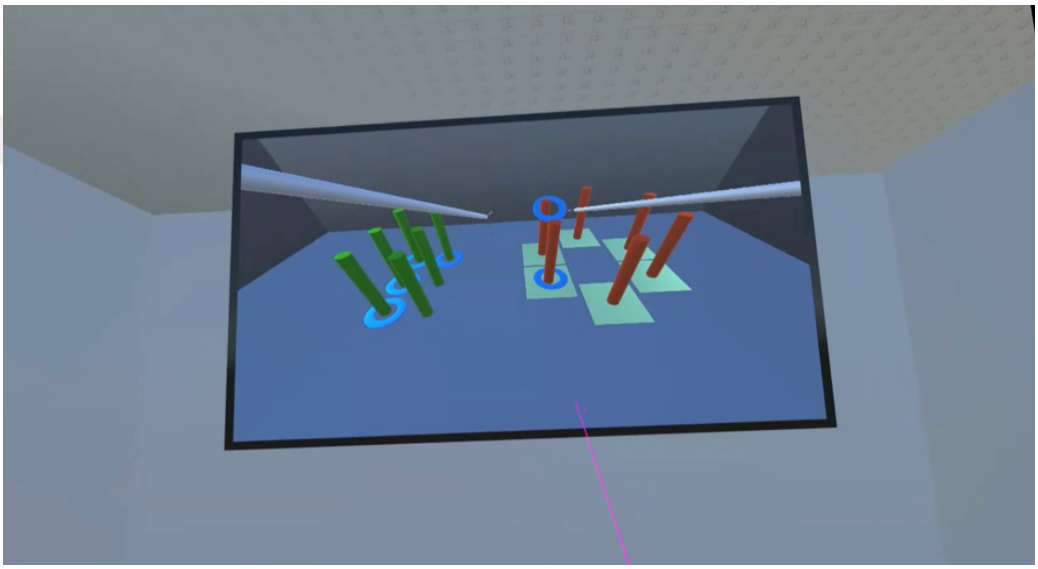


Figure 5: Peg Transfer task from in-game HMD recording.

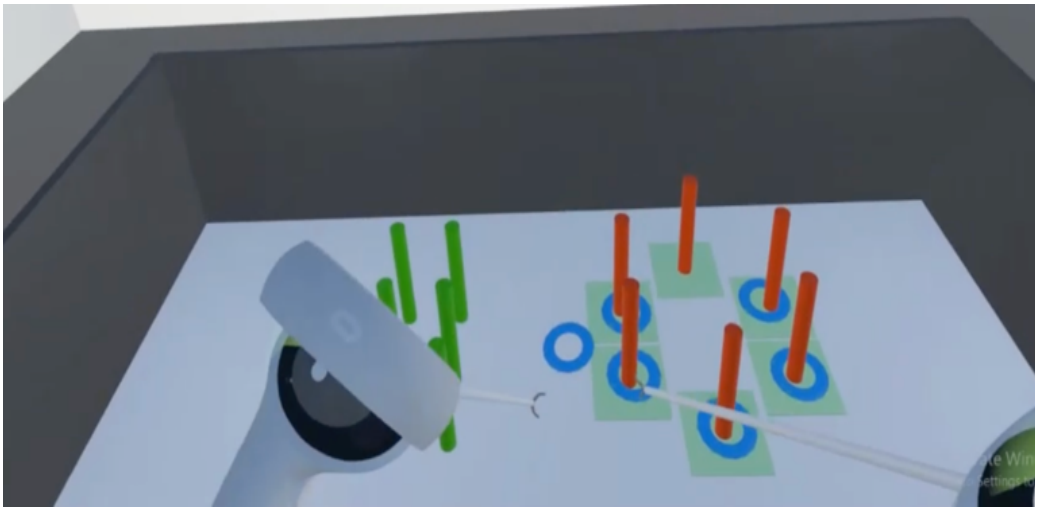


Figure 6: Peg Transfer task from in-game HMD recording top of the box removed to show inside with the naked eye.

Another logical operation that was incorporated in the initial iteration involved the process of quantifying the drops and accurately determining the appropriate

positioning of the toruses. The identification of the object's drop is recognized as an error within the context of the Peg Transfer job. It is essential to discern and eliminate collisions that do not constitute errors. The error in question is related to instances where the user grasps an object but afterward drops it. This can occur either when the force applied exceeds a predetermined threshold or when the timing of transferring the object between the graspers is not properly coordinated. In order to quantify the errors, a script was implemented on the toruses. This script considers whether the objects were grasped prior to the occurrence of the collision event, as well as whether they collided with the success zone or the table. Figure 6 displays five toruses positioned within green success zones, apart from one torus on the table, which is considered an error. Furthermore, the process of determining whether a drop is classified as an error, success, or should be excluded may be observed in Figure 7.

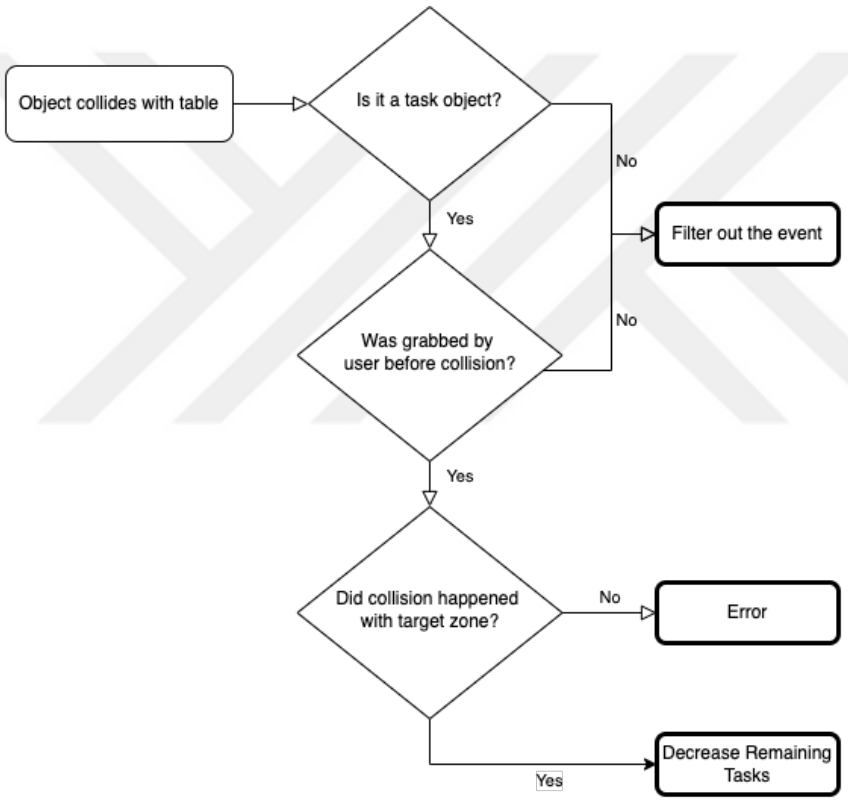


Figure 7: Flowchart of deciding if a collision is a successful move or an error.

3.2 Implementation and Features of the Second Version

The second version is built on top of the existing version, reusing most of the components but changing several aspects of the user interface and physics.

3.2.1. User Interface

The predominant evaluation method of task performance as stated in the literature, is the time to finish the task, combined with some variation of punishment based on errors made on occasion. Therefore, we added the elapsed time for both tasks and the

error count for the Peg Transfer task. Both information is presented below the screen, in a place users can see without changing their head orientation. The displayed elements from the in-game recording can be seen in Figure 8.

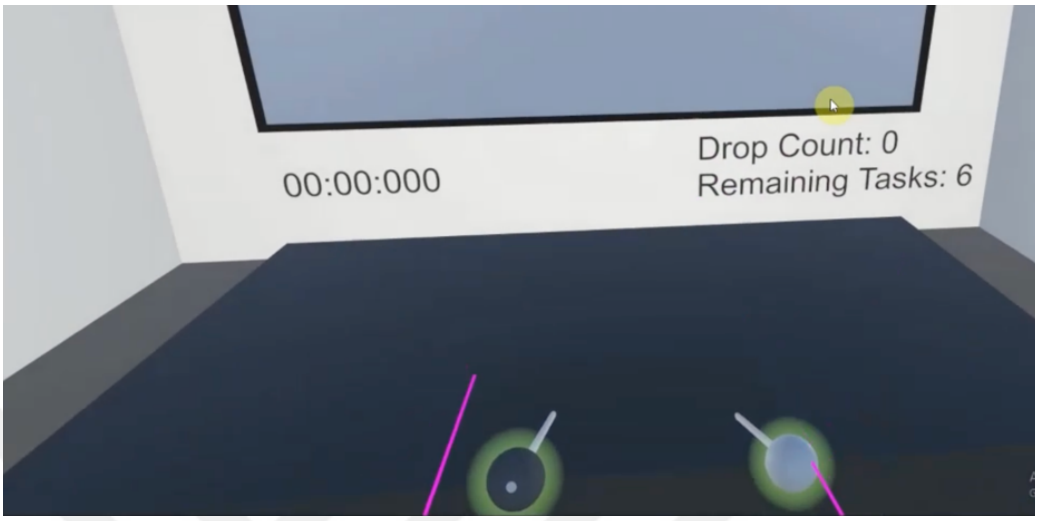


Figure 8: Time elapsed, error count and remaining task indicators added in the second version.

3.2.2. Physics and Interactions

The grabbing mechanism in the first version is improved with feedback from the participants. The ray length and the power threshold to break the joint that held the object attached to the grasper are increased. The top parts of the graspers are also marked with a green glow to indicate the correct holding position. Moreover, the haptic feedback on collision with solid, immovable objects is added to inform the user about the tip of the grasper touching the boundaries of the virtual box.

3.2.3. Tutorials

Before each task, a tutorial that showcases the required motion to achieve the task is added. For the Peg Transfer task, the user is required to correctly transfer one peg. The tutorial for the String Pass also shows the steps to pass a string-like object through one disk. The models used in the String Pass task are shown in Figure 9, the object marked with A, is grabbed with the left grasper to start the tutorial. Then user moves the object through the torus marked with B and grabs with the right grasper. Lastly, the object dropped on the zone marked with C to successfully finish the tutorial.

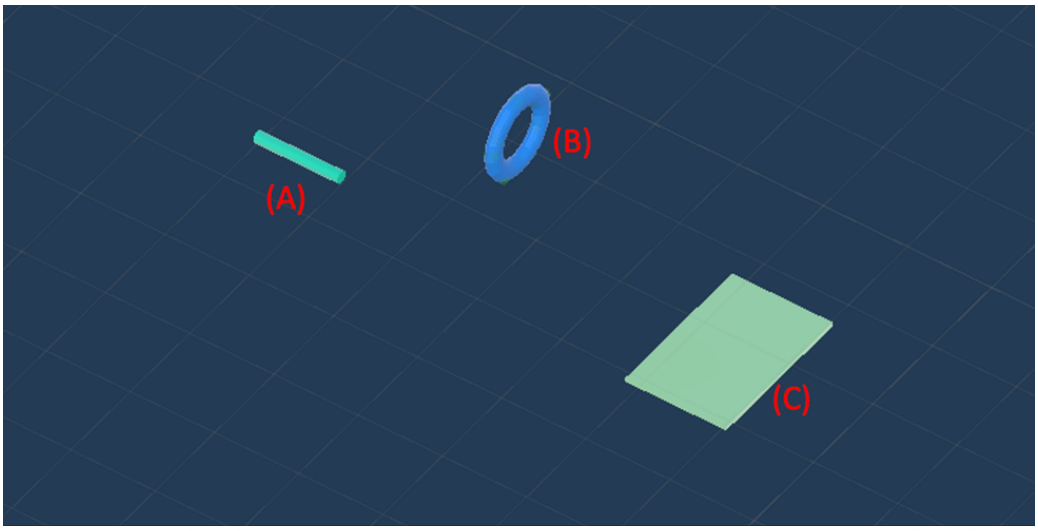


Figure 9: Object models of (A) rigid object for moving, (B) torus to move the object through, and (C) drop zone to put the object to complete the tutorial.

For both tutorials, each step is explained to the user with on-screen text on the right-hand side of the in-game screen. Since the tutorials are started after the task selection from the main menu, the ability to skip the tutorial is also provided with a button. shows the sample scenes from both tutorials added in this version.

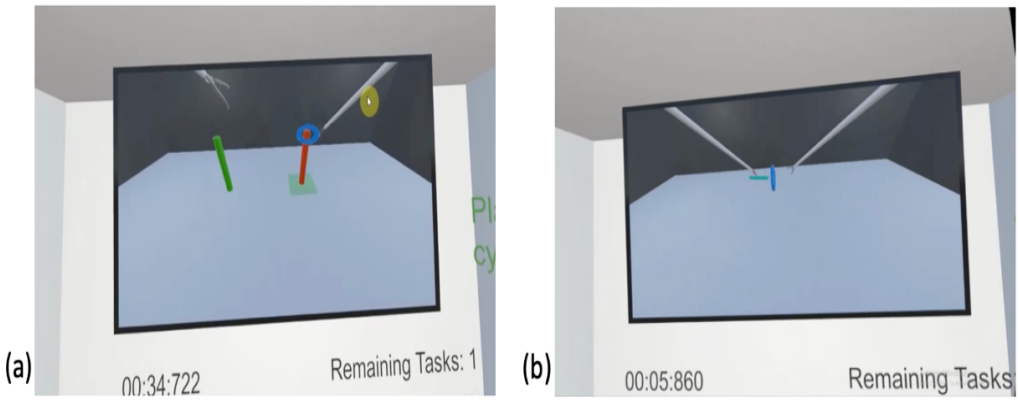


Figure 10: Sample scenes from (a) Peg Transfer tutorial (b) String Pass Tutorial.

3.2.4. Motion Data Collection

The analysis of hand motion data is crucial for assessing the proficiency of users and has been thoroughly explored in earlier studies within the context of both box trainers and computer-assisted trainers. This study aims to analyze the resemblances between the box trainer and our iteration of the virtual reality adaptation. Furthermore, an assessment follows to explore the potential application of VR motion data in assessing the proficiency level of the user. In order to accomplish this objective, a method for automatic data gathering is incorporated into the application. Once the task begins, the user's simultaneous grasp of both graspers triggers the initiation of a listener mechanism and a lister. This listener is responsible for capturing the positional coordinates and rotational orientations of the graspers. The recorded data is then stored

in a comma-separated values (CSV) file, which includes a distinct identifier for each user and the duration of the task. The data files are stored in the local filesystem of the Quest 2 device and subsequently transferred to a computer for analysis following the completion of the experiment. In the following sections of the study, the data that has been gathered will be subjected to a detailed analysis with the objective of revealing patterns, trends, and similarities in hand motions between the two training methodologies. The present analysis aims to provide insights into the suitability and efficacy of motion data as a means of assessing user proficiency. By examining the potential of VR motion data in evaluating user skills, this study seeks to add useful insight into the fields of VR technology and skill assessment methodologies.

CHAPTER 4

EVALUATION

In the study, the VR adaptation is aimed to be evaluated in terms of self-assessed questionnaires, kinematics analysis of the hand motions, and the psychological response of the participants.

4.1 Ethics

The initial Ethical Approval of Research was granted for this study by the Middle East Technical University Human Subjects Ethics Committee before the user tests of the first version in December 2021. After the completion of the first version tests, with the inclusion of physiological data in the second version, the submission of approval is revised and re-submitted. The submission was approved in April 2023.

4.2 Evaluating the First Version

The first version acted as a starting point for our application. The evaluation is based on user experience for this version and aims to find points to improve and build upon. The study utilizes semi-structured interviews with the subject matter experts, surgeons with experience in laparoscopic surgery, as well as higher education students or graduates with no prior experience in this particular medical field for quantitative usability research.

4.2.1. Participants

In order to evaluate the usability with an acceptable confidence level, we aimed to organize a group of volunteers, more than 20 people [65]. The evaluation of the initial version was conducted with a cohort comprising 29 participants, with a mean age of 24.8 ± 3.9. Moreover, 21 participants reported lacking prior experience in VR applications, and 8 participants reported prior experience in VR applications. The group was diverse in terms of previous gaming experience, nine of which categorized themselves as casual gamers, nine core/mid-core, nine non-gamers, and two hardcore gamers. Moreover, we contacted three surgeons who have previous experience in laparoscopic surgery.

4.2.2. Experiment Setup

The experiments involving volunteers are conducted in a direct, in-person manner. The study participants utilized the Meta Quest 2 device to undertake job completion within a controlled setting that allowed for standing and mobility (see Figure 11). In order to determine the prior experience of our participants in virtual reality (VR) and gaming as a whole, individuals are requested to complete the user profile questionnaire prior to engaging in the experiment. Following the administration of the surveys, participants received a comprehensive orientation regarding the laparoscopic procedure, the task structure, and the objective. The allotted time frame for completing the assignment is constrained to a duration of 360 seconds. In the event that the task remains incomplete within this designated time restriction, the user has been promptly notified. Once participants have completed both activities, they are instructed to complete the System Usability Scale (SUS) [66], Gameful Experience Questionnaire (GAMEFULQUEST) [67], Presence Questionnaire (PQ) [68] and the Technology Acceptance Questionnaire (TAM) [69]. Each session, excluding the time to fill out the questionnaires, lasted around 30 minutes.



Figure 11: Sample scene from the test conducted with Meta Quest 2.

The experiments for the surgeon group were conducted in a less constrained manner. During the performance of the tasks, the surgeons who are involved are requested to engage in verbal communication and elucidate their cognitive processes, as well as provide a comprehensive account of their overall experience throughout the testing procedure. The methodology employed in this context is commonly referred to as the "think aloud" technique, which is widely utilized within the field of game user research. Furthermore, at the conclusion of the trial, participants are requested to provide feedback regarding their comprehensive encounter and their perspectives on the utilization of VR in laparoscopic training.

4.3 Evaluating the Second Version

Similar to the first version, the same set of questionnaires is used for quantitative analysis. Different from the first version, this version includes automated motion data collection for the whole group and ECG data from half of the group for physiological response analysis. Half of the group is used for physiological data collection due to the hardships of the collection method and the unavailability of the hardware. Moreover, the preliminary results showed statistical significance with 95% confidence on some features that were expected to differentiate according to skill level, thus the collection of ECG data after 15 participants is concluded.

4.3.1. Participants

Following the enhancements made to the initial iteration, this revised version underwent testing with a distinct cohort of volunteers. The underlying justification for that move was to demonstrate that a different cohort of volunteers would exhibit similar inclinations toward the initial iteration. Moreover, upon doing comparative testing of both iterations, researchers would be able to discern enhanced elements within the program and observe heightened levels of acceptability and happiness among users. In order to fulfill the objective, an additional cohort of 29 individuals volunteered to partake in the experiment. The average age of the second group was 24.78 ± 2.64, ranging from a minimum of 21 to a maximum of 31. Thus, the age distribution of the first and second groups was similar.

4.3.2. Experiment Setup

All participants receive a standardized briefing prior to the commencement of the test, after which they proceed to test the initial iteration of the application. No additional support is offered during the examination, and participants are allotted a maximum of 360 seconds to complete the training task. Following the initial iteration, the initial 15 individuals who took part in the study underwent the necessary preparations for the collection of ECG data. This involved the placement of a wireless ECG transmitter device on the user's head, as well as the positioning of electrodes in appropriate locations to facilitate the recording of ECG signals across three channels. Following the installation of the electrodes, the participant commences the activities. Prior to each task, the participant is required to complete the tutorial part, while recording the rest condition ECG between the tutorial and the task for a duration of one minute. Subsequently, the ECG recording is marked by the end of the rest interval, at which point the subject commences the training task.

4.4 Evaluation Methods

4.4.1. Questionnaires

The System Usability Scale is a method for assessing the perceived usability of a system. The questionnaire includes 10 questions that are applicable to a broad range of systems and are said to be technology agnostic. In the questionnaire, half of the questions are positive whereas the other half are negative and those are separated in odd and even numbers respectively. The questionnaire employs a 5-point Likert Scale

that indicates the level of agreement from strongly disagree to strongly agree. SUS results in a single score between 0 and 100 that expedites comparison. SUS score is calculated based on the following formula:

$$\text{SUS Score} = \left(\sum_{i=1}^{5} (\text{Item}_{2i-1} - 5) + (5\text{-Item}_{2i}) \right) \times 2.5$$

There are alternative methods for interpreting the SUS score. Bangor et al. assign adjectives to score ranges and provide an easy way to interpret the results [70]. In this study, the results of the SUS are interpreted with these adjectives.

Presence is an important aspect of immersive experiences. In this study, the extent of the user's perceived feeling of being in the presented virtual environment is measured with a widely used Presence Questionnaire [68]. Witmer et al. constructed a PQ questionnaire with 28 questions and a 7-point scale. The questions are divided into subtopics, where the first 14 are about the control, eight are about sensory, five questions about distraction, and one is about realism.

Fred Davis, in his work in the late 1980s, deeply investigated how to predict the acceptance of computer technology. Davis first mentions the Technology Acceptance Model (TAM) in 1985 and makes an effort to define a model to provide insight to system designers and implementors, especially in the early development process [71]. Furthermore, in his famous study on usefulness, ease of use, and acceptance, he builds upon his claims and deepens the theory of what affects an information system to be accepted [72]. The Technology Acceptance Model is one of the most widely used questionnaires since its introduction. Venkatesh and Davis provide a set of questions, referred to as TAM2 in their study [73]. We selected a subset of these questions corresponding to usefulness and ease of use. As proposed in their study, questions are answered on a 7-point Likert scale.

4.4.2. Think-Aloud Playtest with Surgeons

Game user research is a domain that connects human-computer interaction, game development, and experimental phycology. Although our simulation is not a game, it shares similarities in terms of interaction medium and having a goal with defined constraints. The feedback from the user is the main focal point for the game user research and can be in the form of qualitative and quantitative. We focus on qualitative approaches that extract information about the experience during play. One of these approaches to figuring out the problems with the subject at hand is the think-aloud technique [74]. It is an intuitive but recently formalized approach that has been used in game development. The user is asked to evaluate the experiment during the play session and the recording is investigated afterward. It is discussed that evaluating the experience during the play prevents the disruptive mental tendencies of producing explanations as opposed to retrospective verbal protocols [75]. In this study, we also benefit from Knoll's overview of how to conduct a think-aloud session in preparation and interaction with the user during the test [76].

The participant group is limited due to reaching out to expert surgeons and arranging test sessions. However, according to Nielsen, after five participants the number of problems found reaches a plateau and a group between three and five people is sufficient to figure a sufficient percentage of the problems [77].

As a result of a verbal protocol, extracting quantitative information from the think-aloud transcripts is important for understanding the underlying patterns and repeating themes. The coding technique from grounded theory is used for this purpose [78]. In our study, our sessions with surgeons took at most 11 minutes and three surgeons participated, thus the findings could be reported by selecting the direct sentences, however using a method that would give insight for using alternative methodologies for analyzing a VR environment. We used the open coding technique in three stages as explained with sample cases and its pitfalls to avoid in Blair's work [79]. The stages consist of coming up with new codes in the open coding stage, connecting the codes into subtopics in the axial coding stage, and coming up with central concepts that correspond to information that is extracted and grounded.

4.4.3. Incorporating Motion Data

Using the motion data and its kinematic analysis is important to underline similarities between the VR adaptation of the box trainer and other similar surgical training methods. To be able to extract the velocity, acceleration, and jerk the raw position and rotation vectors of the tips of the graspers are recorded into a file for each task and marked with a unique identifier. All rows in the recording have the timestamp, grasper name, and task name along with position and rotation vectors. The recording starts when the participant holds both graspers, avoiding the recording of the preparation time at the beginning of the task. Moreover, the recording time starts when the participant touches the first objective object, peg, and string in their respective tasks.

The raw position vectors are then converted into velocity and jerk vectors using the change in position to calculate the velocity for each sample with the below equation.

$$\Delta P_t = P_{t+1} - P_t$$

Then the instantaneous velocity becomes the limit of change in position vector **P** in three-dimensional space while the time difference between samples approaches zero. To find instantaneous velocity, closer to the real value, the sampling rate is set once in each frame.

$$v = \lim_{\Delta t \to 0} \frac{\Delta \boldsymbol{P}}{\Delta t} = \frac{dP}{dt}$$

We use the instantaneous velocities to find average velocity and use vector norm to convert it to average velocity magnitude. Where *t* is the time stamp and *n* is the sample size. Likewise, jerk is calculated by repeating the same process three times to calculate third order derivative of position **P**.

$$\mathrm{v} = \frac{1}{n}\sum_{t=0}^{n}\left\|\frac{d\boldsymbol{P}}{dt}\right\|$$

$$J = \frac{1}{n}\sum_{t=0}^{n}\left\|\frac{d^3\boldsymbol{P}}{dt}\right\|$$

4.4.4. Features of physiological signals

The ECG signals are collected from the 15 participants while playing the second version of the application. Users from this group are informed about the physiological data collection and confirmed that they have no diagnosed cardiological problems. ECG data recording is done after the participant completes the tutorial. The recording consists of two parts, resting and during the play. Participants are asked to sit down and stay still best of their ability until the observer marks the recording with the finish of the rest period by pressing a button located on the top of the recorder device. The resting period is one minute. The recording is saved as two CSV files, the main file with the timestamp and three separate columns of channels corresponding to the signal value at that time. The other file called marker is also in CSV format and has the timestamp of the marking click. The ECG signals are captured using the Mentalab wireless headset, which is equipped with 8-channel ECG and EEG capabilities. The wireless device was affixed to the participant's head using an electroencephalography (EEG) cap that was designed for use with the wireless device. The sampling rate utilized in the recording was 250 Hz.

The raw ECG signals are cleaned and analyzed using Python and The NeuroKit2 Python package [80]. NeuroKit2 package provides a high-level abstraction over the preprocessing and feature extraction phases. There are many ways to extract features from ECG signals, the library works with cleaning and peak detection stages before time domain feature extraction. For the cleaning phase, after experimenting with available methods of cleaning and checking the cleaned signal quality we decided to use the method proposed by Elgendi et al. that uses an 8-20Hz bandpass filter and thresholding according to the expected width of QRS complex on healthy individuals [81]. The library supports various peak detection algorithms, and we selected the algorithm developed by Kalidas et al. due to its desirable properties, (1) the ability to adapt changes in HR and HRV, (2) the ability to adapt changes in signal quality [82]. As we record resting and during the task ECG and want to distinguish between the two, the selected algorithm must be adaptable to the expected change in HR and HRV between those two states by updating its beat detection thresholds. Furthermore, the collision between the HMD of the VR device and the ECG recorder on the head of the participant causes changes in the base voltage of the signal and false peaks, which the algorithm claims to suppress. After the R peaks are identified, mean HR is calculated as the average of beats per minute that is number of R peaks per minute. Furthermore, HRV is calculated mean distance between the R-R peaks.

CHAPTER 5

RESULTS

This chapter presents a statistical analysis conducted on data obtained from questionnaires administered to two groups: one consisting of 29 individuals and the other consisting of 30 individuals. Additionally, qualitative data obtained from professionals in the field on the first version of the questionnaire was also incorporated into the analysis. Additionally, we present the motion and physiological data that were collected during the tasks in the second version. The study comprises two developmental stages: the initial version and subsequent user testing, followed by the refinement of a second version based on the feedback received. The initial cohort of 29 individuals completed the questionnaires subsequent to engaging with the aforementioned iteration, while the subsequent cohort of 30 individuals completed the questionnaires after engaging with both iterations. Our emphasis on the findings of the first version will be on how we identified areas that required improvement. Subsequently, we proceed with an analysis of usability, kinematics, and physiological response as we evaluate the second iteration.

5.1 Feedback for the First Version

5.1.1. Think Aloud Sessions with Surgeons

The think-aloud sessions conducted with the surgeons during the initial version are transcribed into written records and subsequently subjected to analysis using the MAXQDA software, employing coding methodology. The process of open coding involves the identification and marking of significant points within the text. Additionally, the utilization of open and axial coding facilitated the identification of specific topics. Table 1 displays the identified topics along with sample translated segments extracted from the transcripts. The topics of “insufficient feedback,” “wrong angle” and “hard to hold objects” were mentioned in transcripts three, four, and eight times, respectively. Furthermore, the surgeons collectively express the difficulty associated with grasping the objects using the graspers.

Table 1: Most frequent codes and their corresponding segments after analyzing the think-aloud session recordings of surgeons.

	Insufficient Feedback	**Wrong angle**	**Hard to hold objects**
Surgeon 1	Lack of force feedback on VR controllers, it is not possible to stop the user's hand from moving on a collision	The angles do not coincide, the angle I hold the tool and the angle I move inside are not parallel to each other.	It does not grab the object when I close the tips of the grasper.
Surgeon 2		The angle of the camera is wrong, it is making it harder to understand the position of the graspers.	Sometimes I get the object between the tips of the grasper but it does not attach.
Surgeon 3	Lack of tissue feeling, related to the lack of force feedback on VR controllers		I picked it up, but it falls when I try to hold the object.

5.1.2. Questionnaires

In GAMEFULQUEST the questions belonging to four subtopics, "accomplishment", "challenge", "guidance" and "immersion" are investigated among their respective topics. As seen in Figure 12, the below-average score of the responses given to the "Guided" topic is apparent. The questions, "Gives me a sense of being redirected" and "Gives me useful feedback so I can adapt" are two questions that have an average rating below 3 on this topic. Moreover, some of the responses given to the "What is your general opinion about the application?" highlight the problem and show direction on how to improve on it, such as "The task about disks is a familiar image, but the second task is something that should be introduced for those outside the field."

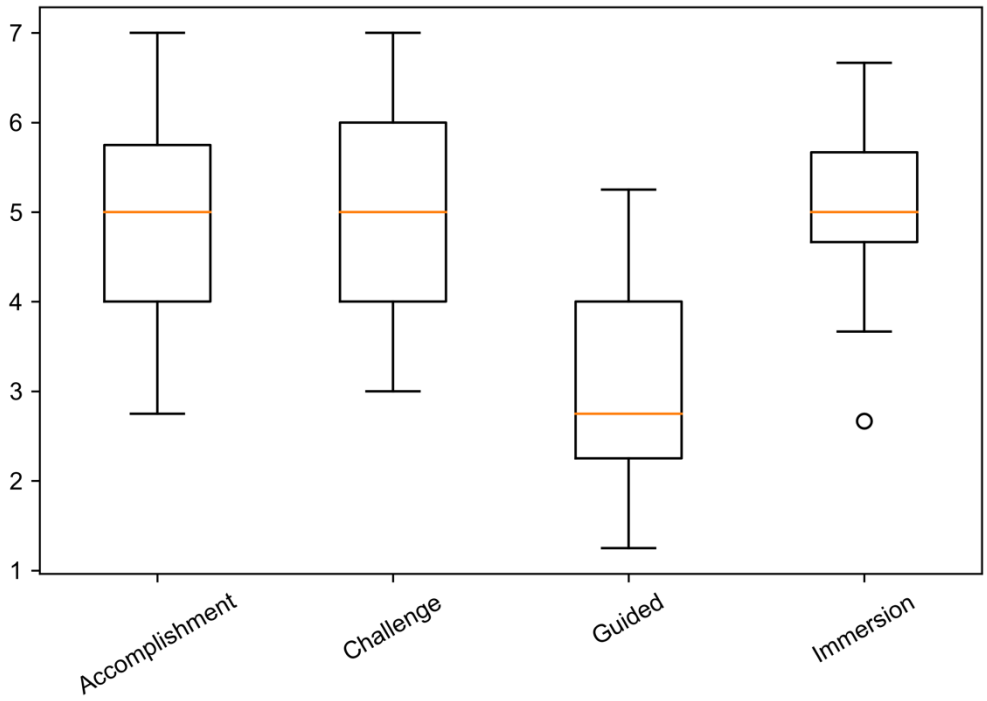


Figure 12: Boxplot of the responses to the GAMEFULQUEST for the first version, grouped by subtopics.

The SUS questionnaire of the first version is important for two reasons, first to see the current stage of the development and understand the patterns in user profile and usability. Moreover, the SUS results of the first version are instrumental in comparing the two versions in the upcoming results. The individual scores of the participants can be seen in Figure 13. Also, it can be seen that 10 users out of 29 gave SUS scores above the "good" threshold and there were two users who evaluated the application as "not acceptable".

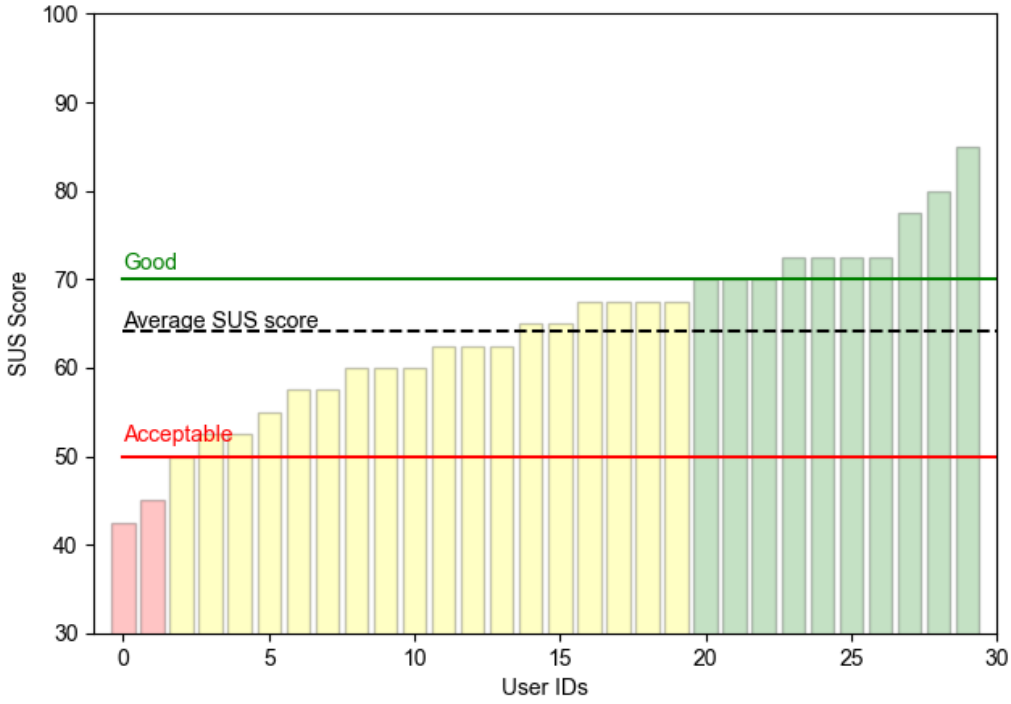


Figure 13: Visualization of SUS scores of the first version per user with "acceptable" and "good" threshold markings.

Another important quantitative data comes from the Technology Acceptance Model questionnaire, to show how the users perceived the usefulness and how easy to use the simulation. In Table 2, six questions on the topic "ease of use" are shown, the mean values of all the questions were above 4 out of 7. Since there is no scoring system for TAM to interpret the results as opposed to SUS, we assess the answers to questions to ascertain what can be improved in the version and use the numerical values to compare it with the second version.

Table 2: TAM Questionnaire mean, and standard deviation of the questions answered for the first version under the topic "ease of use".

Question	**Mean**	**STD**
Learning to use this application is easy for me.	4.97	1.57
I found it easy to get this product to do what I want to do.	4.48	1.47
My interaction with this application is clear and understandable.	4.83	1.28
I found this product flexible to interact with.	4.21	1.35
It was easy for me to become skillful at using this application.	4.93	1.62
I found this application easy to use.	4.79	1.49

5.2 Questionnaires on the Second Version

In the second iteration, participants answered the questionnaires for both versions, making it possible to compare their observations. It is important to note that, the second group that we will present here played both versions before answering the questions, making it easier for them to compare the versions and recognize the changes made in the development of the second iteration.

According to the analysis of the GAMEFULQUEST, the "Guided" topic was the worst performing aspect of the simulation according to the first group with a mean score of 3.00 ± 1.08. The answers of the second group showed a significant increase in the "Guided" topic. The mean of the first version was 3.73 ± 1.45, whereas the mean of the second version was 5.70 ± 0.75. In Figure 14: Boxplot visualization of the GAMEFULQUEST answers in the "Guided" topic of both versions and groups.Figure 14, the comparison of the distribution of the scores between the groups and versions is shown, the difference between the versions is apparent between the first and second iterations of the simulation. Moreover, considering other topics, the difference between the groups and the versions are tested if there is any significant difference. Between the first and second groups, no topic other than "accomplishment" was significantly different than expected. However, in both versions, the mean scores for "accomplishment" were relatively high, with 4.91 ± 1.21 for the first group and 5.57 ± 0.92 for the second group. In a comparison of the two versions of the application, the second group assessed the second version with higher scores in all topics but the "immersion". Moreover, as the determined aspect for improvement, more guidance is appreciated by the users with improvement in mean score from 3.72 ± 1.45 to 5.70 ± 0.75 according to a two-tailed t-test with a p-score less than 0.001.

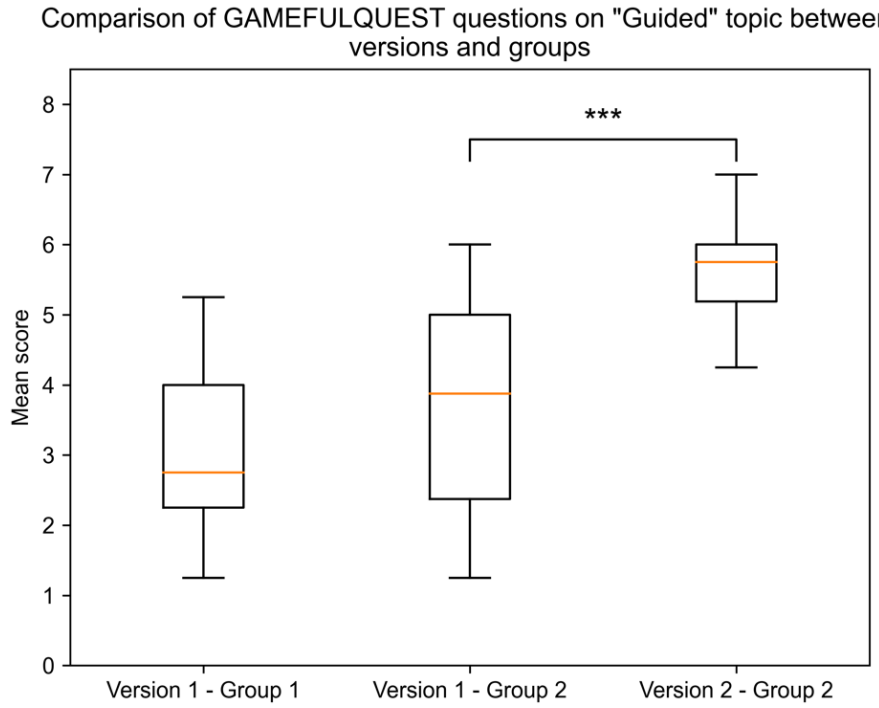


Figure 14: Boxplot visualization of the GAMEFULQUEST answers in the “Guided” topic of both versions and groups.

Table 3: Comparison of versions and groups on GAMEFULQUEST for each topic according to a two-tailed t-test.

	Group 1 vs Group 2 On Version 1		**Version 1 vs Version 2 On Group 2**	
	t	p	t	p
Guided	-2.027	0.059	-5.436	**< 0.001**
Challenge	-0.088	0.930	-2.056	**0.045**
Accomplishment	-2.261	**0.028**	-2.353	**0.023**
Immersion	1.946	0.059	-1.579	0.122
Playfulness	0.267	0.791	-2.028	**0.049**

Moreover, SUS scores when compared between the groups show improvement between the first and second versions. Again, as expected the mean SUS score for the first version scored by the first and second cohort are 64.10 ± 9.89 and 56.96 ± 23.26 respectively. However, the improvement is clear between the first and second versions. The mean score of the second version was 74.03 ± 15.69. The individual scores given by the second cohort for the second version of the game can be seen in Figure 15 for a better comparison with the distribution of the first version.

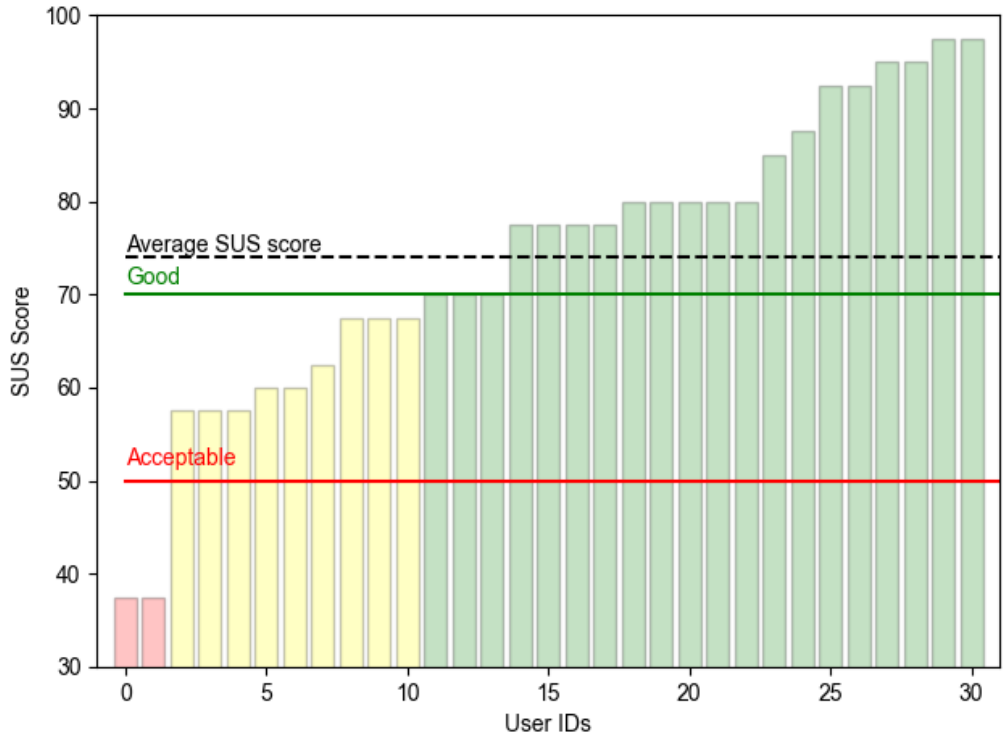


Figure 15: Visualization of SUS scores of the second version per user with “acceptable” and “good” threshold markings.

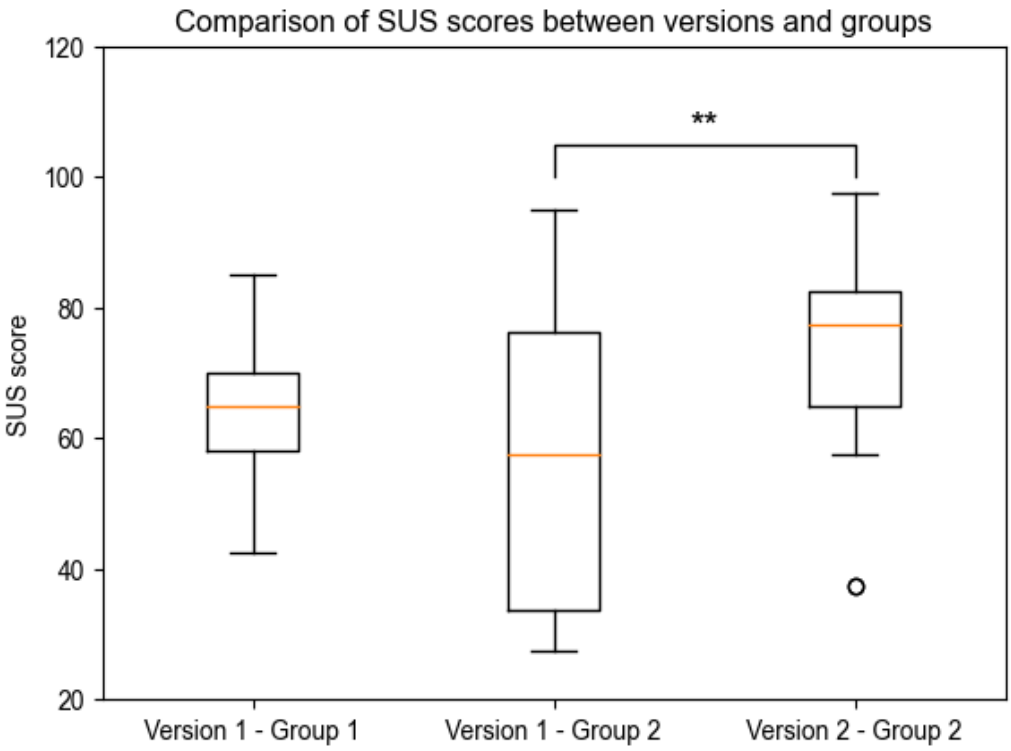


Figure 16: Boxplot visualization of the SUS scores of both versions and groups.

In section 5.1.2, an examination is done based on the findings of the TAM questionnaires, with the objective of identifying potential enhancements for the initial version of the application. This section of the results involved the evaluation of a questionnaire to assess the average scores of different versions in relation to various categories to find whether the groups rated the versions differently. Similar to the findings of the SUS, there was no statistically significant difference observed between the first and second cohorts in relation to the perceived usefulness of the initial version of the application with mean scores of 5.11 ± 0.92 and 5.16 ± 1.4 respectively. However, the results for the second version received 6.33 ± 0.64 out of 7 from the second cohort. A similar case rests on the ease of use topic as well. The first and second groups rated the initial version of the application with 4.7 ± 1.18 and 4.48 ± 1.59. The second cohort rated the second version with 5.61 ± 1.22 out of 7 on ease of use. The box plot visualization in Figure 17 shows the distribution of the scores for each group and version match on the two aforementioned topics.

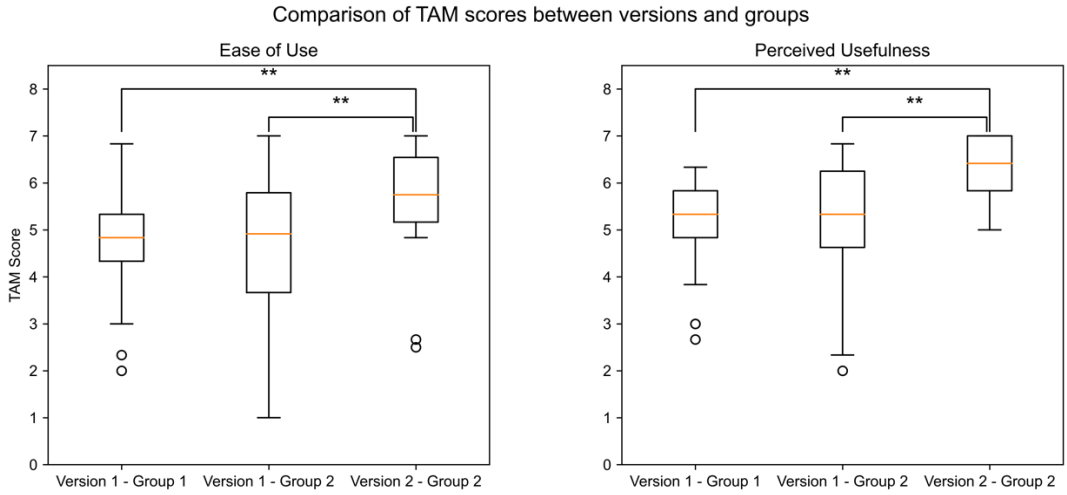


Figure 17: Boxplot visualization of TAM scores of both versions and groups in Ease of Use and Perceived Usefulness topics.

5.3 Kinematic Analysis

Following the rife approach in the literature, the criterion used to define success is the completion time of the tasks. The time to finish the task is capped at 300 seconds for both tasks. Among those who finish the tasks, mean completion times are ~204 seconds and ~257 seconds for Peg Transfer and String Pass respectively. Following the literature, the comparison between the groups, bottom and top half is done with respect to the mean completion times, users who finished the task or exceed the maximum time are labeled with the bottom half and those who achieved to finish the tasks are labeled as the top half. This leaderboard is done separately for the Peg Transfer and String Pass.

The mean velocity and jerk are calculated for both graspers used by the participants and evaluated in terms of successful and unsuccessful groups. The two halves are evaluated to show if there is a significant difference between the movement patterns

of the graspers. An example of the movement throughout the task can be seen in Figure 18.

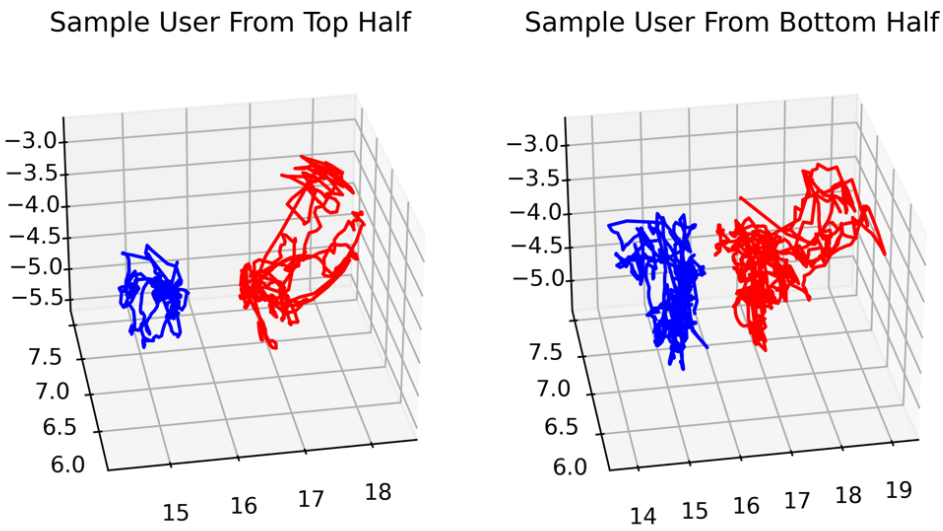


Figure 18: Movement of the grasper tip during the Peg Transfer task for two sample users representing the top half (Left) and bottom half (Right).

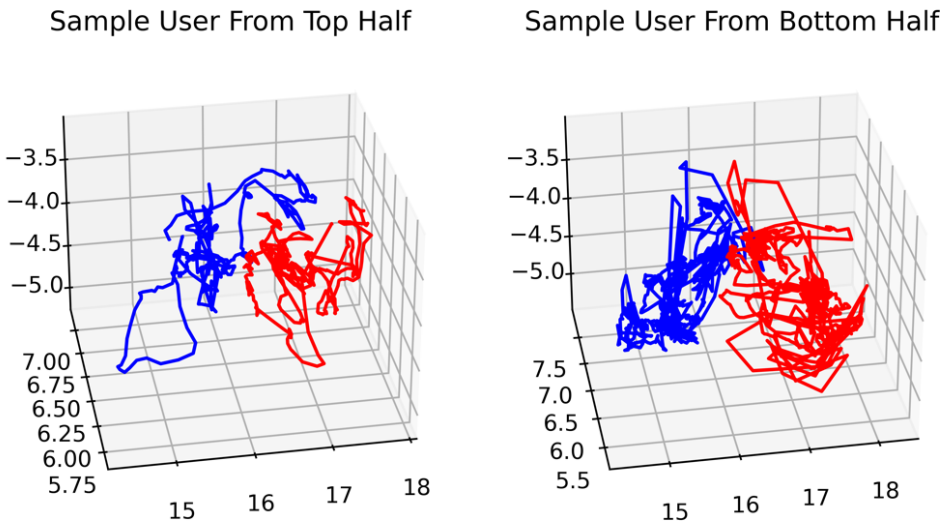


Figure 19: Movement of the grasper tip during the String Transfer task for two sample users representing the top half (Left) and bottom half (Right).

Moreover, an examination is conducted on the instantaneous velocity characteristics of the participants. Due to the absence of limiting hand movements in contrast to doing the task using simulated physical graspers, the velocity exhibits atypical patterns. In order to mitigate the impact of this anomalous motion on kinematic analysis, a rudimentary outlier detection method based on statistical principles is used. This method is designed to identify instantaneous velocity values that exceed the velocity exhibited by the individual user for a given task and hand. Anomaly detection is performed by assessing the probability of a given value being associated with the user's hand for the specified task. For our instantaneous velocity values v, if we assume they belong to a normal distribution X, then we can threshold the values in

terms of multiples of standard deviation by using the Z-score. The Z score is calculated with the below formula, where χ is the sample, μ is the mean and σ is the standard deviation.

$$z = \frac{\chi - \mu}{\sigma}$$

For our case, the sample is the instantaneous velocity ν, and the sample mean, and the standard deviation is calculated from the user's motion data for each task and hand.

$$z = \frac{\nu - \hat{\nu}}{\sigma}$$

This method, also called normalization, enables us to define the outliers in terms of multiples of the standard deviation since we transform our data to some sample with zero mean and a standard deviation of one.

$$\nu_n \leftarrow \frac{\nu - \hat{\nu}}{\sigma}$$

After transforming our samples, we can use confidence intervals for finding the outliers in our sample. Since the probability of observing a Z score greater than 1.96 or less than – 1.96, is less than 5%, we can say observing a sample $\nu \in \nu_n$ with 95% confidence.

This method presupposes that the sample set prior to the transformation follows a normal distribution, and it might be debated whether this assumption is valid. In order to assess the normality of the sample, the Shapiro-Wilk test was employed, as it is recognized as the most robust statistical test for examining the conformity of a dataset to a normal distribution. The velocity data is partitioned into sample sets based on user, task, and hand to test its normality. The results of the Shapiro-Wilk test indicate that all the data conform to a normal distribution, as evidenced by a p-value of greater than 0.05. The outliers detected for two selected users are illustrated in Figure 20. Additionally, an investigation is conducted to determine the frequency at which instantaneous velocity surpasses the 95% confidence interval among the groups segregated into upper and lower halves based on their completion times. When comparing the two groups, significant variation was observed in the occurrence of outliers. The top half of the groups had an average of 13.57 and 11.21 outliers for their left and right hands respectively on the Peg Transfer task. In contrast, the bottom half of the groups had an average of 29.33 and 28.13 outliers for their left and right hands respectively. In a similar vein, it was observed that the top half of the participants' hands produced average outliers of 12.47 and 13.27, while the bottom half had 25.07 and 26.34 on the String Pass test. The frequency of outlier occurrences is illustrated in Figure 14. The groups exhibited a high level of statistical significance in relation to their outlier numbers, as shown by a p-value of less than 0.001.

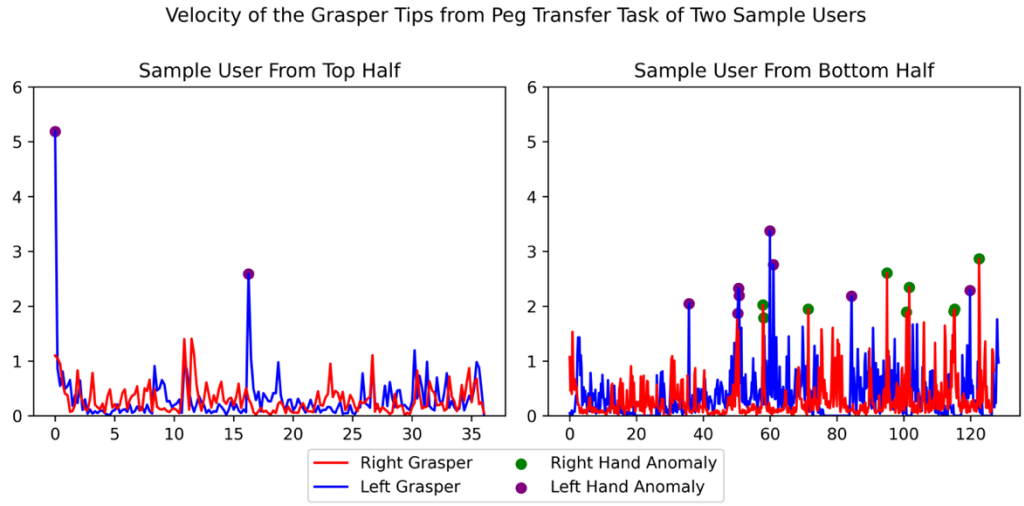


Figure 20: Results of outlier detection for two sample users from each group on the Peg Transfer task.

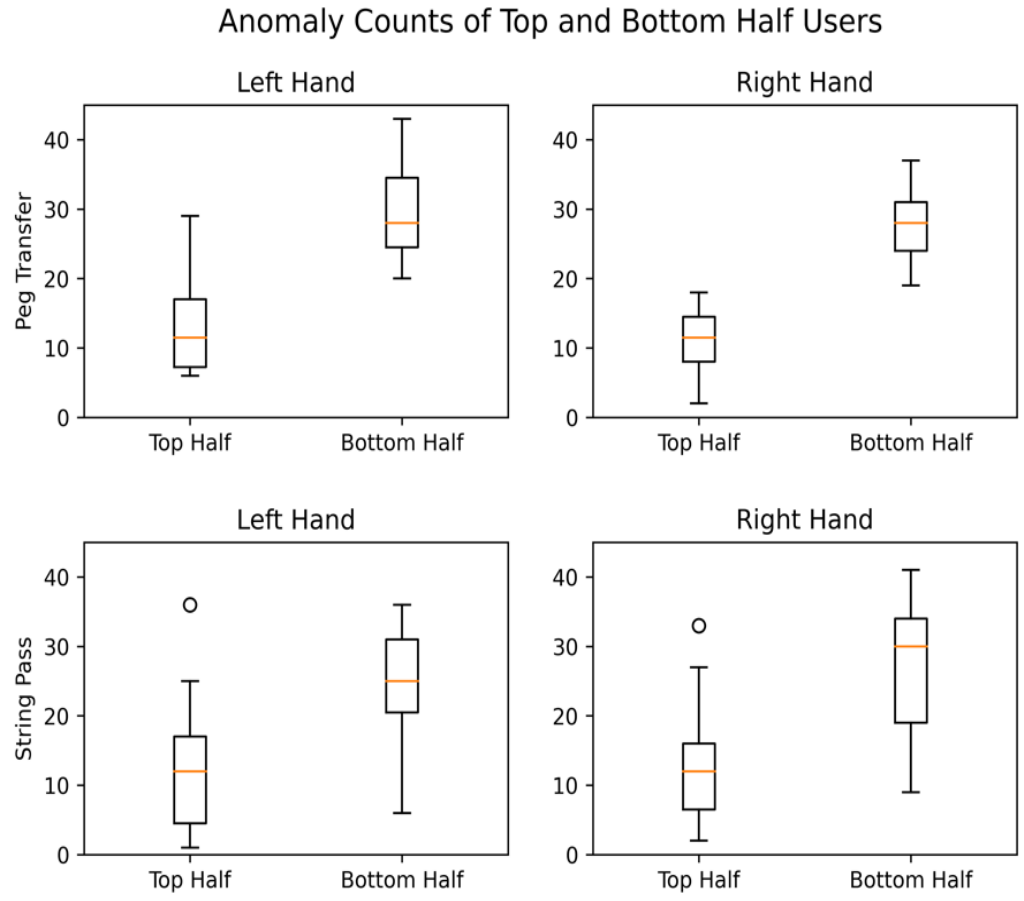


Figure 21: Boxplot visualization of the outlier occurrences during the Peg Transfer and String Pass tasks of each hand.

Distinctive motion features have been examined and identified in both expert and novice participants who utilize simulators. The first characteristic, known as the economy of movement, is quantified by measuring the length of the path required to successfully accomplish a certain task. The motion analysis conducted in our study

reveals significant differences in the lengths of the total pathways observed among each group. Another measure of the proficiency level, as noted in the OSATS, is the examination of the smoothness of movement, which is assessed by the analysis of jerk in our study. The overall distributions of mean jerk and mean path length for each group are shown in Figure 22. The mean path lengths, calculated with the total length of the path of one hand of a user during the task is found to be different between the better and worse performing halves of the users with p value less than 0.001. Nonetheless, mean jerk values did not significantly differ.

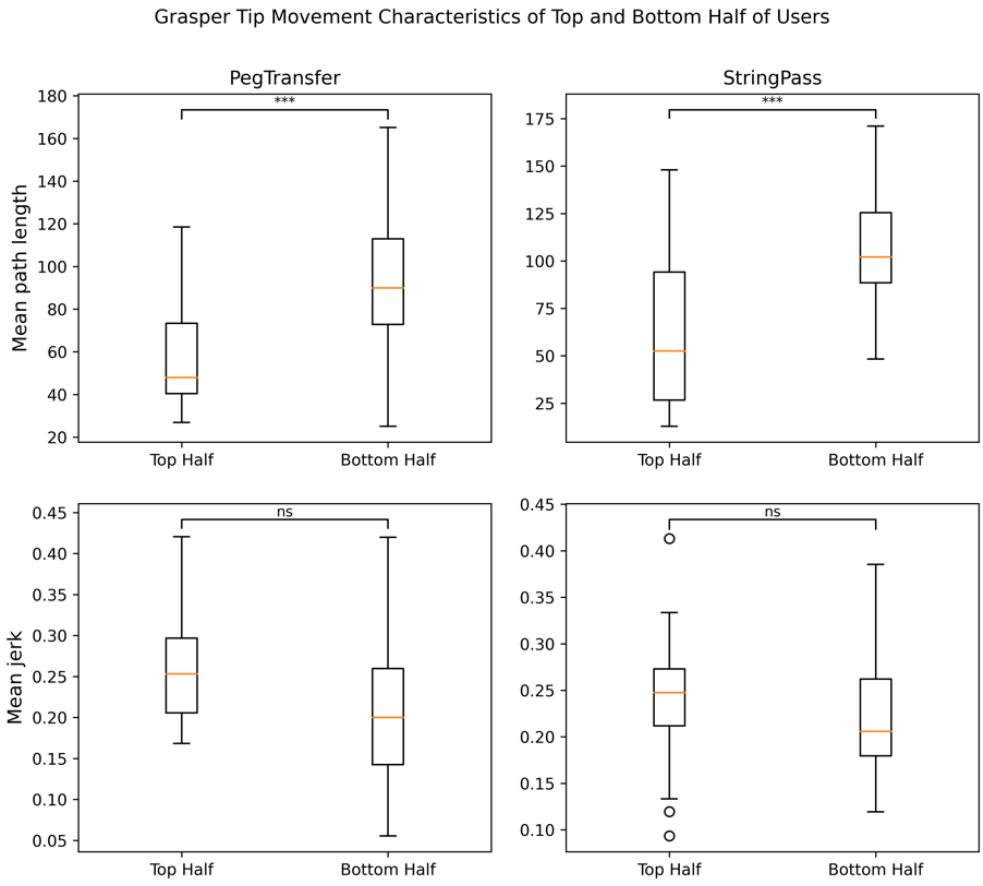


Figure 22: Boxplot visualization of the mean path length and mean jerk values corresponding to the top and bottom half of users for the Peg Transfer and String Pass tasks.

There are other related parameters to hand movement, such as velocity and acceleration all derived from the change of position of hands during the task. Similar to jerk, none is found to be differential in assessing the skill level of the participants.

5.4 ECG Analysis

ECG recordings are saved for each individual user in the form of CSV files. Each user has two CSV files: one containing voltage data and the other indicating the resting phase end and task start markings. Sample results of the ECG cleaning with the NeuroKit2 Python package, including the raw and cleaned signal, detected BPM and all detected heartbeats on top of each other can be seen in a sample user in resting state

is visualized in Figure 23. In certain instances, it has been noted that certain users' ECG recordings exhibit background noise or aberrant voltage values in the default channel utilized for analysis, among the three available channels (see Figure 24). In these instances, ECG features have been extracted from alternative channels. As shown in Figure 25, the quality of the ECG data can be improved by using an alternative channel instead noisy channel shown in Figure 24 for this instance.

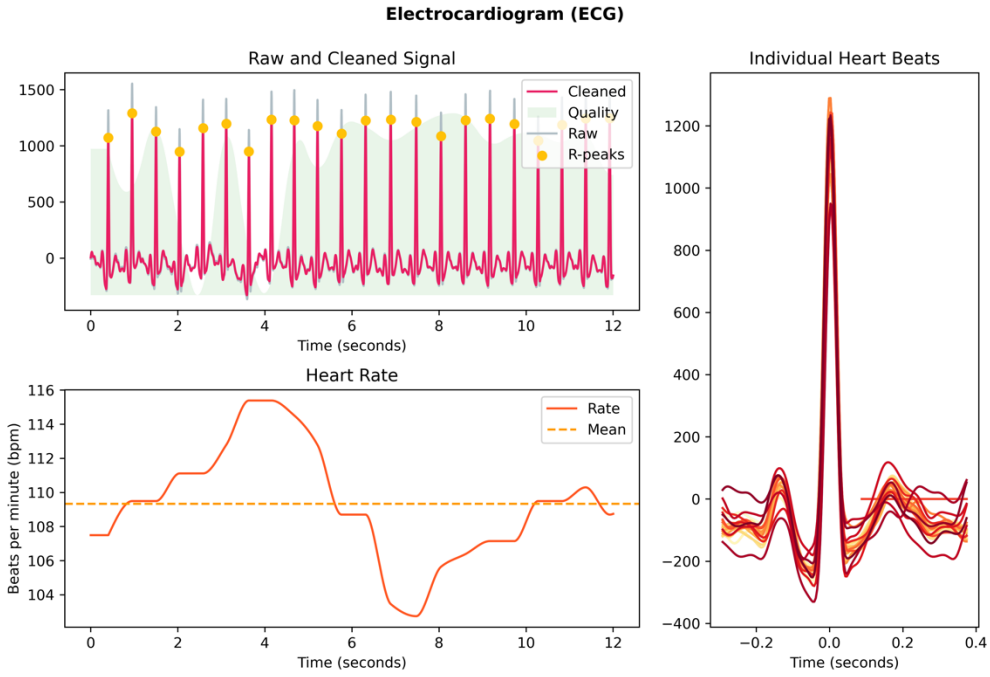


Figure 23: Visualization of ECG recording of a sample user during the initial 12 seconds of resting state.

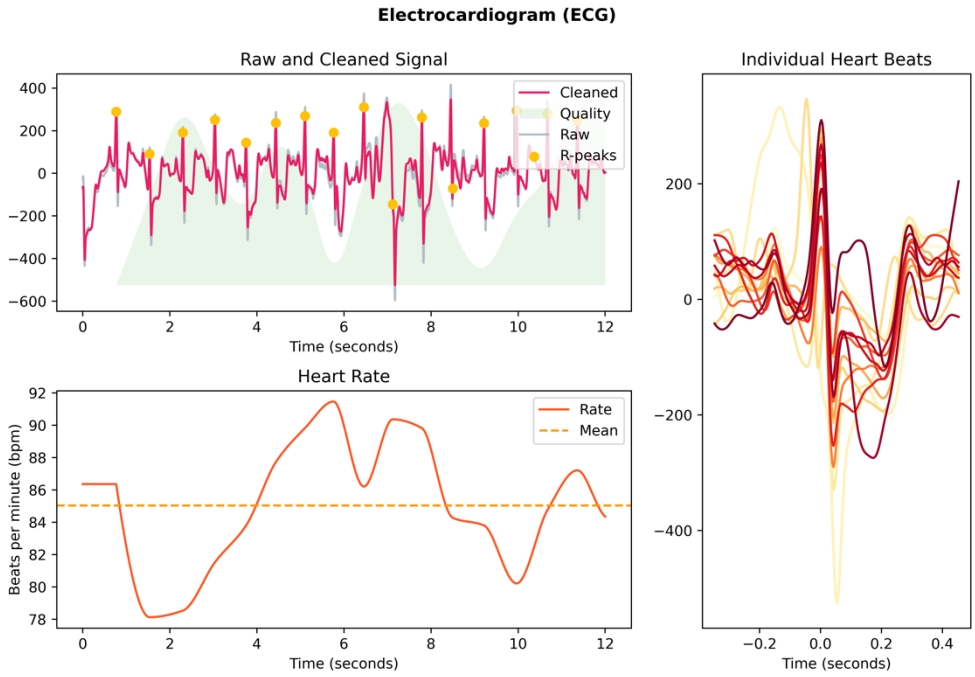


Figure 24: Visualization of ECG recording from channel 5 that has background noise of a sample user during the initial 12 seconds of the Peg Transfer task.

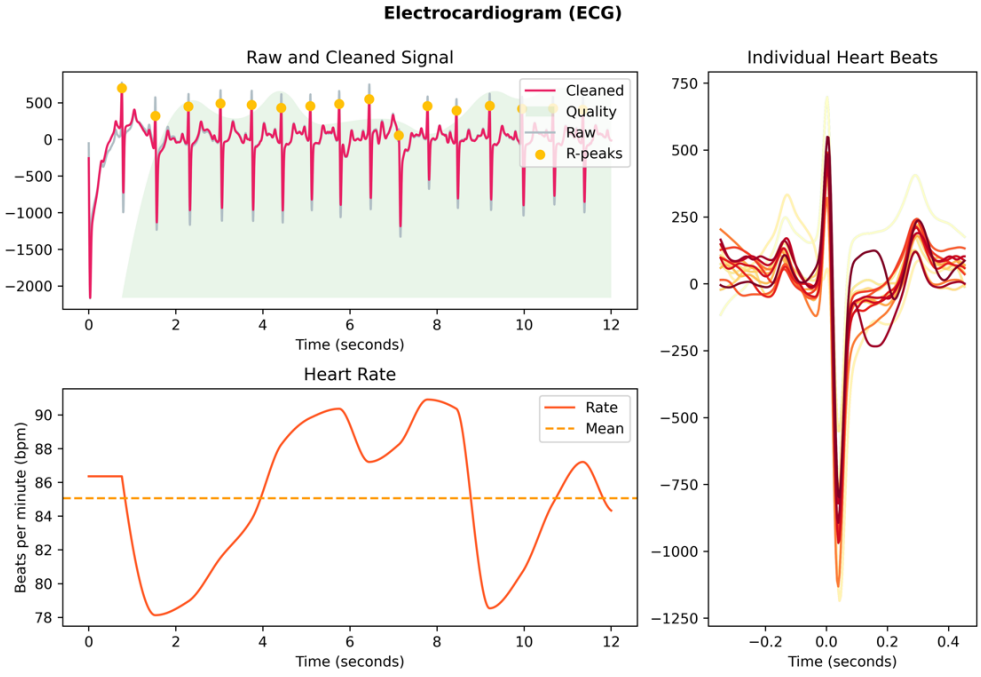


Figure 25: Visualization of ECG recording from channel 1 of a sample user during the initial 12 seconds of the Peg Transfer task.

The evaluation focuses on assessing the participant's heart rate to gain insights into their physiological response when engaging in the procedure within a VR environment. Based on the results of the one-sided t-test, it can be concluded that there exists a statistically significant reduction in the average heart rates of the participants while at rest as compared to when engaged in the task. Nevertheless, there was no statistically significant disparity found between the successful and unsuccessful groups in relation to heart rate during the performance of both tasks. Furthermore, it is important to highlight that the average distance between R-R peaks demonstrates a decrease when comparing the resting and playing conditions, as depicted in Table 4. In the examination of the frequency domain characteristics of HRV during periods of rest and during the task, it was observed that there was no statistically significant distinction in the low-frequency or high-frequency components of the HRV feature.

Table 4: Descriptive statistics of mean HR and mean distance between normal RR peaks (MeanNN) of the participants for each task between resting and playing states.

		Mean	STD	t	p	df
Heart Rate (bpm)						
Peg Transfer	Resting	84.78	10.4	-3.87	< 0.001	13
	Playing	89.11	9.94			
String Pass	Resting	86.88	11.55	-2.18	0.025	13
	Playing	89.22	10.70			
MeanNN (ms)						
Peg Transfer	Resting	717.64	84.42	4.50	< 0.001	13
	Playing	681.58	79.71			
String Pass	Resting	702.08	91.88	2.44	0.030	13
	Playing	682.61	87.91			

CHAPTER 6

DISCUSSION

In our comprehensive investigation, we sought to uncover insights that could provide guidance on the potential applications of VR training in the field of laparoscopy training, that could apply to a broader field of psychomotor training. The design decisions made during the initial and subsequent iterations were based on the data gathered from a total of 59 participants who completed our questionnaires in both phases, as well as input from three expert surgeons specializing in the field of laparoscopy. The objective of enhancing tooling efficiency through improved hardware accessibility necessitates the identification of key points that distinguish it from the current methodology. From this perspective, the examination of motion and the analysis of physiological responses, along with the administration of user experience surveys, provide a deeper understanding of how to develop a system that can be used in training novice surgeons instead of physical box trainers or expensive and inaccessible computer-assisted simulations.

The questionnaires and the think-aloud sessions uncovered the need for:

- More guidance to users, in terms of explaining the required motion patterns and how to achieve the task completion.
- Better feedback from the controllers mimics the feeling of the physical graspers and their interaction.
- Improvements in camera angle and depth perception.
- A better mechanic for holding the objects with the graspers.

The identified issues also highlight the importance of accorporating novice users as they are closer to the target audience for usability study and subject matter experts. Because the camera angle issue was only voiced by the expert surgeons as they had experience with similar trainers and real operations. Moreover, collecting data from 29 participants ensured we could compare the first iteration with the second one after it was developed. Without the subject matter expert opinion, it would be hard to base the hardship of holding the objects, as it would be debatable if the group of

inexperienced participants was biased because of their lack of psychomotor skills or not.

These issues that were identified are addressed in the second iteration. The first improvement was on the angle, the angle is corrected by tilting the camera upwards to provide a more parallel angle between the graspers and the camera with the feedback from the surgeons S1 and S3:

S1: The screen angle is tilted forward. It needs to be straight.

S1: The angles do not coincide; the angle I hold the tool and the angle I move inside are not parallel to each other.

S2: The angle of the grasper is wrong.

The lack of feedback was also mentioned by the surgeons and the participants. The haptic feedback on collision with the sides of the box and the objects are added. The positive effect of haptic feedback is also shown in literature as an enhancing effect on skill acquisition [83]. Furthermore, the mechanism for picking up objects has been modified in order to lessen the difficulty. The size of the collision boxes between the tip of the graspers and the objects is reduced and the ray length for collision detection is increased, enabling users to hold objects without completely putting the object between the opening at the ends of the grasper. The guidance issue was only mentioned by the participants as expected. The surgeons demonstrated a proficient comprehension of the task's objectives and execution, even with limited initial guidance regarding the operational mechanisms of the controller. In order to facilitate the completion of the work, two tutorial portions were incorporated prior to each subtask. With the goal of helping to facilitate the completion of the work, two tutorial sections were incorporated prior to each subtask. These tutorials were designed to provide a step-by-step instructional approach, guiding individuals through the necessary motions required for the successful completion of the task. Specifically, the subtasks involved moving a single peg and passing a short string-like object through a disk.

An analysis was conducted to compare the two groups' experience on the initial version of the job in order to see whether they exhibited similar approaches. This comparison was made despite the fact that the second group answered the questions during the first version after having experienced both versions of the simulation. The results suggest that there was no statistical significance between almost all topics of GAMEFULQUEST but one and SUS scores. Although the mean SUS score of the second group was higher both were in the range of < 70, which is considered acceptable but not good according to Bangor et al. [70]. Nevertheless, the enhancements made in the second iteration resulted in a notable increase in the average SUS score, surpassing the threshold of 70 and entering the realm of satisfactory usability. Despite the potential for heightened optimism towards the second iteration when both versions are observed, the validity of the results and the observed

improvements in usability, as indicated by the SUS questionnaire are backed by a sample size of 30 and a high level of statistical significance with a p-value of less than 0.001. In terms of ease of use and perceived usefulness, the first and second groups rated the initial version similarly, but the tendency of greater acceptance was apparent for the second version. The mean score of 5.61 for ease of use compared to 6.33 on perceived usefulness can be explained by the challenging nature of the tasks. Despite the participants' lack of experience in surgical training, their inclination to provide high scores to the utility questions in the simulation indicates their confidence in its potential to enhance motor abilities. The challenge users have faced is also shown in their responses to the question "Do you have any suggestions?". Two users made comments about the potential benefits of incorporating sound stimuli into the program, which is not currently a feature in the surgical simulation. Also, two other users indicated they want to further use the application with realistic operation settings. Where the 5 of them said they had no suggestions. The question regarding the participants' overall assessment of the application in the second version further illustrates their internal comparison of the two versions and the subsequent inferences they made, which align with our results aimed at enhancing the initial version. The most used words regarding the question can be seen in Figure 26.



Figure 26: Word cloud representation of free text responses of overall opinion on the second version.

We investigated our application of the VR version of surgical training from multiple perspectives. The positive results of the questionnaires show potential; however, the application should also show promise in terms of motion characteristics of other simulators that have been studied in terms of their effectiveness of motor skill acquisition. The utilization of motion analysis and physiological response holds significant value in enhancing the efficacy of training. The preceding research demonstrates how the kinematic attributes of the hands can serve as indicators of an individual's level of skill. It was demonstrated that the skill level of users can be

differentiated based on the analysis of motion data. In the study, the participants were categorized into two groups based on their average completion time for each activity, as none of the individuals had prior laparoscopic expertise. In a manner similar to the simulations employed in laparoscopic training, it has been seen that there exists a significant difference in path length to complete the task between the two groups. Specifically, it has been determined that the upper half of users tend to exhibit a greater degree of efficiency in their execution of tasks, as evidenced by their comparatively shorter path lengths. However, it has been determined that there is no substantial difference in velocity or jerk between the groups. One may argue that the dissimilarity in movement patterns between simulations employing actual graspers and our VR implementation utilizing generic hand trackers is attributable to various factors. Nevertheless, the unconstrained nature of our implementation in terms of hand movement presents novel possibilities for the utilization of alternative methods of motion analysis in the evaluation of skill proficiency. It is contended that the time series nature of the motion data can be examined through the application of time series analysis methodologies. Our investigation demonstrated the presence of outliers in the dataset can be utilized to assess the skill level of the user, which aligns with the findings of Uemura et al [58]. The application of motion data allows the evaluation of the trainee's skills in a more comprehensive and automated fashion within the simulations. The incorporation of an autonomous training capability, which enables the user to get guidance without the need for human intervention, could potentially prove to be a valuable attribute.

In the conducted study, a total of 15 participants were involved in playing the second version of the game while their ECG recordings were obtained. The results of the study did not reveal any comparable instances where the upper and lower halves of the participant groups exhibited notable disparities in terms of heart rate (HR) or heart rate variability (HRV) features, including mean time or standard deviation between RR peaks. Nevertheless, throughout the transition from a state of rest to a state of training, there were notable increases in heart rate (HR) and corresponding decreases in heart rate variability (HRV). The observed elevation in heart rate can be understood from two perspectives: as a physiological response to stress or as a result of increased physical activity. The observed decline in heart rate variability (HRV) aligns with existing previous literature that has examined the association between mental workload and reductions in HRV [64], [84].

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

Virtual reality-based psychomotor training exhibits significant promise across various domains, notably within the realm of surgical education and training. Proficiency in the use of surgical tools and observing through a two-dimensional camera, is essential for performing laparoscopic procedures, in addition to the requisite expertise and precision required for all surgical interventions. Among the myriad challenges inherent in performing surgical procedures on living beings, the commendable dedication exhibited by surgeons in investing countless hours in training to enhance the quality of patient care is truly formidable.

This study introduces a novel approach to enhance the initial phases of laparoscopy training, focusing on the acquisition of psychomotor abilities, particularly dexterity, and precision with a VR adaptation of an existing method of training. The method addresses the challenges associated with operating through a two-dimensional screen and utilizing delicate instruments, while also aiming to achieve proficiency without relying on costly tools and software. In order to provide a thorough understanding, our study examined the process of adapting the proven technique of box trainers in laparoscopy to an immersive virtual world with multiple perspectives. This adaptation was achieved by utilizing common VR equipment without the need for any particular physical gadgets.

In order to address the obstacles and potential drawbacks encountered during the development phase, we executed two iterations and integrated quantitative data with commonly utilized questionnaires within the HCI and VR assessment, as well as feedback from subject matter experts. Following the identification of issues in the initial version, a second iteration was undertaken, wherein the same questionnaires were utilized to assess usability, acceptance, and overall experience. Instead of proposing a novel approach to surgical training, our study centered on the integration of additional benefits and positive aspects of VR training methods. Specifically, we demonstrated two key findings: (1) The successful implementation of motion capture in a VR training application and its utilization for analysis, and (2) the extent to which physiological responses exhibit similarities to established patterns in mental load. Furthermore, it was demonstrated that the application produced by our team achieved a favorable rating in terms of usability, performed at an above-average level in terms of acceptability, and delivered a satisfactory user experience based on self-assessed evaluations.

As elucidated in the preceding literature on psychomotor training, the enhancement of motor abilities necessitates engagement in deliberate and consistent practice. Further investigation should be conducted to assess the efficacy of our application on the acquisition of motor skills through the execution of controlled trials. Furthermore, it is imperative to conduct a thorough examination of the motion analysis techniques employed, namely by conducting experiments with cohorts of skilled surgeons and novices in order to validate the outcomes reported in prior research [34], [36], [38]. Moreover, the integration of motion analysis and physiological signals into VR for laparoscopic surgical training has the potential for further utilization through the implementation of individualized feedback based on the trainee's skill level and physiological state.

REFERENCES

[1] S. V. Kotsis and K. C. Chung, "Application of See One, Do One, Teach One Concept in Surgical Training," *Plast Reconstr Surg*, vol. 131, no. 5, p. 1194, May 2013, doi: 10.1097/PRS.0B013E318287A0B3.

[2] A. Herrmann-Werner *et al.*, "'Best Practice' Skills Lab Training vs. a 'see one, do one' Approach in Undergraduate Medical Education: An RCT on Students' Long-Term Ability to Perform Procedural Clinical Skills," *PLoS One*, vol. 8, no. 9, p. e76354, Sep. 2013, doi: 10.1371/JOURNAL.PONE.0076354.

[3] M. E. Hogg *et al.*, "Training in Minimally Invasive Pancreatic Resections: a paradigm shift away from 'See one, Do one, Teach one,'" *HPB*, vol. 19, no. 3, pp. 234–245, Mar. 2017, doi: 10.1016/J.HPB.2017.01.016.

[4] D. M. Gaba, "The future vision of simulation in health care," *BMJ Qual Saf*, vol. 13, no. suppl 1, pp. i2–i10, Oct. 2004, doi: 10.1136/QSHC.2004.009878.

[5] J. Kalvach, O. Ryska, and M. Ryska, "[Existing laparoscopic simulators and their benefit for the surgeon].," *Rozhl Chir*, vol. 95, no. 1, pp. 4–12, Jan. 2016, Accessed: Aug. 08, 2023.

[6] I. Choy, S. Kitto, N. Adu-Aryee, and A. Okrainec, "Barriers to the uptake of laparoscopic surgery in a lower-middle-income country," *Surg Endosc*, vol. 27, no. 11, pp. 4009–4015, 2013, doi: 10.1007/s00464-013-3019-z.

[7] E. Wilkinson, N. Aruparayil, J. Gnanaraj, J. Brown, and D. Jayne, "Barriers to training in laparoscopic surgery in low- and middle-income countries: A systematic review," *Trop Doct*, vol. 51, no. 3, pp. 408–414, Jul. 2021, doi: 10.1177/0049475521998186

[8] B. Zendejas, R. Brydges, S. J. Hamstra, and D. A. Cook, "State of the evidence on simulation-based training for laparoscopic surgery: A systematic review," *Ann Surg*, vol. 257, no. 4, pp. 586–593, Apr. 2013, doi: 10.1097/SLA.0B013E318288C40B.

[9] B. M. Kyaw *et al.*, "Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration," *J Med Internet Res 2019;21(1):e12959 https://www.jmir.org/2019/1/e12959*, vol. 21, no. 1, p. e12959, Jan. 2019, doi: 10.2196/12959.

[10] J. Pottle, "Virtual reality and the transformation of medical education," *Future Healthc J*, vol. 6, no. 3, p. 181, Oct. 2019, doi: 10.7861/FHJ.2019-0036.

[11] F. Mantovani, G. Castelnuovo, A. Gaggioli, and G. Riva, "Virtual Reality Training for Health-Care Professionals," *http://www.liebertpub.com/cpb*, vol. 6, no. 4, pp. 389–395, Jul. 2004, doi: 10.1089/109493103322278772.

[12] W. R. Sherman and A. B. Craig, *Understanding virtual reality: Interface, application, and design*. Morgan Kaufmann, 2018.

[13] A. R. Luft and M. M. Buitrago, "Stages of motor skill learning," *Mol Neurobiol*, vol. 32, no. 3, pp. 205–216, 2005, doi: 10.1385/MN:32:3:205.

[14] Dave R. H., "Psychomotor levels," in *Developing and writing behavioral objectives*, R. J. Armstrong.Arizona: Educational Innovators Press, 1970.

[15] B. S. Bloom, *Taxonomy of educational objectives : the classification of educational goals, handbook I : cognitive domain*, 1st ed. Longman Group, 1956. Accessed: Aug. 13, 2023.

[16] D. r Krathwohl, B. B. Masia, and B. S. Bloom, *Taxonomy of educational objectives : the classification of educational goals, handbook II : affective domain*. 1964. Accessed: Aug. 13, 2023.

[17] S. R. Chaiken, P. C. Kyllonen, and W. C. Tirre, "Organization and Components of Psychomotor Ability," *Cogn Psychol*, vol. 40, no. 3, pp. 198–226, May 2000, doi: 10.1006/COGP.1999.0729.

[18] E. J. Simpson, *The classification of educational objectives in the psychomotor domain*, vol. 3. Washington, DC: Gryphon House, 1972.

[19] D. R. Krathwohl, "A Revision of Bloom's Taxonomy: An Overview," *Theory Pract*, vol. 41, no. 4, pp. 212–218, Nov. 2002, doi: 10.1207/s15430421tip4104_2.

[20] W. S. Haisted, J. L. Cameron, and F. R. C. S. Eng, "William Stewart Halsted. Our surgical heritage.," *Ann Surg*, vol. 225, no. 5, p. 445, 1997, doi: 10.1097/00000658-199705000-00002.

[21] S. V. Kotsis and K. C. Chung, "Application of See One, Do One, Teach One Concept in Surgical Training," *Plast Reconstr Surg*, vol. 131, no. 5, p. 1194, May 2013, doi: 10.1097/PRS.0B013E318287A0B3.

[22] A. Herrmann-Werner *et al.*, "'Best Practice' Skills Lab Training vs. a 'see one, do one' Approach in Undergraduate Medical Education: An RCT on Students' Long-Term Ability to Perform Procedural Clinical Skills," *PLoS One*, vol. 8, no. 9, p. e76354, Sep. 2013, doi: 10.1371/JOURNAL.PONE.0076354.

[23] and M. S. D. Linda T. Kohn, Janet M. Corrigan, *To Err Is Human. Building a Safer Health System*, vol. 6, no. 3. Washington: National Academies Press, 2000. Accessed: Aug. 08, 2023. [Online].

[24] W. T. M. Mason and P. W. Strike, "Short CommunicationSee one, do one, teach one—is this still how it works? A comparison of the medical and nursing professions in the teaching of practical procedures," *https://doi.org/10.1080/01421590310001605705*, vol. 25, no. 6, pp. 664–666, Nov. 2009, doi: 10.1080/01421590310001605705.

[25] R. Aggarwal *et al.*, "Training and simulation for patient safety," *BMJ Qual Saf*, vol. 19, no. Suppl 2, pp. i34–i43, Aug. 2010, doi: 10.1136/QSHC.2009.038562.

[26] M. M. Hammoud *et al.*, "To the point: medical education review of the role of simulators in surgical training," *Am J Obstet Gynecol*, vol. 199, no. 4, pp. 338–343, Oct. 2008, doi: 10.1016/J.AJOG.2008.05.002.

[27] A. Dhariwal, R. Prabhu, A. Dalvi, and A. Supe, "Effectiveness of box trainers in laparoscopic training," *J Minim Access Surg*, vol. 3, no. 2, pp. 57–63, Apr. 2007, doi: 10.4103/0972-9941.33274.

[28] R. L. Coleman and C. Y. Muller, "Effects of a laboratory-based skills curriculum on laparoscopic proficiency: A randomized trial," *Am J Obstet Gynecol*, vol. 186, no. 4, pp. 836–842, Apr. 2002, doi: 10.1067/MOB.2002.121254.

[29] D. J. Scott *et al.*, "Laparoscopic training on bench models: better and more cost effective than operating room experience?," *J Am Coll Surg*, vol. 191, no. 3, pp. 272–283, Sep. 2000, doi: 10.1016/S1072-7515(00)00339-2.

[30] H. G. Guedes, Z. M. Câmara Costa Ferreira, L. Ribeiro de Sousa Leão, E. F. Souza Montero, J. P. Otoch, and E. Luiz de Almeida Artifon, "Virtual reality simulator versus box-trainer to teach minimally invasive procedures: A meta-analysis," *International Journal of Surgery*, vol. 61, pp. 60–68, Jan. 2019, doi: 10.1016/J.IJSU.2018.12.001.

[31] N. E. Seymour *et al.*, "Virtual Reality Training Improves Operating Room Performance," *Ann Surg*, vol. 236, no. 4, pp. 458–464, Oct. 2002, doi: 10.1097/00000658-200210000-00008.

[32] A. G. Gallagher *et al.*, "Virtual Reality Simulation for the Operating Room," *Ann Surg*, vol. 241, no. 2, pp. 364–372, Feb. 2005, doi: 10.1097/01.sla.0000151982.85062.80.

[33] M. S. Wilson, A. Middlebrook, C. Sutton, R. Stone, and R. F. McCloy, “MIST VR: a virtual reality trainer for laparoscopic surgery assesses performance.,” *Ann R Coll Surg Engl*, vol. 79, no. 6, pp. 403–4, Nov. 1997.

[34] K. R. Van Sickle, D. A. McClusky, A. G. Gallagher, and C. D. Smith, “Construct validation of the ProMIS simulator using a novel laparoscopic suturing task,” *Surgical Endoscopy and Other Interventional Techniques*, vol. 19, no. 9, pp. 1227–1231, Sep. 2005, doi: 10.1007/S00464-004-8274-6.

[35] A. Larsson, “An Open and Flexible Framework for Computer Aided Surgical Training,” *Stud Health Technol Inform*, vol. 81, pp. 263–265, 2001, doi: 10.3233/978-1-60750-925-7-263.

[36] E. M. McDougall *et al.*, “Construct Validity Testing of a Laparoscopic Surgical Simulator,” *J Am Coll Surg*, vol. 202, no. 5, pp. 779–787, May 2006, doi: 10.1016/J.JAMCOLLSURG.2006.01.004.

[37] T. Fritz, N. Stachel, and B. J. Braun, “Evidence in surgical training- A review,” *Innov Surg Sci*, vol. 4, no. 1, pp. 7–13, Mar. 2019, doi: 10.1515/ISS-2018-0026.

[38] A. Chaudhry, C. Sutton, J. Wood, R. Stone, and R. McCloy, “Learning rate for laparoscopic surgical skills on MIST VR, a virtual reality simulator: quality of human-computer interface.,” *Ann R Coll Surg Engl*, vol. 81, no. 4, p. 281, Jul. 1999, Accessed: Aug. 13, 2023. [Online].

[39] A. G. Gallagher and R. M. Satava, “Virtual reality as a metric for the assessment of laparoscopic psychomotor skills: Learning curves and reliability measures,” *Surgical Endoscopy and Other Interventional Techniques*, vol. 16, no. 12, pp. 1746–1752, Dec. 2002, doi: 10.1007/S00464-001-8215-6.

[40] J. D. Mason, J. Ansell, N. Warren, and J. Torkington, “Is motion analysis a valid tool for assessing laparoscopic skill?,” *Surg Endosc*, vol. 27, no. 5, pp. 1468–1477, Dec. 2013, doi: 10.1007/S00464-012-2631-7.

[41] S. Lim, S. Ghosh, P. Niklewski, and S. Roy, “Laparoscopic suturing as a barrier to broader adoption of laparoscopic surgery,” *Journal of the Society of Laparoendoscopic Surgeons*, vol. 21, no. 3. Society of Laparoendoscopic Surgeons, Jul. 01, 2017. doi: 10.4293/JSLS.2017.00021.

[42] D. J. Scott, W. N. Young, S. T. Tesfay, W. H. Frawley, R. V. Rege, and D. B. Jones, “Laparoscopic skills training,” *The American Journal of Surgery*, vol. 182, no. 2, pp. 137–142, Aug. 2001, doi: 10.1016/S0002-9610(01)00669-9.

[43] M. R. Ali, Y. Mowery, B. Kaplan, and E. J. DeMaria, “Training the novice in laparoscopy: More challenge is better,” *Surgical Endoscopy and Other Interventional Techniques*, vol. 16, no. 12, pp. 1732–1736, Dec. 2002, doi: 10.1007/S00464-002-8850-6.

[44] M. Nagendran, K. S. Gurusamy, R. Aggarwal, M. Loizidou, and B. R. Davidson, “Virtual reality training for surgical trainees in laparoscopic

surgery," *Cochrane Database of Systematic Reviews*, vol. 2013, no. 8. John Wiley and Sons Ltd, Aug. 27, 2013. doi: 10.1002/14651858.CD006575.pub3.

[45] A. M. Derossis, G. M. Fried, M. Abrahamowicz, H. H. Sigman, J. S. Barkun, and J. L. Meakins, "Development of a Model for Training and Evaluation of Laparoscopic Skills," *The American Journal of Surgery*, vol. 175, no. 6, pp. 482–487, Jun. 1998, doi: 10.1016/S0002-9610(98)00080-4.

[46] R. Aggarwal, T. Grantcharov, K. Moorthy, J. Hance, and A. Darzi, "A competency-based virtual reality training curriculum for the acquisition of laparoscopic psychomotor skill," *The American Journal of Surgery*, vol. 191, no. 1, pp. 128–133, Jan. 2006, doi: 10.1016/J.AMJSURG.2005.10.014.

[47] R. Aggarwal, P. Crochet, A. Dias, A. Misra, P. Ziprin, and A. Darzi, "Development of a virtual reality training curriculum for laparoscopic cholecystectomy," *British Journal of Surgery*, vol. 96, no. 9, pp. 1086–1093, Aug. 2009, doi: 10.1002/BJS.6679.

[48] G. M. Fried, "FLS assessment of competency using simulated laparoscopic tasks," *Journal of Gastrointestinal Surgery*, vol. 12, no. 2, pp. 210–212, Oct. 2008, doi: 10.1007/S11605-007-0355-0.

[49] M. C. Vassiliou, B. J. Dunkin, J. M. Marks, and G. M. Fried, "FLS and FES: Comprehensive models of training and assessment," *Surgical Clinics of North America*, vol. 90, no. 3, pp. 535–558, Jun. 2010, doi: 10.1016/j.suc.2010.02.012.

[50] M. C. Vassiliou *et al.*, "The MISTELS program to measure technical skill in laparoscopic surgery: Evidence for reliability," *Surgical Endoscopy and Other Interventional Techniques*, vol. 20, no. 5, pp. 744–747, May 2006, doi: 10.1007/S00464-005-3008-Y/TABLES/3.

[51] A. Chellali *et al.*, "Validation of the VBLaST peg transfer task: a first step toward an alternate training standard," *Surg Endosc*, vol. 28, no. 10, pp. 2856–2862, Oct. 2014, doi: 10.1007/S00464-014-3538-2.

[52] A. Buia, F. Stockhausen, N. Filmann, and E. Hanisch, "3D vs. 2D imaging in laparoscopic surgery—an advantage? Results of standardised black box training in laparoscopic surgery," *Langenbecks Arch Surg*, vol. 402, no. 1, pp. 167–171, Feb. 2017, doi: 10.1007/S00423-016-1526-9.

[53] R. Smith *et al.*, "Effect of passive polarizing three-dimensional displays on surgical performance for experienced laparoscopic surgeons," *British Journal of Surgery*, vol. 101, no. 11, pp. 1453–1459, Sep. 2014, doi: 10.1002/BJS.9601.

[54] J. A. Martin *et al.*, "Objective structured assessment of technical skill (OSATS) for surgical residents," *British Journal of Surgery*, vol. 84, no. 2, pp. 273–278, Feb. 1997, doi: 10.1046/J.1365-2168.1997.02502.X.

[55] V. Datta, A. Chang, S. Mackay, and A. Darzi, "The relationship between motion analysis and surgical technical assessments," *The American Journal of Surgery*, vol. 184, no. 1, pp. 70–73, Jul. 2002, doi: 10.1016/S0002-9610(02)00891-7.

[56] K. Ebina *et al.*, "Motion analysis for better understanding of psychomotor skills in laparoscopy: objective assessment-based simulation training using animal organs," *Surg Endosc*, vol. 35, no. 8, pp. 4399–4416, Aug. 2021, doi: 10.1007/s00464-020-07940-7.

[57] A. G. Gallagher *et al.*, "Psychomotor skills assessment in practicing surgeons experienced in performing advanced laparoscopic procedures," *J Am Coll Surg*, vol. 197, no. 3, pp. 479–488, Sep. 2003, doi: 10.1016/S1072-7515(03)00535-0.

[58] M. Uemura *et al.*, "Analysis of hand motion differentiates expert and novice surgeons," *Journal of Surgical Research*, vol. 188, no. 1, pp. 8–13, May 2014, doi: 10.1016/J.JSS.2013.12.009.

[59] Y. Y. Yurko, M. W. Scerbo, A. S. Prabhu, C. E. Acker, and D. Stefanidis, "Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool," *Simulation in Healthcare*, vol. 5, no. 5, pp. 267–271, Oct. 2010, doi: 10.1097/SIH.0B013E3181E3F329.

[60] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Advances in Psychology*, vol. 52, no. C, pp. 139–183, Jan. 1988, doi: 10.1016/S0166-4115(08)62386-9.

[61] S. Kaplan Berkaya, A. K. Uysal, E. Sora Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, "A survey on ECG analysis," *Biomed Signal Process Control*, vol. 43, pp. 216–235, May 2018, doi: 10.1016/J.BSPC.2018.03.003.

[62] A. Rieger, R. Stoll, S. Kreuzfeld, K. Behrens, and M. Weippert, "Heart rate and heart rate variability as indirect markers of surgeons' intraoperative stress," *Int Arch Occup Environ Health*, vol. 87, no. 2, pp. 165–174, Feb. 2014, doi: 10.1007/S00420-013-0847-Z.

[63] A. Omurtag, R. N. Roy, F. Dehais, L. Chatty, and M. Garbey, "Tracking Mental Workload by Multimodal Measurements in the Operating Room," *Neuroergonomics: The Brain at Work and in Everyday Life*, pp. 99–103, Jan. 2019, doi: 10.1016/B978-0-12-811926-6.00016-6.

[64] Z. Zakeri, N. Mansfield, C. Sunderland, and A. Omurtag, "Physiological correlates of cognitive load in laparoscopic surgery," *Sci Rep*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-69553-3.

[65] M. Sandelowski, "Sample size in qualitative research," *Res Nurs Health*, vol. 18, no. 2, pp. 179–183, Apr. 1995, doi: 10.1002/NUR.4770180211.

[66] J. Brooke, “SUS-A quick and dirty usability scale,” in *Usability Evaluation In Industry*, Patrick W. Jordan, B. Thomas, Ian Lyall McClelland, and Bernard Weerdmeester, Eds., CRC Press, 1996, pp. 189–196.

[67] J. Högberg, J. Hamari, and E. Wästlund, “Gameful Experience Questionnaire (GAMEFULQUEST): an instrument for measuring the perceived gamefulness of system use,” *User Model User-adapt Interact*, vol. 29, no. 3, pp. 619–660, Jul. 2019, doi: 10.1007/S11257-019-09223-W/TABLES/9.

[68] B. G. Witmer and M. J. Singer, “Measuring Presence in Virtual Environments: A Presence Questionnaire,” *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 3, pp. 225–240, Jun. 1998, doi: 10.1162/105474698565686.

[69] Patrícia Silva, “Davis’ technology acceptance model (TAM),” in *Information seeking behavior and technology adoption: Theories and trends*, 1989, pp. 205–219.

[70] A Bangor, P Kortum, and J Miller, “Determining what individual SUS scores mean: Adding an adjective rating scale.,” *J Usability Stud*, vol. 4, no. 3, pp. 114–123, 2009.

[71] F. D. Davis, “A technology acceptance model for empirically testing new end-user information systems : theory and results,” 1985.

[72] F. D. Davis, “Perceived usefulness, perceived ease of use, and user acceptance of information technology,” *MIS Q*, vol. 13, no. 3, pp. 319–339, 1989, doi: 10.2307/249008.

[73] V. Venkatesh and F. D. Davis, “A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies,” vol. 46, no. 2, pp. 186–204, Feb. 2000, doi: 10.1287/MNSC.46.2.186.11926.

[74] J. Nielsen, T. Clemmensen, and C. Yssing, “Getting access to what goes on in people’s heads? - Reflections on the think-aloud technique,” *ACM International Conference Proceeding Series*, vol. 31, pp. 101–110, 2002, doi: 10.1145/572020.572033.

[75] H. Kuusela and P. Paul, “A comparison of concurrent and retrospective verbal protocol analysis,” *American Journal of Psychology*, vol. 113, no. 3, pp. 387–404, 2000, doi: 10.2307/1423365.

[76] Tom Knoll, “The think-aloud protocol,” in *Games User Research*, Oxford University Press Oxford, 2018, pp. 189–202.

[77] J. Nielsen, “Estimating the number of subjects needed for a thinking aloud test,” *Int J Hum Comput Stud*, vol. 41, no. 3, pp. 385–397, Sep. 1994, doi: 10.1006/IJHC.1994.1065.

[78] B. G. Glaser and A. L. Strauss, *The Discovery of Grounded Theory*. Routledge, 1999. doi: 10.4324/9780203793206.

[79] E. Blair, "A reflexive exploration of two qualitative data coding techniques," *J Methods Meas Soc Sci*, vol. 6, no. 1, pp. 14–29, Jan. 2015, doi: 10.2458/V6I1.18772.

[80] D. Makowski *et al.*, "NeuroKit2: A Python toolbox for neurophysiological signal processing," *Behav Res Methods*, vol. 53, no. 4, pp. 1689–1696, Aug. 2021, doi: 10.3758/s13428-020-01516-y.

[81] M. Elgendi, M. Jonkman, and F. Deboer, "Frequency bands effects on QRS detection," *BIOSIGNALS 2010 - Proceedings of the 3rd International Conference on Bio-inpsired Systems and Signal Processing, Proceedings*, pp. 428–431, 2010, doi: 10.5220/0002742704280431.

[82] V. Kalidas and L. Tamil, "Real-time QRS detector using stationary wavelet transform for automated ECG analysis," *Proceedings - 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering, BIBE 2017*, vol. 2018-January, pp. 457–461, Jul. 2017, doi: 10.1109/BIBE.2017.00-12.

[83] M. S. R. Prasad, M. Manivannan, G. Manoharan, and S. M. Chandramohan, "Objective Assessment of Laparoscopic Force and Psychomotor Skills in a Novel Virtual Reality-Based Haptic Simulator," *J Surg Educ*, vol. 73, no. 5, pp. 858–869, Sep. 2016, doi: 10.1016/J.JSURG.2016.04.009.

[84] T. Zhou, J. S. Cha, G. Gonzalez, J. P. Wachs, C. P. Sundaram, and D. Yu, "Multimodal Physiological Signals for Workload Prediction in Robot-assisted Surgery," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 9, no. 2, pp. 1–26, Jan. 2020, doi: 10.1145/3368589.

APPENDICES

APPENDIX A

A.1 ETHICAL APPROVAL OF THE FIRST VERSION

Ethical Approval for the first version which does not include the physiological data collection.

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
ueam@metu.edu.tr
www.ueam.metu.edu.tr

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

Sayı: 28620816/ 01 ARALIK 2021

Konu : Değerlendirme Sonucu

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi : İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın Dr. Öğr. Üyesi Elif Sürer ve Dr. Öğr. Üyesi Hasan Onur Keleş

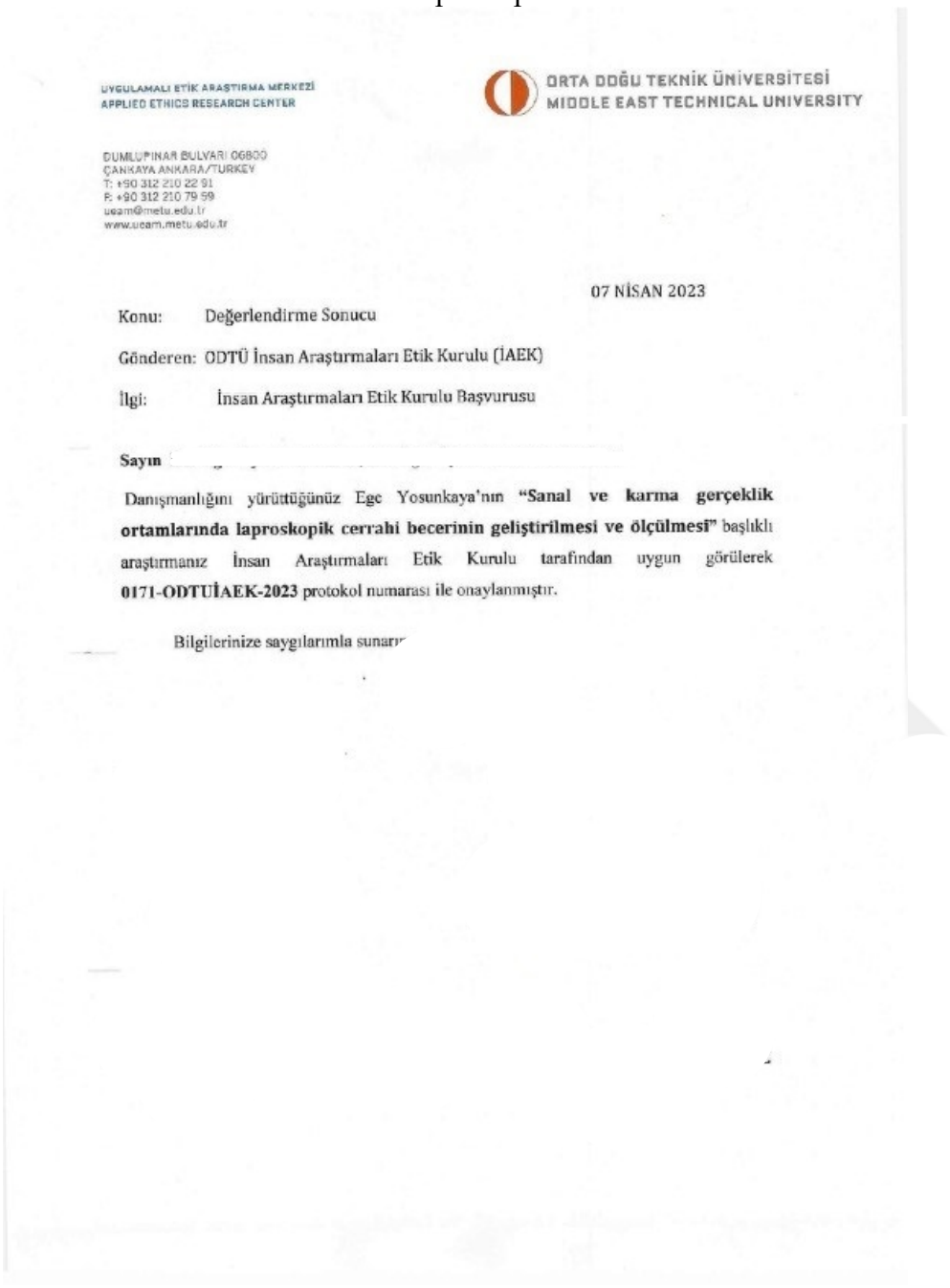
Danışmanlığını yürüttüğünüz Ege YOSUNKAYA'nın "Sanal ve karma gerçeklik ortamlarında laproskopik cerrahi becerinin geliştirilmesi ve ölçülmesi" başlıklı araştırması İnsan Araştırmaları Etik Kurulu tarafından uygun görülmüş ve **454-ODTU-2021** protokol numarası ile onaylanmıştır.

Saygılarımızla bilgilerinize sunarız.

Prof.Dr. Mine MISIRLISOY
İAEK Başkanı

A.2 ETHICAL APPROVAL OF THE SECOND VERSION

Ethical Approval of the second version with the permit to collect physiological data from the participants.



UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
ueam@metu.edu.tr
www.ueam.metu.edu.tr

07 NİSAN 2023

Konu: Değerlendirme Sonucu

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi: İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın

Danışmanlığını yürüttüğünüz Ege Yosunkaya'nın **"Sanal ve karma gerçeklik ortamlarında laproskopik cerrahi becerinin geliştirilmesi ve ölçülmesi"** başlıklı araştırmanız İnsan Araştırmaları Etik Kurulu tarafından uygun görülerek **0171-ODTÜİAEK-2023** protokol numarası ile onaylanmıştır.

Bilgilerinize saygılarımla sunarız

APPENDIX B

B.1 IN-GAME SCREEN SAMPLES FROM PEG TRANSFER TUTORIAL

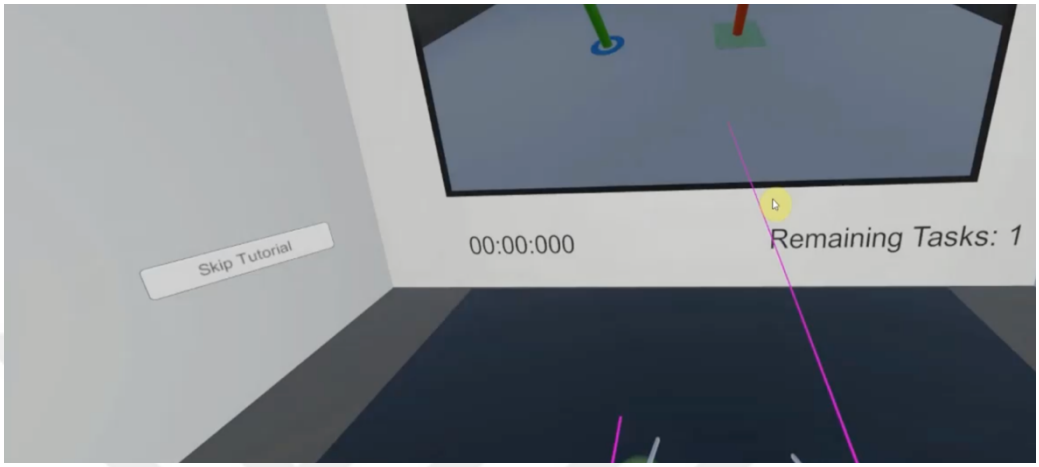


Figure 27: Skip the tutorial button located on the left of the box.

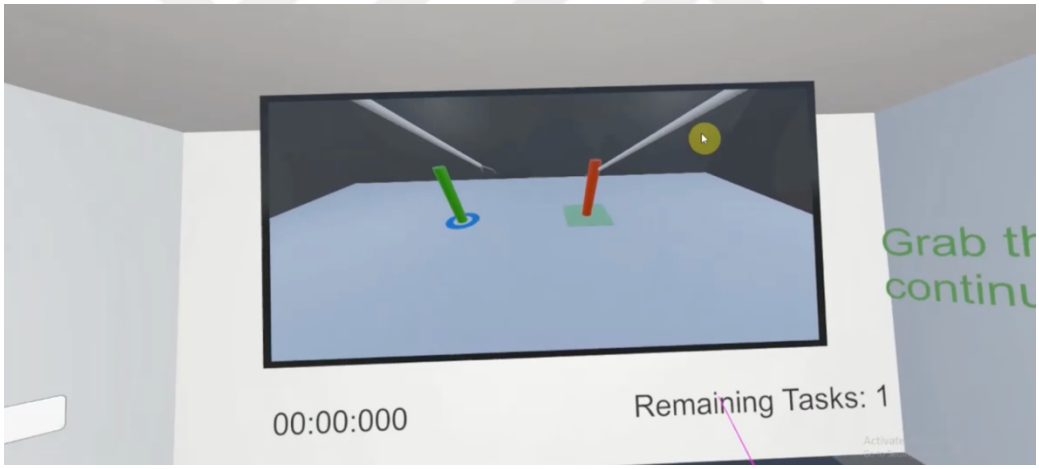


Figure 28: Timer and remaining tasks text components located between the box and the screen.

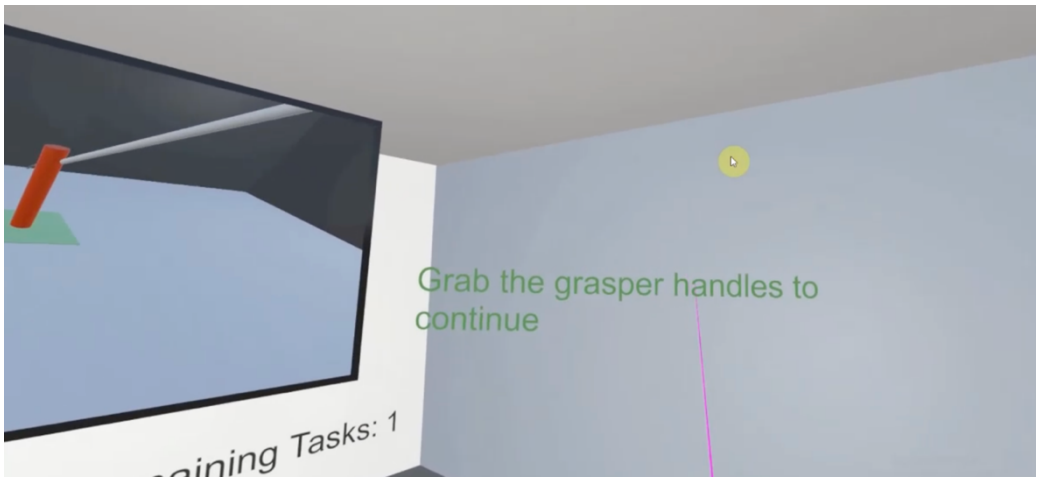


Figure 29: First step of the Peg Transfer tutorial explained with the on-screen text located on the right side of the screen.

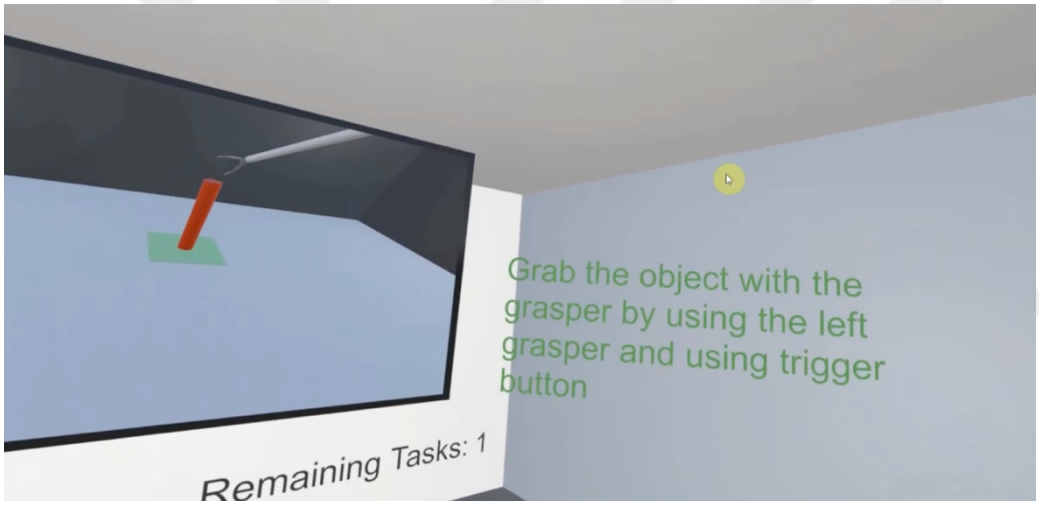


Figure 30: The second step of the Peg Transfer tutorial explained with the on-screen text located on the right side of the screen.

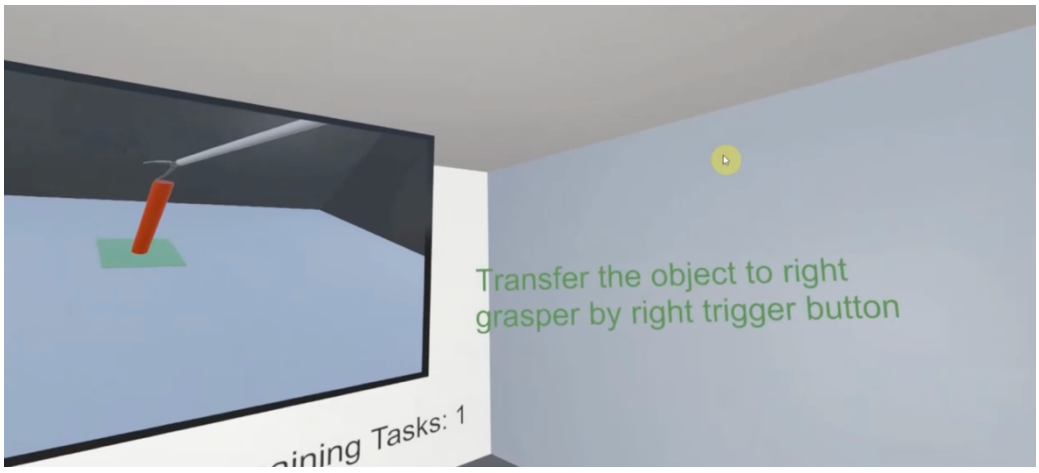


Figure 31: The third step of the Peg Transfer tutorial explained with the on-screen text located on the right side of the screen.

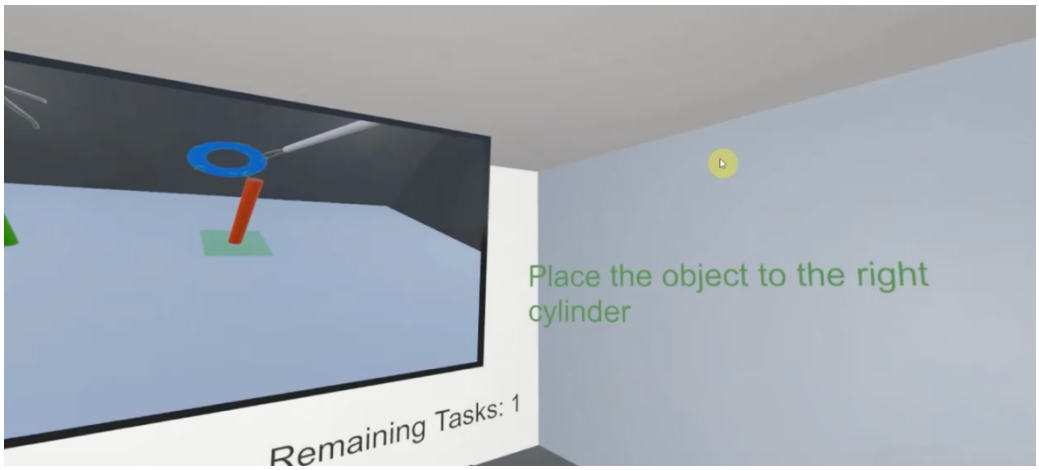


Figure 32: The last step of the tutorial is explained with the on-screen text located on the right side of the screen.



Figure 33: Finished tutorial screen for Peg Transfer tutorial with finish time and the button to continue.

B.2 IN-GAME SCREEN SAMPLES FROM THE STRING PASS TUTORIAL

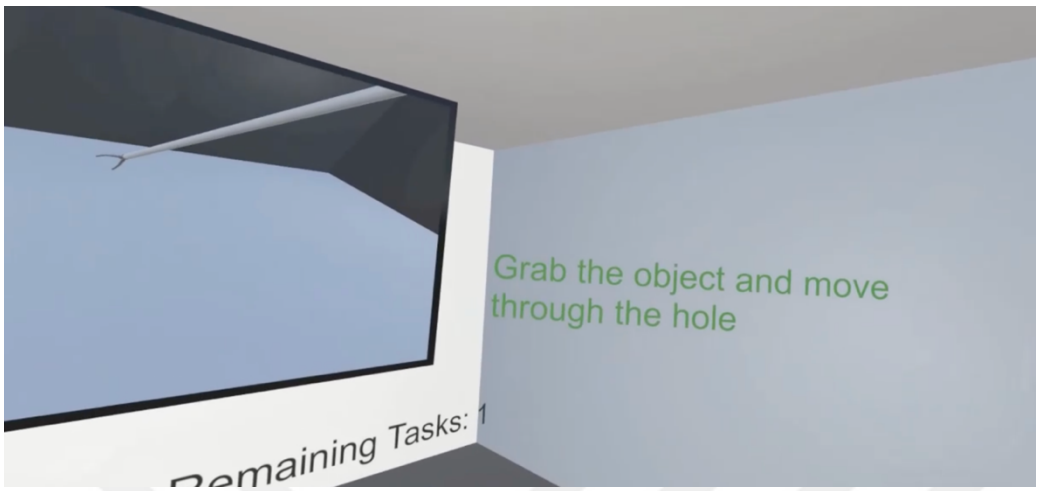


Figure 34: First step of the String Pass tutorial explained with the on-screen text located on the right side of the screen.

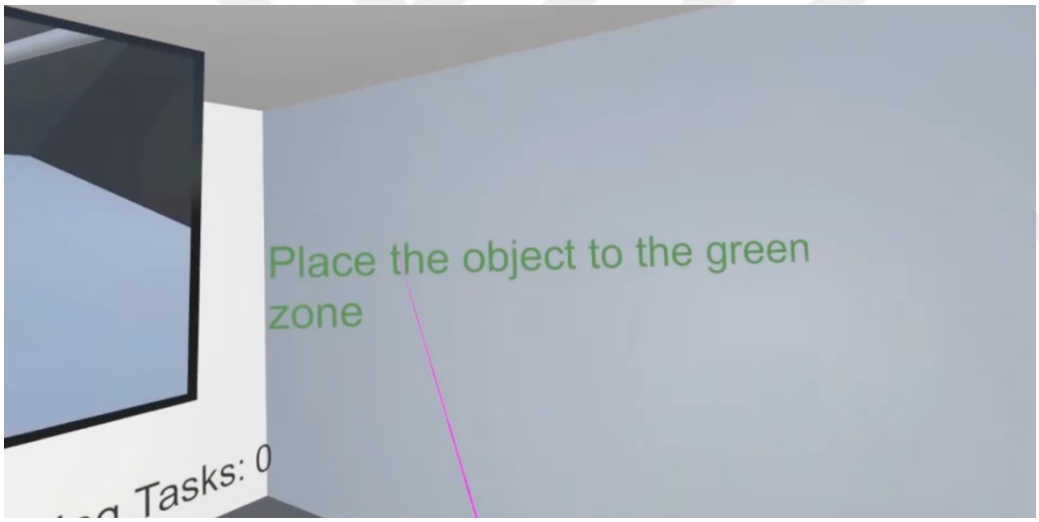


Figure 35: Second step of the String Pass tutorial explained with the on-screen text located on the right side of the screen.

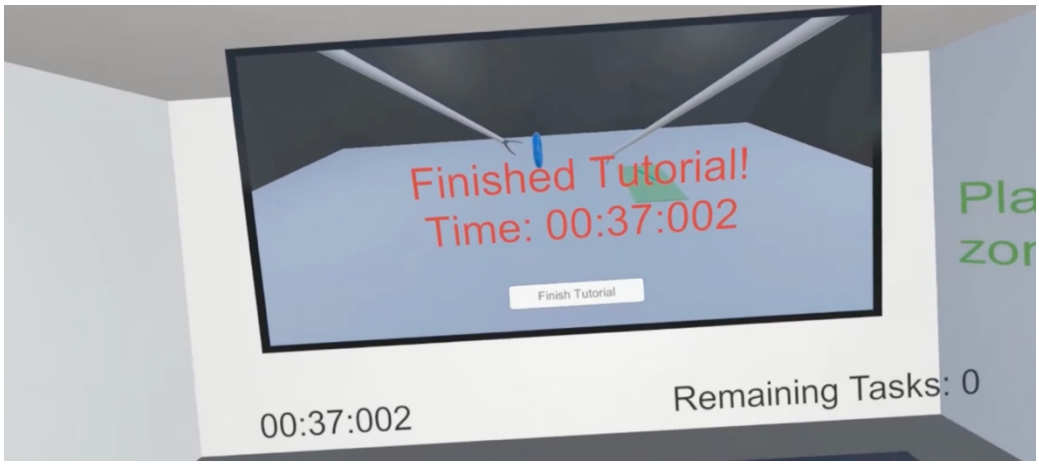


Figure 36: Finished tutorial screen for String Pass tutorial with finish time and the button to continue.