

University of Twente

BSc. Psychology

Bachelor Thesis – Human Factors and Engineering

Department of Cognitive Psychology and Ergonomics

Dr. Marleen Groenier

Dr. Simone Borsci

Usability assessment of mobile applications used for training surgical skills

The effect of domain expertise on user satisfaction

Finished: 20.08.2020

Melina Kowalski

S1940732

COVID-19 Disclaimer

This study was executed during the COVID-19 pandemic, which lead to adjustments for planning the data collection method. The study design had to be conducted remotely, to comply with social distancing measures, which is why personal contact between the researchers and the participants was not possible.

Abstract

Introduction: With the common use of minimally invasive surgery, traditional training approaches such as the master-apprenticeship model cannot fully account for the new skills needed (Darzi & Munz, 2004). Technical tools may train psychomotor skills without the risks of real life surgery (Darzi & Munz, 2004). Although professional simulators may have a higher accessibility and resemblance to real life surgery, they are costly. Therefore, employing simulations in smartphone applications could be a more cost-efficient alternative (Darzi & Munz, 2004).

Objectives: Systems should fulfil the criteria of effectiveness, efficiency and satisfaction (Borsci, Federici & Lauriola, 2009). To measure if domain expertise influences the satisfaction criterion the research question “How does the level of expertise affect the user satisfaction of a skill training application (SimuSurg) compared to a knowledge based application (Touch Surgery)” was asked.

Methods: The System Usability Score, the Net Promoter Score and an additional question concerning application preference were used to measure user satisfaction. All participants tested both applications. To measure the usability of the applications, tasks had to be completed remotely. To analyse the data regression analyses were run to determine the predictive value of domain expertise on user satisfaction.

Results: The results of the regression analyses showed that there is no significant effect of domain expertise on user satisfaction. Touch Surgery scores were higher on both, the SUS scores and NPS scores. However, SimuSurg was preferred over Touch Surgery by intermediates and experts. Therefore, Touch Surgery was rated as more usable, while SimuSurg scored higher on preference ratings.

Conclusion: Domain expertise had no significant effect on user satisfaction when comparing SimuSurg to Touch Surgery within this dataset. SimuSurg was preferred over Touch Surgery although the SUS and NPS scores indicated a higher usability for Touch Surgery. Comments by experts indicated that SimuSurg could be more valuable for teaching than Touch Surgery if the application would be improved, which could explain preference and usability discrepancies.

Contents

Introduction	- 5 -
Serious gaming	- 5 -
Training with smartphone applications	- 6 -
User Satisfaction.....	- 6 -
System usability scale and net promoter score	- 7 -
Research Question	- 8 -
Methods	- 8 -
Design	- 8 -
Participants	- 8 -
Materials	- 9 -
Tasks	- 12 -
Procedure	- 13 -
Data Analysis.....	- 16 -
Results	- 17 -
User Satisfaction.....	- 17 -
Regression Analysis	- 18 -
Discussion	- 20 -
Relationship of domain expertise on user satisfaction	- 20 -
Strengths and Limitations	- 22 -
Recommendations	- 22 -
Conclusion.....	- 23 -
References	- 24 -
Appendices	- 26 -

Introduction

Due to technological advances in the medical field that require different surgical techniques and decreasing opportunities to practice on patients, the demands for training aspiring surgeons have changed as well. Skills that were previously not as prominent and the amount of information that medical staff is required to master creates a high demand for applications that assist them in training. Since minimally invasive surgery (MIS) is one of the most commonly used ways to conduct surgery nowadays, traditional training methods may not fully account for the new skills surgeons need (Darzi & Munz, 2004).

MIS is a method in which the surgeon can remotely control the instruments through a small incision and views the process through a camera that is inserted as well, therefore not being in direct contact with the patient's body (Cao, MacKenzie, & Payandeh, 1996). Compared to traditional open surgery, MIS leaves the patient with reduced pain and cosmetically less visible wounds, which has changed the perception of operations from being considered a major risk that will affect the patient's life to a manageable procedure (Darzi & Munz, 2004). This process requires numerous psychomotor skills the surgeon needs to train, such as handling the instruments while viewing them through a camera, which changes the perspective (Gallagher & Smith, 2003). Concluding, there is a high demand for a training approach that accounts for a change in required skills and that reduces stress for surgeons in training.

Serious gaming

Although there are many benefits of MIS, the technical approach poses challenges for the surgeons. Especially, at the beginning of the laparoscopic approach, higher complication rates could be observed compared to open surgery (Darzi & Munz, 2004; Gallagher & Smith, 2003). The traditional master-apprentice model, in which an expert trains the surgeons through the monitoring of their progress cannot fully account for the more complex demands of MIS (Darzi & Munz, 2004). One problem of the master-apprentice model is that the training process is highly subjective, which makes comparisons between trainees and evaluations more difficult. Additionally, the surgeon is only exposed to training in actual practice within real-life surgery, which limits the amount of practice and variability of patients the trainees are exposed to (Darzi & Munz, 2004).

Simulator-based training could account for increased comparability, as training sessions can be standardized and learner progression can be monitored through the recorded data. Significant improvements in areas that pose a challenge for successfully conducting minimally invasive surgery have been found after trainees were exposed to simulators (Darzi & Munz,

2004). While using simulators to train prospective surgeons could be an effective training method, they can be expensive and possible exposure to them may be limited.

Simulations that resemble real-life situations, such as MIS surgery can often be found in serious games. Serious games are games with a serious purpose, meaning that the game is intended to provide the user with more than mere enjoyment, such as an educational purpose. Especially, in medical professions, serious games are used often (Gentry et al., 2019). Serious games can be developed on, for instance, smartphone applications, which makes them more accessible than traditional training methods or simulators. The increased accessibility is also due to cost efficiency. In the case of smartphone applications, the users only require a smartphone and often do not need to invest in the program themselves. The context of gaming can lead to more creative ways of teaching skills and knowledge (Gentry et al., 2019) and the games can be used for longer and more frequent training intervals than simulators that are located in educational institutions. In educational programs, the use of serious games also decreases the time needed by supervisors, lecturers, or other staff to monitor the training (Gentry et al., 2019). However, questions and discussions may not be resolved as easily. Especially, because problems cannot be solved immediately, the usability of serious games applications has to be sufficient to avoid performance or satisfaction problems.

Training with smartphone applications

One medium for training professionals and students in the medical field is with applications on mobile phone devices. Smartphones are owned by a large number of people, which makes them highly accessible (Lumsden, Byrne-Davis, Mooney, & Sandars, 2015). Additionally, users are usually comfortable with their devices, which means that the device itself does not have to be learned. Other benefits of using smartphones for training purposes in clinical medicine include an increased efficiency concerning time spent as well as being able to access sources that can be updated (Lumsden, Bryne-Davis, Mooney, & Sandars, 2015). Using interactive smartphone applications to train medical skills could improve the training process when compared to traditional methods such as viewing a surgical procedure (Sparrow, Liu, & Wegner, 2011).

User Satisfaction

Serious games and simulators may provide higher accessibility and cost-effectiveness, especially, when applied to smartphone applications. However, the possible benefits only apply if the system is usable. Usability defines how easy to use a system is, especially, when looking at its functions, the structure of the system, and the contents presented (Casaló, Flavián, & Guinalú, 2008). For instance, an application that users cannot find the desired functions in,

may not be considered usable. Usability also involves factors such as the degree of control users have about the system and that they know what exactly they are doing and in what part of the application they are being located in. Users should also be able to make the application function in a timely manner, meaning that the time spent on a task and looking for a task is kept to a minimum and that the actions needed to get somewhere are as quick as possible.

Borsci, Federici and Lauriola (2009) summarize that usability can be defined in three criteria: effectiveness, efficiency and satisfaction. The effectiveness is composed of the ability to complete a task and the quality of the result. Efficiency refers to the effort that has to be put in to perform a task and satisfaction covers the subjective opinion about the system.

System usability scale and net promoter score

Interactive smartphone applications could prove to be a helpful tool in assisting the training of medicine. However, to use these applications to train skills and relevant information to future surgeons, devices have to be usable. One of the main ways to measure the usability of an application is to assess user satisfaction (Borsci, Federici, Bacci, Gnaldi, & Bartolucci, 2015), which covers the satisfactory criterion of usability mentioned above. The system usability scale (SUS) is a standardized questionnaire, which is used in several publications, which makes it highly comparable (Lewis, & Sauro, 2018). Additionally, it is reliable and valid (Borsci, Federici, Bacci, Gnaldi, & Bartolucci, 2015).

The scale consists of ten items and participants are asked to indicate their opinion on a five-point scale (from strongly agree to strongly disagree). The result is either a unidimensional score that reflects the usability of the system or a bi-dimensional score (Borsci, Federici, Bacci, Gnaldi, & Bartolucci, 2015). To compute bi-dimensional scores, two out of the ten questions make up an own SUS score that refers to the learnability of the system, while the remaining questions represent the actual usability. Therefore, instead of one unidimensional score that summarises the overall usability, two scores (learnability and usability) are computed.

Another questionnaire that is frequently used is the net promoter score. The net promoter score is concerned with the appropriateness of an application (Sasmito, & Nishom, 2019). It is a one-item questionnaire in which the subject is asked how likely they are to recommend the system. Based on both questionnaires, it is possible to measure user satisfaction and more generally, the usability of an application, which ultimately reflects on if an application will be used frequently.

It has to be noted that there is a contextual overlap between the two questionnaires, as SUS scores were found to predict the NPS scores by 39% (Sauro, 2012).

Research Question

Interactive training on smartphone applications makes learning skills and information accessible and has the possibility of assisting students effectively. However, there may be differences between knowledge-based applications compared to ones focused on training psychomotor skills. As the target group for medical training apps is mainly students as their satisfaction with applications is decisive for the widespread use, it is important to assess how the level of expertise influences the user preferences.

Professionals, who already possess the skills necessary for surgery and have sufficient knowledge may not require such training devices. Additionally, lay people with little to no experience with medical training could lack basic skills and knowledge to profit from smartphone training applications. However, a comparison between different levels of expertise could show how already existing skills and knowledge affect user satisfaction. Therefore, the research question is “How does the level of expertise affect the user satisfaction measured with the SUS and NPS of a skill training application (SimuSurg) compared to a knowledge-based application (Touch Surgery)”.

In addition to the main research question, the secondary aim is to estimate to what extent the SUS scores predict the NPS scores for the two applications within this dataset.

Methods

Design

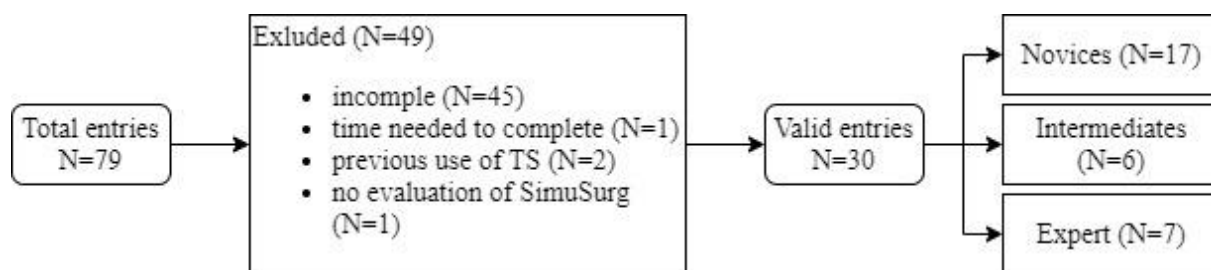
The study design was a within-subject, online-based usability test, which was conducted remotely due to the COVID-19 pandemic. The applications were randomised to avoid bias due to the order of the tasks. The aim was to assess how expertise levels affect the user satisfaction of a knowledge-based application (Touch Surgery) compared to a skill-based application (SimuSurg). The ethics committee of the University of Twente approved the design of the study (request no. 200884).

Participants

The total number of participants were 79, from which 49 data entries had to be removed (see Figure 1).

Figure 1

Participant flowchart



Divided by groups, there were 17 novices ($M_{\text{age}} = 21.8$, $SD = 2.31$), six intermediates ($M_{\text{age}} = 24$, $SD = 0.63$), and seven experts ($M_{\text{age}} = 36.14$, $SD = 7.4$).

Participants from the novice group were required to be active students with no medical background. The intermediate group's participants had to be active students with a minimum of one completed year in a medicine study programme. Furthermore, the experts were required to have at least three years of experience in a medical profession in which surgery was involved (e.g. surgeons, surgical assistants, nurses). Additionally, all participants were expected to have sufficient English proficiency and own a smartphone able to run both applications, as well as a personal computer or tablet.

The participants were gathered through convenience sampling. Participants' responses were excluded if the proof of completion through screenshots or eyewitness accounts could not be provided. Furthermore, the data was excluded if the questionnaire was incomplete or if they finished in less than twenty minutes. Due to the scope of the study, any time below twenty minutes was assumed to be unreasonable.

Materials

SimuSurg

SimuSurg was developed by the Royal Australasian College of Surgeons (Royal Australasian College of Surgeons, 2020.), as stated in the information section in the application, aimed at the familiarisation with instruments needed for MIS, movements and controlling the instruments and object manipulation, for instance, cutting and grasping. The application provided simulated situations that were divided into three levels, namely, beginner, intermediate, advanced, and expert. Each stage consisted of six levels with increased difficulty. In order to start a new level, the previous ones had to be completed. The levels were focused on using the tools in a three-dimensional space, including the camera perspective. The tools had to be operated with the thumbs. Every level had a specific task that needed to be completed, such as cutting or grasping. After a level was completed, the time and rating were shown to the user.

Touch Surgery

Touch Surgery was developed by Digital Surgery Ltd (Digital Surgery Ltd, 2020.), aimed at providing an accurate and valid training application to learn and rehearse surgical procedures. The application provided a number of procedures that were divided into phases. To complete a procedure, the user had to complete a learning phase in which the medical procedure was simulated, while written descriptions of the shown stages were given. Some procedures also required the user to swipe on the screen to forward the procedure. At the end of the learning phases, users were able to test the information they received in the test phases by answering multiple-choice questions.

Informed consent

An informed consent was used as an agreement of participants that they read the information brochure (see Appendix A), which described the procedure and the aim of the study and the participants' rights, as well as the researcher's contact information. Participants also received an invitation letter (see Appendix B), in which the requirements were listed.

Demographic questionnaire

In addition to the informed consent, a demographic questionnaire was included, in which information about age, gender, the current occupation, their smartphone model and the frequency of playing video and mobile games were to be answered (see Appendix C).

Instructions

Due to the outbreak of Covid-19, the study had to be conducted remotely. The study set-up included several stages, which is why a number of instruction sections (see Appendix D) had to be included to explain the details of the study as extensively as possible. Therefore, there were instructions about how to set-up the two applications Touch Surgery and SimuSurg, in which the installation process was explained. Furthermore, there were instructions on how to take a screenshot for the proof of completion and complementary instructions were given on how to upload the screenshots. Additionally, a description of how to uninstall the applications and delete the Touch Surgery account was provided at the end as well.

System usability scale and net promoter score

The questionnaires that followed the assessment of the Touch Surgery and the SimuSurg applications were the System Usability Scale and the Net Promoter Scale. The System Usability

Scale is a ten-item questionnaire that was asked after the users had to interact with each application. It has to be noted that one word in the System Usability Scale was changed. The word “system” was replaced by “application” to make the questionnaire easier to understand by lay people, who may not recognize the term “system” for an application. The following equation (1) was used to calculate the SUS scores:

$$\text{SUS Score} = [(SUS1 - 1) + (5 - SUS2) + (SUS3 - 1) + (5 - SUS4) + (SUS5 - 1) + (5 - SUS6) + (SUS7 - 1) + (5 - SUS8) + (SUS9 - 1) + (5 - SUS10)] * 2.5 \quad (1)$$

To interpret the SUS scores, the Curved Grading Scale (CGS) by Lewis and Sauro (2018) was used. The CGS was used to transform the SUS scores into grades ranging from A+ to an F, which made systems more comparable to others. The SUS scores were transformed into grades by assigning the SUS scores the following way:

- Grade F (0–51.7)
- Grade D (51.8–62.6)
- Grade C– (62.7–64.9)
- Grade C (65.0–71.0)
- Grade C+ (71.1–72.5)
- Grade B– (72.6–74.0)
- Grade B (74.1–77.1)
- Grade B+ (77.2–78.8)
- Grade A– (78.9–80.7)
- Grade A (80.8–84.0)
- Grade A+ (84.1–100)

The one-item Net Promoter Score questionnaire was also asked after each application was used. Participants rated the question on a scale from 0 (“not at all likely”) to 10 (extremely likely). The rating was then divided into 3 categories:

- Detractors (0-6)
- Passive (7-8)
- Promoters (9-10)

The following equation (2) was used to calculate the NPS score:

$$\text{NPS} = \text{Promoter\%} - \text{Detractor\%} \quad (2)$$

In addition to the SUS and the NPS, the participants answered which application they preferred to be able to compare the two applications directly. The participants were able to answer with “Touch Surgery”, “SimuSurg”, “Both”, and “None”. The last two options were included for participants with equal preferences. The goal was to see if they clearly preferred one application to the other. If none of the applications was of interest or if both applications were equally satisfying this was seen as valuable to this research as well.

Post-task questionnaire

In addition to the two standardized questionnaires used, there were a number of post-task questions that were asked as well. Participants had to indicate whether they had used the application before, if they completed all described tasks successfully and if they encountered any technical problems. In case the user wanted to give any additional comments, they were able to include them as well.

End-of-survey questionnaire

After both applications were tested, the end-of-survey questionnaire was presented to the participants. They were asked which application they preferred by indicating their preference for one of the two applications, none of the applications or both. Afterwards, they were able to give additional comments.

The study was conducted on the Qualtrics platform, which is where all questionnaires, instructions, and the file upload section for submitting the proof of completion were positioned. Participants were asked to directly upload their screenshots from their personal computers to the Qualtrics platform after they were requested to.

Tasks

The tasks themselves differed depending on the application (see Figures 2 and 3).

Figure 2

SimuSurg task flowchart

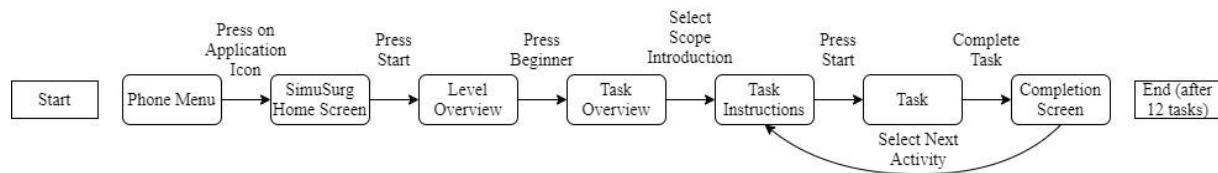
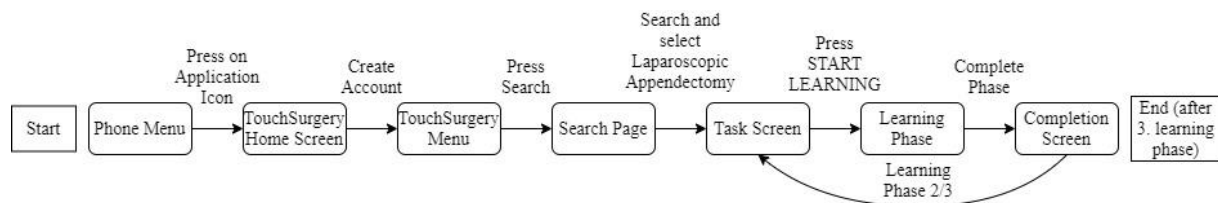


Figure 3

Touch Surgery task flowchart



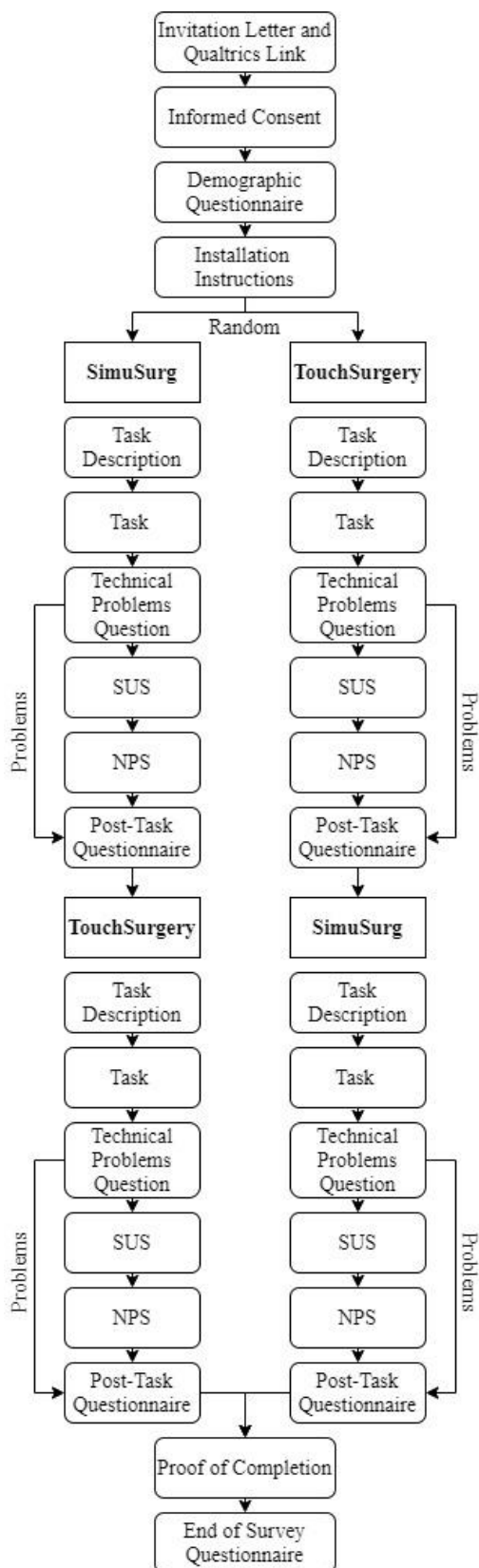
As mentioned above, the order in which the applications were shown was assigned randomly. Both applications, Touch Surgery and SimuSurg had a number of tasks that users needed to complete in order to indicate user satisfaction. The page in which the tasks were described started with the aim, followed by a description of which parts participants were required to complete and when to take a screenshot. Twelve tasks had to be completed in the SimuSurg application. The tasks included six tasks at the beginner level and six tasks at the intermediate level. Compared to the SimuSurg tasks, the Touch Surgery tasks were longer, which is why participants only needed to complete three tasks in the learning phase of the application. Therefore, the time spent on both applications was assumed to be similar.

Procedure

The researchers started by sending the invitation letter and the link to the Qualtrics questionnaire (see Figure 4).

Figure 4

Procedure flowchart



After reading the invitation letter and opening the link to the questionnaire, the participants read and agreed to the informed consent. The participants were then asked to fill out the demographic questionnaire, indicating their age, profession, frequency of playing video and mobile games,

etc. After continuing, they were presented with the instructions to download both apps. The instruction page included instructions for downloading SimuSurg and Touch Surgery from the Google Play Store and the App Store. The participants downloaded the applications.

The next page of the questionnaire was randomly assigned to participants. They either received the instructions for the tasks on SimuSurg or on Touch Surgery. The Touch Surgery page included the descriptions on how to create an account in the Touch Surgery application and which tasks were to be completed. The researchers also gave additional instructions on how to take screenshots on common smartphone models and gave an example screen of what the screenshot was supposed to look like. After reading the instructions for the tasks, the participants created the account, searched for the “Laparoscopic Appendectomy” task and completed the three learning phases. After completion, they took a screenshot of the screen at the end of the third learning phase and continued with the questionnaire.

Participants that were presented with the SimuSurg application first, received the same screenshot instructions and an example screenshot for this application. They read that they had to complete twelve tasks at the beginner and intermediate levels. After participants read the instructions, they continued with opening the app and completing the twelve tasks. For the proof of completion, they took a screenshot of the menu page, which showed that all tasks in the intermediate level were completed.

After the first application’s tasks were finished, participants had to answer whether they had experienced technical difficulties that made it impossible to finish the tasks in the application or if they indicated that they were not able to run the application, the questionnaire ended. If participants were able to run the application they continued with the System Usability Scale questionnaire, in which they were able to evaluate the usability of the application. Next, they were presented with the Net Promoter Score, which is one question where participants were asked how likely they are to recommend the application. The next page included several questions in which they were asked if there were any technical issues, if they were able to complete the tasks in thirty minutes or less and if they had used the application before. Additional comments on the application could be given on the same page as well.

After the post-task questionnaire was completed, participants were asked to complete the tasks of the second application. After reading the instructions and completing the tasks, they would take a screenshot and fill out the same questionnaire as for the first application. If both application’s tasks and the questionnaires were completed, participants were asked to upload their proof of completion. The upload page included instructions on how to transfer the screenshots from the phone to the personal computer and how to upload them to the Qualtrics

page. After both screenshots were successfully handed in, the end of survey questionnaire was presented to the participants, in which they indicated which application they preferred and if they had any additional comments.

The last instruction page that was presented was the instructions on how to uninstall the SimuSurg and Touch Surgery applications and how to delete the Touch Surgery account. Deleting and uninstalling the accounts were optional. The participants reached the last page that confirmed that they finished the survey after pressing continue and the response was recorded.

Data Analysis

Descriptive statistics were used to calculate the mean age, gender, SUS and NPS scores to summarize the results and estimate the average trend of the population. Furthermore, Cronbach's Alpha was computed on the SUS questionnaires to check the reliability of the data. To measure user satisfaction, the SUS and NPS scores were calculated for every group separately and interpreted with the help of the CGS (Sauro, & Lewis, 2018).

To answer the research question of how domain expertise affects user satisfaction of the two applications, linear regression analyses were computed for the SUS scores, the NPS scores and the question concerned with the preference between the applications. Therefore, the independent variable "group" was separately tested with the dependent variables "SUS score for SimuSurg", "SUS score for Touch Surgery", "NPS scores for SimuSurg", "NPS scores for Touch Surgery", and "application preference". Furthermore, the relation between the SUS and the NPS was explored. In addition to the exploration of the relationship between the SUS and the NPS, the data was explored with a combined group of intermediates and experts, to test whether experience within the medical field in general, would have an effect on user satisfaction. Additionally, the bi-dimensionality of the SUS scores was explored by calculating the SUS scores for the learnability items (questions 4 and 10) and the usability items (the remaining questions) separately.

The assumptions for linear regression analysis were tested before continuing with the analysis (see Appendix E). The linearity of the residuals was given for all continuous dependent variables. The assumption of having no auto-correlation was met for most variables. There is an auto-correlation indicated by the Durbin-Watson test measure of 0.6 for the SUS score variable for SimuSurg. As measures below one should be seen as concerning, any significant results including this variable have to be interpreted with caution. Concluding, a simple linear regression analysis could be used.

Results

User Satisfaction

SUS Scores

The SUS scores for SimuSurg and Touch Surgery show that the novice and the expert group perceived Touch Surgery as more usable (see Table 1). Solely the grades based on the intermediates ratings showed no difference if compared based on the CGS by Lewis and Sauro (2018). Although there were no differences in the grading of the two applications in the intermediate group, the mean score for Touch Surgery was higher than the average SUS for SimuSurg. Based on the grades and the mean scores, Touch Surgery can be considered as more usable.

The highest contrast between scores could be observed in the novice group, in which SimuSurg was rated a D, while Touch Surgery scored an A-. The expert group’s scores showed a similar variation. Although SUS scores between groups differed from each other, the conclusion concerning the overall usability based on SUS scores does not differ based on group differences.

Table 1

Group and Application SUS score means, standard deviations and grades

Application/ Group	SimuSurg			Touch Surgery		
	Mean	SD	Grade	Mean	SD	Grade
Novice	60.73	21.30	D	79.11	13.40	A-
Intermediate	67.50	14.57	C	70.41	13.36	C
Expert	66.42	18.64	C	77.50	14.93	B+

NPS scores and application preference

All NPS scores in all groups were negative, meaning that there were more detractors than promoters within all groups. This means that none of the groups would have recommended either of the two applications to a friend or a colleague (see Table 2). It can be noted that the highest NPS was found in the Novice group for the Touch Surgery application. While the NPS score for Touch Surgery in the Novice group is almost acceptable, the remaining scores cannot be seen as high enough to consider either application as acceptably usable. Especially, the NPS score of -71 for SimuSurg in the Expert group is an indication for serious usability issues within the Touch Surgery application.

Table 2

NPS scores for SimuSurg and Touch Surgery divided by groups

Application/ Group	NPS	
	SimuSurg	Touch Surgery
Novice	-59	-11
Intermediate	-50	-50
Expert	-71	-57

The results for the additional one-item question about the application preference were based on the four options SimuSurg, Touch Surgery, both, and none. Participants were given the option to answer both or none to account for equal preferences between the applications. The Novice group’s results showed that Touch Surgery was preferred over SimuSurg, which was in line with the SUS and the NPS results (see Table 3). However, both the intermediate and the expert group indicated a preference for SimuSurg over Touch Surgery, which contradicted the SUS and the NPS scores. The intermediate group rated both applications equally in the SUS and the NPS, while 66.7 per cent indicated a preference for SimuSurg over Touch Surgery. Additionally, the expert group’s results for this item showed a preference for SimuSurg over Touch Surgery, while they rated Touch Surgery more positively in the SUS and the NPS.

Table 3

Application preference results in percentages for SimuSurg, Touch Surgery, both applications and none of the applications divided by groups

Application/ Group	SimuSurg	Touch Surgery	Both	None
Novice	23.5%	58.8%	11.8%	5.9%
Intermediate	66.7%	33.3%	0.0%	0.0%
Expert	42.9%	28.6%	14.3%	14.3%

Regression Analysis

Predictive value of the influence of domain expertise on SUS, NPS, and application preference

A simple linear regression model was applied to test the predictive value of domain expertise on the SUS, the NPS and the question concerning the application preference. The regression analysis was executed on the three groups and the SUS and NPS scores for both applications separately, therefore running a total of five regression analyses. The results show that none of the analyses showed a significant predictive value (see Table 4).

Table 4

Regression Analysis results for the predictive value of groups on SimuSurg, Touch Surgery and Application Preference

Dependent / Independent	SimuSurg		Touch Surgery		Application Preference					
	SUS		NPS		SUS		NPS			
	Coeff.	Sig.	Coeff.	Sig.	Coeff.	Sig.	Coeff.	Sig.		
Group	3.22	.45	-.03	.95	-1.57	.61	-.72	-1.28	-.97	0.58

Exploration of dimensionality and dataset alterations

Based on the concept of the bi-dimensionality of the SUS questionnaire (Borsci, Federici, Bacci, Gnaldi, & Bartolucci, 2015) additional regression analyses were run. There was an indication of bi-dimensionality within this dataset (Amariei, 2020); however, the sample was found to be inadequate, which is why the bi-dimensionality of the SUS questions have to be interpreted with caution. To test whether domain expertise had an effect on user satisfaction of the two applications, the SUS total scores were divided into the variables “usability dimension” and “learnability dimension”. According to research conducted by Borsci et al. (2015), the learnability dimension consisted of items four and ten of the SUS questionnaire. Therefore, the remaining items measure the usability dimension. To separate the scores two separate SUS total scores were computed.

Furthermore, the intermediate and expert groups were combined to test whether any experience in the medical field would yield significant results. There was one significant effect between domain expertise and the usability dimension of the SUS in SimuSurg (see Table 5). An increase of the group score by one results in an increase of 2.55 of the total SUS score within the usability dimension. However, as mentioned, the results about the bi-dimensionality of the SUS questions within this dataset were inconclusive and the significant effect appeared in the SUS total score for SimuSurg, which was the dependent variable in which auto-correlation with a Durbin-Watson score of 0.6 was found. Therefore, it can be assumed that the data for this

variable were invalid, which is why the results of this regression analysis had to be interpreted with caution.

Table 5

Results of data exploration based on the assumption of bi-dimensionality

Dependent/Independent	SimuSurg				Touch Surgery			
	Usability dimension		Learnability dimension		Usability dimension		Learnability dimension	
	Coeff.	Sig.	Coeff.	Sig.	Coeff.	Sig.	Coeff.	Sig.
Group	2.55	.03	-.24	.69	-1.09	.25	-.52	.32

SUS and NPS

In addition to the regression analyses between groups and usability scores, a regression analysis was performed on the SUS and the NPS to test the predictive value of the SUS on the NPS. A significant regression equation was found for the SUS on the NPS for SimuSurg. In SimuSurg the NPS was equal to $-.286 + .085$ times SUS score. Additionally, a significant predictive value was found for the SUS on the NPS for Touch Surgery. Therefore, in Touch Surgery the NPS was equal to $-.005 + .079$ times SUS score. This finding confirms the past findings that the SUS score was able to predict the NPS score to some extent.

Discussion

Relationship of domain expertise on user satisfaction

To answer the research question of how domain expertise affects the user satisfaction of a skill training application (SimuSurg) compared to a knowledge-based application (Touch Surgery) the SUS scores were compared to the NPS scores. Additionally, a simple linear regression analysis was run to test whether domain expertise predicted the total scores for the SUS and the NPS.

The results of the regression analyses showed that the domain expertise was not able to predict the outcome of the SUS or the NPS questionnaire; therefore, it can be assumed that the expertise level had no significant effect on the user satisfaction in either of the two applications in this sample. Several explanations could account for domain expertise not having a predictive effect on user satisfaction. In a study in which insulin pumps' usability based on domain expertise was tested, the results showed that novices experienced difficulties based on knowledge deficits, while experts had to invest more effort into completing the tasks (Liu,

Osvalder & Karlsson, 2007). Furthermore, the researchers state that novices struggled to learn the new system, while experts had to adapt their mental models to the system that was inconsistent with the previously used ones. For this study, this could mean that usability ratings were similar, but the reasons for the similar scores may be different.

When comparing the SUS scores with the NPS scores, minor differences could be observed. Novices and experts rated Touch Surgery as more usable when looking at the SUS scores, while intermediates ratings were a “C” for both applications. The same conclusion can be drawn from the NPS scores, which indicates that Touch Surgery can be seen as more usable, however, a third question was used for direct comparison between the two applications. When asked about the preference between SimuSurg and Touch Surgery, only the novice group rated Touch Surgery as more preferable. Both intermediates and experts indicated a preference for SimuSurg over Touch Surgery.

These results are contrasting to the ones from the SUS and the NPS, which indicated that the preference question may not have solely measured the usability of the two applications. A possible explanation for these results can be found in the answers of participants in the additional comments. To summarize, the comments about Touch Surgery show that experts thought that the content was too easy for them, but that the application itself was more usable than SimuSurg. In general, there were multiple comments that were mentioning that SimuSurg does not represent real-life skills, but that the application would have a high potential for being useful in educational settings if the content was adjusted to resemble real-life surgery.

Therefore, a possible explanation of the preference question results is that the experts thought that SimuSurg was more entertaining to use, but did not represent realistic skills needed for training. Touch Surgery was more realistic and easier to use, however, they thought that it was too easy, which is why they preferred SimuSurg. Therefore, the preference question may have been aimed at measuring a combination of the effectiveness of the applications and their usability. Hendrix, Bellamy-Wood, McKay, Bloom, and Dunwell (2018) discuss the relationship between the difficulty and the player’s enjoyment. Accordingly, it is essential to match the challenge to the players’ experience level. As Touch Surgery was found to be too easy by a portion of the expert group, this could be an explanation for their preference for SimuSurg. However, the effectiveness of the educational aspect of SimuSurg and Touch Surgery is beyond the scope of the study.

Strengths and Limitations

Due to the outbreak of Covid-19, the study design had to be executed remotely. As a result, technical difficulties and problems with instruction formulations could not be corrected through a conversation with participants, which could be why the dropout rate was higher than expected as well. Additionally, some participants may not have been able to partake, for instance, because they did not have access to a smartphone that is connected to the Google Play Store or the App Store. Secondly, the proof of completion is not as reliable as direct observation of the participant's participation, which is why it is possible that participant's data were included although they did not complete the tasks.

Another limitation is that the group criteria may have included or excluded the wrong criteria. Technical medicine students of the University of Twente practice MIS skills on simulators that are designed for educational training. Therefore, the intermediates may have the same or more training for skills needed in MIS than participants from the expert group who have had little experience with surgery in the three or more years that they worked in a medical profession.

While the remote approach in this study could have led to a higher dropout rate, misunderstandings and decreased opportunities to partake in this study due to a lack of required personal devices, there are strengths as well. By collecting the data remotely, more possible participants were reached as the geographical limitations did not apply. Additionally, participants were able to partake while staying at home, therefore, reducing the time and effort needed to participate.

Recommendations

Based on the results of the preference question and the comments of the participants, future research may focus on redesigning SimuSurg to make the skill training more representative of real life. Additionally, it would be interesting to see the results of a study with a larger sample size or with different groups, such as students with and without simulator-based training and how those factors influence user satisfaction.

This study was aimed at examining the relationship between domain expertise and user satisfaction. However, satisfaction is only partly responsible for the overall usability of a system. Considering the usability criteria by Borsci, Federici and Lauriola (2009), effectiveness and efficiency should be considered as well. To measure the effectiveness of the two applications tested, the quality of the results should be considered, for instance, by looking at differences between the time needed to complete the task between groups. For the efficiency of

SimuSurg and Touch Surgery differences between the amount of effort different groups have to invest in reaching a goal could be estimated as well.

Conclusion

Based on the SUS and NPS scores of SimuSurg and Touch Surgery, it was found that domain expertise does not affect user satisfaction of a knowledge-based application compared to a skill training application, which is in line with studies that focused on predicting usability based on domain expertise in the medical field. However, a significant predictive effect of the SUS scores on NPS scores was found, meaning that the SUS scores were able to predict a significant amount of the NPS scores. Overall, Touch Surgery has a higher usability compared to SimuSurg, although both applications cannot be considered of high usability. By asking participants about their preference, between the two applications it was found that experts preferred SimuSurg over Touch Surgery, even if Touch Surgery's usability was rated higher, which could be due to a mismatched relationship between the difficulty of Touch Surgery's content to the expert's expertise.

References

- Amariei, A. L. (2020). *Usability assessment of medical training applications – Exploring the dimensionality of the System Usability Scale* [Unpublished Bachelor's thesis]. University of Twente.
- Borsci, S., Federici, S., Bacci, S., Gnaldi, M., & Bartolucci, F. (2015). Assessing user satisfaction in the era of user experience: Comparison of the SUS, UMUX, and UMUX-LITE as a function of product experience. *International Journal of Human-Computer Interaction*, 31(8), 484-495.
- Borsci, S., Federici, S., & Lauriola, M. (2009). On the dimensionality of the System Usability Scale: a test of alternative measurement models. *Cognitive processing*, 10(3), 193-197.
- Cao, C., MacKenzie, C., & Payandeh, S. (1996). Task and motion analyses in endoscopic surgery. In *Proceedings ASME Dynamic Systems and Control Division* (pp. 583-590).
- Casaló, L., Flavián, C., & Guinalú, M. (2008). The role of perceived usability, reputation, satisfaction and consumer familiarity on the website loyalty formation process. *Computers in Human behavior*, 24(2), 325-345.
- Darzi, A., & Munz, Y. (2004). The impact of minimally invasive surgical techniques. *Annual Review of Medicine*, 55(1), 223-237. doi: 10.1146/annurev.med.55.091902.105248
- Digital Surgery Ltd. (2020). *Touch Surgery* (6.20) [Mobile App]. App Store. <https://apps.apple.com/de/app/touch-surgery-surgical-videos/id509740792>
- Gallagher, A. G., & Smith, C. D. (2003). Human-factors lessons learned from the minimally invasive surgery revolution. In *Seminars in Laparoscopic Surgery* (Vol. 10, No. 3, pp. 127-139). Sage CA: Thousand Oaks, CA: Sage Publications.
- Gentry, S. V., Gauthier, A., Ehrstrom, B. L. E., Wortley, D., Lilienthal, A., Car, L. T., ... & Car, J. (2019). Serious gaming and gamification education in health professions: systematic review. *Journal of medical Internet research*, 21(3), e12994.

- Hendrix, M., Bellamy-Wood, T., McKay, S., Bloom, V., & Dunwell, I. (2018). Implementing adaptive game difficulty balancing in serious games. *IEEE Transactions on Games*.
- Liu, Y., Osvalder, A. L., & Karlsson, M. (2007). User's Expertise Differences When Interacting with Simple Medical User Interfaces. In *Symposium of the Austrian HCI and Usability Engineering Group* (pp. 441-446). Springer, Berlin, Heidelberg.
- Lumsden, C. J., Byrne-Davis, L. M. T., Mooney, J. S., & Sandars, J. (2015). Using mobile devices for teaching and learning in clinical medicine.
- Sasmito, G. W., & Nishom, M. (2019). Usability Testing based on System Usability Scale and Net Promoter Score. In *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* (pp. 540-545). IEEE.
- Sauro, J. (2012). Predicting net promoter scores from system usability scale scores. Retrieved July, 26, 2020.
- Lewis, J. R., & Sauro, J. (2018). Item benchmarks for the system usability scale. *Journal of Usability Studies*, 13(3).
- Royal Australasian College of Surgeons. (2020). *SimuSurg* [Mobile App]. App Store.
<https://apps.apple.com/de/app/simusurg/id1174517345>
- Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *science*, 333(6043), 776-778.

Appendices

Appendix A: Information brochure

This usability-test represents a part of the project “Usability assessment of mobile applications used for training surgical skills”. Your contribution will be used to evaluate two apps which are aiming to teach basic surgical skills. The goal of the study is to see how easy to use those apps are for different target groups. To achieve this, we want to receive some input from you, as an end-user. In this usability-test we will look at how you perceive the two apps and how you evaluate them. For us, the data which you will provide will be used in the writing process of our Bachelor’s Theses. The benefit for you is experiencing two apps through which you can train surgical skills and learn surgical procedures.

During this session, you will have to perform tasks and answer questions:

- Firstly, we will ask for background information;
- Secondly, the actual usability-test will start. You will have to complete tasks in both apps. After each task, you will have to answer questions and upload proof of completion;
- Thirdly, you will receive questions about the session.

Below you can find some information about your rights and about the way in which your information will be handled:

- This session will take approximately 45 minutes. There is a limit of 30 minutes to successfully complete a phase, after which you can abort it and mention it in the questionnaire.
- You are free to withdraw yourself from this study at any given time, without providing a reason.
- For validation purposes, we will ask you to make screenshots to prove that you completed the tasks and upload them in the received form. Those screenshots should not contain any information that could be used to identify yourself.
- Your answers will be anonymized, safely stored, and accessed just by the members of the research team. If you decide at a later date that you do not agree with your data being used in the study, you can contact one of the researchers and ask for your answers to be removed without providing a reason.
- The applications you are going to test might use your personal data (e.g. device information).
- The Touch Surgery application will require you to create an account.

- The Touch Surgery application uses realistic depictions of medical procedures. Those depictions might be disturbing. If you do not feel comfortable with those depictions, you are advised to stop using the app and inform one of the researchers.

If you need further information about the research, before, during, or after the session, you can contact one of the researchers:

- Alexandru-Lucian Amariei (e-mail: a.amariei@student.utwente.nl);
- Melina Marie Kowalski (e-mail: m.m.kowalski@student.utwente.nl);
- Christof Schulz (e-mail: c.schulz-2@student.utwente.nl).

If you have questions about your rights as a research participant or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee of the Faculty of Behavioural, Management and Social Sciences at the University of Twente by ethicscommittee-bms@utwente.nl.

Consent form

1. I have read and understood the study information dated 03/06/2020, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.
2. I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.
3. I understand that taking part in the study involves:
 - Providing some basic information about myself to the researchers' team;
 - Testing two applications for training surgical skills;
 - Completing and answering to the best of my ability to the questionnaires I will receive during the session;
 - The applications I am going to use might also make use of some of the information I provide (e.g. results of the simulation).
4. I understand that information I provide will be used as input for evaluating two medical training applications and subsequently writing reports (Bachelor's Theses) about them.

5. I understand that personal information collected about me that can identify me, such as my age, gender or profession, will be anonymized and not be shared beyond the study team.
6. I give permission for the answers in the questionnaires that I provide to be archived in University of Twente student theses repository, so it can be used for future research and learning.

Appendix B: Invitation letter

Dear [],

We are three students from the psychology program of the University of Twente, Melina, Christof and Alexandru, and we are currently doing our bachelor's theses. The aim of our project is to test the **usability of mobile applications (serious games) used for training surgical skills**. We will look at how you perceive two apps and how you evaluate them. The goal of the study is to see how easy to use two applications are for different target groups. To achieve this, we want to receive some input from you as an end-user.

For us, the data which you will provide will be used in the writing process of our bachelor's theses and to inform the educational program about the usefulness of these kind of apps, for example for Endoscopic Skills. The benefit for you is to experience two applications through which you can train your surgical skills and learn about surgical procedures.

To complete the study, please make sure that you have a mobile device (smartphone) and a desktop computer or laptop available. You will have to test the applications on your mobile device and fill out a survey on the PC/laptop. The study will take approximately 45 minutes. If you have any questions about participating in the study, do not hesitate to send us an email! Click on this link to participate in the study:

https://utwentebbs.eu.qualtrics.com/jfe/form/SV_39K1J0TeCFmUydT

Kindest regards,

The research team

Melina Kowalski, Christof Schulz, Alexandru Amariei

Appendix C: Demographic Questionnaire

1. How old are you?
2. What gender do you identify with?
 - Male
 - Female
 - Other
3. What is your current occupation?
 - Student (without a medical background)
 - Senior medical student (final year of studying in a medical related field)
 - Medical professional (e.g. surgeon or nurse, with at least 2 years of expertise in the field)
 - Other:
4. What smartphone, brand and model (e.g. iPhone X) are you going to use for testing the apps?
5. How often do you play mobile games (e.g. on smartphone or tablet)?
 - I regularly play video games (more than 3 times per week)
 - I sometimes play mobile games (1-3 times per week)
 - I rarely play mobile games (1-3 times per month)
 - I never play mobile games
6. How often do you play video games (e.g. on console or personal computer)?
 - I regularly play video games (more than 3 times per week)
 - I sometimes play video games (1-3 times per week)
 - I rarely play video games (1-3 times per month)
 - I never play video games

Appendix D: Instruction Pages

SimuSurg task description

This stage should take approximately 15 minutes. If you find yourself not able to successfully complete the task within 30 minutes, you can abort the task and mention it in the questionnaire.

Please read the instructions carefully and do not be afraid to take a second look in case you encounter a problem!

Task: Open the SimuSurg app. Press "**Start**". Now, press on "**Beginner**" and click on the first level named "**Scope introduction**". After looking at the instructions for the level, press "**Start**" once again. If you complete a level successfully, press "**Next activity**" and start the next level. Don't worry if you fail a level, you can simply re-try until you manage to solve it. **Please stop once you solved level no. 12, called "Irrigation Introduction" (in the "Intermediate stage")**. After completing the 12th level please take a screenshot.

Please do not forget to take a screenshot of the completion screen, after finishing the 12th level (Irrigation Introduction, in the Intermediate stage, seen in the bottom left corner of the screen). You can find instructions on how to do that below.

If you encounter a problem during this stage, please send an email to a.amariei@student.utwente.nl or a WhatsApp message at

After you completed the stage and answered the question at the bottom, you may proceed to the next section.

Touch Surgery task description

This stage should take approximately 15 minutes. If you find yourself not able to successfully complete the phase within 30 minutes, you can abort it and mention it in the questionnaire.

Please read the instructions carefully and do not be afraid to take a second look in case you encounter a problem!

Account set-up: To set up the account you will need to open the application and press on "**Create an account**". Fill in your email address and choose a password. Now you have to

tick the first box to agree to the EULA, terms of agreement and privacy policy. The second box has to be ticked as well, to confirm that you are at least 18 years old. Now that you accepted the two necessary requirements, you can click on "**Create Account**" again. Press the "Find your procedures" to continue. You are now asked to fill in your first and last name and press "**Confirm**". You should see a page that asks for your profession. There are several options given to you, but you may also press "other/none of the above" at the bottom if none of them apply to you. Now, you will be asked what your main interests are. You can choose whatever you like or select one at random if none of them appeal to you. Your choice will not influence this research. After you chose your interests, you will be asked to indicate your secondary interest. Again, you can choose what you like or select one at random. You should be seeing the home screen of the application now.

Task: On the bottom of the page, you should see multiple icons. Please press the magnifying glass at the bottom of the page. If you press the correct icon you should be on a page with the search function on the top. Type in "**Laparoscopic Appendectomy**" in the search field. You should see a task with that name in the search results. If you press the task you should see a page with the option "**START LEARNING**". There are three learning and three testing phases. Please only complete the three learning phases. When you press "**START LEARNING**", the first learning phase should start. After finishing it you will see the options to exit, proceed with learning phase 2, or with testing phase 1. Please select "**Learn Phase 2**". After completing the second phase, you will have to repeat the same procedure to advance to the last stage, namely press on "**Learn Phase 3**". After completing the third learning please, please take a screenshot.

Note: You do not have to complete the tests for each phase in this training course. We ask you to focus solely on the learning aspect of the course.

Please do not forget to take a screenshot of the completion screen, after finishing the 3rd learning phase (Appendectomy). You can find instructions on how to do that below.

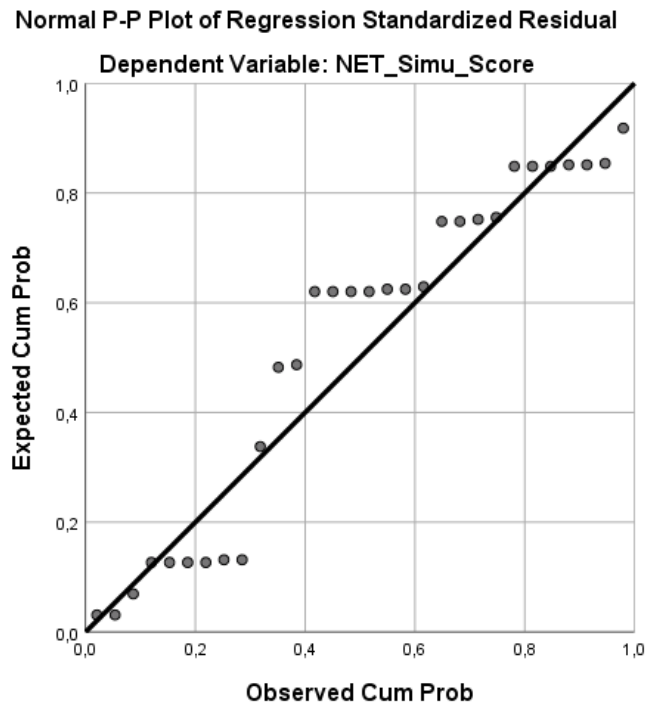
If you encounter a problem during this stage, please send an email to a.amariei@student.utwente.nl or a WhatsApp message at

After you completed this stage and answered the question at the bottom, you may proceed to the next section.

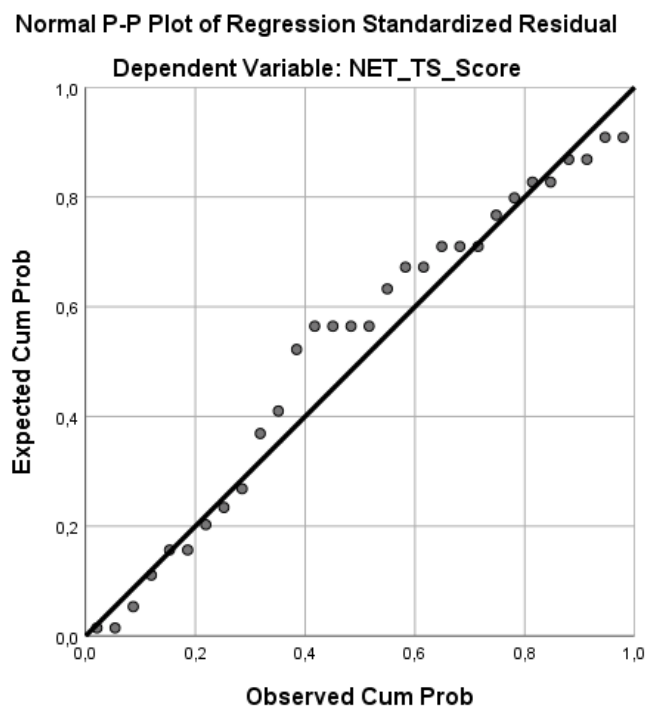
Appendix E: Assumption Testing

Linearity of Residuals

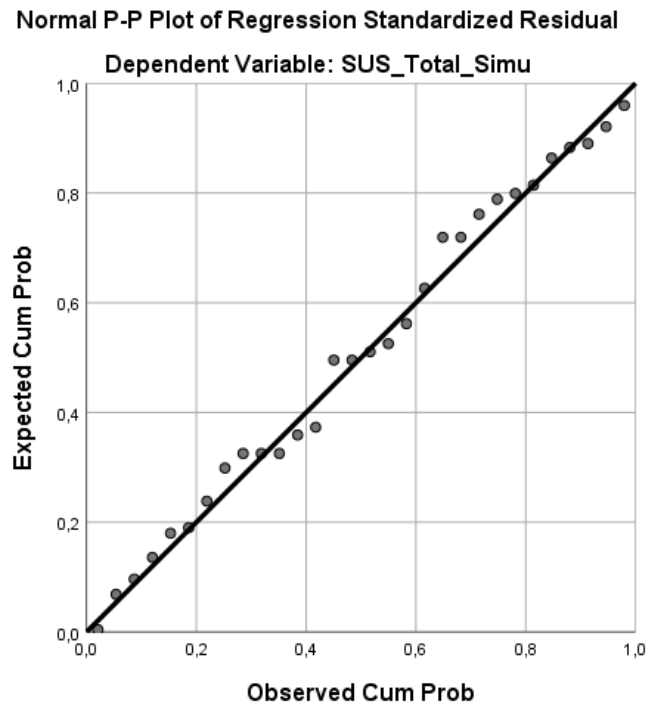
SimuSurg NPS



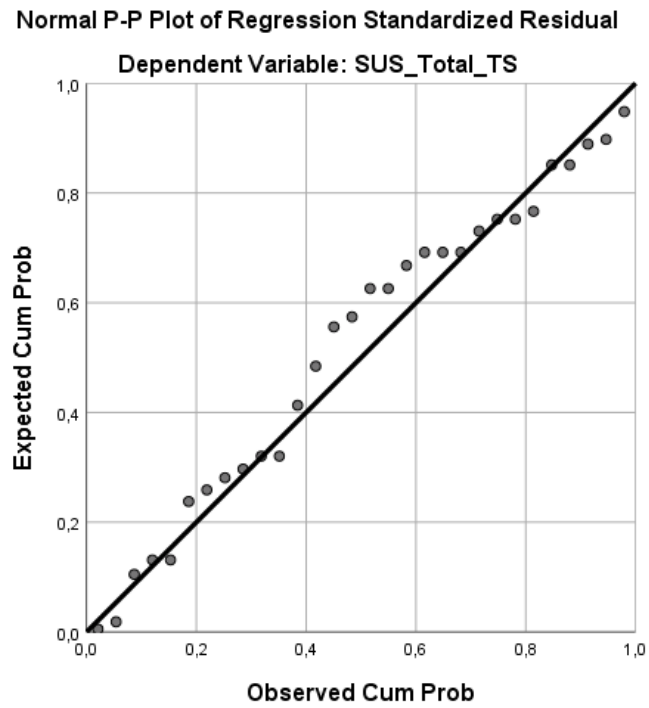
Touch Surgery NPS



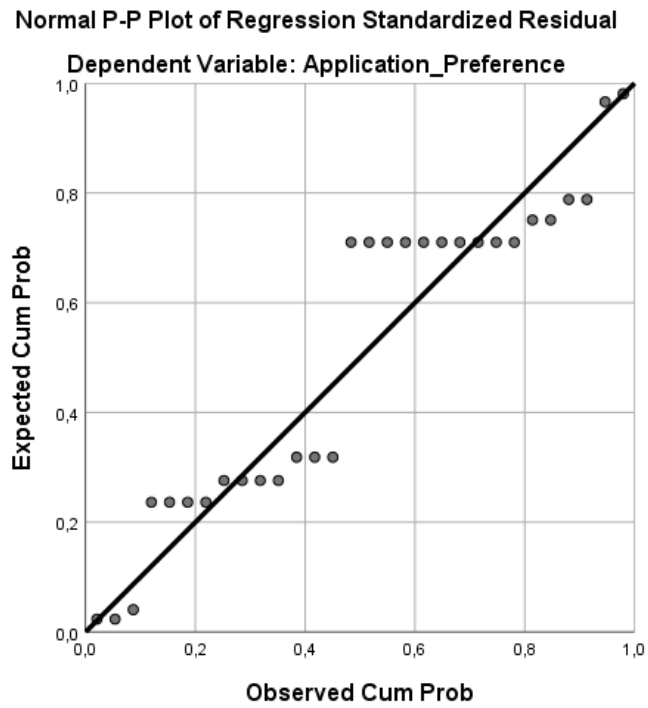
SimuSurg SUS Total Score



Touch Surgery SUS Total Score



Application Preference



Auto-correlation

Variable	Durbin-Watson Score
SimuSurg SUS Total Score	0.65
Touch Surgery SUS Total Score	1.66
SimuSurg NPS	1.54
Touch Surgery NPS	1.31
Application Preference	1.58