Elen Collaço

Impact of immersion on the performance of haptic virtual reality for inferior alveolar nerve anesthesia training

São Paulo

2019

Elen Collaço

Impact of immersion on the performance of haptic virtual reality for inferior alveolar nerve anesthesia training

Ph. D. Thesis presented to the Escola Politécnica, Universidade de São Paulo, to obtain the degree of Doctor of Science.

Concentration area: Computer Engineering

Advisor: Prof. Dr. Romero Tori

São Paulo

2019

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 19 de dezembro de 2019

Assinatura do autor: 

Assinatura do orientador: 

Catalogação-na-publicação

Collaço, Elen
Impact of immersion on the performance of haptic virtual reality for inferior alveolar nerve anesthesia training / E. Collaço -- versão corr. -- São Paulo, 2019.
103 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1.REALIDADE VIRTUAL 2.AVALIAÇÃO DE DESEMPENHO 3.ANESTESIA DENTÁRIA I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II.t.

Elen Collaço

Impact of immersion on the performance of haptic virtual reality for inferior alveolar nerve anesthesia training

Ph. D. Thesis presented to the Escola Politécnica, Universidade de São Paulo, to obtain the degree of Doctor of Science.

Aprovado em: 02 de dezembro de 2019.
Prof. Dr. Romero Tori - POLI/USP
Prof. Dr. Roseli de Deus Lopes - POLI/USP
Prof. Dr. Mary Caroline Skelton Macedo - FO/USP
Prof. Dr. Antonio Valerio Netto - UNIFESP
Prof. Dr. Alexandre Cardoso - UFU

São Paulo
2019

Dedicated to my father Robeto, my mother Isabel, and my sister Cíntia.

Acknowledgements

Firstly, I would like to thank Prof. Dr. Romero Tori who supervised this work, and who introduced virtual reality research to me. I have been fascinated ever since. I am very grateful for the knowledge you shared with me about virtual reality and human-computer interaction. Also, I am very grateful for having the opportunity to work on VIDA Odonto simulator.

Secondly, I would like to thank my informal co-supervisor, Prof. Dr. Elisabeti Kira, with whom I have had fruitful collaboration since 2018. I am very grateful for the knowledge you have shared about statistical methods and the helpful advice you have provided on machine learning techniques. I have enjoyed our many discussions, which kept me motivated.

I thank Prof. Dr. Maria Aparecida Machado for collaborations concerning the dental part of this study. Also, I thank her for providing support to conduct the pilot experiment in Bauru School of Dentistry of the USP.

I thank Prof. Dr. Mary Caroline Skelton from Dentistry School of USP for intermediations and suggestions. I thank Prof. Dr. Oswaldo Crivello Junior for helping make room for the experiment in the Oral and Maxillofacial Surgery Department, from Dentistry School of the USP.

I thank my labmate Lucas Sallaberry, with whom I have worked closely all this years.

I thank Philippe Bertrand for introducing me to embodiment illusion concepts, and to the virtual reality system The Machine to Be Another. I really enjoyed our work and collaborations.

I thank Anna Carolina Muller Queiroz for helpful comments and suggestions on syringe handling and avatar embodiment questionnaires.

I would like to thank all dental students from Bauru and from São Paulo Dental School at USP who took part in my experiment. I thank the INTERLAB team for helping me to conduct my experiments.

A sincere expression of thanks to Samsung OCEAN for their continued support as well as the loan of the notebooks, tablets, smartphones, HMD and GEAR VR devices for this study.

I also thank 3D Systems® from Brazil for donating one haptic device for the study.

This research was supported by research project grants from FAPESP (Process 2016/26290-3).

Abstract

Collaço, E., **Impact of immersion on the performance of haptic virtual reality for inferior alveolar nerve anesthesia training**. 102 p. Ph.D. Thesis – Escola Politécnica, Universidade de São Paulo, São Paulo, 2019.

The procedure for the administration of anesthesia to the inferior alveolar nerve is one of the most stressful in dental training. Although most studies have made efforts to develop Virtual Reality (VR) dental training, the majority have used non-immersive technologies. However, the effects of immersion, in both the explanation and training phases, on the performance of the dental anesthesia process, remains unclear. Therefore, in this work, we conducted an experimental study to evaluate the impact of immersion on the performance of haptic VR for inferior alveolar nerve anesthesia training. The immersive haptic VR, named VIDA Odonto®, was developed in our research group, and this work contributed with that development. The experiment involved 163 clinical students, divided into 4 groups under different combinations of immersion and haptic feedback. Their performance was evaluated in terms of execution time and the measurements related to the needle insertion. Moreover, we developed and tested a machine learning method for automatic evaluation of the dentistry student's performance, that is an innovative methodology for needle insertion assessment in VR-based dental training. Also, the participants were asked to report their syringe handling, simulator sickness and embodiment experience through questionnaires. Results indicated that groups receiving immersive explanations and/or immersive training showed more accuracy and confidence in administering the anesthesia. Participants perceived a high sense of the virtual hand ownership and realism of the haptic feedback when handling the syringe. Moreover, the proposed experimental design can be applied to any other HMD-based VR simulators that involves performance analysis. These results bring us one step closer to understanding the impact of head-mounted displays (HMDs) on virtual anesthesia training before their adoption at scale, and provide insights in how needle insertion performance can be assessed automatically by machine learning.

Keywords: Virtual reality, immersion, HMD, performance evaluation, dental anesthesia, needle insertion, machine learning, inferior alveolar nerve, haptic feedback.

Resumo

Collaço, E., **Impact of immersion on the performance of haptic virtual reality for inferior alveolar nerve anesthesia training**. 102 p. Tese de doutorado – Escola Politécnica, Universidade de São Paulo, São Paulo, 2019.

O procedimento anestésico do nervo alveolar inferior é um dos mais estressantes no treinamento odontológico. Embora alguns estudos tenham sido feitos para desenvolver o treinamento em realidade virtual (RV), a maioria utiliza tecnologias não imersivas. Portanto, os efeitos da imersão, nas fases de explicação e de treinamento, no desempenho da anestesia dentária, ainda são desconhecidos. Dessa forma, neste trabalho, foi realizado um estudo experimental para avaliar o impacto da imersão no treinamento da anestesia do nervo alveolar inferior. Foi utilizado um simulador em RV imersivo (com uso de óculos de RV) e um dispositivo háptico no formato de seringa anestésica. O simulador em RV usado nesse estudo, denominado VIDA Odonto®, foi desenvolvido pelo nosso grupo de pesquisa e este trabalho contribuiu para esse desenvolvimento. O experimento envolveu 163 estudantes clínicos, divididos em 4 grupos sob diferentes combinações de imersão e retorno háptico. O desempenho no procedimento de anestesia foi avaliado em termos de tempo de execução e medidas relacionadas à inserção da agulha. Além disso, foi desenvolvido e testado um método de aprendizado de máquina para a avaliação automática do desempenho do estudante de odontologia, o qual é uma metodologia inovadora para avaliação da inserção de agulha no treinamento odontológico baseado em RV. As experiências dos participantes em relação à manipulação da seringa, aos efeitos colaterais do simulador e à percepção do *avatar embodiment* foram auto-avaliadas por meio de questionários. Os resultados indicaram que os grupos que receberam explanação imersiva e/ou treinamento imersivo demonstraram mais precisão e confiança na aplicação da anestesia. Os participantes relataram sensação forte de propriedade da mão virtual e realismo do retorno háptico ao manusear a seringa. Além disso, o desenho experimental proposto pode ser aplicado a qualquer outro simulador de RV imersivo que envolva análise de desempenho. Esses resultados nos aproximam um pouco mais da compreensão do impacto dos dispositivos de RV imersiva no treinamento em anestesia virtual antes de serem adotados em grande escala e, além disso, fornecem informações sobre como o desempenho da inserção de agulha pode ser avaliado automaticamente pelo aprendizado de máquina (*machine learning*).

Palavras-chave: realidade virtual, imersão, HMD, avaliação de desempenho, aprendizagem de máquina, anestesia dentária, inserção da agulha, nervo alveolar inferior, retorno háptico.

List of Figures

List of Tables

List of abbreviations and acronyms

VIDA Odonto	Virtual Interactive Distance-learning on Anatomy Odontology
VR	Virtual Reality
HMD	Head-Mounted Display
AR	Augmented Reality
3D	Three-dimensional
2D	Two-dimensional
ANOVA	Analysis of Variance
PCA	Principal Component Analysis
GAM	Generalized Additive Models
Q-Q plot	Quantile-quantile Plot
KMO	Kaiser-Meyer-Olkin
Full	Full immersion
NH	No Haptic Feedback
NE	Non-immersive Explanation
NT	Non-immersive Training
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
FPR	False Positive Rate
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
DOF	Degree-of-Freedom

HMM	Hidden Markov Modeling
SSQ	Simulator Sickness Questionnaire
Interlab	Laboratório de Tecnologias Interativas

List of symbols

$t(k)$	score of Student-t distribution with k degree of freedom
z	score of Normal distribution, coordinate of Z-axis
$\chi^2(k)$	score of Chi-square distribution with k degree of freedom
T^2	Hotelling's statistics
F	F (Fisher-Snedecor) distribution
$F_{p,n-p}(\alpha)$	upper 100α-th percentile of F distribution with p and $n-p$ degrees of freedom
α	significance level, percentile of a distribution
γ	threshold setting, equivalent to $1-\alpha$
μ_X	true mean of X-coordinate insertion point
μ_Y	true mean of Y-coordinate insertion point
$\overline{x}$	sample mean of X-coordinate insertion point
$\overline{y}$	sample mean of y-coordinate insertion point
n	sample size
p	dimension of a random vector
r	proportionality constant of the ellipse radii
$\boldsymbol{S}$	sample covariance matrix
$\boldsymbol{L}$	lower triangular matrix of Cholesky decomposition
θ	generic angle, $0 \leq \theta \leq 2\pi$

Contents

1 Introduction

The procedure for the administration of anaesthesia to the inferior alveolar nerve is one of the most stressful in dental training (BRAND et al., 2010; KARY et al., 2018). In our previous work we developed an immersive VR simulator, named VIDA Odonto®[1] (TORI et al., 2016; TORI et al., 2018), for training the administration of anesthesia. This simulator shows a complete, realistic and anatomically correct head of a teenage male model for the purposes of these procedures, in a dental room (Figure 1a). Other VR simulators for training dental procedures are found in the literature, and in studies related to their use in educational contexts (TOWERS et al., 2019). However, few are full immersive, and to date we have not found reports on the immersive impacts of haptic VR using HMD for anesthesia training.

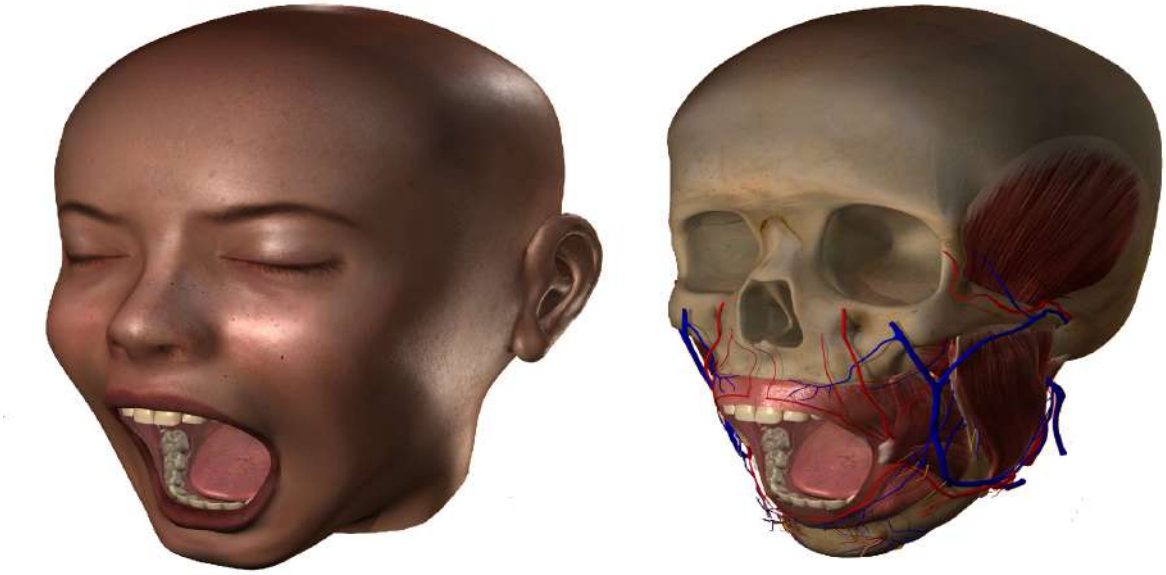
Moreover, even when using systems that allow tracking movement data, such as HMDs, hand controllers and haptic devices, learning assessments usually rely on multiple choice questions or subjective evaluations (QUEIROZ et al., 2019), (CORRÊA et al., 2018). Considering the anesthesia procedure, assessing students' performance is difficult and has been done subjectively, mainly based on patients' reports and professors' observations. When implementing training programs at scale, the subjective performance assessment method is costly, time consuming and doesn't give precise feedback to the student. Therefore the development of a methodological strategy for assessing students' performance when using movement tracking devices is very important to improving surgical accuracy and clinical outcomes.

Immersion is an important concept associated with VR systems, particularly HMD-based VR. Slater (2003) defines immersion as an objective characteristic of the system: "*The more that a system delivers displays (in all sensory modalities) and tracking that preserves fidelity in relation to their equivalent real-world sensory modalities, the more that it is 'immersive'*".

Although HMD has become more affordable, few studies in the dental context using HMD have been conducted. For example, dental school residents showed superior self-confidence and knowledge levels (both measured by questionnaires) after using HMD when watching a 360º video of maxillofacial surgery, compared to those using less immersive technology (PULIJALA et al., 2018). In our previous research (TORI et al., 2016; TORI et al., 2018) we showed the simulator viability when using HMD in the preclinical training phase. Participants were immersed in a virtual dental room (Figure 1b), and were asked to perform the inferior alveolar nerve anesthesia holding a cylindrical object. Their movements

[1] VIDA (Virtual Interactive Distance-learning on Anatomy) Odonto (Odontology).

were tracked using a Leap Motion (infrared sensor device) and neither immersion nor haptic feedback effects on training performance were investigated.



(a)

(b)

Figure 1 – The VIDA Odonto® patient's head and the dental room 3D models. **(a)** The patient's head model 3D, and **(b)** The VIDA Odonto® dental room.

The majority of the studies in dentistry have used less immersive technology, such as 2D or 3D monitors, coupled with haptic technology for performing drill procedures (STEINBERG et al., 2007; KOLESNIKOV et al., 2009; RHIENMORA et al., 2011; BAKR; MASSEY; ALEXANDER, 2015; BAKKER et al., 2010; MIRGHANI et al., 2018; ZHENG et al., 2007; CHEN; SUN; LIAO, 2018; TSE et al., 2010; DIEGO et al., 2012) or for administering inferior alveolar nerve anesthesia (CORRÊA; NUNES; TORI, 2014; POYADE; LYSAKOWSKI; ANDERSON, 2014). A large number of current dental simulators use real-time systems that track the movement of an integrated dental tool. This information is rendered on a 2D or 3D monitor using 3D models of the dental

arches and the manipulations are carried out in empty space (e.g. Simodont®, Virteasy® and Voxel-Man Dental®) or in a simulated tooth arrangement in a phantom human head (e.g. DentSim®). Therefore, HMD acceptance and the effects of immersion, in both the explanation and training phases, on the performance for dental anesthesia, remains unclear.

In order to investigate how immersion and haptic cues impact students' performance of a dental anesthesia procedure, an experimental study was carried out. In this study 163 students, divided in 4 groups, performed a training in the VIDA Odonto® simulator, under different combinations of immersion and haptic feedback. The study design consisted of 2 Phases (Explanation and Training) and 4 experimental groups, with different combinations of immersion and haptics. Table 1 (Section 3.2) shows the configuration of each group. Then, by statistically analysing the captured training data, the participants' performance was evaluated in terms of execution time and insertion point measurements (coordinates, accuracy, needle angle, needle depth). Also, subjective measures about the syringe handling, simulator sickness and embodiment experience were taken through questionnaires.

We hypothesized that the groups submitted to both explanation and training in immersive conditions would (A) show higher accuracy and require less time to perform the anesthesia procedure, and would (B) have more sense of the virtual hand ownership illusion and sense of controlling the syringe than groups submitted to combined condition in one of the phases. Concerning the machine learning method we hypothesized that the needle insertion point could be assessed automatically by the Hotelling's T^2 statistical method.

1.1 Motivation

The anesthesia administration procedure for the inferior alveolar nerve is one of the most stressful in dental training (BRAND et al., 2010; KARY et al., 2018). The main difficulty relates on the proximity of the needle tip to the mandibular foramen at the moment of the injection into the pterygomandibular region. Despite the large number of efforts made in order to improve training in dental education, the immersive impacts of haptic VR using HMD on the performance of anesthesia administration training, (especially on the inferior alveolar nerve block) have not been reported to date in the literature. The closest approaches have considered a less immersive technology, such as 2D or 3D monitors (TOWERS et al., 2019). Moreover, assessing a student's performance is difficult, and it is limited to measurement of task execution time and number of errors, or a subjective rating evaluation by an expert. In our previous work we developed the immersive VR VIDA Odonto® simulator for training the administration of the anesthesia (TORI et al., 2016; TORI et al., 2018). In this work, we conducted an experimental study to evaluate the impact of immersion on the performance of the administering of inferior

alveolar nerve anesthesia in the VIDA Odonto® simulator. We analyze and interpret the variables measures according to their types. Moreover, we propose a criterion to classify the individual student's performance according to her/his puncture coordinate point by means of machine learning. These results bring us one step closer to understanding what the impact of HMDs are on learning before their adoption at scale, and insights in how the needle insertion performance can be assessed automatically.

1.2 Objectives

The primary objectives of this research were (a) implementing a haptic immersive VR environment, and (b) evaluating the impact of the immersion on the performance of inferior alveolar nerve anesthesia training through a statistical experimental design and its analyses.

The secondary objective was to develop and validate a machine learning method to classify dental students' performance on needle insertion into "succeeded" or "failed".

1.3 Novel Contributions

Having reviewed the state-of-the-art dental VR simulators and relevant techniques to assess performance, our contributions can be summarized as follows.

- We implemented a state-of-the-art immersive haptic VR simulator, and analysis performance for inferior alveolar nerve anesthesia training. It extends previous works developed by our research group Tori et al. (2016) and Tori et al. (2018);
- We showed that the insertion point of the participants in the immersed groups were closer, on average, to the reference insertion point, and the immersed groups performed the anesthetic procedure faster.
- We developed and tested a machine learning method based on T^2-statistic so that a dental student's performance is classified as "successful" or "failed". This method can be applied to any other needle insertion task.
- We implemented an experimental design to evaluate the impact of immersion on inferior alveolar nerve anesthesia performance. However, there is no validated experimental design in the literature to test the impact of immersion on motor task performance. Therefore, we encourage the application of the experimental design implemented in this study in any other HMD-based VR simulators that involves performance analysis.
- The sample size ($n = 163$) is unusually large, providing statistically powerful results.

- While most of the questionnaire analyses in the literature are conducted only using basic statistics, we applied aggregation methods that may have revealed important insights on subjective evaluations as we have shown on syringe handling and avatar embodiment;

- We collected and recorded, via the Unity game engine platform, the dataset we used for the objective response variables: task execution time; insertion point coordinates; insertion accuracy (distance from the insertion point from the reference point); needle angle (deviation of the needle from reference direction); and needle depth. Also, positions and rotations of the syringe (corresponding to the stylus device's orientation and tip position) were recorded for future needle's trajectory analysis (full access to our data set is provided in a repository which DOI is: 10.17632/dpczr6rns9.1).

1.4 Organization of the chapters

Chapter 2 describes the state-of-the-art of dental surgery simulators and the machine learning methods for assess performance tracked by movement devices. A identification of gaps in research and the importance of the theme topic are also provided.

Chapter 3 describes in detail the VIDA Odonto® haptic VR simulator, the experimental design, the objective measures tracked by the haptic device, and the subjective response variables collected from questionnaires. Moreover, the statistical methods and analyses are described according to the variable types, and the participant sample are provided.

Chapter 4 presents the statistical results for each of the measured variables: task execution time; needle insertion accuracy; needle insertion point; needle depth and angle; syringe handling; avatar embodiment; and simulator sickness. Moreover, it present results for the ROC curve and machine learning based on prediction ellipse. This also provides references to further discussion detail.

Chapter 5 provides explanation of the results, suggests reasons for the results, and comments on whether a result was expected or unexpected. Moreover, it provides identification of the limitations.

Chapter 6 presents the significance of the findings, and suggestions for further research.

2 Literature Review

Dental surgery VR training simulators, to train motor skills, have been studied and reported in literature. Moreover, several machine learning methods for assessing surgical performance have drawn tremendous research interest in the medical context. However, few studies have been reported in dentistry training. This chapter briefly describes the state-of-the-art dental VR training, and the commonly applied objective and subjective evaluations. Moreover, the common machine learning methods on surgical performance related somehow to the needle insertion task are provided.

2.1 HMD-based dental training

Immersion is an important concept associated with VR systems, particularly HMD-based VR. According to Slater (SLATER, 2003), the level of immersion depends on objective features of the system and not on users' perception: "*The more that a system delivers displays (in all sensory modalities) and tracking that preserves fidelity in relation to their equivalent real-world sensory modalities, the more that it is 'immersive'*". Immersion is objective and measurable, meaning that one system can have a higher level of immersion than another (CUMMINGS; BAILENSON, 2016). The level of the immersion takes into account some components such as: field of view (FOV), the display size and resolution, stereoscopy, head tracking, frame rate, and realism of lighting. Nowadays HMD seems to be a better way to provide fully immersive VR systems, since we can use haptic device preventing users to "pass through" virtual objects, and perceiving their sense of presence (MESTRE, 2017).

Towers et al. (2019) reported in a recent review on dental education simulators that VR has been used as an umbrella term to describe a number of technologies. VR dental simulators can be constructed using full immersive technology, such as HMD device, or using a less immersive technology, such as computer monitors. In general, the intraoral parts are modeled inside a computer and the dental instrument is tracked by a haptic device. Moreover, the haptic device provides a sense of touch, allowing the students to feel exactly how different tissues and anatomical parts should feel. In this context, VR with haptics has been widely used in dental simulators.

Computer monitors are conventionally 2D, but some are able to simulate 3D models via stereoscopic glasses. The majority of studies have reported the use of 3D glasses (as described in Section 2.2). One publication reported an HMD-based immersion VR dental training (CHEN et al., 2003). A virtual patient was created, and was produced inside a virtual environment, whereby a dental mirror and hand pieces (provided by haptic device)

could be operated using HMD technology. The experiment was conducted with dental beginners and experts. The participants were asked to perform caries extraction from virtual tooth. The authors analyzed the forces applied between beginners and experts.

The VR Dental Anesthesia Simulator[1] is an initiative by the faculties of dentistry and medicine of the University of Alberta, Canada. Using the HTC Vive® controllers, students can virtually practice the local and the inferior alveolar nerve anesthesia procedures. However, it provides low realism, since the needle insertion is made holding the HMD controllers, and there is no haptic feedback, resulting in differences when compared to lifelike training anesthesia. This simulator was the only full immersive found to date. Unfortunately there is currently no literature that has reported a validation of this system.

Taking a slightly different approach, one study using a haptic augmented reality (AR) headset was found (RHIENMORA et al., 2010a). The system allows dental students to practice surgery in the correct postures as in the actual environment by combining 3D tooth and tool models with the real-world view using a headset AR tool. Within the AR environment, the system displays kinematic feedback, (using a haptic device) such as force utilization, in three axes of each procedure stage for a crown preparation, and tool/mirror movement pattern. Results from an initial evaluation showed that the simulator is promising for supplemental training.

2.2 VR in dental training

A recent systematic review (JODA et al., 2019) reported a low inclusion of studies of VR training in dentistry. From only 16 studies included, the majority were used for educational motor skill training ($n = 9$), maxillofacial surgical protocols ($n = 5$), investigation of human anatomy ($n = 1$), and dental phobia ($n = 1$). Moreover, a paucity of immersive VR using HMD devices can be observed.

The majority of the found studies in dentistry have used less immersive technology, such as 2D or 3D monitors, coupled with haptic technology (TOWERS et al., 2019). Based on the findings, two dental performances were identified and classified as: (1) drill training using commercial simulators (e.g. Simodont®, Virteasy®, Voxel-Man Dental® and DentSim®) (as shown in 2.2.1), and (2) anesthesia training using a prototype developed in a research project (as shown in 2.2.2).

2.2.1 Drill training

Many haptic VR simulators are based on tooth drilling training, and the major dental simulators in the market are Simodont®, Virteasy®, IDEA® and DentSim® (ROY; BAKR;

[1] https://cognitiveprojections.ca/

GEORGE, 2017). A brief insight into these VR simulators and others reported in literature is provided below. Figure 2 shows various types of dental drill simulators. They are described throughout this section.

The DentSim® [2] (developed by Denx Ltd and commercialized by Image Navigation Ltd., USA) simulator consists of a head manikin with plastic teeth, dental instruments (handpieces, air, water, light, and suction) and one computer. There are infrared emitters attached to the dental instrument and inside the head manikin, in which the infrared camera allows tracking of their position and orientation in space. The monitor displays the teeth being prepared by the student providing real-time feedback. Moreover, the final tooth preparation can be compared to the ideal preparation programmed by the faculty.

(a) DentSim®[2] **(b) Simodont®[3]** **(c) Voxel-Man®[6]**

(d) PerioSim® **(e) VirTeasy®[4]**

Figure 2 – The drill dental simulators.

Source – Figure d was extrated from (STEINBERG et al., 2007), and the remaining figures were extracted from the company website referenced by footnote.

The Nissin Simodont® [3] Dental Trainer (formerly Moog) contains a touch monitor user interaction with the courseware and a 3D display viewer with 3D glasses. The 3D display provides high-resolution images of teeth, and dental instruments. The student manipulates the physical instrument in an empty space underneath the 3D display viewer. The speed of the virtual instrument can be controlled using a real foot pedal. The system has different virtual instruments for the preparation of the tooth, and the students can select the cases from planned courses. Moreover, the system provides haptic feedback that simulates the differences in the tissues in the tooth.

The VirTeasy® [4] Dental (developed by HRV Simulation, France) is a simulator very

[2] https://image-navigation.com/home-page/dentsim/
[3] https://www.simodontdentaltrainer.com/
[4] https://www.virteasy.com/

similar to Simodont® technology. The most notable difference is that the 3D visualization monitor is larger and a patient's head was modeled without internal layers.

The IDEA®[5] Dental (Individual Dental Education Assistant, USA) is a haptic simulator that uses a Phantom haptic device produced by SensAble Technologies® (current Geomagic Touch). This unit provides a 3D animation of "dentoform" on the screen that allows the student to feel dental enamel, dentin, and pulp tissue. The simulator measures and records task time, percentage of desired material removed, and deviation from the assigned drilling task, reflecting the level of accuracy, and a score is displayed on the screen.

The PerioSim®(STEINBERG et al., 2007) has been developed at the University of Illinois at Chicago, College of Dentistry (COD). The simulator has one computer, a haptic device by SensAble Technologies® and one 3D monitor used with 3D glasses. An additional monitor is used to display instructions for the students.

Unlike the simulators presented above, the The Voxel-Man®[6] (developed by Voxel-Man Group, Germany) simulator enables bi-manual haptic interaction, and the teeth are displayed on a 3D monitor (3D glasses). The high resolution tooth models have been derived from real teeth by microtomography. The system allows students to obtain real-time feedback, and objective evaluation of their performance.

Overall, it was possible to observe that none of the simulators described in here are fully immersive. Moreover, it was found that the realism of the intraoral model only provides the upper and/or lower dental arches with a full complement of teeth and their surrounded gingival tissue.

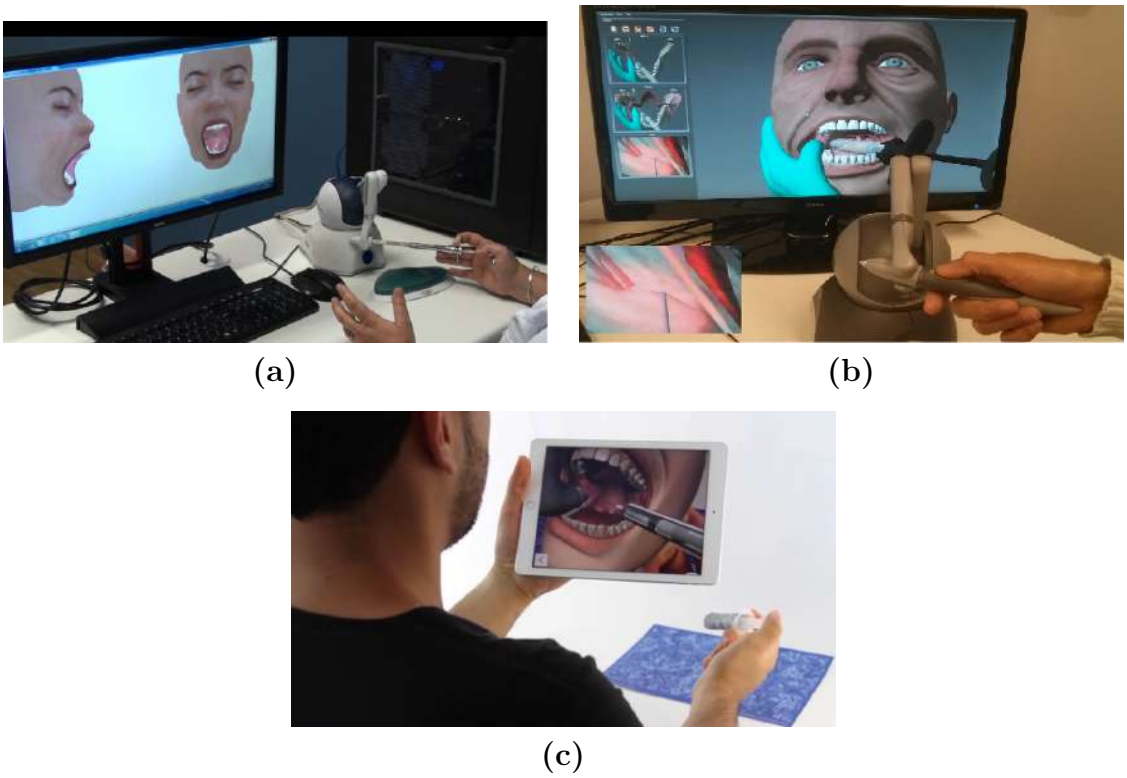
2.2.2 Anesthesia training

The development of haptic VR simulators for dental anesthesia training was not widely explored. No studies were found until now combining full immersive and haptic technology.

Previous authors (CORRÊA; NUNES; TORI, 2014) and (POYADE; LYSAKOWSKI; ANDERSON, 2014) have implemented a haptic training system based upon an accurate anatomical model for the inferior alveolar nerve anesthesia training. Corrêa, Nunes e Tori (2014) showed its validation through questionnaires, and 25 participants took part in the study (11 beginners, 12 at intermediate level, and 2 experts). The questionnaire was evaluated in terms of appearance of the virtual models and characteristics of the haptic device. The authors modified the haptic device to achieve realism by replacing the device stylus with a real syringe as shown in Figure 3a. Poyade, Lysakowski e Anderson (2014) presented the workflow of the system (Figure 3b); however, formal assessment in order to validate the training effect is still pending. These studies were the only ones found in the

[5] http://www.ideadental.com/
[6] https://www.voxel-man.com/

literature using haptic technology for dental anesthesia. However, these systems are not immersive, and the intraoral and the facial surface model are provided by a 2D monitor.



(a) **(b)**

(c)

Figure 3 – The anesthesia dental simulators. **(a)** modified haptic device for dental anesthesia **(b)** haptic device for dental anesthesia **(c)** app system for dental anesthesia.

Source – Figure (a) was extracted from (CORRÊA; NUNES; TORI, 2014), (b) from (POYADE; LYSAKOWSKI; ANDERSON, 2014), and (c) from (PEREIRA et al., 2016).

Pereira et al. (2016) developed an app system for smartphones and tablets named "Dental Anesthesia Simulator" for inferior alveolar nerve training at a dental school in Brazil (Figure 3c). The training can be done using the smartphone's camera or through VR mobile glasses. The student can practice holding a real syringe tracked by a marker-based augmented reality that can be easily obtained and printed from the system. This is a low cost solution; however, it only provides the needle puncture location training.

2.3 VR evaluations in dentistry

A recent systematic review in medical and dental applications on haptic interaction for needle insertion training can be found in Corrêa et al. (2018). The authors reported that the experimental studies used to investigate and validate the needle insertion simulators are based on subjective tests. Moreover, another recent study of immersive environments using HMD on learning performance (QUEIROZ et al., 2019) showed that assessing student's

performance is difficult and has been done subjectively. In the following subsections, the studies using objective and subjective VR evaluations are described.

2.3.1 Objective evaluations

By adding movement tracking devices to the simulators (such as haptic devices), objective evaluation becomes possible. However, even using such devices, several studies found for evaluating dental VR simulators are limited to the measurement of task execution time, number of errors measured by the simulator, or a subjective rating evaluation by an expert. Most of these studies have shown significant statistical differences among samples, using the parametric test ANOVA or a non-parametric test (e.g. Mann-Whitney U test or Kruskall Wallis test). The studies in this section are described based on the aim, subjects, measure of performance, statistical technique, and conclusion from the authors.

Jasinevicius et al. (2004) performed an experiment to compare the efficacy of the DentSim® with a non-computer assisted simulation system. The objective was to determine whether there were differences between the two systems in crown preparation and the amount of faculty instruction time. The experts evaluated the preparations and assigned a rating using a scale varying from 1 (not clinically acceptable) to 4 (excellent). ANOVA was applied to the preparation ratings and to student-faculty interactions time. The authors found that the dental simulator aided in decreasing faculty time in instruction, but no differences were found in preparation ratings.

Al-Saud et al. (2017) investigated motor skill acquisition using the Simodont® in terms of task completion (%), drill time (seconds), leeway and container errors scores (%). ANOVA was performed on mean scores between groups. The authors found significant differences between groups in overall performance. They concluded that dental motor skills in novice trainees are best optimized through a combination of instructor and the VR simulator.

Ben-Gal et al. (2013) led an experiment in order to determine whether the haptic simulator IDEA® could provide a reliable and valid assessment of manual dexterity. The following task parameters were recorded: (i) time to completion (ii) accuracy (iii) number of trials to successful completion and (iv) score provided by the simulator. ANOVA was used to test the significance of between group differences. The authors found that the simulator was able to differentiate between non-professionals and dental students, or non-professionals and dentists.

2.3.2 Subjective evaluations

The subjective response variables usually came from questionnaires based on Likert scales and/or interviews.

Steinberg et al. (2007) designed a study to evaluate whether faculty considered PerioSim® realistic and useful for training through questionnaires. Subjects used a seven-point rating scale with 7 on the scale defined as "very realistic" and 1 as "not realistic". The authors stated that although the system has some limitations on realism of the oral structures, the simulator would help students develop tactile skills.

Bakr, Massey e Alexander (2015) have investigated the haptic realism of Simodont® dental trainer. Twenty four dental students participated in the study. Participants were offered a trial session on the simulator and were asked to evaluate different aspects of this virtual simulator by completing a post-experimental questionnaire on a five point Likert scale. The authors concluded that all students valued and appreciated the additional educational benefits of the simulator, and showed that all students agreed that Simodont® should be used in conjunction with other traditional educational methods.

2.4 Machine learning on surgical performance

Machine learning is a branch of artificial intelligence, and a powerful non-parametric supervised or non-supervised learning algorithm used in pattern recognition (VAPNIK, 2013). According to the recent systematic review in Dias, Gupta e Yule (2019), the machine learning in surgical performance assessment mainly involves motion-tracking technologies (e.g., infrared cameras, wearable devices or robotic arms). The data recorded by these technologies is used to feed the machine learning algorithms in order to recognizing patterns and make predictions from the data. Dias, Gupta e Yule (2019) also reported that the majority performance assessment studies are based on medical surgery. Also, we observed from this review that: (a) the majority studies have developed statistical methods to classify surgical skill level as novice or expert, and (b) there is a scarcity of studies in dental performance assessment.

A few research groups have developed a dental VR simulator using machine learning. Rhienmora *et al.*(RHIENMORA et al., 2010b; RHIENMORA et al., 2011) has reported a VR non-immersive dental simulator, and how they analyzed crown preparation data using K-means and Hidden Markov Modeling (HMM). A 3D tooth model was acquired from a patient's volumetric tooth data (a 23-year-old male). The phantom haptic device produced by SensAble Technologies® was used to provide force feedback and data recording. They simulated tooth surface exploration and cutting with a cylindrical burr by utilizing a surface displacement technique. A cohort of 10 participants were recruited (5 novices and 5 experts). The authors concluded that the HMM is an effective tool for classifying a particular operator as novice-level or expert-level. Also, they reported that future experiments will be conducted with more participants to train the machine learning.

Beyond the anesthesia procedure there are several tasks which involve the needle

insertion, such as palpation (which assists the needle insertion task), laparoscopy and arthroscopy (involving needle insertion at the suture stage), and vascular procedures (which start with needle insertion) (CORRÊA et al., 2018). Due to the scarcity of studies found in dental context, the studies reported below have some relationship to the needle insertion task using machine learning techniques.

In the medical context, Leong et al. (2007) have shown that K-means and HMM methods combined can be used to learn models of laparoscopy motion trajectories, even in a group of subjects with mixed abilities, with no target classification of technical skill levels. The proposed model learned the trajectories, and then ranked the subjects in terms of consistencies to the trajectories. To obtain the trajectory a Polaris infrared tracker (Northern Digital[7], Inc., Waterloo, Ontario) with 6 DOF was attached to the tip of the laparoscopic instrument. Eleven subjects were recruited for the study (9 medical students and 2 practicing surgeons).

Fard et al. (2016) developed a predictive framework for objective skill assessment based on the trajectory movement of the surgical robot arms during a suturing task performed by novice and experts surgeons. They compared two frequently used machine learning techniques, the Logistic regression and the Support Vector Machine. The authors reported that the results presented could form a basis for decision support tools that automatically evaluate surgeons' dexterity and provide more personalized skill assessment and online feedback to trainees based on their performance. Also, they reported that more studies should be conducted with a larger number of participants.

Mazomenos et al. (2016) showed evaluation of surgical performance in vascular surgery using two clustering algorithms: K-means and Expectation maximization. The objective was to categorize the participants according to their experience level and manipulation patterns of surgical tools. A cohort of 12 participants separated into two groups (novices, experts) performed on a phantom silicon model using an endovascular robotic platform. The authors concluded that the discrimination of experience level and types of instruments was achieved with the generated motion features using the clustering algorithms. Moreover, they observed that the expert surgeon manipulates the tool in a more efficient, economical way than the novice, resulting in smoother trajectory.

Overall, it was possible to observe that the majority studies considered a low number of participants to obtain the dataset for the machine learning, in which the use of a large dataset would expected to be. Moreover, the surgery is a complex task and involves different manipulation of the instruments (e.g. anesthesia, suturing, and knot tying). Therefore, these differences must taken into account to develop machine learning for dental needle insertion performance.

[7] https://www.ndigital.com/

3 Materials and Methods

3.1 The VIDA Odonto® VR simulator

The VIDA Odonto® (TORI et al., 2016; TORI et al., 2018) consists of a VR training system used for the administration of anesthesia to the inferior alveolar nerve. The anesthesia procedure is performed by inserting the tip of the needle into the soft tissue (penetrable surface) surrounding the jaw. The system displays a complete virtual dental office, including a teenage patient (a ten year old male child) on a dental chair. The patient's head is a complete anatomical model, including the mandibular soft tissue and the skull, containing the most relevant parts of the nervous, circulatory and muscular systems. This simulator (Figure 4) includes an HMD (stereo Oculus Rift CV1) with a field of view of 110 degrees, 1080 x 1200 per eye resolution display and head tracking used on participants in immersive conditions. A full HD TV was also used to provide the 3D virtual environment for participants in non-immersive conditions. The virtual environment was implemented using the Unity 3D 5.5.2f1 game engine. Both virtual dental office and patient were designed in 3D Maya Studio. The simulator ran at 90 frames per second in the HMD, using an Intel i5-9400F CPU, with 2.9 GHz, 16 GB RAM, and an NVIDIA GeForce RTX 2070, running on Windows 10. All the instructions were previously recorded and the participants had no interaction with the experimenter.

Figure 4 – The VIDA Odonto® haptic VR simulator.

Source – Author.

The simulator includes three essential parts that were adapted to this study: haptic feedback module, procedure replay and explanation module. The workflow of the simulator including the haptic feedback module and procedure replay are illustrated in Figure 5.

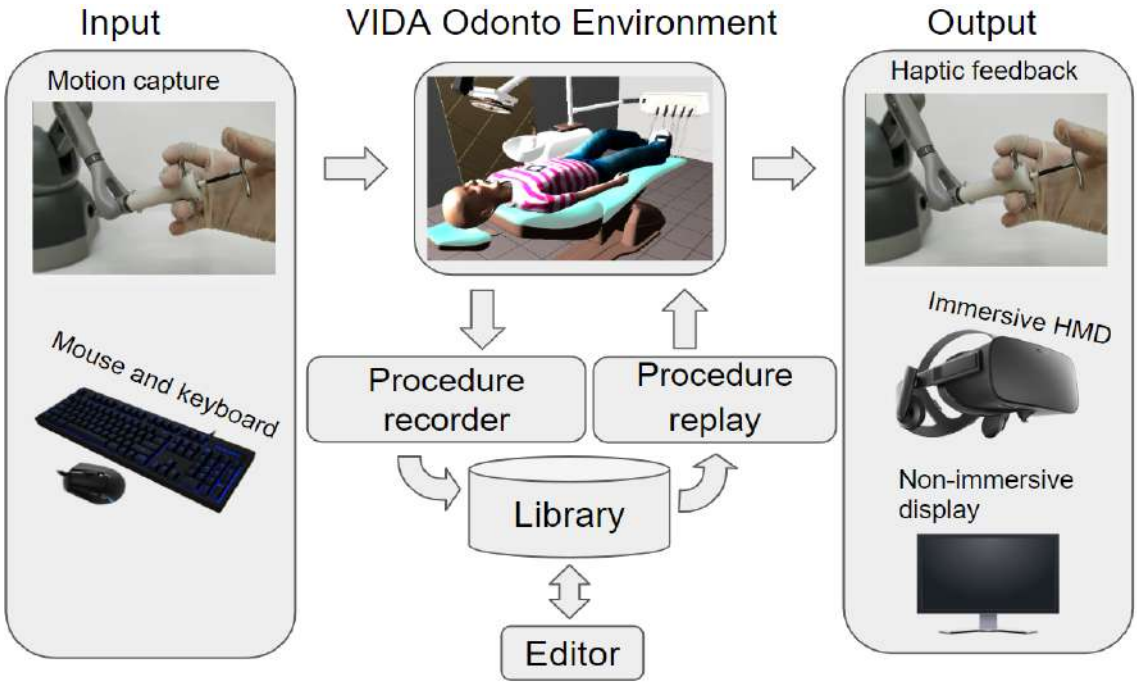


Figure 5 – The system architecture of VIDA Odonto® simulator.

Source – Author.

3.1.1 Explanation module

Before executing the training, participants watched a video explanation, according to the instructor's point of view, on how the procedure had to be executed (see Section 3.2 for experimental details and see Appendix F for dental instructor's storytelling explanation of anesthesia). They were able to see the patient's head with an open mouth. A blue transparent glove between the participant and the patient's head represented the dental professor's hand stretching the patient's cheek. The anatomical parts allowed an enhanced explanation of the process, with detailed visualizations during the explanation stage. It provided relevant information from the anatomical point of view that could not be offered in traditional training conducted in the real preclinical practice. The 6DOF VR head tracking allows the participants to freely look around the dental room and observe in detail the inner part of the patient's mouth.

3.1.2 The inferior alveolar nerve anesthesia procedure module

There are several methods reported in literature to administer an inferior alveolar nerve anesthesia (KHALIL, 2014). The technique consists of the insertion of the needle into pterygomandibular region in order to delivery the anesthetic solution in pterygomandibular space until it achieves the mandibular foramen, a region where the inferior alveolar nerve (vein and artery) is located (MALAMED, 2004). According to Gokhale et al. (2019) the success of the anesthesia of the inferior alveolar nerve is one of the most difficult tasks for inexperienced dentists, especially in children. The novice dentists may encounter difficulty in identifying the anatomical landmarks which are very useful in applying the anesthesia.

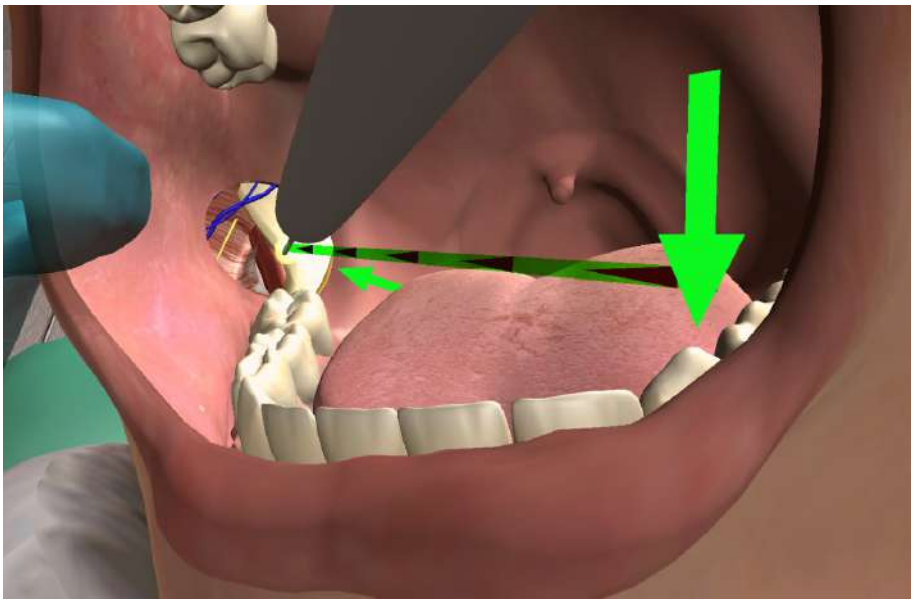


Figure 6 – Intraoral photograph of the virtual patient from VIDA Odonto VR environment when administering the inferior alveolar nerve anesthesia. The needle is inserted on the right side oral cavity showing key anatomical parts of the exposed mandible. The green arrow indicates the landmarks and the green and black dashed line indicates the needle trajectory.

Source – Author.

Several intraoral landmarks can be used as a reference to guide the clinician when administering the anesthesia (KHALIL, 2014). The procedure implemented in this study consists in use as a landmark the tip of the canine tooth (green arrow shown in Figure 6) of the side opposite to the area to be anesthetize. Starting at the point of the canine, the needle travels through the oral mucosa following approximately 2 cm behind the last erupted tooth (green and black dashed line shown in Figure 6). Once the correct needle position and angulation have been defined, the needle penetration should be a maximum of 2/3 of the length of the needle into the tissue.

3.1.3 Haptic feedback module

This simulator includes one haptic device (Geomatic Touch® from 3D Systems Enterprises). This device supports 6-degree-of-freedom (6 DOF) positional and rotational sensing input and 3 DOF force feedback output. This haptic module can render forces up to 3.3 N. The position and rotation is sampled at a rate of 62,5 per second (once every 16ms). To increase the tactile realism so that the participants could feel as if they were holding a real syringe, we used the metal handle of an actual syringe attached to the haptic mechanical arm using a 3D printed adapter (as shown in Figure 7). As participants were not expected to perform abrupt movements with their fingers while holding the syringe. A simple rigid non-animated gloved hand model was used, giving the participants more embodiment cues during the procedure.

Figure 7 – Modified haptic device by replacing the device stylus with a real syringe.

Source – Author.

The physical collision of the tip of the needle with the 3D objects added to the scene, such as soft tissue, teeth and bones, was only calculated for objects tagged as “touchable”. The collision parameters are determined by the attributes from the Unity 5 Haptic Plugin for Geomagic OpenHaptics. The “stiffness” attribute defines how hard the surface of an object feels. For the bone and teeth, we set up a higher “stiffness” of 1, in order to simulate a hard material, while the soft tissue had a “stiffness” of 0.8, giving a flexible tactile feeling. The “pop through” attribute indicates the cutting forces in terms of tissue resistance until it is ruptured (as shown in Figure 8). Soft tissues had a “pop through” parameter of only 0.05, meaning the needle of the syringe could penetrate it very easily. On the other hand, teeth and bone had a “pop through” of 0, defining them as impenetrable objects. The frictional force is due to the sliding of the needle inside the tissue (as shown in Figure 8). The “static friction” parameter was 0.5, while the “dynamic friction” was 0.41. The

collision parameters were manually refined and validated by a committee of dental experts. When the needle tip collides with the surface tissue and cuts it, the insertion starts and the penetration movement is constrained by the initial insertion direction (no rotation is allowed).

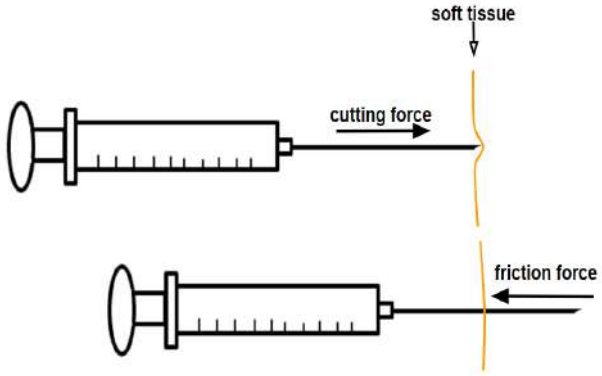


Figure 8 – Forces involved during a needle insertion on soft tissue.

Source – Author.

3.1.4 Procedure replay module

Via the replay module developed, anyone can visualize a student's previously recorded performance, in which the system loads the training scene, and the anesthesia procedure using the HMD or watching in a 2D monitor, in first-person perspective. The anatomical layers, such as the superficial skin, can be removed in order to better visualize the nervous, circulatory, and muscular systems. The person can also change his/hers viewing position, play, pause, rewind, or fast-forward the recorded procedure.

3.2 Experimental design

The purpose of the experiment was to analyze how the immersion condition and haptic cues impact the participant's performance through a set of response variables. The anesthetic simulation procedure consists of two subsequent phases: explanation and training, that can be executed under immersive or non-immersive conditions. When under immersive condition, the participants wear the HMD and under non-immersive condition, the participants visualize the explanation or perform the training watching a full HD TV screen. Moreover, the haptic feedback can be turned on or off during the training phase. A single factor Between-Groups Experimental Design was employed. The factor had four levels (groups): Full, NE, NT, and NH groups. The Full group had immersion in all phases and haptic feedback set on. In the other three groups we disabled only one of immersion or haptic feedback, as follows: Non-immersive Explanation (NE), Non-immersive Training (NT) and No Haptic Feedback (NH). The Table 1 shows the configuration of the four

experimental groups depending on under which condition the phase was performed, and on the status of the haptic feedback.

Table 1 – Group configuration of the experimental design

Group	Phase		Haptic Feedback
	Explanation	Training	
Full	immersive	immersive	yes
NE	**non-immersive**	immersive	yes
NT	immersive	**non-immersive**	yes
NH	immersive	immersive	**no**

Each participant was assigned, according to their order of arrival at the experimental session, to one of four groups. Participants were guided through the study by hypothesis-blind experimenters, who were trained in the VIDA Odonto® simulator. All participants attended three different sessions: (1) Pre-experiment, (2) Experiment, (3) Post-experiment.

1. Pre-experiment

   This session were composed of two Pre-experiment stages:

   - **Pre-experiment I**: Participants completed a computer-based Pre-experiment questionnaire, including what year of Dental School they are in, demographic questions, as well as questions to determine previous experience with virtual reality. Immediately after filling in the Pre-experiment questionnaire, participants had to wear a disposable hygiene eye mask and a real dentistry glove to perform the anesthesia. They were then led to the VR lab. Participants were invited to sit on a swivel chair next to the experimental bench containing an HMD, a haptic device and a TV. Before putting the HMD on, each participant heard a pre-recorded set of instructions displayed on the TV screen and played through the speakers in the lab room, and they were asked to follow the instructions.
   - **Pre-experiment II**: Participants received practice trials to become familiar with the haptic device and the HMD. This stage is important to prevent possible differences in task performance due to differences in their familiarity with the device. During the haptic device familiarization, participants were asked to look at the set of red spheres displayed in the monitor, and they were instructed to insert the haptic needle into the spheres presented at preset positions in 3D space of an empty virtual environment. After the haptic device familiarization, the participants were required to wear the HMD. The participants had the opportunity to adjust the HMD so they could find the best fit using a test image to guarantee the needle would be in focus during the experiment. Right after

the HMD adjustment, the participants were immersed in the VIDA Odonto® simulator and were instructed to look sideways, up and down of the dental room in order to familiarize themselves with the environment, and especially to see the patient, the dental sink and the professor avatar standing on the other side of the dental chair.

2. Experiment

   The experiment session was composed of two phases and includes the status of the haptic feedback.

   - **Explanation phase**: Participants experienced the instructor's explanation. The NE group refers to the participants receiving the anesthesia explanation from the dental instructor in a passive way on full HD TV screen. Participants from NT, NH and Full groups wore the HMD. They had head's agency, and observed the anesthesia explanation from the instructor's perspective in immersive condition. Participants were instructed to follow a single injection straight line towards the right side of the mandible by using anatomical reference points that are easily observed during the anesthesia procedure (for storytelling explanation of anesthesia see Appendix F). In all conditions it was explicitly showed the 3D model of the patient's right side exposed mandible, with nerves and arteries, in order to better visualize the inferior alveolar nerve region. Moreover, all participants had the opportunity to watch the explanation a second time
   - **Training phase**: Participants were asked to administer the inferior alveolar nerve block anesthesia, located on the right side of the mandible using the syringe coupled to the haptic device. Participants were previously instructed to say out loud "I am ready", in order to immediately record the positions and rotations of the syringe (corresponding to the stylus' device's orientation and tip position). The participants in the NT group performed the anesthesia using the full HD TV screen, in a non-immersive condition. Participants in the NE, NH and, Full group performed the anesthesia wearing the HMD, in immersive condition. They were also instructed to clearly announce that they have finished or wanted to give up the training session.
   - **Status of Haptic feedback**: During the training session, the haptic feedback in the syringe could be disabled or not. The participants in the NH group had a non-haptic feedback when the tip of the needle reached the patient's soft tissue. Full, NE and NT groups experienced the haptic feedback when the tip of the needle reached the patient's soft tissue. After the participant inserted the tip of the needle into the soft tissue, they heard instructions that the experiment had ended and they were asked to remove the HMD. Afterwards they were

guided by an assistant to another room to answer the online post-experiment questionnaire.

3. Post-experiment

   In a reserved room participants were invited to complete a questionnaire related to the avatar embodiment, the syringe handling, and the simulator sickness they experienced. They were also asked to write down their thoughts related to the experience.

3.2.1 Participant sample

The participants were 163 students from a School of Dentistry. They were assigned, according to their order of arrival at the experimental session, to one of four groups: Full ($n = 42$), NH ($n = 41$) , NE ($n = 40$) and NT ($n = 40$). They were from third- to sixth-year clinical students with some experience in anesthesia technique (121 women and 42 men; mean $\pm$ SD age is 24 $\pm$ 4.13, with a range from 20 to 54 years old). Among participants:

- 146 declared be right-handed, 12 left-handed and 5 ambidextrous;
- 36 declared be from third-year, 68 from fourth-year, 42 from fifth-year and 17 from sixth-year;
- 72 declared had normal vision and do not use eyeglasses or lenses, 13 declared difficult eyesight on close-up and 78 eyesight on far-away.

Participants were recruited through flyers posted around the building and by announcements in classrooms at the School of Dentistry from the University of São Paulo (USP). Recruitment methods and materials were in compliance with the rules of the School of Dentistry. Ethics approval was obtained from the Ethics Committee of the University of São Paulo (CAAE: 79294617.9.0000.5561), and the study was performed according to the committee's guidelines and regulations. All participants gave written informed consent before taking part in the study. The sample sizes in each group were calculated based on estimates from a prior pilot experiment, with a statistical power analysis of 80%.

3.3 Quantitative measures and questionnaires

Using the haptic device capability of accurately tracking, in real time, the syringe trajectory, and needle tip position in a reference plane, we were able to collect and record in the Unity platform the quantitative measurements we use as the objective response variables: task execution time; insertion point coordinates; insertion accuracy (distance from the insertion point from the reference point); needle angle (deviation of the needle

from reference direction); and needle depth. The subjective response variables came from questionnaires about the haptic use, the simulator sickness symptoms, and the embodiment experience. The statements on these questionnaires are presented in Appendix G. The complete statistical analyses are in Appendix A.4 (for the objective measurements) and in Appendix C, D, and E (for the questionnaires). In the following, we describe in detail these records.

Needle insertion point, needle angle, and needle depth. We recorded kinematic data (horizontal, vertical and depth position, respectively X, Y and Z coordinates, and rotation around each of the axes) from the haptic device stylus tip at 16ms sample rate, which represents a 60Hz frame during training session. The syringe corresponds to the stylus' orientation and tip position. We used a plane as a reference placed on the virtual surface of the 3D tissue representation. The reference plane is positioned at the target point and is perpendicular to the reference injection direction. We recorded the needle tip trajectory, position, and needle angulation performed by the anesthetic dental expert and we use it as a reference located in the center of the plane. The recording starts when the participant says loudly "I am ready" and ends when the participant starts to remove the needle from the tissue. The actual needle length value is 2.5 cm and the needle depth is recorded by calculating the amount of the needle that was inserted from the tip of the needle up to the tissue surface. This depth value is calculated by the haptic device, and it is normalized between 0 and 1. Therefore, the maximum value of needle depth is 1. By using the center of the plane as the reference for the insertion accuracy, the angle of the needle is calculated by measuring how much the participant deviates from the target in both angle and position distance.

Task Execution Time to perform anesthesia. The time (in seconds) beginning when the participant said loudly "I am ready" (to start the anesthesia procedure) and ended when the participant started to remove the needle from the tissue or if the participant declared the procedure was completed.

Syringe handling Questionnaire. Participants were asked to rate 10 questions about their experience with haptic feedback. Questions were scored on a 7-point Likert-scale, where 1 means the least agreement and 7 the most agreement. Some questions were adapted from Ullrich, *et al.* (ULLRICH; KUHLEN, 2012) and some were created by the researchers. This questionnaire served as a tool to check whether participants' perceptions of the realism of the feedback, the functioning of the syringe and the tactile sensitivity impact of their performance. The questionnaire scale of reliability was good, with a Cronbach's Alpha=.84.

Avatar Embodiment Questionnaire (GONZALEZ-FRANCO; PECK, 2018). Participants were asked to rate 11 questions scored on a -3 to +3 using a 7-point Likert-scale, where -3 means the least agreement and +3 the most agreement. Questions of their em-

bodiment experience were explored, including their feeling of body ownership, agency and motor control, the location of the body and external appearance. In this study, participants wore a dentistry glove and the embodiment was modulated by a virtual gloved white hand. The original survey of Gonzalez, *et al.*(GONZALEZ-FRANCO; PECK, 2018) contained 25 items and we excluded 14 questions as they were not applicable to our experiment. The questionnaire scale of reliability was strong with Cronbach's Alpha=.90.

Simulator Sickness Questionnaire (SSQ) (KENNEDY et al., 1993). Participants were asked to rate on a 0 (none) to 3 (severe) using a 4-point Likert-scale 15 questions concerning their discomfort in the VR simulator. The questions were extracted from the translation and cross-cultural Brazilian version (CARVALHO; COSTA; NARDI, 2011) based on the original SSQ (KENNEDY et al., 1993). The SSQ aims to analyse factors concerning side effects such as: (a) oculomotor symptoms (i.e., eyestrain and difficulty concentrating), (b) disorientation (i.e., dizziness and vertigo), and (c) nausea (i.e., belching and abdominal discomfort). This questionnaire served as a tool to identify whether these factors were induced by the VIDA Odonto® haptic VR simulator.

3.4 Statistical methods and analysis

The variables associated with the measurements in the previous section were statistically analysed according to their variable types. Primary, descriptive analyses were done for all variables. The complete statistical analyses of the objective measures are in Appendix A and the analyses of the questionnaires are in Appendix C, D, E. In this section we describe the basics of each inferential analysis, and give the references to the techniques applied. All statistical analyses were performed using the software R (Version 3.6.1). The approach for plotting ellipses was implemented in Matlab (Version R2015a).

3.4.1 Survival Analysis for task execution time

The anesthesia execution time was studied considering survival analysis techniques based on Kaplan-Meier estimates for the probability of task completion and on the Cox proportional hazards model to statistically compare the groups. The tests were based on the one-sided Wald methodology (KLEIN; MOESCHBERGER, 2006), considering Full group as reference for inference purposes. Since we verified that there was no difference between Full and NH groups, we considered a reduced Cox model by joining the Full and NH groups, considering them as the new reference group, and compared them against the NE and NT groups. The comparison of the Cox model coefficients were done using Wald test. The Cox model can provide estimates of the task execution rate for each group.

3.4.2 ANOVA for insertion accuracy, needle angle and needle depth

Considering the needle insertion responses (insertion accuracy, needle angle, and needle depth), an ANOVA (Analysis of Variance) was applied to group comparison with respect to insertion accuracy and needle angle. Since the distribution of insertion accuracy data were asymmetric, a logarithmic transformation was applied. Normality and homoscedasticity for the transformed data were checked based on Shapiro-Wilk and Bartlett's tests respectively. The log transformation was not able to normalise the needle depth; therefore, the non-parametric Kruskal-Wallis test was considered for this response (MONTGOMERY, 2017; NETER et al., 1996).

Due to problems with the Unity 5 Haptic Plugin for Geomagic OpenHaptics, we had some missing observations on the needle depth variable. Therefore, the final sample for needle depth analysis consisted of 154 participants: Full ($n = 37$), NH ($n = 41$), NE ($n = 37$) and NT ($n = 39$).

3.4.3 Bivariate Normal confidence and prediction ellipses for insertion points

The analysis of the needle insertion position, (X, Y)-coordinate, data was based on the Hotelling's T^2 statistic, which requires a multivariate normal assumption, in our case, it is a bivariate normal. The normality assumption of the (X, Y)-coordinate of the insertion point was validated through Q-Q plots, histograms, and the Shapiro-Wilk test. The Hotelling's T^2 statistic is the Mahalanobis distance multiplied by a constant (the constant depends on the variable being standardized) and it has, up to a constant, a F distribution with degrees of freedom p and $n - p$, where p is the number of variables; in our case $p = 2$.

We applied this statistic to two situations: one situation is constructing a confidence region for the mean (μ_X, μ_Y), and the other is constructing a prediction region for an individual (future) observation. We considered the confidence level of 95% for the confidence region and 75% for the prediction region. The confidence and prediction regions are defined as the Mahalanobis distance being less than or equal to a constant r^2. Since the Mahalanobis distance is a quadratic form, with the inverse of the sample covariance matrix $\boldsymbol{S}$, the confidence region has an ellipsoid shape. The ellipse is centered at $(\overline{x}, \overline{y})$. The orientation of the major and minor axes are the eigenvalues of $\boldsymbol{S}$ and their lengths (their radii) are $r-$proportional to the square root of the eigenvalues, respectively.

The confidence region ellipse for the true mean coordinates is centered at the sample mean coordinates and the radii are based on a constant that multiply the square root of the eigenvalues. In this case, the constant is $r = \left(\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)\right)^{\frac{1}{2}}$, with $F_{p,n-p}(\alpha)$ being the upper $(100\alpha) - th$ percentile of the F distribution with p and $(n - p)$ degrees of freedom.

For the prediction ellipse of a new future observation, the constant multiplying the

square root of the eigenvalues is $r = \left(\frac{p(n^2-1)}{n(n-p)} F_{p,n-p}(\alpha)\right)^{\frac{1}{2}}$, with $F_{p,n-p}(\alpha)$ being the upper $(100\alpha) - th$ percentile of the F distribution with p and $(n-p)$ degrees of freedom. Thus, the prediction ellipse is larger than the confidence ellipse (JOHNSON; WICHERN et al., 2002). Details of the method are shown in Appendix B.

3.4.3.1 A note on plotting the ellipses

According to the confidence and prediction regions described above, we would need to compute the eigenvectors and eigenvalues of the covariance matrix $\boldsymbol{S}$ to plot the ellipses. An alternative way of plotting is to decompose the covariance matrix $\boldsymbol{S}$ as $\boldsymbol{L}\boldsymbol{L}^{-1}$, with $\boldsymbol{L}$ being the lower triangular matrix of the Cholesky decomposition (RENCHER, 2005)). Applying this decomposition, the Mahalanobis distance becomes a norm, in R^2, of a bidimensional vector. Therefore, its geometric form is defined as a circle of radius r. The circumference of radius r is defined by its polar coordinates $(r\cos\theta, r\sin\theta)^T$, for $\theta \in (0, 2\pi)$. Therefore, the ellipses can be plotted by $(\overline{x}, \overline{y})^T + r\boldsymbol{L}(\cos\theta, \sin\theta)^T$, for $\theta \in (0, 2\pi)$. We found this method of drawing ellipses, not only mathematically elegant, but also computationally faster. Details of the method are shown in Appendix B.

3.4.4 Receiver operating characteristic curve for machine learning

In order to validate the statistical method for machine learning, the receiver operating characteristic (ROC) curve analysis was considered (KRZANOWSKI; HAND, 2009). The ROC curve is a graph that illustrates the performance of a binary classification method at various thresholds settings.

In this work the classification method is defined by the prediction ellipse based on Hotelling's T^2, so that a dental student was classified as "succeeded" if his/hers insertion point had lain inside the prediction ellipse, or "failed" if the insertion point had lain outside the ellipse. A threshold setting is a value of $\gamma = 1 - \alpha$ corresponding to the coefficient of the prediction ellipse. We provide the basic theory for Hotelling's T^2 in Appendix B

A dental professor evaluated each dental student with respect to the anesthesia administration using the replay module (see Section 3.1.4). Each student was then labeled as "succeeded" or "failed" according to the professor's evaluation. We considered these labels as the correct classification. The professor's evaluations and the classification method allow us to construct the confusion matrix given in Table 2.

Table 2 – The Confusion Matrix

Classified by the method	Correct classification		
	Succeeded	Failed	Total
Succeeded	True positive (TP)	False positive (FP)	TP + FP
Failed	False negative (FN)	True negative (TN)	FN + TN
Total	TP + FN	FP + TN	n

Three evaluation metrics are derived from the matrix in Table 2: the sensitivity (proportion of true positives), the specificity (proportion of true negatives), and the accuracy (proportion of correct predictions). These metrics are used to measure the T^2-statistic's performance as a classifier, and they are defined by

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.1}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \tag{3.2}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} = \frac{\text{TP} + \text{TN}}{n} \tag{3.3}$$

The axes of a ROC graph are the false positive rate (x-axis) and sensitivity (y-axis). Note that the false positive rate (FPR) is defined as 1 - Specificity. Therefore, the ROC curve reveals the tradeoff between sensitivity and specificity. Since a perfect classification method would yield 100% sensitivity (no false negatives) and 100% specificity (no false positives), the best threshold setting is the one which, at the same time, maximizes the sensitivity and minimizes the FPR. This setting can be chosen by examining the ROC curve so that it gives the highest point closest to the y-axis.

The area under the curve is a metric on how good a classifier is. For that we can reduce the ROC curve to a scalar value using the area under the curve (AUC) (KRZANOWSKI; HAND, 2009). A perfect classifier would have an AUC of 1.0. For example, if the curve is way up to the left, then this means that the classifier for some γ perfectly labeled every observation from the test data, and its AUC would be 1.0 (Example 1 in the Figure 9). If the curve is a straight line from the bottom left to the top right, then the classifier does no better then a random guess, and its AUC would be 0.5 (Example 3 in the Figure 9). However, most classifiers will appear as jagged behavior when the γ is varied as Example 2 shown in the Figure 9. Therefore, different classification methods can be compared quantitatively by the total area under their receiver operator characteristic curves using

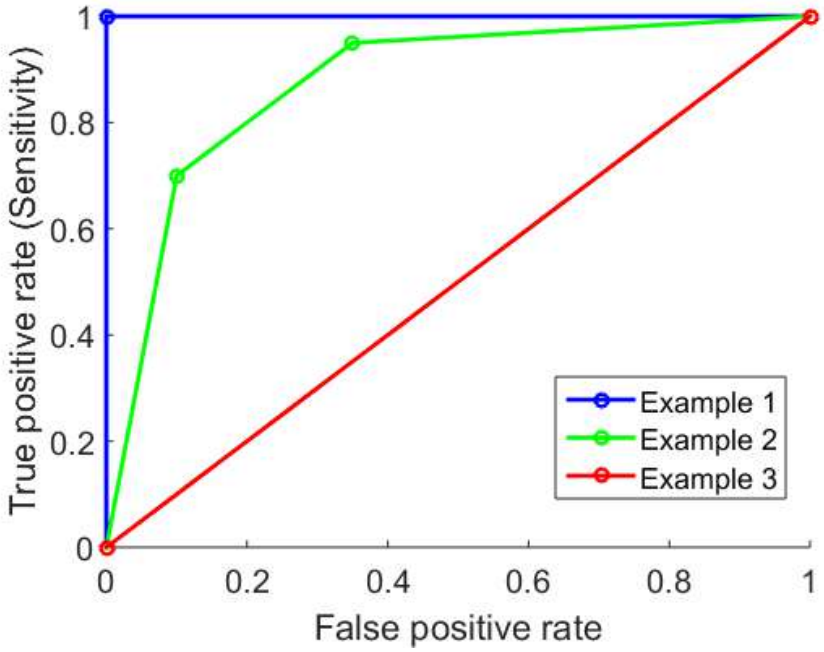


Figure 9 – Illustration of a receiver operating characteristic (ROC) curve.

Source – Author.

trapezoidal integration. Intuitively, the AUC estimates the probability that an expert chosen at random will score better than a randomly selected novice. Higher AUCs are more useful distinguishers, with an AUC of 1.0 being ideal and 0.5 no better than pure chance (WATSON, 2014).

In machine learning the idea is to use the data to construct the classification rule. Then, having settled on a rule, we seek to know how effective it will be in assigning future objects to classes. According to Krzanowski e Hand (2009) in principle, one could simply apply the rule to the data (training set), and examine the accuracy with which it classifies those objects. However, it would be unwise. Since the classification rule has been constructed using the training set it is optimized for those data. Therefore, the machine learning approach is based on splitting the available data into two sets: a training set and a test set, the former for choosing the rule and the latter for assessing its performance.

Given the dental professor's evaluation for each dental student's performance, we were able to take part of the sample the succeeded observations for machine learning. This sample, randomly chosen, is named training sample. The training sample was used to train the Hotelling's T^2 statistical method. Therefore, our machine learning method was trained based on the professor's criterion. The trained method was then tested using the remaining sample. The procedure can be described by listing the basic steps involved in the machine learning computation:

Step 1: Randomly select the training sample (a subset of the "succedded" observations according the professor's criterion) of size n; and compute the sample mean vector, the sample covariance matrix;

Step 2: For a fixed value of the threshold setting $\gamma = 1 - \alpha$, plot the prediction ellipse (according to the description in the Subsection 3.4.3, and expression elliptical contour B.9 - Appendix B). Recall that, for this case, $p = 2$;

Step 3: Verify if each observation of the test sample has lain inside the ellipse ("succeeded") or outside the ellipse ("failed"); and obtain the values of the confusion matrix;

Step 4: Fix another value of the threshold setting $\gamma = 1 - \alpha$ and go back to steps 2 and 3. Update the confusion matrix. Repeat for each different value of γ;

Step 5: Plot the ROC curve, and calculate the AUC;

Step 6: Given the ROC curve, choose an appropriate γ value according to its sensitivity and specificity;

Step 7: Plot the prediction ellipse with the chosen γ value. At this point we say that the system has been trained.

Step 8 For automatic evaluation, a new observation of needle insertion coordinates will be evaluated as "succeeded" or "failed" if the point lies inside or outside the prediction ellipse, respectively.

In order to test the effectiveness of the machine learning method, we apply the trained classification method to each observation of the test sample and obtain its confusion matrix. The sensitivity, specificity, and accuracy defined by the expressions (3.1), (3.2), and (3.3), respectively, are the performance metrics.

3.4.5 Principal Component Analysis and Factor Analysis for the Syringe Handling and Avatar Embodiment questionnaires

Principal component analysis (PCA) and Factor analysis using varimax rotation was applied to investigate the questionnaire responses (JOHNSON; WICHERN et al., 2002). The Bartlett's test and the Kaiser-Meyer-Olkin (KMO) analysis confirmed the suitability of the data for PCA and Factor Analysis. The number of factors is determined by analysing the scree plot and the total variance explained for the components. The factor scores for each participant is computed by multiplying the loadings for the participant's response. The residuals of the data were non-normally distributed, and log transformation did not normalised the data; thus a non-parametric test (Kruskal-Wallis) was applied on factor scores in comparing the groups. When differences among groups were observed we performed Bonferroni multiple comparisons to determine which means scores differ.

3.4.6 Generalised Additive Model for Simulator Sickness questionnaire

Given the large number of zeros (meaning no simulator sickness symptoms felt), we considered the number of symptoms felt per participant, which is a count data. We applied a zero-adjusted negative binomial model of generalised additive models (GAM)(WOOD, 2017) to investigate the total number of symptoms of SSQ. This model had a good fit to the residuals analysis, including the worm plot. Moreover, since the link function for the zero probability is logit, we can compute the percentage of zeros estimated by the model.

4 Results

In this section we present results on the following quantitative measured variables: task execution time, insertion accuracy (distance from the insertion point to the instructor reference point), insertion point coordinates, needle angle (deviation of the needle from reference direction), and needle depth. The results of the avatar hand embodiment, the syringe handling, and the simulator sickness questionnaires are presented next. A result is considered statistically significant if its p-value is less than 0.05. Details of the statistical methods applied are in Section 3.4, and the complete statistical analyses are in Appendix A (for the objective measurements), and in Appendix C, D, E (for the questionnaires). The T^2-statistic for machine learning is described in Appendix B.

4.1 Task execution time

In order to compare the groups according to the time to complete the anesthesia procedure, Kaplan-Meier estimates for the probability of task execution against time were plotted for the four groups, as shown in Figure 10. The shadowed areas around the curves represent pointwise 95% confidence intervals. From this figure we note that the immersive groups (Full and NH) behave somewhat similarly, with higher estimated probabilities of task completion.

The participants in the NT group took longer to complete the anesthesia procedure, with a noticeable difference compared to the other three groups. Apparently, the NE group could also be distinguished from NH and Full groups. For statistically testing the behavior seen in Figure 10, the Cox proportional hazards model was fit, and the Wald test considering Full group as reference was used for inference purposes. The results indicated no difference between the immersive groups (Full and NH) (one-sided $\chi^2(1) = 2.23, p = 0.1354$). Further analysis showed there is significant difference of the immersive groups (Full and NH) to NE (one-sided $\chi^2(1) = 8.48, p = 0.0036$) and to NT (one-sided $\chi^2(1) = 76.80, p < 0.0001$). Therefore the combined immersive groups (NE and NT) took significantly longer to execute the procedure. See Appendix A.1 for more details of this analysis.

It was notable that those in the NT group, in which the training was performed in non-immersive environment, took much longer to administer the anesthesia. Although, the execution time may not have an influence on performance measures (insertion accuracy, needle angle and depth), it may be indicative that the student feels more confident or secure during the anesthesia administration using the immersive technology.

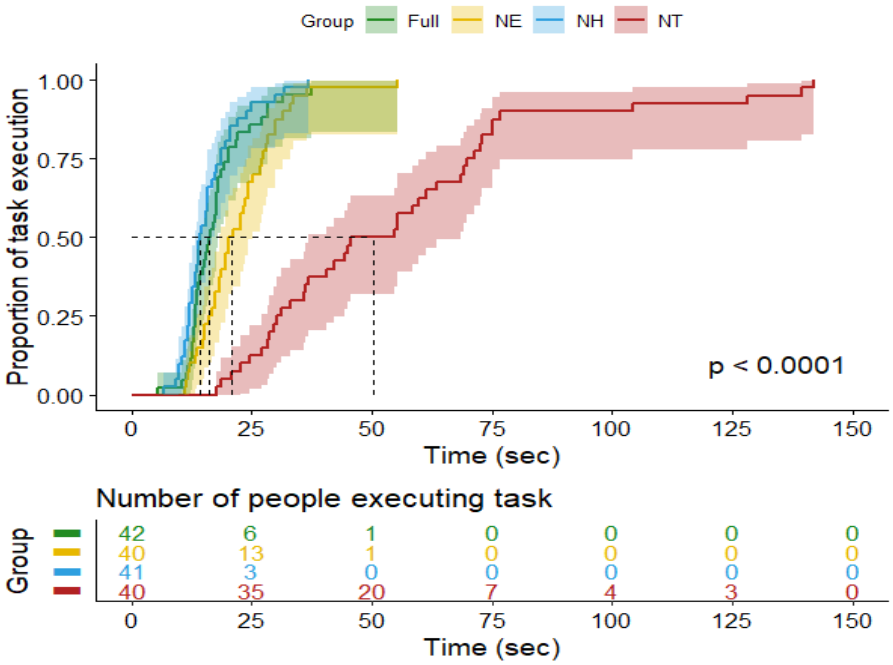


Figure 10 – Kaplan-Meier estimates curves of task execution time for each group with respective pointwise 95% confidence interval (shadowed area), followed by the number of people executing the procedure per time by group.

Source – Author.

4.2 Needle insertion accuracy

Considering the instructor's insertion point as the target, the insertion accuracy is the distance from the insertion point to this target point. Descriptively comparing the groups through the statistics summary in Table 5 of Appendix A.2, we noted that the immersive groups (Full and NH) presented lower mean values on the insertion accuracy than the combined immersive groups (NE and NT), meaning that the immersive groups were closer to the target, and therefore more accurate. In order to statistically test this finding, we performed a one-way ANOVA to the logarithm of the insertion accuracy; transformation was done so that the normality and the homoscedasticity assumptions were met. The ANOVA and contrast analyses results indicated that NH group is not different from Full (two-tailed $t(81) = 1.260$, adjusted $p = 0.4507$), but NE (two-tailed $t(80) = 3.987$, adjusted $p < 0.001$) and NT (two-tailed $t(80) = 2,494$, adjusted $p = 0.0367$) groups behaved differently from Full with respect to the mean log insertion accuracy.

Therefore, we confirmed that the groups training the anesthesia under immersive conditions (Full and NH) were more accurate than the groups submitted to combined immersive conditions (NE and NT).

4.3 Insertion point coordinates

Taking the coordinates of the insertion point onto the reference plane allows us to visualise the data in two dimensions. The reference plane is positioned at the target point and is perpendicular to the reference injection direction. The students' insertion point coordinates were plotted and coloured according to their group (Figure 11 and 12). Taking the bivariate data we generated two elliptical curves using Hotelling's T^2 statistic (JOHNSON; WICHERN et al., 2002). The center of the ellipses was $(\overline{x}, \overline{y}) = (0.1736, 0.0560)$, represented by the red star. The data can be considered to have a bivariate normal distribution for the x axis (Shapiro Wilk, $W = 0.99272, p = 0.5897$) and the y axis (Shapiro Wilk, $W = 0.99463, p = 0.8184$). The innermost ellipse is a 95% confidence region for the (population) mean insertion point coordinates (μ_X, μ_Y), as shown in Figure 11. We can note from this figure that the reference insertion point, as defined by the instructor (represented by black star in the figure), is inside the confidence region.

In order to provide an evaluation for future observations of insertion, we need to set a monitoring region. Each new observation is expected to lie in a prediction region. By doing so, we propose a machine learning method to classify future observations for the anesthetic procedure. The larger the prediction percentile (defined by $(1 - \alpha)$) is, the larger the prediction region that will be defined. As an example, we constructed a 75% prediction region for future insertion point coordinates, corresponding to the second and outermost ellipse in Figure 11 and 12. Any future student's observation inside the prediction ellipse would be classified as "succeeded" with respect to the insertion point coordinates evaluation. Points outside the prediction ellipse represents potential variation far from the reference insertion point and it would be classified as "failed" or, moreover, it would be used to determine whether immediate action should be taken about the anesthetic procedure during the student's training. Moreover, from Figure 11 and 12, it appears that participants in combined condition have more difficulties assessing the target insertion point, since most of the points outside the prediction ellipse are from combined groups (NE and NT). Details of the confidence region and prediction region construction are in Subsection 3.4.3, and the plotting of ellipses are in Subsection 3.4.3.1. Details of the machine learning developed area in Subsection 3.4.4. More details of T^2-statistic are in Appendix B.

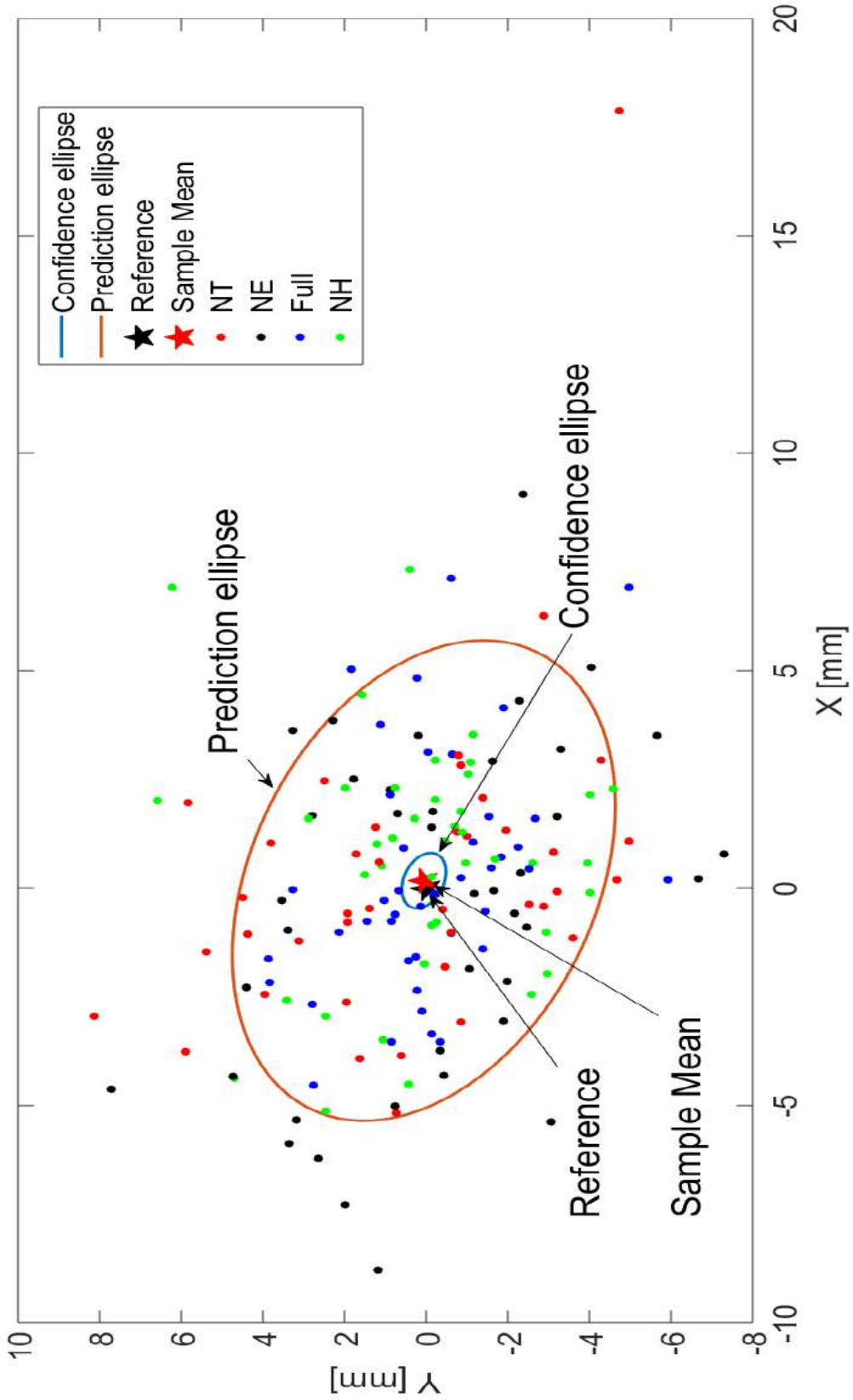


Figure 11 – Scatter plot of insertion point data with a 95% confidence ellipse for the mean and 75% prediction ellipse for individual observations.

Source – Author.

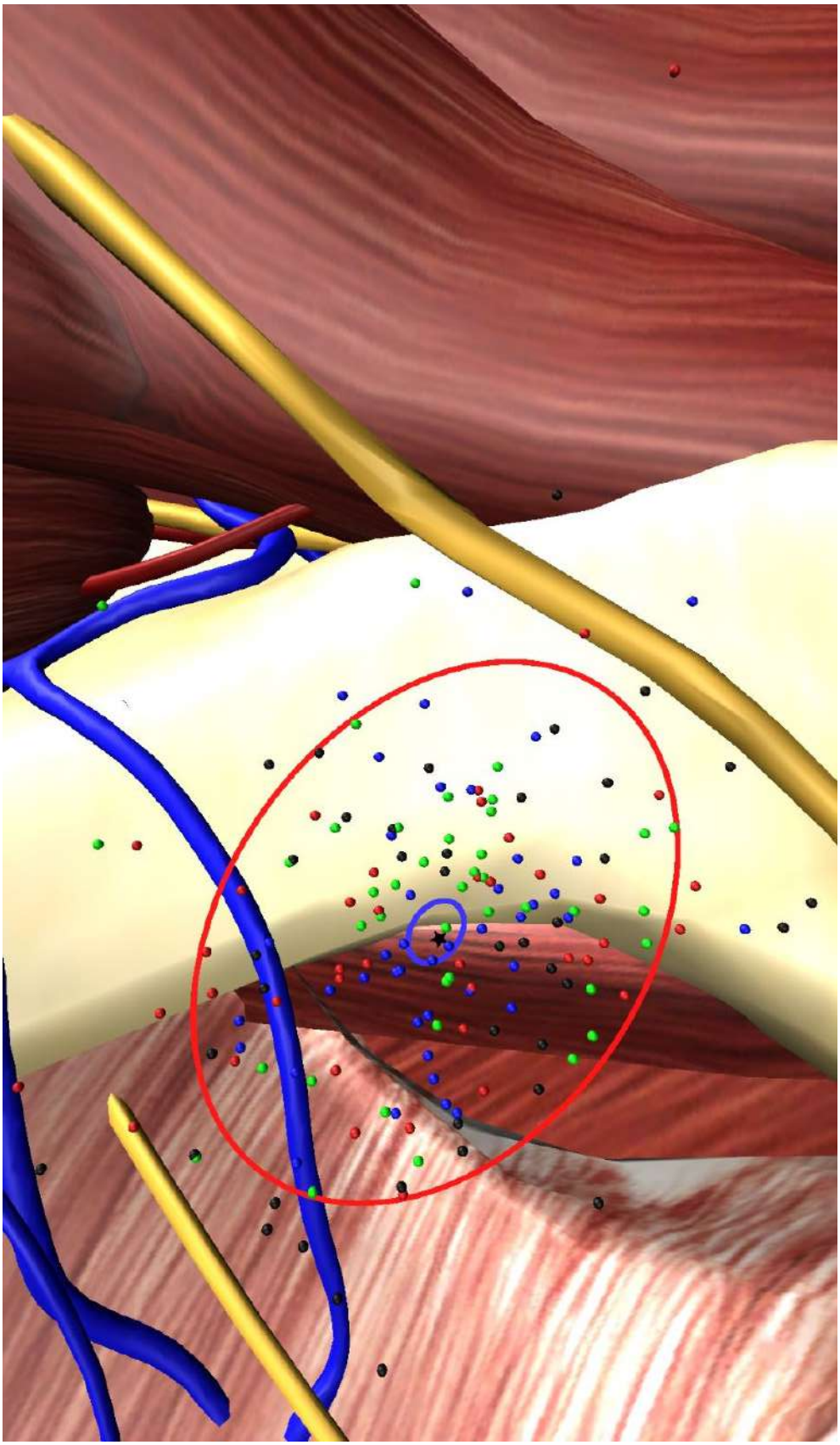


Figure 12 – The insertion points with confidence and prediction ellipses projected inside the virtual patient's mouth.

Source – Author.

4.3.1 ROC curve and machine learning based on prediction ellipse

In order to plot a ROC curve and use it as a machine learning technique, we need to take part of the sample and used it to train our classification method based on the prediction ellipse. The remaining sample is used to test the predictive power of the "trained" method.

From 163 students, the professor classified 105 as succeeded, and 58 as failed. Therefore, among the succeeded 105 we have chosen randomly 50 observations to train the method. The remaining 113 observations were used to test the method. We assumed that the professor's classification is the correct one. A participant, whose insertion point had lain inside the prediction ellipse, was classified as succeeded by our method. Otherwise, his/hers classification was failed. Based on these two classifications, the confusion matrix (described in Subsection 3.4.4) was obtained.

Varying the threshold setting γ ($\gamma = 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90$) we were able to plot the ROC curve as shown in Figure 13. In this figure, each red circle corresponds to a threshold setting, in increasing order.

The area under the curve (AUC) is 0.90, meaning that the Hotelling's T^2 method proposed is a good classifier, and may be used to appropriately discriminate (as successful or failed) the student's performance related to needle insertion coordinates.

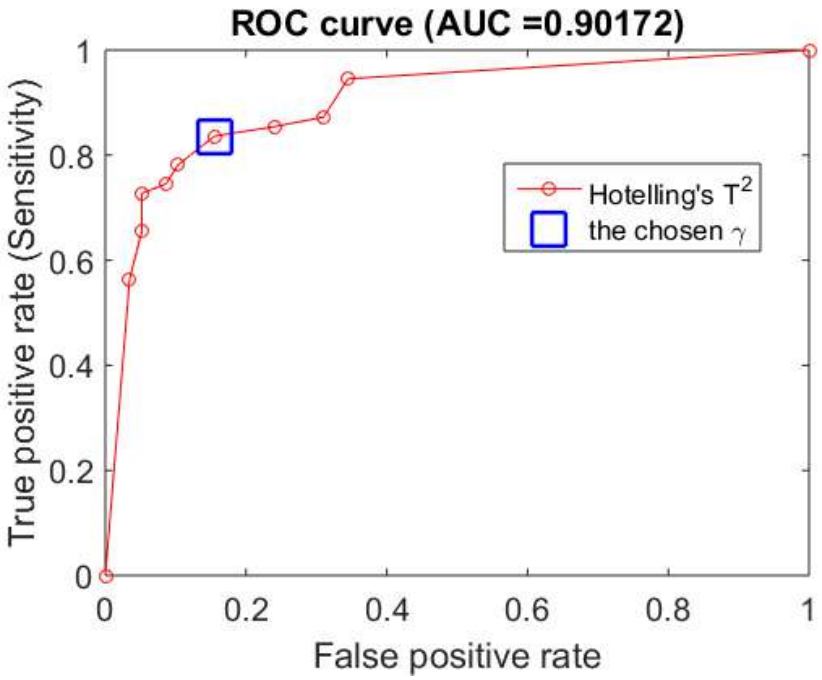


Figure 13 – The receiver operating characteristic (ROC) curve of the Hotelling's T^2 with some fixed classification threshold γ values and its AUC value.

Source – Author.

By examining the ROC curve, we have chosen $\gamma = 0.75$, as pointed out in the Figure 13 by the blue box. Based on this coefficient, a 75% prediction ellipse was computed. This

ellipse and the remaining 113 observations were plotted in the Figure 14. We can notice that, out of 55 succeeded (blue) observations, 46 were classified as succeeded (inside the ellipse) by our method, yielding sensitivity (true positive rate) of 83.6%. On the other hand, 9 (red) were misclassified classified as succeeded, resulting a false positive rate of 15.5%, equivalent to specificity of 84.5%. Also, out of 58 failed (red) observations, 49 were classified as failed (outside the ellipse) by our method, and 9 (blue) were misclassified as succeeded.

Table 3 – The Confusion Matrix based on $\gamma = 0.75$ for the test sample.

Classified by the method	Correct classification		
	Succeeded	Failed	Total
Succeeded	46 (83.6%)	9 (15.5%)	55
Failed	9 (16.4%)	49 (84.5%)	58
Total	55 (100.0%)	58 (100.0%)	113

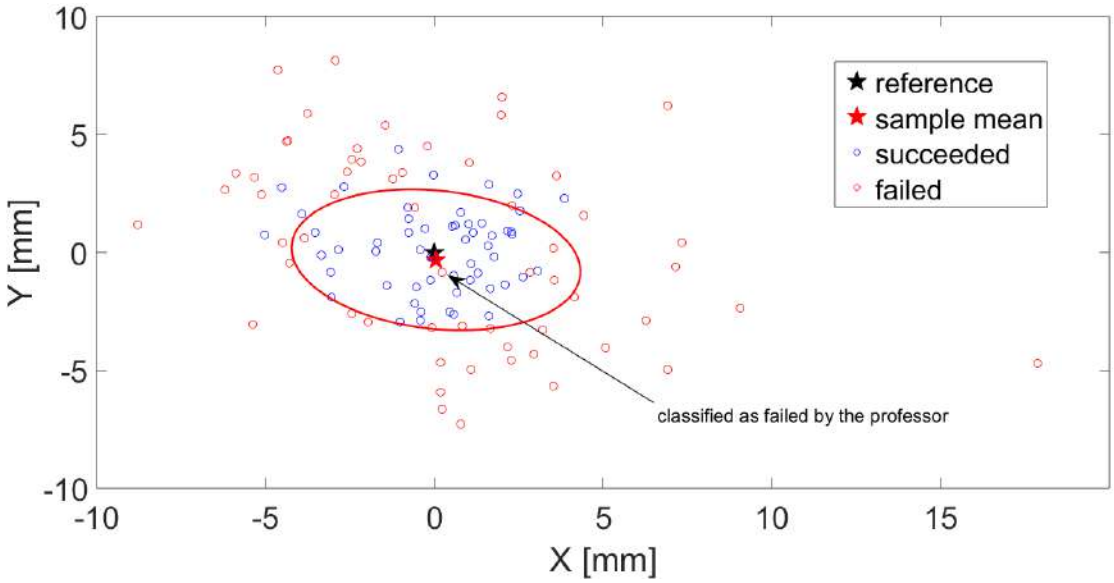


Figure 14 – Plot of a 75% prediction ellipse based on training data, and scatter plot of the insertion point of the test sample. The blue observations indicate "successful", and the red observations indicate "failed" according the professor's evaluation for each student's needle insertion.

Source – Author.

The confusion matrix (Table 3) allowed the computation of the performance metrics (given by expressions (3.1), (3.2), and (3.3)) of the T^2-statistic classifier based on 75% prediction ellipse, for the remaining (test) sample of size 113.

Since the machine learning presented accuracy of 84%, sensitivity of 83.6%, and specificity of 84.5%, we concluded that the T^2-statistic performed well as a classifier for the

student's needle insertion point. Therefore, the machine learning method was validated by the test sample.

4.4 Needle Angle and Depth of penetration

The descriptive summaries of needle angle, in Table 8 of Appendix A.4, indicated that the groups had similar behaviour with respect to the needle angle, and this data could be considered as normally distributed. Performing a one-way ANOVA for needle angle, we confirmed that there was no significant difference among the groups ($F(3, 159) = 0.908, p = 0.439$). The normality and homoscedasticity assumptions were met.

The normalized needle depth calculated by the haptic device lies between 0 and 1. The needle depth data presented nine missing observations due to a technical problem described in Section 3.4.2. The descriptive analysis (Appendix A.3) showed that the mean and median values of needle depth are similar for all groups, and the needle depth data have an asymmetric distribution. Since normality and homoscedasticity assumptions were not met by the log transformation, nonparametric analysis was performed and revealed no differences among the groups (Kruskal-Wallis (two-tailed): $\chi^2(3) = 4.1684, p = 0.2438$), corroborating the descriptive results.

4.5 Syringe Handling

The Figure 15 illustrates the mean score, and respective standard error, for each questionnaire item by group, and allowed us to notice that in all items the mean score is 4 or above, recalling that 4 is the center of this 7-point Likert scale. Therefore, in general, there was a moderate to strong agreement with the statement of this questionnaire. The mean scores are higher at items Q1 (*I noticed the tactile stimulus upon penetrating the soft tissue*), Q3 (*A device with tactile feedback upon resting against the soft tissue is useful*), and Q10 (*The tactile sensation of resting the needle on the soft tissue helped my performance*), meaning strong agreement in the statements related to tactile realism in all groups, including the NH group whose haptic feedback was set off.

The NT group has a lower mean score in 7 of the 10 items, specially Q5 (*It was easy to control the syringe*) and Q6 (*The syringe was stable*), meaning that this group felt slightly more difficulty in controlling the syringe than other groups.

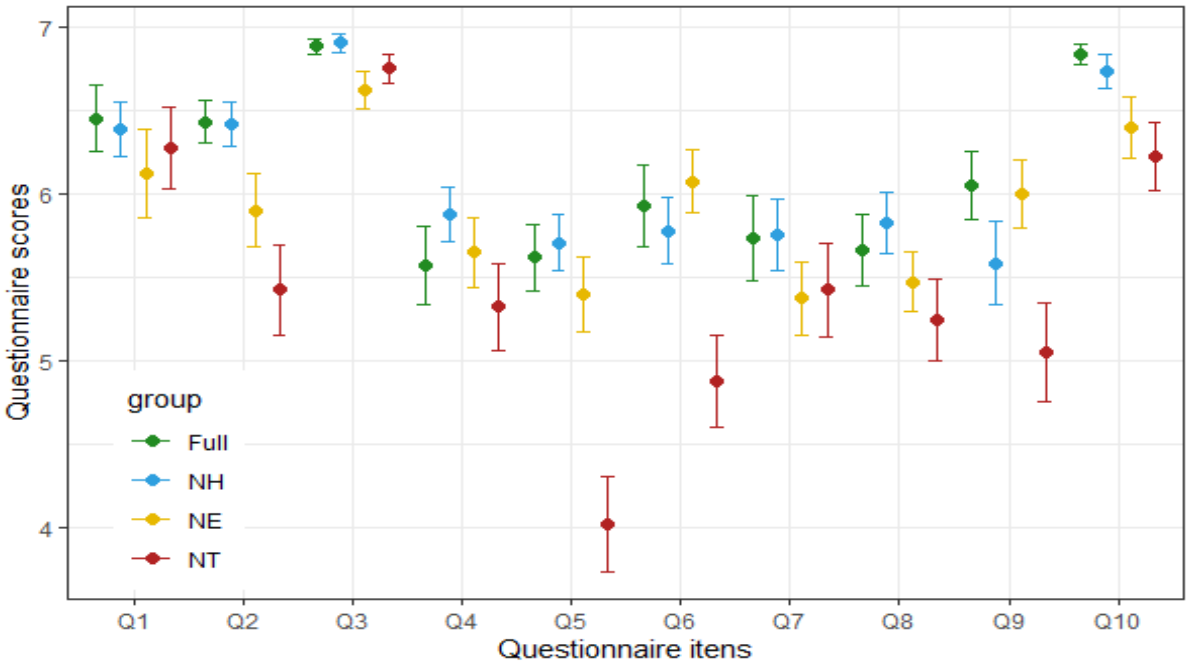


Figure 15 – Mean and standard error bar for each item of the Syringe Handling Questionnaire by group.

A principal component analysis (PCA) technique (JOHNSON; WICHERN et al., 2002) was used to jointly analyse the responses of all items of the Syringe Handling Questionnaire. Three factors were extracted. Factor 1 aggregated the items related to the tactile sensation, therefore we named it “tactile realism”. Factor 2 aggregated the items related to controlling the syringe, therefore we named it “syringe control”, and Factor 3 was termed “ease of performance”.

We found no significant difference among groups for Factor 1 (Kruskal-Wallis $\chi^2(3) = 3.472, p = 0.3245$). This result confirmed that the NH group had similar behaviour to other groups with respect to the *tactile realism*, even though this group had the haptic feedback turned off.

Factor 3 (*ease of performance*) comprised a borderline result (Kruskal-Wallis $\chi^2(3) = 8.042, p = 0.04515$), and concluded that we do not have enough evidence to confirm any difference among groups.

For Factor 2 (*syringe control*), there was a significant difference among groups (Kruskal-Wallis $\chi^2(3) = 16.897, p = 0.00074$), and Bonferroni multiple comparisons yielded that the NT group differs from Full (two-sided $z = -3.54, p = 0.0004$), and NT differs from NH group (two-sided $z = -2.903, p = 0.0037$). We pointed out that the items grouped in Factor 2 are the same items where group NT presented lower mean values compared to other groups, as seen in the Figure 15, confirming that the participants in NT group had more difficulty controlling the syringe. Details of the questionnaire analyses are in Appendix C.

4.6 Avatar Embodiment

The Avatar Embodiment Questionnaire is based on a 7-point Likert scale ranging from -3 to +3. The Figure 16 illustrates that questions Q2, Q3, Q6, Q7, and Q9 had mean scores below zero in all groups meaning strong disagreement. These statements are related to the feeling that the virtual hand belong to someone else.

On the other hand, the remaining statements Q1, Q4, Q5, Q8, Q10, and Q11, related to the virtual hand ownership, presented high mean scores, in general, with the mean score of the NT group being descriptively lower. This initial analysis suggests that increased level of immersion produces high sense of virtual hand ownership, since the Full, NH, and NE groups had higher scores than NT.

A principal component analysis (PCA) (JOHNSON; WICHERN et al., 2002) was applied to jointly analyse the Avatar Embodiment Questionnaire. Two factors were extracted and items were grouped into: Factor 1 *lack of agency control* and Factor 2 *virtual hand ownership*.

The Kruskal-Wallis test results showed that there was a significant difference among groups for Factor 1 ($\chi^2(3) = 16.907, p = 0.0007$). Further analysis using Bonferroni multiple comparisons showed that, for Factor 1, NT differs from the other three groups (Full: two-sided $z = -3.036, p = 0.0024$, NE: two-sided $z = -2.948, p = 0.0016$, NH: two-sided $z = -2.727, p = 0.0032$), and the groups Full, NE, and NH behave similarly ($p = 1.0000$). There was a significant difference for Factor 2 ($\chi^2(3) = 9.287, p = 0.02571$),

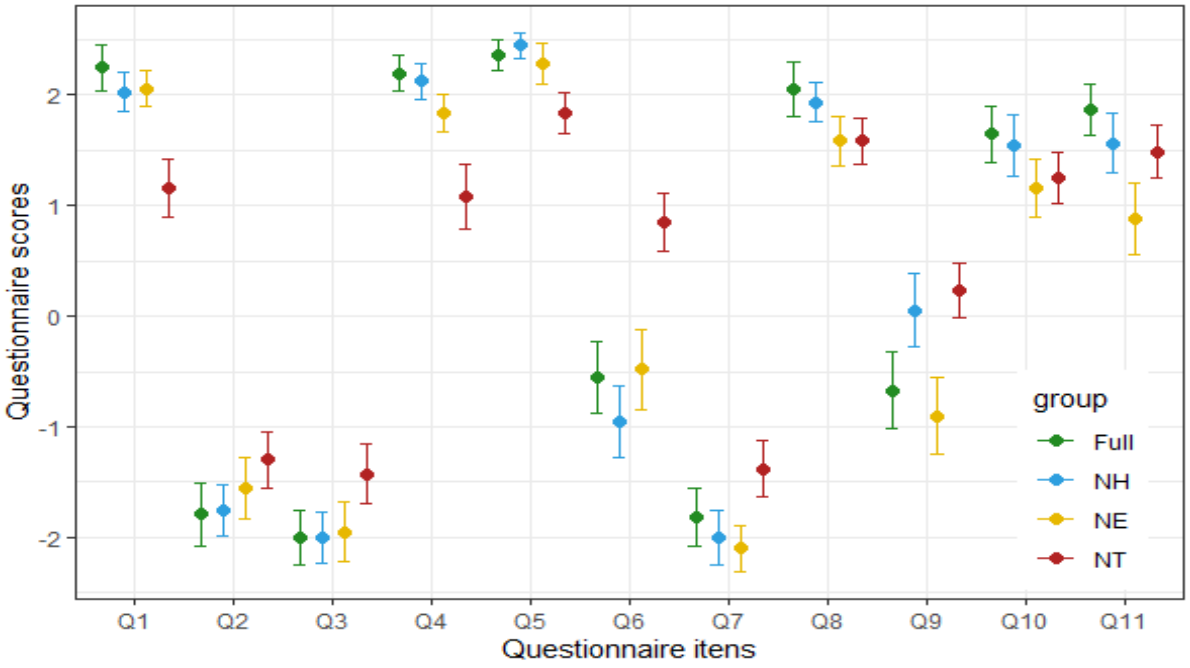


Figure 16 – Mean and standard error bar for each item of the Avatar Embodiment Questionnaire by group.

and Bonferroni multiple comparisons detected a difference only between NT and Full groups (two-sided $z = 2.264, p = 0.0236$).

These inferential results corroborate that the participants in the NT group had felt less *lack of agency control* than the other groups.

Detailed analyses of this questionnaire are in Appendix D.

4.7 Simulator Sickness

The Simulator Sickness Questionnaire (SSQ) (KENNEDY et al., 1993) results showed that there was a very little incidence of simulator sickness and no participant asked to terminate their participation in the experiment. We observed a majority of "none" side effects (87%), "slight" (9%), "moderate" (3%), and "severe" (1%) symptoms concerning the 15 questions.

As shown by Table 1 in Supplementary Material 4, higher positive scores were noted only on four of the questionnaire items concerning the oculomotor side effect: eyestrain, difficulty focusing, "Fullness of the head", and blurred vision.

Given the majority of zeros (none side effects), we analysed the number of simulator sickness symptoms each participant felt. This being a count data, we applied the zero-adjusted negative binomial model of GAM (Generalized Additive Models) technique (Wood, 2017). Results revealed that the zero frequency scores of the NT group differs significantly

from Full (two-tailed $t(157) = 2.083,\ p = 0.03885$). For the other groups there were no differences from the Full. The model estimates the percentage of zeros of NT group is 0.356 which is more than double the percentage of zeros in the Full group (0.167), similar to the descriptive results. Detailed analyses are in Appendix E.

5 Discussion

This study extends our preliminary studies proposed to train the student in administering inferior alveolar nerve anesthesia in the immersive VR simulator VIDA Odonto® (TORI et al., 2016; TORI et al., 2018). We had shown its viability using an infrared sensor device (Leap Motion) attached to the HMD for tracking a cylindrical object mimicking the syringe's needle (TORI et al., 2016; TORI et al., 2018). However, the infrared sensor presented instabilities and low precision. Moreover, the tip of the needle would sometimes be occluded by the back of the hand, and may fall outside the controller's field of view. This limitation made it extremely hard to accurately detect the syringe's trajectory. Another limitation was the absence of force feedback. In order to tackle those problems we included in VIDA Odonto® simulator a haptic device, which is also a very precise movement tracker. With the possibility to capture precise data we could use the haptic VR Simulator to test the effects of immersion on performance tracked by a haptic device that takes into account the haptic perceptions. The experiment was designed to evaluate the impact of visual immersion in both explanation and training phases, and of haptic force feedback on inferior alveolar nerve anesthesia performance. The variables associated with the measures were statistically analysed according to their variable type, and Hotelling's T^2 method was proposed for machine learning. The analysis showed clear performance advantages for immersed groups on both explanation and training phases compared to the combined immersed groups.

Compared to the combined immersive groups, when one of the phases were in non-immersive conditions, the insertion point of the participants in the immersed groups were closer, in mean, to the reference insertion point, and also the immersed groups performed the anesthetic procedure faster, meaning that the participants in the immersed groups had more accuracy, and felt more confidence in administering the anesthesia. These results were expected because the HMD allowed immersive groups to experience a free angle of view and hand positioning inside the virtual patient's mouth. These conditions are closer to the real environment and allowed them to better explore the patients' mouth and to better plan the hand movement needed to perform the procedure. Also, the HMD appears to have allowed better depth cues about the needle orientation in the z-direction both in explanation and training phases.

As expected, the 95% confidence ellipse for the (population) mean insertion point coordinates constructed with the students insertion data contains the instructor reference insertion point coordinate, meaning that, in general, the students followed the instructor's explanation of the needle tip movement towards the insertion target. Also, T^2-statistic revealed good performance for machine learning in classifying the students into "successful"

and "failure" for needle insertion point.

Concerning the needle angle, all groups had similar performance, as expected, since the anesthetic procedure is based on a straight movement of the syringe. The main difficulty in administering the anesthesia relies on the proximity of the needle tip to the mandibular foramen at the moment of the injection into the pterygomandibular region, which can be solved if the instructor's trajectory has been followed by the students.

Another important finding is that those who performed under non-immersive training condition reported feeling slightly more difficulty in controlling the syringe than those in immersive training groups. This result was expected since, usually, when students use a monitor-based simulation, they have to keep looking at the monitor while training the movement for anesthesia administration, and this kind of simulation offers less realism than the immersive condition.

We also investigated the avatar embodiment effects (GONZALEZ-FRANCO; PECK, 2018), a paradigm derived from the classical rubber hand illusion (BOTVINICK; COHEN, 1998) whereby subjects in an experiment reported visual and tactile sensations when experiencing a rubber hand as belonging to themselves. The same principles have been demonstrated within the VR, when the appearance and the agency of the avatar may enhance the embodiment illusion (ARGELAGUET et al., 2016; KONDO et al., 2018). In our study the immersive groups reported feeling the virtual hand as being their own hand even though the hand model was a non-animated gloved hand. Considering that the hand movements required for the dental technique applied in this study, were slow and controlled, it was sufficient for the virtual non-animated hand ownership illusion. Those in the NT (non-immersive training) group reported lower scores on the *virtual hand ownership* factor suggesting that lower levels of immersion in the training phase negatively impact the perception of motor control over the avatar hand. The literature provides evidence that the projected hand, when synchronized with the real hand, is beneficial in motor tasks. For example, in the biopsy procedure Benyahia et al. (2015) and Nguyen, Lakhal e Chellali (2015) showed that the non-animated virtual hand increased the participant's feeling of accuracy when manipulating the virtual syringe. The authors also argued that the virtual hand increases the depth perception providing additional spatial cues to the trainers in the virtual environment. Moreover, Raz, Weiss e Reiner (2008) found that the haptic feedback greatly contributes to the inducement of the virtual hand, when both the real and the virtual hand moved exactly at the same time during a brushing task. Therefore, we conjectured that the projected hand can contribute to a higher level of immersion, but in some circumstances when there are several changes in finger positions, the perception of the projected hand may be altered, and the combination of maximum synchronization using a technology that tracks with accuracy hand's movements should be considered.

The Simulator Sickness Questionnaire results showed that there was a very little incidence of simulator sickness, and the few incidences were reported by those in immersed groups concerning the oculomotor factor. This is not surprising because symptoms of simulator sickness are often similar to those of motion sickness, but affect a smaller proportion of the population and are usually much less severe (KENNEDY et al., 1993). Moreover, the results reveal that the percentage of zeros scores (no side effects) of the NT group is more than twice the percentage of zeros of Full group.

Contrary to our expectations, no significant differences among groups were observed in the needle depth. We expected that participants in NH group would present difficulty in penetration depth of the needle, due to the lack of cutting forces when inserting the needle in the tissue and the frictional force during the needle sliding inside the tissue being penetrated. Also, the analyses of the Syringe Handling Questionnaire revealed no significant differences among groups in *tactile realism* factor. Thus, even though the haptic feedback was off for the participants in the NH group, they manifested agreement in favor of the tactile realism factor similar to the other groups with haptic feedback. These results suggest that, in our study, the immersion vision dominated the touch perception and possibly created an illusion of haptic feedback. Previous studies about visual-haptic perception have been conducted when judging the size, shape and position of the objects. For example, Rock e Victor (1964), participants have grasped a square object when looking at it through an optical distorting lens that made it appear rectangular. The participants reported the object being rectangular, and the authors concluded that the shape perception was determined almost completely by vision. Moreover, they argued that, in particular cases, "*the vision is so powerful in relation to touch that the very touch experience itself undergoes a change*". However, in some circumstances the perception may be clearly affected by haptics (ERNST; BANKS, 2002; FAIRHURST et al., 2018). Therefore, in our particular case, we conjectured that having a real syringe in their hands, provided by the tangible interface, had highly influenced the students' visual perception over their haptic perception. Hence participants on the NH group were able to perform similarly to participants that received haptic feedback on the depth of the needle penetration, and had reported feeling the *tactile realism*.

The literature review in Chapter 2 revealed that there is a paucity of immersive dental anesthesia training studies using HMD, as evidenced from a recent systematic review of VR training in dental medicine (JODA et al., 2019). There are only two studies[1] [2] in the literature using haptic VR simulator for dental anaesthesia. However, these systems are

[1] CORRÊA, C. G.; NUNES, F. d. L. dos S.; TORI, R. Virtual reality-based system for training in dental anesthesia. In: SPRINGER. International Conference on Virtual, Augmented and Mixed Reality, 2014. p. 267–276

[2] POYADE, M.; LYSAKOWSKI, A.; ANDERSON, P. Development of a haptic training simulation for the administration of dental anaesthesia based upon accurate anatomical data. In: Conference and Exhibition of the European Association of Virtual and Augmented Reality. EuroVR2014, 2014

not fully immersive since the anatomical model is displayed on a 2D monitor. The only full immersive simulator found until now is the VR Dental Anesthesia Simulator[3] developed by the University of Alberta. However, it provides low tactile realism, since the needle insertion is made by holding HMD traditional hand controllers, that are shaped very differently from a syringe and do not allow the hand motor skill training required for the procedure. Moreover, there is no literature that has reported a validation of this system. The majority of the found studies was for drill training using Simodont®, Virteasy® and Voxel-Man Dental® in which the students manipulate the physical instrument in an empty space underneath the 3D display viewer or in a simulated teeth arrangement attached to a phantom human head, while the feedback information is displayed on a 2D or 3D monitor.

Considering the performance assessment, a systematic review in medical and dental applications of haptic interaction for needle insertion training developed by Corrêa et al. (2018) showed that most studies are based on subjective tests, and most of the experiments involved a range from 10 to 50 participants. Another recent systematic review on learning performance in a general scope using HMD (QUEIROZ et al., 2019) showed that assessing student's performance is difficult and has been done mainly subjectively. The authors also identified a paucity of machine learning methods applied in immersive learning performance. Moreover, according to a recent systematic review in Dias, Gupta e Yule (2019) the machine learning applied on performance assessments are mainly based on medical surgery. It was however, observed from this review that there is a scarcity of studies in dental performance assessment. For example, Rhienmora et al. (2011) developed a VR non-immersive dental simulator using hidden Markov models (HMMs) to analyze crown preparation data. However, studies in immersive dental anesthesia simulators remain unclear.

Our VR simulator and method to assess performance is different from the majority of the studies in the literature in several ways. First, in our experiment the sample size is unusually large, providing (statistically) powerful results. Also unusual, most of our data consists of objective variables recorded from movement data tracking provided by the system (full access to our data set is provided). Both objective and subjective variables were analysed using appropriate statistical methods. Some of the statistical techniques applied are advanced models that have been developed in the past few decades. Second, we proposed and tested a machine learning method for automatic evaluation of the dentistry student's performance, that is an innovative methodology for needle insertion assessment in VR-based dental training. Third, our simulator is fully immersive, and we evaluated immersion impacts using HMD on both explanation and training phases. The operative field is projected directly on student's eyes. Therefore, students can operate completely within the virtual dental room, and attention does not need to be alternated between the

[3] https://cognitiveprojections.ca/

operative field and the feedback information displayed on a 2D/3D monitor.

However, we identify some limitations of this study. The haptic feedback was only based on the cutting force and the frictional to inserting the needle to the proper depth. If the needle were to be rotated involving 6DOF (translation and rotation) inside the penetrable tissue, for example if the indirect technique of anaesthesia administration was implemented, then this should lead to haptic stimuli dominating visual stimuli. We conjecture that, in this situation, a group training with the haptic feedback set off would be different from other groups with respect to haptic realism and needle depth.

Currently, during preclinical dental anesthesia training, the students used to practice on human subjects as learning models (KARY et al., 2018). This traditional training method faces ethical and patient safety issues. Thus, it is important to investigate the immersion impact on learning to improve current dental training and reduce the need for human subjects for dental practice training. As HMD technology becomes more price competitive, training in an immersive simulator may be a safe platform that allows anatomical variation and physiological replication models (CORRÊA et al., 2018). Our analysis provide robust evidence that immersion positively impacts the performance in inferior alveolar anesthesia training, on both explanation and training phases. These results bring us one step closer to understanding what the impact of HMDs are on virtual anesthesia training before their adoption at scale, and insights into how needle insertion anesthesia performance can be assessed automatically by machine learning.

6 Conclusion and Future Work

In this work, we conducted an experimental study to evaluate the impact of immersion on the performance of the administering of inferior alveolar nerve anesthesia in the immersive VR simulator VIDA Odonto®. The variables associated with the performance's measurements were statistically analysed according to their variable type, and Hotelling's T^2 method was proposed for machine learning.

In conclusion, immersed groups on both explanation and training phases showed clear performance advantages compared to the combined immersed groups.

The insertion point of the participants in the immersed groups were closer, on average, to the reference insertion point, and also the immersed groups performed the anesthetic procedure faster.

We implemented an experimental design to evaluate the impact of immersion on inferior alveolar nerve anesthesia performance. However, there is no validated experimental design in the literature to test the impact of immersion on motor task performance. Therefore, we encourage the application of the experimental design implemented in this study in any other HMD-based VR simulators that involves performance analysis.

In our experiment the sample size ($n = 163$) is large, providing (statistically) powerful results. Also unusual, most of our data consists of objective variables recorded from movement data tracking provided by the system.

This study provides evidence that the dental students' needle insertion point can be assessed by using the Hotelling's T^2 statistical method, and we consider this a novel application for machine learning. For needle insertion point, we set a monitoring region for future observations based on the prediction ellipse. Based on interviews with experts we decided to first implement the machine learning for insertion point coordinates because they reported the needle insertion point to be one of the most important factors for the success of the direct anesthesia technique;

Another important finding is that those who performed under non-immersive training conditions reported feeling slightly more difficulty in controlling the syringe than those in the immersive training groups.

The subjective evaluation suggested that immersive vision dominated the touch perception and possibly created an illusion of haptic feedback. Participants also perceived a high sense of the virtual hand ownership when handling the syringe.

Future research should consider: a depressable syringe plunger; a record of the fluid volume deposited and velocity of application; animated virtual patient; animated virtual

left hand for stretching the cheek; anatomical variation; gamification; and HMD with no wires.

Concerning the machine learning, futures works should consider:

- using the dataset recorded from this work to propose and develop new machine learning methods. New methods can be compared quantitatively, with the Hotelling's T^2 developed in this work, by the total area under their receiver operator characteristic curves.
- temporal series methods for the needle's trajectory;
- the joint behaviour of multiple variables (needle insertion point, needle angle, and needle depth). This should provide better results if the anesthesia administration performance advantages is influenced by more than one variable.
- providing formative online feedback to preclinical dental students during their practice on the VIDA Odonto® simulator, using the machine learning developed in this work.

With further development, the machine learning used in this work could become a viable and a low cost tool for objective assessment of needle insertion accuracy. It should facilitate the objective evaluations of the anesthesia technique at scale. Moreover, the method presented in this work is viable for any other needle insertion task.

Bibliography

AL-SAUD, L. M. et al. Feedback and motor skill acquisition using a haptic dental simulator. *European Journal of Dental Education*, Wiley Online Library, v. 21, n. 4, p. 240–247, 2017. Cited on page 11.

ARGELAGUET, F. et al. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In: IEEE. *2016 IEEE Virtual Reality (VR)*. [S.l.], 2016. p. 3–10. Cited on page 43.

BAKKER, D. et al. Transfer of manual dexterity skills acquired on the simodont, a dental haptic trainer with a virtual environment, to reality. a pilot study. *Bio-algorithms and Med-systems*, v. 6, n. 11, p. 21–24, 2010. Cited on page 2.

BAKR, M.; MASSEY, W.; ALEXANDER, H. Can virtual simulators replace traditional preclinical teaching methods: a students' perspective. *Int. J. Dent. Oral Health*, v. 2, n. 1, 2015. Cited on pages 2 and 12.

BEN-GAL, G. et al. Testing manual dexterity using a virtual reality simulator: reliability and validity. *European Journal of Dental Education*, Wiley Online Library, v. 17, n. 3, p. 138–142, 2013. Cited on page 11.

BENYAHIA, S. et al. Designing the user interface of a virtual needle insertion trainer. In: ACM. *Proceedings of the 27th Conference on l'Interaction Homme-Machine*. [S.l.], 2015. p. 18. Cited on page 43.

BOTVINICK, M.; COHEN, J. Rubber hands 'feel'touch that eyes see. *Nature*, Nature Publishing Group, v. 391, n. 6669, p. 756, 1998. Cited on page 43.

BRAND, H. S. et al. Effect of a training model in local anesthesia teaching. *Journal of dental education*, Am Dental Educ Assoc, v. 74, n. 8, p. 876–879, 2010. Cited on pages 1 and 3.

CARVALHO, M. R. d.; COSTA, R. T. d.; NARDI, A. E. Simulator sickness questionnaire: tradução e adaptação transcultural. *J Bras Psiquiatr*, v. 60, n. 4, p. 247–252, 2011. Cited on page 23.

CHEN, L.-Y. et al. A dental training system using virtual reality. *Proceedings 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation. Computational Intelligence in Robotics and Automation for the New Millennium (Cat. No.03EX694)*, v. 1, p. 430–434 vol.1, 2003. Cited on page 6.

CHEN, X.; SUN, P.; LIAO, D. A patient-specific haptic drilling simulator based on virtual reality for dental implant surgery. *International journal of computer assisted radiology and surgery*, Springer, v. 13, n. 11, p. 1861–1870, 2018. Cited on page 2.

CORRÊA, C. G.; NUNES, F. d. L. dos S.; TORI, R. Virtual reality-based system for training in dental anesthesia. In: SPRINGER. *International Conference on Virtual, Augmented and Mixed Reality*. [S.l.], 2014. p. 267–276. Cited on pages 2, 9, and 10.

CORRÊA, C. G. et al. Haptic interaction for needle insertion training in medical applications: The state-of-the-art. *Medical engineering & physics*, Elsevier, 2018. Cited on pages 1, 10, 13, 45, and 46.

CUMMINGS, J. J.; BAILENSON, J. N. How immersive is enough? a meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, Taylor & Francis, v. 19, n. 2, p. 272–309, 2016. Cited on page 6.

DIAS, R. D.; GUPTA, A.; YULE, S. J. Using machine learning to assess physician competence: A systematic review. *Academic Medicine*, LWW, v. 94, n. 3, p. 427–439, 2019. Cited on pages 12 and 45.

DIEGO, J. P. S. et al. Researching haptics in higher education: The complexity of developing haptics virtual learning systems and evaluating its impact on students' learning. *Computers & Education*, Elsevier, v. 59, n. 1, p. 156–166, 2012. Cited on page 2.

ERNST, M. O.; BANKS, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, Nature Publishing Group, v. 415, n. 6870, p. 429, 2002. Cited on page 44.

FAIRHURST, M. T. et al. Confidence is higher in touch than in vision in cases of perceptual ambiguity. *Scientific reports*, Nature Publishing Group, v. 8, n. 1, p. 15604, 2018. Cited on page 44.

FARD, M. J. et al. Machine learning approach for skill evaluation in robotic-assisted surgery. *arXiv preprint arXiv:1611.05136*, 2016. Cited on page 13.

GOKHALE, N. S. et al. Alternative approaches for inferior alveolar nerve technique in children: A review. *Acta Scientific Dental Sciences*, v. 3, p. 10–16, 2019. Cited on page 16.

GONZALEZ-FRANCO, M.; PECK, T. C. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, Frontiers, v. 5, p. 74, 2018. Cited on pages 22, 23, and 43.

JASINEVICIUS, T. R. et al. An evaluation of two dental simulation systems: virtual reality versus contemporary non-computer-assisted. *Journal of dental education*, Am Dental Educ Assoc, v. 68, n. 11, p. 1151–1162, 2004. Cited on page 11.

JODA, T. et al. Augmented and virtual reality in dental medicine: A systematic review. *Computers in biology and medicine*, Elsevier, 2019. Cited on pages 7 and 44.

JOHNSON, R. A.; WICHERN, D. W. et al. *Applied multivariate statistical analysis*. [S.l.]: Prentice hall Upper Saddle River, NJ, 2002. v. 5. Cited on pages 25, 28, 32, 38, and 39.

KARY, A. L. et al. Preclinical local anesthesia education in dental schools: A systematic review. *Journal of dental education*, Am Dental Educ Assoc, v. 82, n. 10, p. 1059–1064, 2018. Cited on pages 1, 3, and 46.

KENNEDY, R. S. et al. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, Taylor & Francis, v. 3, n. 3, p. 203–220, 1993. Cited on pages 23, 40, and 44.

KHALIL, H. A basic review on the inferior alveolar nerve block techniques. *Anesthesia, essays and researches*, Wolters Kluwer–Medknow Publications, v. 8, n. 1, p. 3, 2014. Cited on page 16.

KLEIN, J. P.; MOESCHBERGER, M. L. *Survival analysis: techniques for censored and truncated data*. [S.l.]: Springer Science & Business Media, 2006. Cited on pages 23 and 56.

KOLESNIKOV, M. et al. Periosim: Haptic virtual reality simulator for sensorimotor skill acquisition in dentistry. In: IEEE. *2009 IEEE International Conference on Robotics and Automation*. [S.l.], 2009. p. 689–694. Cited on page 2.

KONDO, R. et al. Illusory body ownership of an invisible body interpolated between virtual hands and feet via visual-motor synchronicity. *Scientific reports*, Nature Publishing Group, v. 8, n. 1, p. 7541, 2018. Cited on page 43.

KRZANOWSKI, W. J.; HAND, D. J. *ROC curves for continuous data*. [S.l.]: Chapman and Hall/CRC, 2009. Cited on pages 25, 26, and 27.

LEONG, J. J. et al. Hmm assessment of quality of movement trajectory in laparoscopic surgery. *Computer Aided Surgery*, Taylor & Francis, v. 12, n. 6, p. 335–346, 2007. Cited on page 13.

MALAMED, S. F. *Handbook of local anesthesia*. [S.l.]: Elsevier Health Sciences, 2004. Cited on page 16.

MAZOMENOS, E. B. et al. Catheter manipulation analysis for objective performance and technical skills assessment in transcatheter aortic valve implantation. *International journal of computer assisted radiology and surgery*, Springer, v. 11, n. 6, p. 1121–1131, 2016. Cited on page 13.

MESTRE, D. R. Cave versus head-mounted displays: Ongoing thoughts. *Electronic Imaging*, Society for Imaging Science and Technology, v. 2017, n. 3, p. 31–35, 2017. Cited on page 6.

MIRGHANI, I. et al. Capturing differences in dental training using a virtual reality simulator. *European Journal of Dental Education*, Wiley Online Library, v. 22, n. 1, p. 67–71, 2018. Cited on page 2.

MONTGOMERY, D. C. *Design and analysis of experiments*. [S.l.]: John wiley & sons, 2017. Cited on page 24.

NETER, J. et al. *Applied linear statistical models*. [S.l.]: Irwin Chicago, 1996. v. 4. Cited on page 24.

NGUYEN, D. V.; LAKHAL, S. B.; CHELLALI, A. Preliminary evaluation of a virtual needle insertion training system. In: IEEE. *2015 IEEE Virtual Reality (VR)*. [S.l.], 2015. p. 247–248. Cited on page 43.

PEREIRA, L. A. P. et al. Criação, desenvolvimento, aplicação e validação de um simulador computadorizado de realidade virtual para o ensino e treinamento de bloqueio do nervo alveolar inferior. Doctoral dissertation. Universidade Estadual de Campinas - UNICAMP, 2016. Cited on page 10.

POYADE, M.; LYSAKOWSKI, A.; ANDERSON, P. Development of a haptic training simulation for the administration of dental anaesthesia based upon accurate anatomical data. In: *Conference and Exhibition of the European Association of Virtual and Augmented Reality*. [S.l.]: EuroVR2014, 2014. Cited on pages 2, 9, and 10.

PULIJALA, Y. et al. Effectiveness of immersive virtual reality in surgical training—a randomized control trial. *Journal of Oral and Maxillofacial Surgery*, Elsevier, v. 76, n. 5, p. 1065–1072, 2018. Cited on page 1.

QUEIROZ, A. C. M. et al. Immersive virtual environments and learning assessments. In: SPRINGER. *International Conference on Immersive Learning*. [S.l.], 2019. p. 172–181. Cited on pages 1, 10, and 45.

RAZ, L.; WEISS, P. L.; REINER, M. The virtual hand illusion and body ownership. In: SPRINGER. *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. [S.l.], 2008. p. 367–372. Cited on page 43.

RENCHER, A. C. *A review of "Methods of Multivariate Analysis, "*. [S.l.]: Taylor & Francis, 2005. Cited on page 25.

RHIENMORA, P. et al. Haptic augmented reality dental trainer with automatic performance assessment. In: *IUI*. [S.l.: s.n.], 2010. Cited on page 7.

RHIENMORA, P. et al. A virtual reality simulator for teaching and evaluating dental procedures. *Methods of information in medicine*, Schattauer GmbH, v. 49, n. 04, p. 396–405, 2010. Cited on page 12.

RHIENMORA, P. et al. Intelligent dental training simulator with objective skill assessment and feedback. *Artificial intelligence in medicine*, Elsevier, v. 52, n. 2, p. 115–121, 2011. Cited on pages 2, 12, and 45.

ROCK, I.; VICTOR, J. Vision and touch: An experimentally created conflict between the two senses. *Science*, American Association for the Advancement of Science, v. 143, n. 3606, p. 594–596, 1964. Cited on page 44.

ROY, E.; BAKR, M. M.; GEORGE, R. The need for virtual reality simulators in dental education: A review. *The Saudi dental journal*, Elsevier, v. 29, n. 2, p. 41–47, 2017. Cited on page 8.

SLATER, M. A note on presence terminology. *Presence connect*, Citeseer, v. 3, n. 3, p. 1–5, 2003. Cited on pages 1 and 6.

STEINBERG, A. D. et al. Assessment of faculty perception of content validity of periosim©, a haptic-3d virtual reality dental training simulator. *Journal of Dental Education*, Am Dental Educ Assoc, v. 71, n. 12, p. 1574–1582, 2007. Cited on pages 2, 8, 9, and 12.

TORI, R. et al. Treinamento odontológico imersivo por meio de realidade virtual. In: *Brazilian symposium on computers in education (simpósio brasileiro de informática na educação-sbie)*. [S.l.: s.n.], 2016. v. 27, n. 1, p. 400. Cited on pages 1, 3, 4, 14, and 42.

TORI, R. et al. Vida odonto: Virtual reality environment for dental training. *Brazilian Journal of Computers in Education*, v. 26, n. 02, p. 80, 2018. Cited on pages 1, 3, 4, 14, and 42.

TOWERS, A. et al. A scoping review of the use and application of virtual reality in pre-clinical dental education. *British dental journal*, Nature Publishing Group, v. 226, n. 5, p. 358–366, 2019. Cited on pages 1, 3, 6, and 7.

TSE, B. et al. Design and development of a haptic dental training system-haptel. In: SPRINGER. *International conference on human haptic sensing and touch enabled computer applications*. [S.l.], 2010. p. 101–108. Cited on page 2.

ULLRICH, S.; KUHLEN, T. Haptic palpation for medical simulation in virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, IEEE, v. 18, n. 4, p. 617–625, 2012. Cited on page 22.

VAPNIK, V. *The nature of statistical learning theory*. [S.l.]: Springer science & business media, 2013. Cited on page 12.

WATSON, R. A. Use of a machine learning algorithm to classify expertise: analysis of hand motion patterns during a simulated surgical task. *Academic Medicine*, LWW, v. 89, n. 8, p. 1163–1167, 2014. Cited on page 27.

WOOD, S. N. *Generalized additive models: an introduction with R*. [S.l.]: Chapman and Hall/CRC, 2017. Cited on page 29.

ZHENG, G. et al. Computer-assisted preoperative planning and surgical navigation system in dental implantology. In: IEEE. *2007 6th international special topic conference on information technology applications in biomedicine*. [S.l.], 2007. p. 139–142. Cited on page 2.

Appendix

APPENDIX A – Statistical analyses for the objective measurements

In this supplementary material the variables associated with the measurements were statistically analyzed according to their variable type. The results for the task execution time are in Section A.1, for the insertion accuracy (distance from insertion point to the reference point) and for the insertion point coordinates are in Section A.2, for the needle depth are in Section A.3, and for the needle angle (deviation of the needle from reference direction) are in Section A.4. The adopted significance level is 0.05.

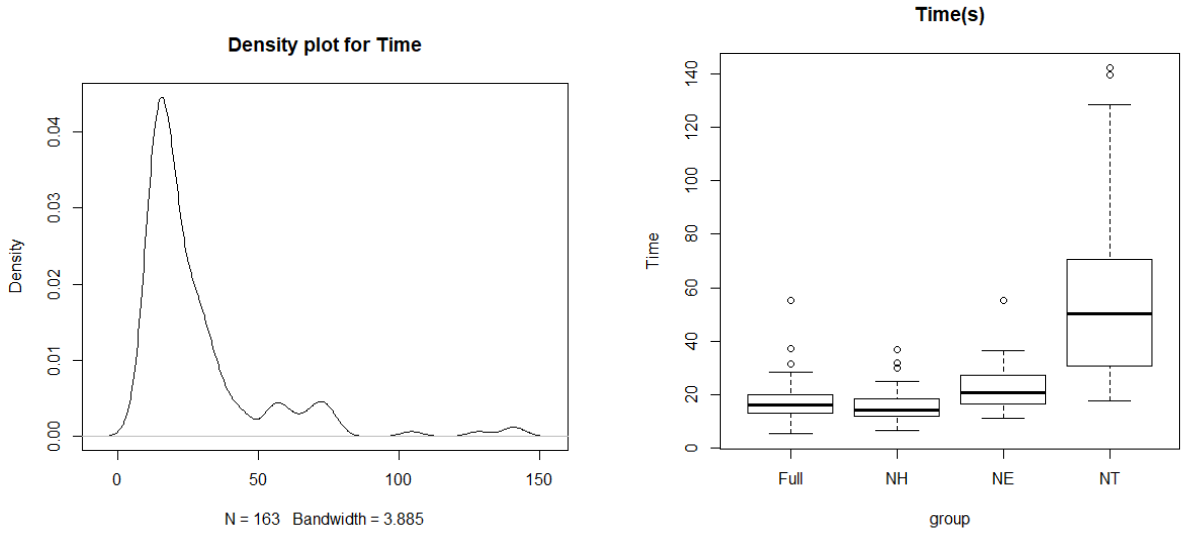
A.1 Task execution time

The Table 4 shows the descriptive statistics of the task execution time by group. The respective boxplot is in Figure 17.

Table 4 – Descriptive statistics of the task execution time in seconds by group.

Group	mean	std dev	min	1st quartile	median	3rd quartile	max
Full	18.37	8.49	5.38	13.31	16.30	19.99	55.45
NH	16.14	6.23	6.77	11.93	14.27	18.57	36.75
NE	22.43	8.57	11.12	16.68	20.72	27.23	55.27
NT	55.38	30.99	17.58	31.01	50.25	70.06	142.00

From the Table 4 and Figure 17 we noted that the participants in group NT took longer, on average, to complete the procedure and the variability in this group is much higher compared to the others. Also, we can see, by Figure 17, that the distribution of the time, all groups together, is asymmetric.



(a) **(b)**

Figure 17 – Density and boxplot by group for task execution time. **(a)** Density **(b)** Boxplot

The Figure 10 in Results section presents the Kaplan-Meier estimates for the probability of task execution against time (KLEIN; MOESCHBERGER, 2006).

By this figure and the descriptive results, the immersive (NH and Full) groups behave somewhat similarly, with higher estimated probabilities of task completion. The participants in the NT group took longer to complete the anesthesia procedure, with noticeable difference compared to the other three groups. Apparently, the NE group could also be distinguished from NH and Full groups.

For statistically testing the behavior described, the Cox proportional hazards model was fitted considering Full group as reference. The corresponding Wald test showed that there were significant differences among groups (one-sided $\chi^2(3) = 79.3, p < 0.00001$). The results comparing each group with the reference (Full group) yielded that the NH group is similar to the Full (one-sided $\chi^2(1) = 2.23, p = 0.1354$) and for the NE group we obtained one-sided $\chi^2(1) = 3.52, p = 0.0606$, which is a borderline result. Therefore, we fitted a reduced Cox model considering as reference the Full and NH groups together. The results given in Table 5 indicate that the immersive groups (Full and NH together) are significantly different from the NE group (one-sided $\chi^2(1) = 8.48, p = 0.0036$), and also different from the NT group (one-sided $\chi^2(1) = 76.801, p < 0.00001$), as expected.

Table 5 – Wald test results for Cox Model with (Full and NH) groups as reference for task execution time.

Group	coef	exp(coef)	SE(coef)	$\chi(1)$	p-value
NE	-0.5674	0.5672	0.1947	8.480	0.00359
NT	-2.3110	0.0992	0.2637	76.808	2e-16

The estimates from the Cox model indicate that the probability of task completion, for any given time, of the NE group is 56.7% of the probability for the Full and NH groups. The probability of task completion, for any given time, of the NT group is 9.9% of the probability for the Full and NH groups.

A.2 Needle insertion accuracy

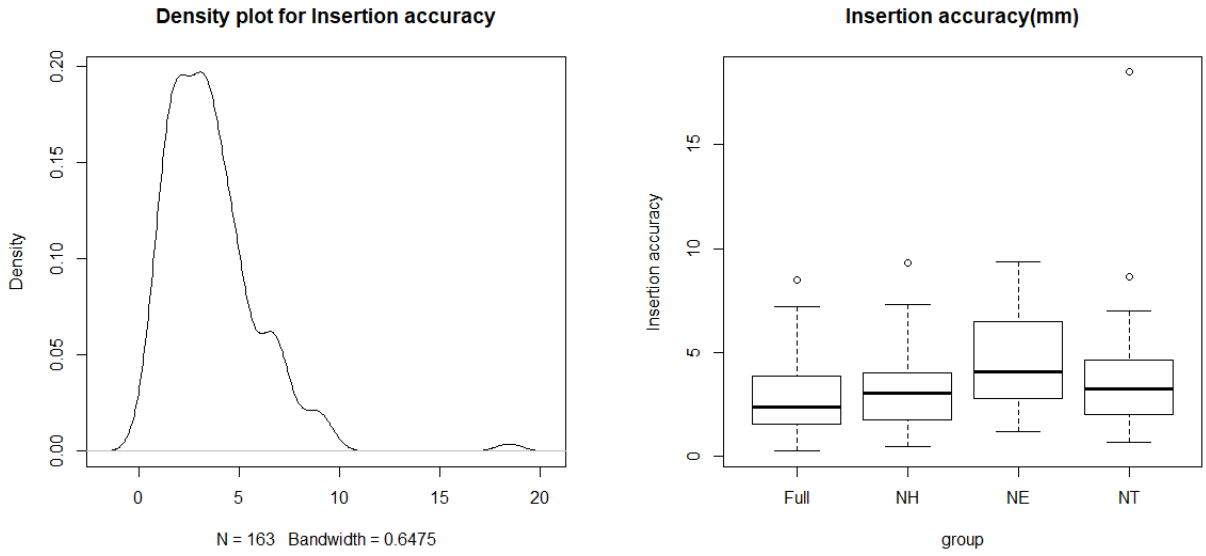
The Table 6 shows the descriptive statistics of the needle puncture accuracy (to a reference point) by group. The respective boxplot is in Figure 18.

Table 6 – Descriptive statistics of the needle insertion accuracy in millimetres by group.

Group	mean	std dev	min	1st quartile	median	3rd quartile	max
Full	2.86	1.83	0.25	1.55	2.39	3.80	8.49
NH	3.25	1.92	0.50	1.74	3.04	4.03	9.30
NE	4.54	2.21	1.18	2.85	4.06	6.47	9.36
NT	3.93	2.97	0.68	2.03	3.26	4.65	18.48

We noticed by Table 6 and the boxplot of Figure 18 that the NE group had a higher mean and median puncture accuracy, meaning that the participants in this group had less precision with respect to the reference puncture point. One participant in NT group presented an outlier value for the puncture accuracy. Also, the puncture accuracy data is asymmetric, therefore, for inferential purposes we consider the logarithm transformation of the puncture accuracy.

The result of the ANOVA for the log puncture accuracy indicated that there is significant difference among groups ($F(3, 159) = 5.81, p = 0.00086$). We considered contrasts to detect the differences between groups. The results indicated that the NH group does not differ from the Full group ($t(81) = 1.26$, adjusted $p = 0.4507$). There are significant differences between Full and NT group ($t(80) = 2.494$, adjusted $p = 0.0367$) and between Full and NE group ($t(80) = 3.987$, adjusted $p < 0.001$). Table 7 presents a summary of the contrast testing.



(a) **(b)**

Figure 18 – Density and boxplot by group for needle insertion accuracy. **(a)** Density **(b)** Boxplot

Table 7 – Contrast results for the log puncture accuracy

Contrast testing	Contrast estimate	Standard Error	$t-$value	adjusted $p-$value
NE−Full = 0	0.5613	0.1408	3.987	< 0.001
NH−Full = 0	0.1763	0.1399	1.260	0.4507
NT−Full = 0	0.3512	0.1408	2.494	0.0367

The log puncture accuracy data satisfied the normality (Shapiro-Wilk, $W = 0.98538$, $p = 0.08491$) and homocedasticity (Bartlett $K^2 = 4.3209, p = 0.2288$) assumptions.

These results confirmed our descriptive analysis, indicating that the non-immersive (NT and NE) groups had less precision and accuracy with respect to the reference puncture point, since their mean and standard deviation are significantly higher than those of the immersive (Full and NT) groups.

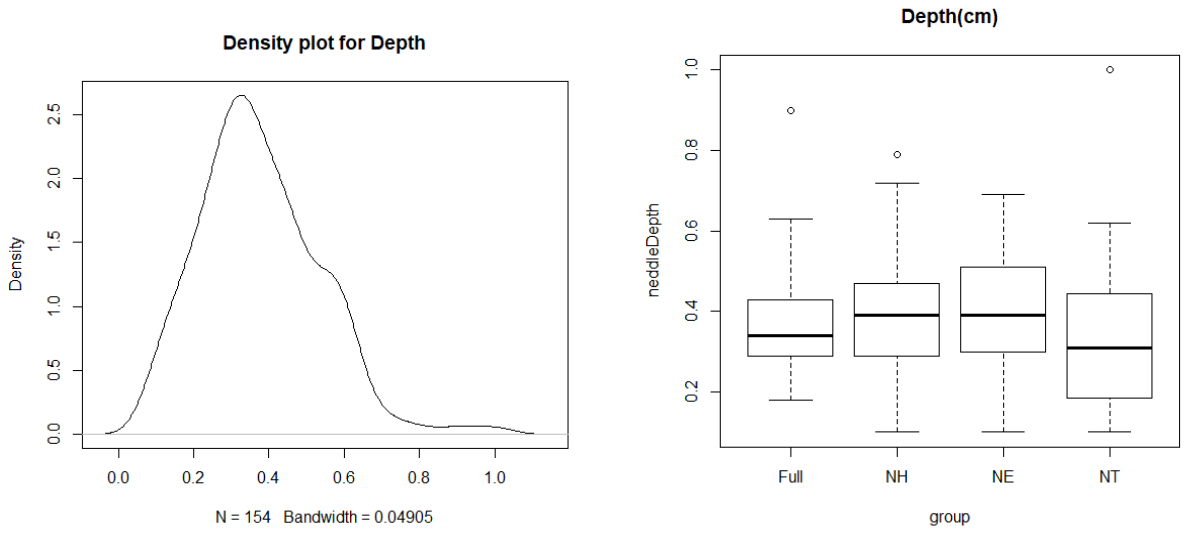
A.3 Depth of the penetration of the needle

The actual needle length value is 2.5 cm and the needle depth is recorded by calculating the amount of the needle that was inserted from the tip of the needle up to the tissue surface. This depth value is calculated by the haptic device, and it is normalized between 0 and 1. Therefore, the maximum value of needle depth is 1.

By Table 8 and the boxplot in Figure 19 we noticed that the descriptive statistics of all groups are very similar, not considering the few outlier values. Also, by the density plot of Figure 19 we can see that the distribution of the needle depth is asymmetric.

Table 8 – Descriptive statistics of the needle depth by group.

Group	mean	std dev	min	1st quartile	median	3rd quartile	max
Full	0.37	0.14	0.18	0.29	0.34	0.43	0.90
NH	0.40	0.16	0.10	0.29	0.39	0.47	0.79
NE	0.40	0.14	0.10	0.30	0.39	0.51	0.69
NT	0.34	0.19	0.10	0.19	0.31	0.45	1.00



(a) **(b)**

Figure 19 – Density and boxplot by group for needle depth. **(a)** Density **(b)** Boxplot

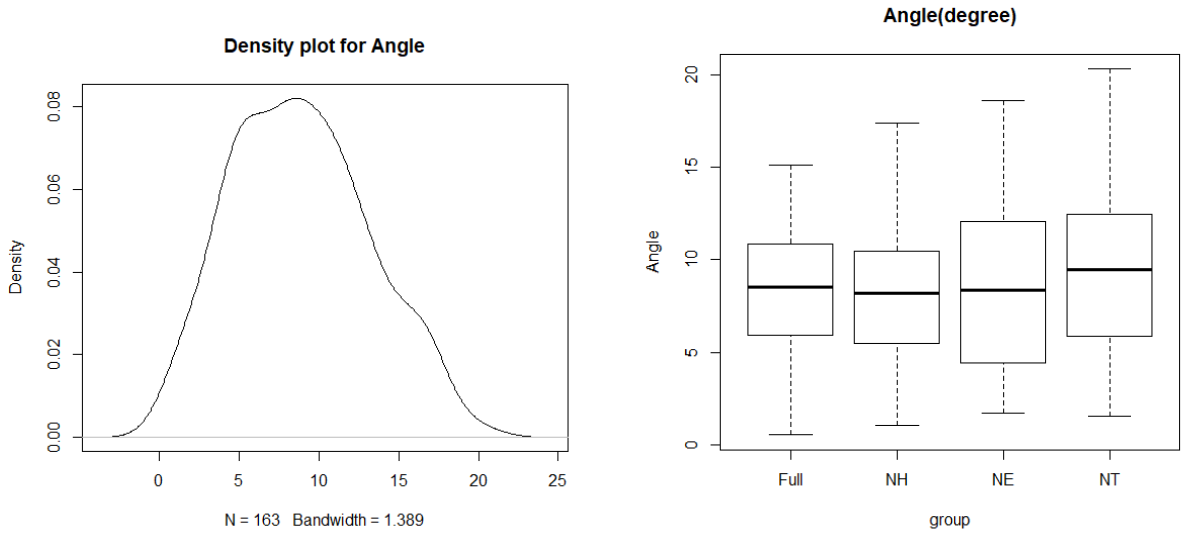
Since the log transformation was not able to stabilize the variances (Bartlett $K^2 = 11.002$, $p = 0.01171$) and the normality test yielded a borderline result (Shapiro-Wilk $W = 0.98423, p = 0.0765$), we decided to apply the nonparametric Kruskal-Wallis test and the results indicated that there is no difference among groups median ($\chi^2(3) = 4.168, p = 0.2438$). This result is compatible with the descriptive analysis.

A.4 Needle angle

The Table 9 shows the descriptive statistics of the needle angle by group. The respective boxplot is in Figure 20.

Table 9 – Descriptive statistics of the needle angle in degrees by group.

Group	mean	std dev	min	1st quartile	median	3rd quartile	max
Full	8.61	3.61	0.52	5.97	8.51	10.83	15.12
NH	8.33	4.06	1.01	5.48	8.20	10.46	17.41
NE	8.84	4.85	1.71	4.55	8.37	10.83	15.12
NT	9.81	4.53	1.55	6.02	9.47	12.38	20.32



(a) **(b)**

Figure 20 – Density and boxplot by group for needle angle. **(a)** Density **(b)** Boxplot

In Table 9 and the boxplot in Figure 20 we noticed that all groups have similar mean and median needle angle values. Even the dispersion measures seem to be similar. The data distribution seemed somewhat symmetric.

The ANOVA result indicated that there is no mean difference among groups ($F(3, 159) = 0.908, p = 0.439$), corroborating the descriptive analysis.

The angle data satisfied the normality (Shapiro-Wilk, $W = 0.9839, p = 0.05569$) and homoscedasticity (Bartlett $K^2 = 3.8847, p = 0.2742$) ANOVA assumptions.

APPENDIX B – Confidence Region and Prediction Region

Let $(X_1, \ldots, X_p)^T$ be a random vector with multivariate Normal distribution with mean $\boldsymbol{\mu} = (\mu_{X_1}, \ldots, \mu_{X_p})^T$ and covariance matrix $\boldsymbol{\Sigma}$.

Let $\boldsymbol{X}_{(n\times p)} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)^T$, with $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})$ for $i = 1, \ldots, n$, be a random sample of $(X_1, \ldots, X_p)$. Then sample mean and sample covariance matrix are, respectively,

$$\overline{\boldsymbol{X}}_{(p\times 1)} = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{X}_j = (\overline{X}_1, \ldots, \overline{X}_p)^T \quad \text{and} \quad \boldsymbol{S}_{(p\times p)} = \frac{1}{n-1}\sum_{j=1}^{n}(\boldsymbol{X}_j - \overline{\boldsymbol{X}})(\boldsymbol{X}_j - \overline{\boldsymbol{X}})^T$$

and recall that $E(\overline{\boldsymbol{X}}) = \mu \quad \text{and} \quad Var(\overline{\boldsymbol{X}}) = \frac{\boldsymbol{\Sigma}}{n}$

B.1 Confidence Region for the mean of a multivariate Normal variable

When $\boldsymbol{\Sigma}$ is unknown, the (standardized) distribuition of $\overline{\boldsymbol{X}}$ is the Hotelling's T^2 statistic given by

$$T^2 = (\overline{\boldsymbol{X}} - \boldsymbol{\mu})^T \left(\frac{\boldsymbol{S}}{n}\right)^{-1} (\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \sim \frac{p(n-1)}{(n-p)} F_{p,n-p} \tag{B.1}$$

where $F_{p,n-p}$ denotes the $F-$Snedecor distribution with p and $n-p$ degrees of freedom.

Therefore,

$$P\left(n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})^T \boldsymbol{S}^{-1} (\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \leq \frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)\right) = 1 - \alpha \tag{B.2}$$

where $F_{p,n-p}(\alpha)$ is the upper $(100\alpha) - th$ percentile of a $F-$distribution with p and $n-p$ degrees of freedom.

In fact, (B.2) is used to define a confidence region for the mean $\boldsymbol{\mu}$, as follow.

After the sample $\boldsymbol{x}$ has been observed, we can compute $\overline{\boldsymbol{x}}$ and $\boldsymbol{S}$, so that a $100(1-\alpha)\%$ confidence region for the mean $\boldsymbol{\mu}$ of a multivariate Normal distribution is the $p-$dimensional ellipsoid determined by

$$\left\{\boldsymbol{\mu} \in R^p : \; n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})^T \boldsymbol{S}^{-1} (\overline{\boldsymbol{x}} - \boldsymbol{\mu}) \leq \frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)\right\} \tag{B.3}$$

B.2 Prediction Region for an individual value of a multivariate Normal variable

For obtaining a control region for a future observation $\boldsymbol{W}$ of a multivariate Normal distribution we consider the variable $(\boldsymbol{W} - \overline{\boldsymbol{X}})$, Note that $E(\boldsymbol{W} - \overline{\boldsymbol{X}}) = 0$ and its covariance matrix is given by $Var(\boldsymbol{W} - \overline{\boldsymbol{X}}) = Var(\boldsymbol{W}) + Var(\overline{\boldsymbol{X}}) = \boldsymbol{\Sigma} + \frac{1}{n}\boldsymbol{\Sigma} = \frac{n+1}{n}\boldsymbol{\Sigma}$ since the new observation $\boldsymbol{W}$ is independent of $\overline{\boldsymbol{X}}$.

The (standardised) distribution of $(\boldsymbol{W} - \overline{\boldsymbol{X}})$ is also a Hotelling's T^2 statistic,

$$T^2 = (W - \overline{\boldsymbol{X}})^T \left(\frac{n+1}{n}\boldsymbol{S}\right)^{-1} (W - \overline{\boldsymbol{X}}) \sim \frac{p(n-1)}{(n-p)} F_{p,n-p} \tag{B.4}$$

where $F_{p,n-p}$ denotes the $F-$Snedecor distribution with p and $n-p$ degrees of freedom.

Similarly the confidence region (B.3) for the mean, after a sample $\boldsymbol{x}$ has been observed, a prediction region for a future individual observation $\boldsymbol{w}$ is given by

$$\left\{\boldsymbol{w} \in R^p : \ \frac{n}{n+1}(\boldsymbol{w} - \overline{\boldsymbol{x}})^T \boldsymbol{S}^{-1} (\boldsymbol{w} - \overline{\boldsymbol{x}}) \leq \frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)\right\} \tag{B.5}$$

B.3 Plotting the confidence and prediciton regions for bivariate normal $(p = 2)$

The confidence region (B.3) for the mean $\boldsymbol{\mu}$ and the prediction region (B.5) of a new observation $\boldsymbol{w}$ are of the form

$$\left\{\boldsymbol{a} \in R^2 : (\boldsymbol{a} - \overline{\boldsymbol{x}})^T \boldsymbol{S}^{-1} (\boldsymbol{a} - \overline{\boldsymbol{x}}) \leq r^2\right\} \tag{B.6}$$

that means, the regions are ellipses centered at $\overline{\boldsymbol{x}}$ and the axes are defined by the eingenvectors of the matrix $\boldsymbol{S}$, and the half-lenghts of the ellipses are proportional to the respective eingenvalues. A hypothetical example is plotted in Figure 21. The center is at $\overline{\boldsymbol{x}}$, and the major and minor axes are represented by pink and green arrows, respectively.

An alternative way of plotting these regions is to consider the Cholesky decomposition of the matrix $\boldsymbol{S}$: $\boldsymbol{S} = \boldsymbol{L}\boldsymbol{L}^T \implies \boldsymbol{S}^{-1} = (\boldsymbol{L}^{-1})^T \boldsymbol{L}^{-1}$ where $\boldsymbol{L}$ is a (unique) lower triangular matrix, also known as Cholesky factor.

Therefore, (B.6) can be written as

$$\begin{aligned} & (\boldsymbol{a} - \overline{\boldsymbol{x}})^T \boldsymbol{S}^{-1} (\boldsymbol{a} - \overline{\boldsymbol{x}}) \leq r^2 \\ \implies \quad & [\boldsymbol{L}^{-1}(\boldsymbol{a} - \overline{\boldsymbol{x}})]^T \underbrace{[\boldsymbol{L}^{-1}(\boldsymbol{a} - \overline{\boldsymbol{x}})]}_{b} \leq r^2 \qquad \text{(B.7)} \\ \implies \quad & \boldsymbol{b}^T \boldsymbol{b} \leq r^2 \qquad \text{(B.8)} \end{aligned}$$

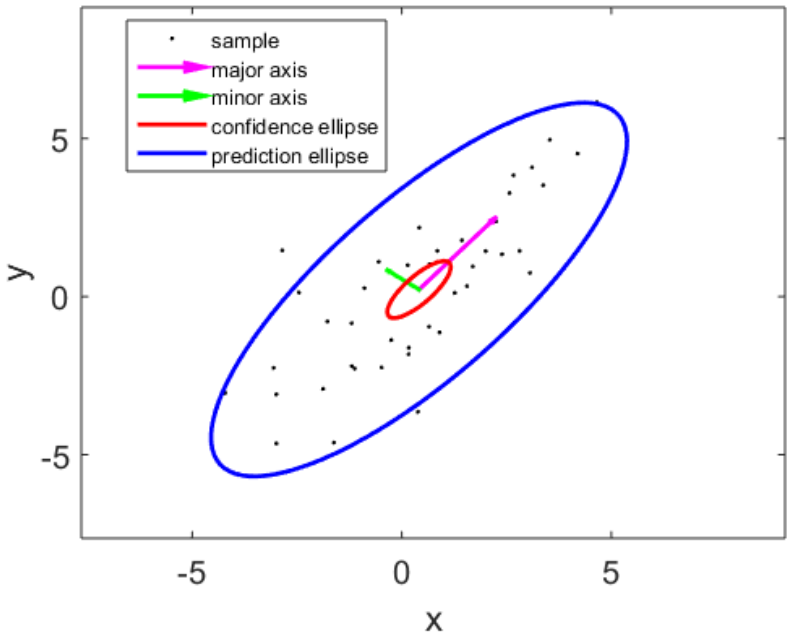


Figure 21 – Illustration for 90% confidence and prediction ellipses ($p = 2$).

Source – Author.

which defines a circle of radius r centered at origin.

The circumference of this circle can be drawn through the coordinates $(x, y)^T$ with $x = r\cos\theta$ and $y = r\sin\theta$ for $\theta \in [0, 2\pi]$

So, the elliptical contour is given by

$$\boldsymbol{a} = \overline{\boldsymbol{x}} + r\boldsymbol{L}\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix} \quad for \quad \theta \in [0, 2\pi] \tag{B.9}$$

Therefore, the elliptical contour of the confidence region for the mean $\boldsymbol{\mu}$, with $(1-\alpha)$ confidence, is given by (B.9) with $r = \frac{2(n-1)}{n(n-2)}F_{2,n-2}(\alpha)$.

And the elliptical contour of the prediction region for an individual observation $\boldsymbol{w}$, with $(1-\alpha)$ confidence, is given by (B.9) with $r = \frac{2(n-1)(n+1)}{n(n-2)}F_{2,n-2}(\alpha)$

APPENDIX C – PCA and Factor Analysis for Syringe Handling

The Syringe Handling Questionnaire consists of 10 items with a 7-point Likert scale ranging from 1 = strongly disagree to 7 = strongly agree. The items are presented in Table 10, since we will refer to the questions as we go through the analysis. The adopted significance level is 0.05.

Table 10 – Syringe Handling questions

Question	Description
Q1	I noticed the tactile stimulus upon penetrating the soft tissue
Q2	I felt confident administering the anesthesia
Q3	A device with tactile feedback upon resting against the soft tissue is useful
Q4	The administration of the anesthesia seemed real
Q5	It was easy to control the syringe
Q6	The syringe was stable
Q7	The tactile sensation of resting the needle on the soft tissue seemed real
Q8	I felt as if I was really administering anesthesia to the patient when handling the syringe
Q9	The syringe helped my performance
Q10	The tactile sensation of resting the needle on the soft tissue helped my performance

The Table 11 presents the mean score, and respective standard error, for each questionnaire item by group, and allowed us to notice that in all items mean score is 4 or above and that 4 is the center of this Likert scale. Therefore, in general, there was a moderate to strong agreement with the statements.

The mean scores are higher at items Q1, Q3, and Q10, meaning strong agreement in these question statements related to tactile feedback in all groups.

Contrary to our expectations, the participants of the NH group, whose haptic feedback was turned off, reported high mean score on questions Q1, Q3, and Q7 related to the feeling of the tactile feedback.

As expected, the participants of the NT group has a lower mean score in 7 of the 10 items, specially on Q5 and Q6, reflecting that this group felt a little bit more difficulty in controlling the syringe.

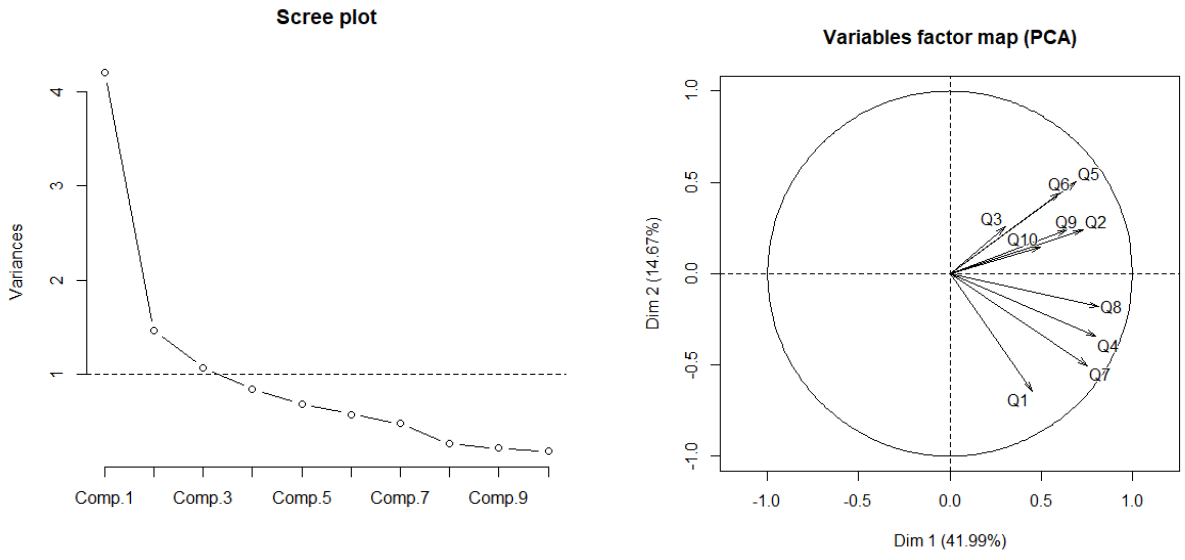
Table 11 – Mean (and standard error) for each Syringe Handling question by group

Question	Full	NH	NE	NT
Q1	6.45 (0.20)	6.39 (0.16)	6.13 (0.27)	6.28 (0.24)
Q2	6.43 (0.13)	6.42 (0.13)	5.90 (0.22)	5.43 (0.27)
Q3	6.88 (0.05)	6.90 (0.06)	6.63 (0.11)	6.75 (0.09)
Q4	5.57 (0.23)	5.88 (0.17)	5.65 (0.21)	5.33 (0.26)
Q5	5.62 (0.20)	5.71 (0.17)	5.40 (0.22)	4.03 (0.29)
Q6	5.93 (0.24)	5.78 (0.20)	6.08 (0.19)	4.88 (0.28)
Q7	5.74 (0.25)	5.76 (0.21)	5.38 (0.22)	5.43 (0.28)
Q8	5.67 (0.22)	5.83 (0.18)	5.48 (0.18)	5.25 (0.25)
Q9	6.05 (0.20)	5.59 (0.25)	6.00 (0.20)	5.05 (0.29)
Q10	6.80 (0.06)	6.70 (0.10)	6.40 (0.18)	6.23 (0.20)

To analyze all the questionnaire items together, we applied a Principal Component Analysis. The Bartlett's test ($\chi^2(45) = 705.242, p < 0.0001$) indicated that questionnaire responses were correlated highly enough to provide a reasonable analysis via PCA, and also the Kaiser-Meyer-Olkin (KMO) overall value being 0.78 (bigger than 0.70) confirmed that the data was suited for Factor Analysis. Also, none of the questions were eliminated because all items have a measure of sampling adequacy (MSA) higher than 0.5, as seen in Table 12.

Table 12 – Kaiser-Meyer-Olkin factor adequacy for the Syringe Handling Questionnaire PCA

Overall MSA = 0.78	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
MSA for each item	0.63	0.84	0.71	0.86	0.71	0.70	0.76	0.81	0.88	0.78



(a) Scree plot **(b) Biplot**

Figure 22 – Scree plot and Biplot for the Syringe Handling Questionnaire

The Scree plot in Figure 22 presents an elbow between *Comp. 2* and *Comp. 3*. Also, by the total variance explained criterion given by Table 13, only the first three components have the proportion of variance greater than 10%, therefore we decided to keep three components. The Figure 22 (biplot) shows how these first two components aggregated the questionnaire items.

Table 13 – Variance explained for the components of the Syringe Handling Questionnaire PCA.

	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5
Variance or eigenvalue	4.199	1.467	1.067	0.845	0.692
Percentage of Variance Explained	41.99	14.67	10.70	8.45	6.92
Cumulative percentage of Variance Explained	41.99	56.66	67.36	75.81	82.73

A varimax rotation (23) was performed to see whether the rotated factor loadings would provide additional insights. The varimax rotation loadings for three factors solution are displayed in Table 14, along with the specific variances. The rotated loadings indicated that questions Q1, Q4, Q7, and Q8, load highly on Factor 1, representing the items related to tactile sensation, thus this factor might be called *Tactile Realism*. The Factor 2 was more close to questions Q5, Q6, and Q2, related to control of the syringe, therefore we named this factor as *Syringe Control*, and the Factor 3 aggregated items Q9 and Q10 related to the performance, thus we termed it as *Ease of Performance*.

Table 14 – Factor loadings for the Syringe Handling Questionnaire PCA

Question	*Tactile realism* Factor 1	*Syringe control* Factor 2	*Ease of performance* Factor 3
Q1. I noticed the tactile stimulus upon penetrating the soft tissue	0.60		
Q7. The tactile sensation of resting the needle on the soft tissue seemed real	0.91		
Q4. The administration of the anesthesia seemed real	0.63		0.50
Q8. I felt as if I was really administering anesthesia to the patient when handling the syringe	0.58		0.57
Q5. It was easy to control the syringe		0.95	
Q6. The syringe was stable		0.68	
Q2. I felt confident administering the anesthesia		0.42	0.57
Q9. The syringe helped my performance			0.60
Q10. The tactile sensation of resting the needle on the soft tissue helped my performance			0.33
Q3. A device with tactile feedback upon resting against the soft tissue is useful			
Variance=eigenvalue	2.03	1.75	1.65
Proportion of Variance Explained	0.20	0.17	0.17
Cumulative Proportion of Variance Explained	0.20	0.38	0.54

Note: Factor loadings less than 0.3 are not displayed.

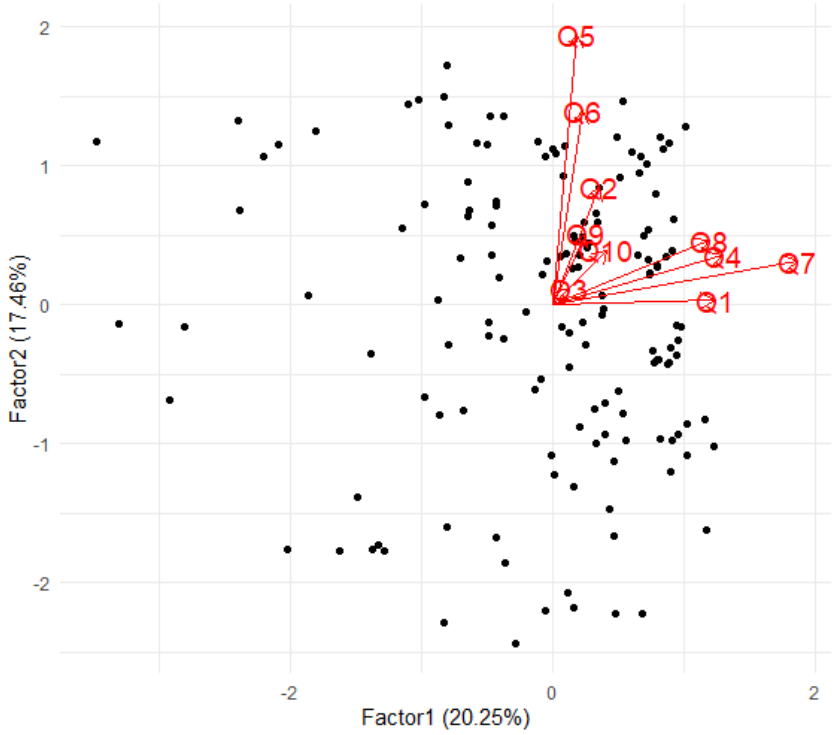


Figure 23 – Factor 1 and Factor 2 biplot for the rotated PCA of the Syringe Handling Questionnaire.

Table 15 shows the median scores, and respective interquartile range (IQR), of the factors by groups, and Figure 24 displays the boxplot of the factors scores for each group. We noticed that the immersive groups (Full and NH) have slightly higher median scores for all factors. The group NT has lower median score and higher dispersion compared to the other groups, for all factors, specially for Factor 2.

Table 15 – Factors median (and IQR) score by group for the Syringe Handling Questionnaire PCA.

Factor	Group			
	Full	NH	NE	NT
Factor 1. Tactile Realism	2.223 (0.565)	2.166 (0.430)	2.023 (0.395)	2.015 (0.518)
Factor 2. Syringe Control	1.943 (0.342)	1.874 (0.252)	1.831 (0.312)	1.561 (0.673)
Factor 3. Ease of Performance	2.274 (0.391)	2.230 (0.345)	2.143 (0.318)	2.012 (0.667)

In order to compare the groups, we applied a non-parametric test on factor scores, since the score data were non-normally distributed, and log transformations did not normalize the data.

The Kruskal-Wallis rank sum test results yielded no significant difference among groups for Factor 1 *Tactile Realism* ($\chi^2(3) = 3.472, p = 0.3245$), meaning that, with respect to the tactile realism, all four groups had similar responses. Therefore, although the participants of NH group did not have the haptic feedback, they responded similarly to the other groups, who had haptic feedback.

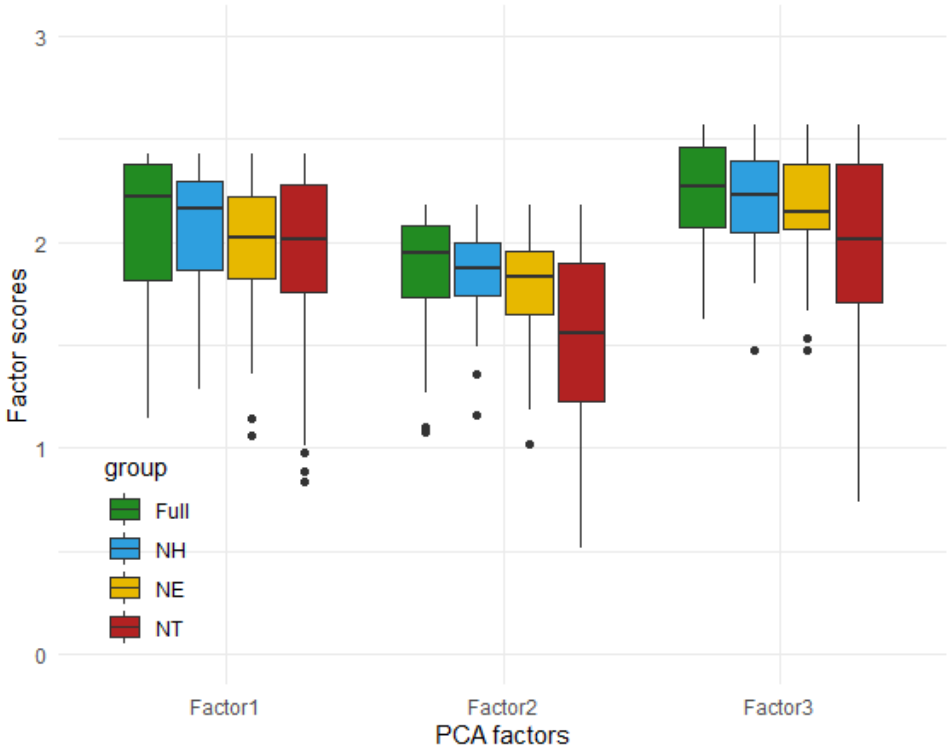


Figure 24 – Boxplot of the factor scores by group of the PCA for the Syringe Handling Questionnaire.

The test result for Factor 3 *Ease of Performance* was a borderline result ($\chi^2(3) = 8.042, p = 0.04515$), and multiple comparisons were not able to detect significant difference among groups, with only one borderline difference between NT and Full groups ($p = 0.0250$, with adjusted $p = 0.050$). So, we concluded that the Ease of Performance was not different among groups.

A significant difference among groups were detected for Factor 2 *Syringe Control* ($\chi^2(3) = 16.897, p = 0.00074$). Further analysis using Bonferroni multiple comparisons indicated difference between NT and Full groups (two-sided $z = -3.54, p = 0.0004$), and between NT and NH groups (two-sided $z = -2.903, p = 0.0037$).

By the estimates of Table 15 we confirmed that the participants in the NT group had a little bit more difficulty controlling the syringe compared to the immersive groups (Full and NH).

APPENDIX D – PCA and Factor Analysis for Avatar Embodiment

The Avatar Embodiment Questionnaire consists of 11 questions with a 7-point Likert scale ranging from -3 = strongly disagree to $+3$ = strongly agree. The items are presented in the Table 16, since we will refer to the questions as we go through the analysis. The adopted significance level is 0.05.

Table 16 – Avatar Embodiment questions

Question	Description
Q1	I felt as if the virtual hand was my hand
Q2	It felt as if the virtual hand I saw was someone else
Q3	It seemed as if I might have more than one hand
Q4	It felt like I could control the virtual hand as if it was my own hand
Q5	The movements of the virtual hand were caused by my movements
Q6	I felt as if the movements of the virtual hand were influencing my own movements
Q7	I felt as if the virtual hand was moving by itself
Q8	I felt as if my hand was located where I saw the virtual hand
Q9	I felt as if my (real) hand were drifting toward the virtual hand or as if the virtual hand were drifting toward my (real) hand
Q10	It felt as if my (real) hand were turning into an 'avatar' hand
Q11	At some point it felt as if my real hand was starting to take on the posture or shape of the virtual hand that I saw

Table 17 presents the mean score, and respective standard error, for each question by group, and allowed us to visualise that for the items Q1, Q4, Q5, Q8, Q10, and Q11, related to virtual hand ownership, the mean score was higher than 0 (in a -3 to +3 scale). On the other hand, for the items Q2, Q3, Q6, and Q7, related to the feeling that the virtual hand belong to someone else, the mean score was lower than 0.

Table 17 – Mean (and standard error) for each Avatar Embodiment questions by group.

Question	Full	NH	NE	NT
Q1	2.24 (0.21)	2.02 (0.18)	2.05 (0.16)	1.15 (0.26)
Q2	-1.79 (0.28)	-1.76 (0.23)	-1.55 (0.28)	-1.30 (0.25)
Q3	-2.00 (0.24)	-2.00 (0.23)	-1.95 (0.27)	-1.43 (0.27)
Q4	2.19 (0.16)	2.12 (0.16)	1.83 (0.17)	1.08 (0.29)
Q5	2.36 (0.14)	2.44 (0.12)	2.28 (0.18)	1.83 (0.18)
Q6	-0.55 (0.32)	-0.95 (0.32)	-0.48 (0.36)	0.85 (0.26)
Q7	-1.81 (0.26)	-2.00 (0.24)	-2.10 (0.21)	-1.38 (0.25)
Q8	2.05 (0.25)	1.93 (0.18)	1.58 (0.22)	1.58 (0.21)
Q9	-0.67 (0.34)	0.05 (0.33)	-0.90 (0.34)	0.23 (0.25)
Q10	1.64 (0.26)	1.54 (0.28)	1.15 (0.26)	1.25 (0.23)
Q11	1.86 (0.23)	1.56 (0.27)	0.88 (0.32)	1.48 (0.24)

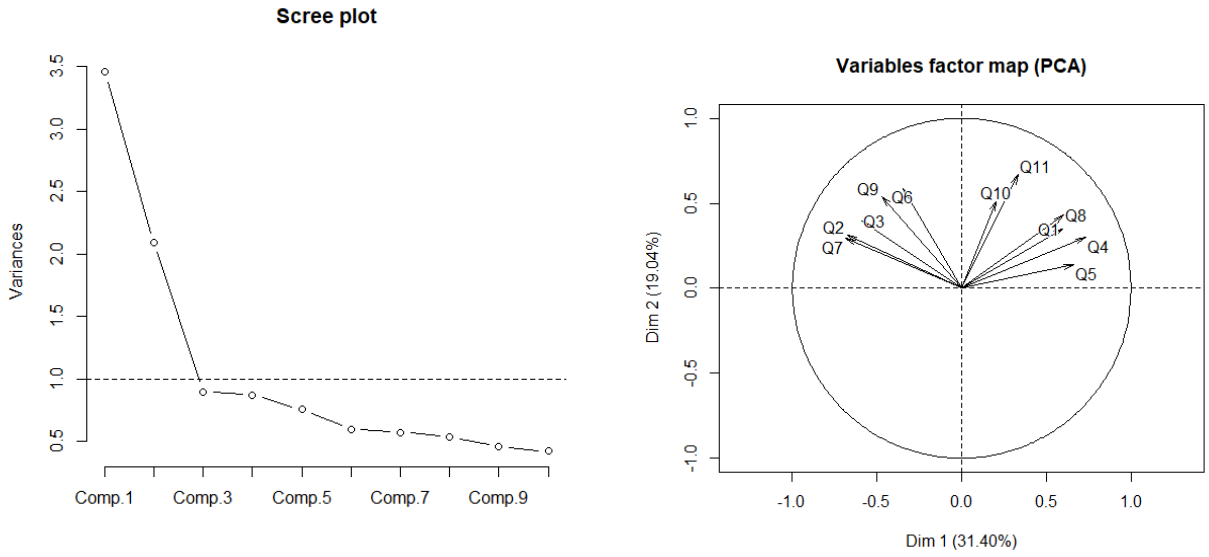
Also, the NT group differs from the other groups, having a lower mean score, especially on items with high scores Q1 (*I felt as if the virtual hand was my hand*), Q4 (*It felt like I could control the virtual hand as if it was my own hand*), and Q5 (*The movements of the virtual hand were caused by my movements*), and higher scores on items with general low mean scores Q2, (*It felt as if the virtual hand I saw was someone else*), Q3, (*It seemed as if I might have more than one hand*), Q6 (*I felt as if the movements of the virtual hand were influencing my own movements*), and Q7 (*I felt as if the virtual hand was moving by itself*).

Therefore, descriptively, the sense of virtual hand ownership was predominant and strong in all groups, with less intensity in the NT group.

To jointly analyze the questionnaire items, we performed a Principal Component Analysis. The Bartlett's test ($\chi^2(55) = 470.151, p < 0.0001$) indicates that variables are correlated highly enough to provide a reasonable analysis for PCA. Since the Kaiser-Meyer-Olkin (KMO) is 0.79 (larger than 0.70) the data is suited for Factor Analysis. Also, no question was eliminated because all items have MSA higher than 0.5, as seen in Table 18.

Table 18 – Kaiser-Meyer-Olkin factor adequacy for the Embodiment Questionaire PCA.

Overall MSA = 0.79	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
MSA for each item	0.77	0.79	0.84	0.80	0.79	0.76	0.81	0.84	0.75	0.78	0.71



(a) Scree plot **(b) Biplot**

Figure 25 – Scree plot and Biplot for the Avatar Embodiment Questionnaire

The Scree plot in Figure 25 presents an elbow at *Comp. 3*, but the eigenvalues of the *Comp. 3* and of the remaining components are smaller than 1. Also, by the total variance

explained criterion given by Table 19, only the first two components have a proportion of variance greater than 10%. Therefore, we decided to consider two components.

Table 19 – Variance explained for the components of the Avatar Embodiment Questionnaire PCA.

	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5
Variance or eigenvalue	3.452	2.093	0.898	0.795	0.755
Percentage of Variance Explained	31.40	19.04	8.18	7.95	6.87
Cumulative percentage of Variance Explained	31.40	50.44	58.62	66.58	73.45

The first two components, which explain 50% of the total variance, have an interesting subject-matter interpretation. The Figure 25 (biplot) shows how these first two components aggregated the questionnaire items.

A varimax rotation (Figure 26) was performed to see whether the rotated factor loadings would provide additional insights. The varimax rotation loadings for two factors solution are displayed in Table 20, along with the specific variances. The biplot and the rotated loadings indicated that Q2, Q3, Q6, Q7, and Q9 questions load highly on the Factor 1. Factor 1 appears to represent the feeling that control was led by someone else over one's own hand, this factor might be called *no hand ownership illusion*.

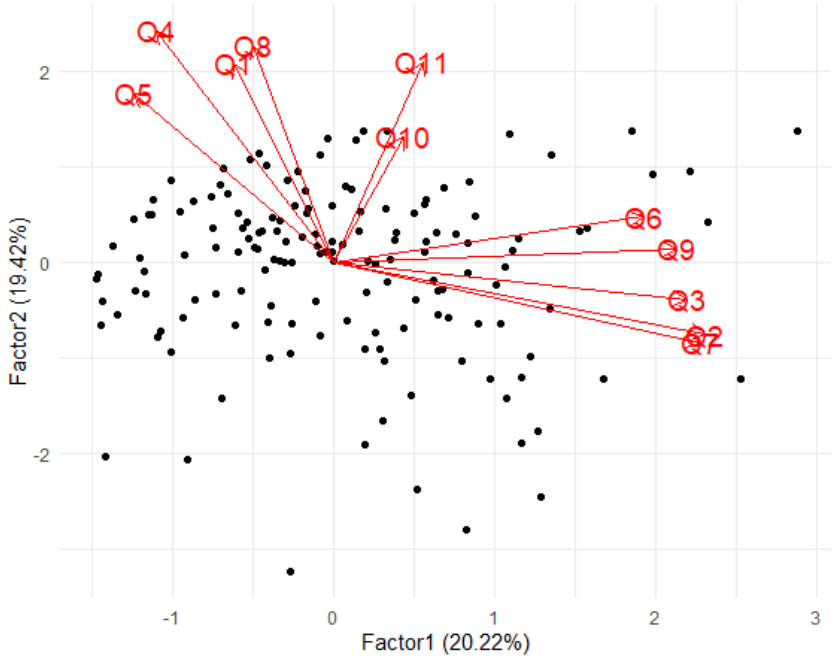


Figure 26 – Varimax rotation with draws eigenvectors (red arrows) and participants (black dots) for Avatar Embodiment Questionnaire.

Table 20 – Rotated loadings for the Avatar Embodiment Questionnaire.

Question	Lack of Agency Control Factor 1	Virtual Hand Ownership Factor 2
Q2. It felt as if the virtual hand I saw was someone else.	0.66	
Q3. It seemed as if I might have more than one hand.	0.62	
Q6. I felt as if the movements of the virtual hand were influencing my own movements.	0.55	
Q7. I felt as if the virtual hand was moving by itself.	0.65	
Q9. I felt as if my (real) hand were drifting toward the virtual hand or as if the virtual hand were drifting toward my (real) hand.	0.61	
Q1. I felt as if the virtual hand was my hand.		0.59
Q4. It felt like I could control the virtual hand as if it was my own hand.	-0.31	0.69
Q5. The movements of the virtual hand were caused by my movements.	-0.35	0.51
Q8. I felt as if my hand was located where I saw the virtual hand.		0.65
Q11. At some point it felt as if my real hand was starting to take on the posture or shape of the virtual hand that I saw.		0.60
Q10. It felt as if my (real) hand were turning into an 'avatar' hand.		0.38
Variance=eigenvalue	2.22	2.14
Proportion of Variance Explained	0.20	0.19
Cumulative Proportion of Variance Explained	0.20	0.39

Note: Factor loadings less than 0.3 are not displayed.

Table 21 shows the median, and respective interquartile range, of the factors scores and Figure 27 displays the boxplot of the factors scores for each group. Note that the NT group differs from the other groups, having a higher median score for Factor 1, and having a lower median score for Factor 2.

Table 21 – Factors median (and IQR) by group for the Avatar Embodiment Questionnaire.

Factor	Group			
	Full	NH	NE	NT
Factor 1. Lack of agency control	-0.601 (0.484)	-0.578 (0.513)	-0.621 (0.729)	-0.330 (0.393)
Factor 2. Virtual hand ownership	0.731 (0.269)	0.743 (0.396)	0.626 (0.305)	0.573 (0.535)

In order to compare the groups, we applied a non-parametric test on factors scores, since the score data were non-normally distributed, and log transformations did not normalize the data.

The Kruskal-Wallis rank sum test results showed that there was a significant difference among groups for Factor 1 *Lack of Agency Control* ($\chi^2(3) = 16.907, p = 0.0007$). Further analysis using Bonferroni multiple comparisons showed that, for Factor 1, NT differs from all other three groups (Full: two-sided $z = -3.036, p = 0.0024$, NE: two-sided $z = -2.948, p = 0.0016$, NH: two-sided $z = -2.727, p = 0.0032$), and the groups Full, NE, and NH behave similarly ($p = 1.0000$). We concluded that the NT group had a higher median score, with more agreement with items related to sensing the virtual hand as somebody else's hand, although, all groups had reported mean and median scores up to 0

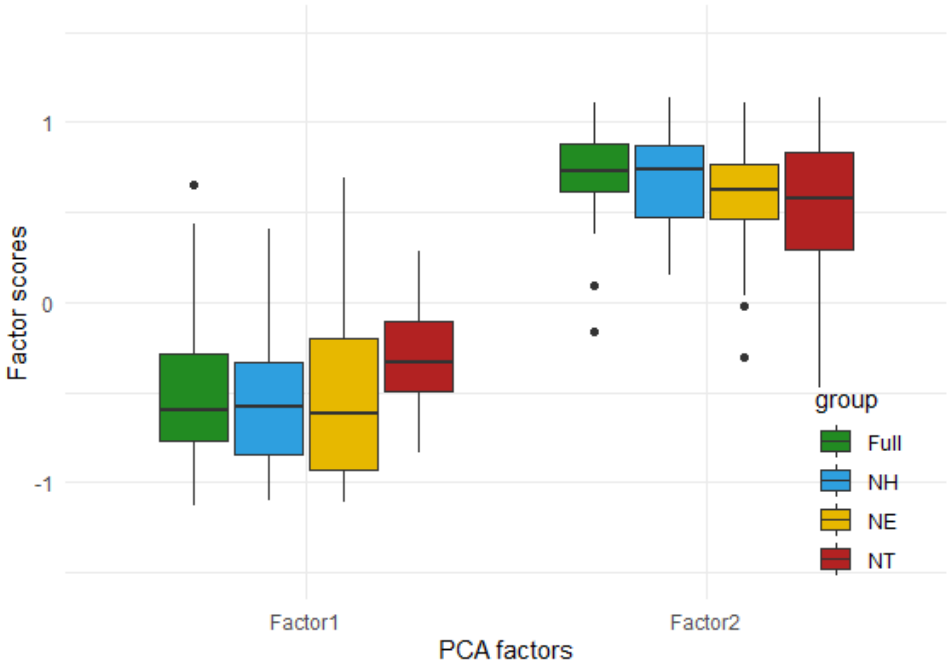


Figure 27 – Boxplot of the factor scores of the PCA by group for the Avatar Embodiment Questionnaire.

(in a -3 to +3 scale) per question grouped in Factor 1.

There was significant difference for Factor 2 *Virtual Hand Ownership* ($\chi^2(3) = 9.287, p = 0.02571$), and Bonferroni multiple comparisons detected difference only between NT and Full groups (two-sided $z = 2.264, p = 0.0236$).

Therefore, as described earlier, the participants in the NT group had felt less *Virtual Hand Ownership* than the other groups.

APPENDIX E – GAM for Simulator Sickness

Participants were asked to rate 15 questions concerning their discomfort in the VR simulator using a 4-point Likert-scale: 0 (none), 1 (slight), 2 (moderate), and 3 (severe).

The descriptive results showed that there was very little incidence of simulator sickness and no participant asked to terminate their participation in the experiment before concluding the anesthetic procedure. We observed a majority of "none" side effects (87%), "slight" (9%), "moderate" (3%), and "severe" (1%) symptoms concerning the 15 questions (not weighted from Kennedy et al's formula Kennedy et al (1993)). The "severe" score was noted concerning the oculomotor factor, in which the most prevalent symptoms reported were eyestrain, difficulty focusing and blurred vision. Only one participant from Full group reported eye strain symptoms. Three participants from NT group reported difficulty focusing symptom. Seven participants reported blurred vision (among participants four were from Full, two from NT and one from NE group).

Given the low frequency of moderate and severe symptoms felt, we decided to consider, for each questionnaire item, if the participant felt the symptom (value 1) or did not feel the symptom (value 0).

Table 22 presents, for each item of the questionnaire, the number of participants who felt no symptoms (zero frequency) of simulator sickness, and the respective percentage with respect to its group. Note that the percentage of zero symptoms were lower than 85% only on four of the questionnaire items (Q4, Q5, Q10, and Q11). These items concern the oculomotor side effect: eyestrain, difficulty focusing, "Fullness of the head", and blurred vision.

Table 22 – The zero frequency (and percentage on each group) for each item of the SSQ by group.

SSQ item	Group							
	Full		NH		NE		NT	
	$n = 42$		$n = 41$		$n = 40$		$n = 40$	
Q1. General discomfort	40	(95%)	38	(93%)	36	(90%)	33	(83%)
Q2. Fatigue	41	(98%)	40	(98%)	38	(95%)	36	(90%)
Q3. Headache	37	(88%)	38	(93%)	37	(93%)	34	(85%)
Q4. Eyestrain	24	(57%)	26	(63%)	34	(85%)	31	(78%)
Q5. Difficulty focusing	27	(64%)	28	(68%)	31	(78%)	25	(63%)
Q6. Salivation increase	39	(93%)	40	(98%)	40	(100%)	38	(95%)
Q7. Sweating	39	(93%)	41	(100%)	39	(98%)	40	(100%)
Q8. Nausea	41	(98%)	41	(100%)	38	(95%)	39	(98%)
Q9. Difficulty concentrating	38	(91%)	38	(93%)	40	(100%)	36	(90%)
Q10. “Fullness of the head”	28	(67%)	28	(68%)	32	(80%)	29	(73%)
Q11. Blurred vision	16	(38%)	21	(51%)	22	(55%)	21	(53%)
Q12. Dizziness eyes	39	(93%)	40	(98%)	37	(93%)	36	(90%)
Q13. Vertigo	40	(96%)	39	(95%)	38	(95%)	37	(93%)
Q14. Stomach awareness	42	(100%)	41	(100%)	40	(100%)	40	(100%)
Q15. Burping	42	(100%)	41	(100%)	40	(100%)	40	(100%)

Since the percentage of zero symptoms in each question is high, we decided to analyze the total number of symptoms felt, as seen in Table 23.

Table 23 – Frequency (and percentage per group) of the total number symptoms felt per participant by group.

Number of symptoms felt	Group							
	Full		NH		NE		NT	
0	7	(16.7%)	12	(29.3%)	13	(32.5%)	15	(37.5%)
1	11	(26.2%)	10	(24.4%)	11	(27.5%)	7	(17.5%)
2	8	(19.0%)	6	(14.6%)	10	(25.0%)	5	(12.5%)
3	4	(9.5%)	6	(14.6%)	3	(7.5%)	4	(10.0%)
4	7	(16.7%)	4	(9.8%)	0	(0.0%)	2	(5.0%)
5	2	(4.8%)	1	(2.4%)	1	(2.5%)	1	(2.5%)
6	1	(2.4%)	1	(2.4%)	1	(2.5%)	3	(7.5%)
7	2	(4.8%)	0	(0.0%)	1	(2.5%)	1	(2.5%)
8	0	(0.0%)	1	(2.4%)	0	(0.0%)	1	(2.5%)
10	0	(0.0%)	0	(0.0%)	0	(0.0%)	1	(2.5%)

We noticed in Table 23 that participants of the immersive groups (Full and NH) seem to have felt more symptoms than the other groups, since they presented a lower percentage of zero symptoms (16.7% and 29.3%, respectively) compared to NE and NT groups, that had non-immersive condition in one of the phases. Even though one participant of the NT group had felt 10 symptoms (the maximum value of all groups) out of 15 total, this group presented a higher percentage of zero symptoms (37.5%).

The total number of symptoms felt is a count data with a high number of zeros, therefore we applied the Zero-Adjusted Negative Binomial Type I (called ZANBI) model of the

Generalized Additive Models (GAM) technique. The adopted significance level is 0.05.

The results summary is in Table 24 and they reveal that the zeros frequency scores of NE and NH groups do not differ from the Full group ($p = 0.10206$ and $p = 0.17837$, respectively, two-tailed $t(157)$). However, NT group differed significantly from Full (two-tailed $t(157) = 2.083,\ p = 0.03885$).

Table 24 – Output summary of ZANBI model fit with logit link function for the probability of zero.

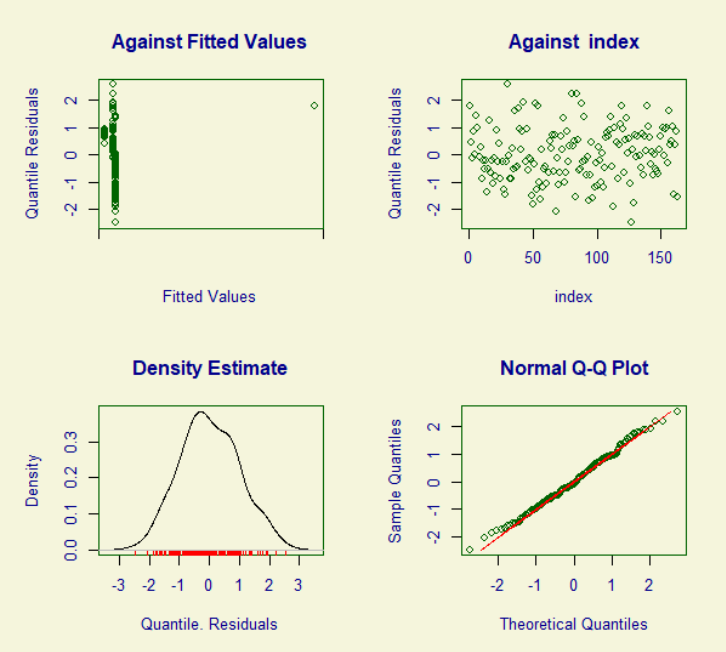
	Estimate	Std. Error	$t-$ value	two-tailed $p-$value
Full (intercept)	-1.6094	0.414	-3.887	0.00015
NH	0.7270	0.538	1.352	0.17837
NE	0.8786	0.534	1.645	0.10206
NT	1.0986	0.527	2.083	0.03885

The ZANBI model estimates that the percentage of zeros of NT group is 0.356 which is more than double the percentage of zeros of Full group (0.167), since

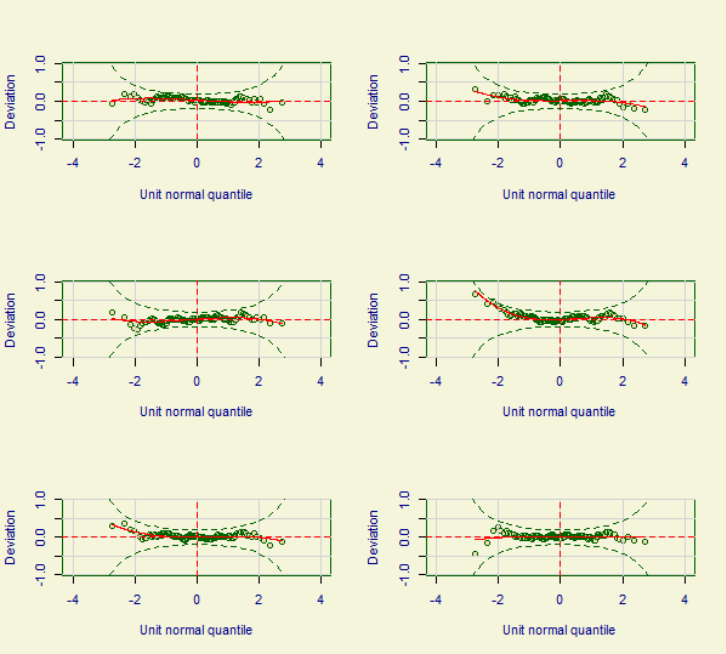
$$P(Full = 0) = \frac{e^{-1.6094}}{1 + e^{-1.6094}} = 0.16667 \quad \text{and} \quad P(NT = 0) = \frac{e^{-1.6094+1.09861}}{1 + e^{-1.6094+1.0986}} = 0.35622$$

which corroborate the descriptive analysis.

The ZANBI model fit was validated by the residuals analysis including the worm plots (Figure 28).



(a) Residuals



(b) Worm plot

Figure 28 – Residuals analysis and worm plots of ZANBI model fit for the SSQ total symptoms score

APPENDIX F – Instructions given to participants and translations

This supplementary material contemplate the instructions given to participants during the experiment.

Pre-experiment instructions	
ENGLISH	PORTUGUESE
Hello, thank you very much for participating in this experiment. Your participation will be very important to evaluate the dental anesthesia simulator Remember, you are not being evaluated, the simulator is. Therefore, you don't have to worry about your performance, but it is very important that you follow the guidelines and perform the procedure with the utmost care. <Haptic Device Familiarization> It's time to familiarize yourself with the simulator. Note that you can control the syringe that appears on the screen. Now, touch the numbered balls in sequence 1 through 4. Say out loud when you are ready to begin. <HMD Adjustment> Observe the two balls. Please tell us if you think that these balls are the same distance from you or is one ball is closer than the other. In the latter case tell us which ball is closer.	Olá, muito obrigado por colaborar com este experimento. A sua participação vai ser muito importante para avaliar o simulador de anestesia odontológica. Lembrando, não é você quem está sendo avaliado, é o simulador! Portanto, não precisa ficar preocupado com o seu desempenho, contudo é muito importante que siga as orientações e realize o procedimento com a máxima atenção. <Haptic Device Familiarization> É hora de se familiarizar com o simulador. Observe que você consegue controlar a seringa que aparece na tela. Agora, acerte as bolas numeradas na sequência de 1 a 4. Avise em voz alta quando estiver pronto para começar. <HMD Adjustment> Observe estas duas bolas. Nos diga, por favor, se você acha que essas duas bolas estão à mesma distância de você ou se existe uma bola mais perto do que a outra. Nesse caso, nos diga qual a bola mais perto.

Pre-experiment instructions	
ENGLISH	PORTUGUESE
<HMD Familiarization> You are now in the virtual world. It's important that you feel comfortable, so feel free to adjust the position of the chair, the height of the seat, or anything else that will help your movement and positioning. <small pause> Now, we are going to familiarize you with interaction in this world. To start, turn your head to the right <long pause> ... now turn to the left <long pause> now, move your head and look up, <long pause> ... move your head and look down. <long pause>	<HMD Familiarization> Você agora está no mundo virtual. É importante que você se sinta confortável, portanto fique à vontade para ajustar a posição da cadeira, a altura do assento, ou qualquer outra coisa que ajude na sua movimentação e posicionamento. <pequena pausa> Agora, vamos fazer você se familiarizar com a interação nesse mundo. Para começar, vire a sua cabeça para a direita <pausa longa>...agora, vire para a esquerda <pausa longa>...agora, mova sua cabeça para cima <pausa longa>...e para baixo.<pausa longa>
<HMD with Haptic Device Familiarization> Our assistant is going to place the syringe in your hand. Please, pierce the numbered balls in the sequence 1 through 4. Let us know, by saying aloud, when you're ready to begin. <VIDA Odonto Environment Familiarization> Now you will go to the Dental Office at the patient's side. Look to the right and wave to the instructor next to the dental chair ... <pause> ... Now look to the left until you see the dental sink. <long pause> Now, go back and look into the patient's mouth. <long pause> Please touch the white dot inside the mouth. Let me know by saying aloud, when you're ready to start.	<HMD with Haptic Device Familiarization> O nosso assistente agora irá colocar a seringa na sua mão. Por favor, acerte as bolas numeradas na sequência de 1 à 4. Nos avise em voz alta que você está pronto para começar. <VIDA Odonto Environment Familiarization> Agora, você irá para o consultório odontológico, ao lado do paciente. Olhe para a direita e acene para o professor que está ao lado da cadeira odontológica...<pausa>...Agora, olhe para a esquerda até encontrar a pia odontológica. <pausa longa > Agora, volte a olhar para dentro da boca do paciente. <pausa longa > Por favor, acerte o ponto branco dentro da boca. Avise em voz alta quando estiver pronto para começar.

Pre-experiment instructions	
ENGLISH	PORTUGUESE
Now, look in the direction of the arrow. You can move your head to get a better view. Locate the white spot the arrow is pointing to. Let me know by saying out loud that you're ready to start.	Agora, olhe na direção da seta. Você pode movimentar bem sua cabeça para ter uma visão clara. Acerte o ponto branco indicado pela seta. Avise em voz alta quando estiver pronto para começar.

Table 25 – Pre-experiment instructions

Experiment instructions	
ENGLISH	PORTUGUESE
<immersive condition> You are now going to watch the procedure being performed by an instructor. During the explanation, you are going to be able to move your head at will in order to better observe what you want to see. Position your head so that you can look into the patient's mouth. Tell us, by saying aloud, if you're ready to watch the instructor's instructions. <non-immersive condition> You are now going to watch on TV the administration process being done by the instructor. Tell us out loud when you are ready to start watching the video. <expert anesthesia instructor narrative> Hello, I am professor Maria Aparecida of the USP Dental School. You are going to be trained in the direct technique for administering anesthesia to the inferior alveolar nerve.	<immersive condition> Você vai presenciar agora a aplicação do procedimento feita por uma professora. Durante a explicação, você vai poder movimentar sua cabeça à vontade para observar melhor o que quiser. Posicione sua cabeça de forma que consiga enxergar a boca do paciente. Nos diga em voz alta se está pronto para assistir a instrução da professora. <non-immersive condition> Você irá assistir agora em uma TV a aplicação do procedimento feita por uma professora. Nos diga em voz alta quando estiver pronto para assistir ao vídeo. <narrativa da professora especialista em anestesia> Olá! Eu sou a professora Maria Aparecida da Faculdade de Odontologia da USP. Você irá treinar a técnica direta da anestesia do nervo alveolar inferior.

Experiment instructions	
ENGLISH	PORTUGUESE
For this procedure, use the tip of the canine tooth of the side opposite to the area you want to anesthetize. Starting at the point of the canine follow this trajectory to approximately 2 cm behind the last erupted tooth. Once the needle insertion point is defined, it is time to insert the needle into the tissue. Pay attention to insert a maximum of 2/3 of the length of the needle in the case of this patient that you are treating virtually. Now you're going to watch again the video of the procedure without interruption.	Para este procedimento, use como referência a ponta do canino do lado oposto a região onde quer anestesiar. A partir da ponta do canino siga essa trajetória aproximadamente até dois centímetros atrás do último dente irrompido. Uma vez definido o ponto de introdução da agulha, agora é hora de introduzir a agulha no tecido. Preste atenção para introduzir no máximo 2/3 do comprimento da agulha no caso deste paciente que você está atendendo virtualmente. Agora você irá assistir de novo o vídeo do procedimento sem interrupção.
<storyteller voice> Would you like to hear the Professor's instructions again?	<voz do narrador> Você gostaria de ouvir as instruções da professora novamente?
Now you are going to perform the administration of anesthesia as the instructor explained previously.	Agora, você vai realizar a aplicação da anestesia conforme a professora explicou anteriormente.
Our assistant is going to place the syringe in your hand.	Nosso assistente irá colocar a seringa na sua mão.
Let us know when you're ready to start. Also let us know when you believe you have completed the procedure or if you've given up.	Nos avise quando estiver pronto para começar. Avise também quando considerar que conseguiu completar o procedimento ou se desistiu da tentativa.

Table 26 – Experiment instructions

Post-experiment instructions	
ENGLISH	PORTUGUESE
Very good, now you can carefully remove the syringe from the patient's mouth. You'll now be taken by one of our assistants to the post-experiment room.	Muito bem, agora você pode retirar com cuidado a seringa da boca do paciente. Você será conduzido agora por um dos nossos instrutores para a sala pós-experimento.

Table 27 – Post-experiment instructions

APPENDIX G – Questionnaires, scales and translations

This supplementary material contemplates all three questionnaires (Haptic Feedback, Avatar Embodiment, and Simulator Sickness) in both English and Portuguese.

Haptic Feedback. 7-point Likert-scale, where 1 means the least agreement and 7 the most agreement.	
ENGLISH	PORTUGUESE
1. I noticed the tactile stimulus upon penetrating the soft tissue.	1. Eu percebi um estímulo tátil ao penetrar a agulha no tecido mole.
2. I felt confident administering the anesthesia.	2. Eu me senti confiante ao administrar a aplicação da anestesia.
3. A device with tactile feedback upon resting against the soft tissue is useful for improving the realism of administering the anesthesia.	3. Um dispositivo com sensação tátil ao encostar no tecido mole é útil para melhorar o realismo da aplicação da anestesia.
4. The administration of the anesthesia seemed real.	4. A aplicação da anestesia parecia real.
5. It was easy to control the syringe.	5. Foi fácil controlar a seringa.
6. The syringe was stable.	6. A seringa era estável.
7. The tactile sensation of resting the needle on the soft tissue seemed real.	7. A sensação tátil de encostar a agulha no tecido mole parecia real.
8. I felt as if I was really administering anesthesia to the patient when handling the syringe.	8. Eu senti como se eu estivesse realmente aplicando a anestesia no paciente ao manusear a seringa.
9. The syringe helped my performance.	9. A seringa ajudou minha performance.
10. The tactile sensation of resting the needle on the soft tissue helped my performance.	10. A sensação tátil ao encostar no tecido mole ajudou minha performance.

Table 28 – Haptic Feedback Questionnaire.

Avatar Embodiment. 11 questions scored on a -3 to +3 using a 7-point Likert-scale, where -3 means the least agreement and 3 the most agreement.	
ENGLISH	PORTUGUESE
1. I felt as if the virtual hand was my hand.	1. Eu senti como se a mão virtual fosse a minha mão.
2. It felt as if the virtual hand I saw was someone else.	2. Parecia que a mão virtual que eu vi era a mão de outra pessoa.
3. It seemed as if I might have more than one hand.	3. Parecia que eu tinha mais de uma mão.
4. It felt like I could control the virtual hand as if it was my own hand.	4. Parecia que eu podia controlar a mão virtual como se fosse minha própria mão.
5.The movements of the virtual hand were caused by my movements.	5. Os movimentos da mão virtual foram causados pelos meus movimentos.
6. I felt as if the movements of the virtual hand were influencing my own movements.	6. Eu senti como se os movimentos da mão virtual influenciassem os meus próprios movimentos.
7. I felt as if the virtual hand was moving by itself.	7. Eu senti como se a mão virtual estivesse movendo sozinha.
8. I felt as if my hand was located where I saw the virtual hand.	8. Eu senti como se a minha mão estivesse localizada na mão virtual que eu via.
9. I felt as if my (real) hand were drifting toward the virtual hand or as if the virtual hand were drifting toward my (real) hand.	9. Eu senti como se a minha mão real estivesse flutuando em direção à mão virtual ou como se a mão virtual estivesse flutuando em direção à minha mão real.
10. It felt as if my (real) hand were turning into an 'avatar' hand.	10. Era como se a minha (real) mão estivesse se tornando uma mão 'avatar'.
11. At some point it felt as if my real hand was starting to take on the posture or shape of the virtual hand that I saw.	11. Em algum momento senti como se a minha mão real estivesse começando a assumir a postura ou a forma da mão virtual que eu vi.

Table 29 – Avatar Embodiment Questionnaire.

Simulator Sickness. 15 questions rated on a 0 (none) to 3 (severe) using a 4-point Likert-scale.	
ENGLISH	PORTUGUESE
1. General discomfort	1. Mal-estar
2. Fatigue	2. Cansaço
3. Headache	3. Dor de cabeça.
4. Eyestrain	4. Vista cansada
5.Difficulty focusing	5. Dificuldade de concentração
6. Salivation increase	6. Aumento de salivação
7. Sweating	7. Sudorese
8. Nausea	8. Náusea
9. Difficulty concentrating	9. Dificuldade de concentração
10. “Fullness of the head”	10. A “cabeça pesada”
11. Blurred vision	11. Visão embaçada
12. Dizziness eyes open/ Dizziness eyes close	12. Tontura com os olhos abertos e/ou fechados
13. Vertigo	13. Vertigem
14. Stomach awareness	14. Desconforto estomacal
15. Burping	15. Arroto

Table 30 – Simulator Sickness Questionnaire.