Augmenting Dummies for Training Central Venous Catheterization

Helena Catarina Margarido Mendes

helena.mendes@tecnico.ulisboa.pt

Thesis to obtain the Master of Science Degree in Biomedical Engineering

Supervisors: Prof. Dr. Daniel Simões Lopes and Cátia Isabel Andrade Botelho Costa

1 Instituto Superior Técnico, Universidade de Lisboa
2 Visualization and Intelligent Multimodal Interfaces, INESC-ID Lisboa
3 Hospital da Luz Learning Health, Lisboa

October 2019

Abstract—Medical simulation based on augmented reality (AR) is a promising tool to improve medical training, particularly for needle-based interventions. However, limited research has been conducted in new solutions for training central venous catheterization (CVC). The purpose of this work is to compare the AR application developed with the conventional method to train CVC, using a dummy of the upper torso and neck. The proposed system (named PIÑATA) was implemented using AR targets placed in the needle and the dummy to display internal anatomical structures and additional information about the position and orientation of the needle. Observation and co-design sessions were conducted to collect the requirements, obtain feedback from the users and improve the prototype. A total of 18 potential users - attending specialists and medical residents - performed needle insertion tasks for CVC with PIÑATA and with the conventional training system. The performance was objectively measured by the task completion time and the number of errors. A correlation between the task completion time of the two training methods was found, suggesting the concurrent validity of PIÑATA. An inherent difference in the task completion time ($p = 0.040$) and in the number of errors ($p = 0.036$) between novices and experts proved the construct validity of the new tool. The qualitative answers of the participants also suggest its face and content validity, a high acceptability rate and a medium perceived workload. The overall results show that the AR tool proposed can complement the conventional training of CVC.

Keywords—Medical training, augmented reality, central venous catheterization, needle placement, professional assessment.

1 Introduction

For over a century, medical education stayed unchanged, following the principle “See one, do one, teach one”: medical trainees start as passive learners and it is progressively given to them more responsibility and autonomy through repeated practice until they become competent practitioners. However, some studies suggest that this system inadequately prepares trainees for entrance into surgical practice, stating that trainees do not gain enough autonomy during residency. [1] Also, medical error is still one of the leading causes of death. [2] Therefore, it becomes clear the importance of invest in medical education and of making training methods more efficient and reliable, because patients will ultimately benefit. [3]

With the recent technological innovation, other approaches to the apprenticeship model of medical professionals appeared, including surgical simulation. Simulators provide a safe way to improve surgeons’ skills, through repeated practice, outside the operating apparatus, and allow to measure and assess their technical skills. [4], [5] However, the conventional simulators (physical simulators) still have some limitations: in the case of the simple plastic mannequins, the reduced functionality and non-reusability in case of destructive tasks; in the case of the more advanced, the high complexity needed to mimic medical scenarios which implies a high cost. Technologies as virtual reality (VR) and augmented reality (AR) bring potential opportunities for improving medical simulation. [4] Optical-see through (OST)-AR, particularly, can reduce the complexity of the physical simulators replacing some of its embedded electronic technology by AR content and using that content to complement the training process. [6]

OST-AR is an integrated technique of image processing, where real objects and virtual (computer-generated) objects are combined in a real environment. [7] An OST-AR system consists at least in three components: (i) a tracking component, responsible to calculate the three dimensional (3D) location and orientation of the camera in real time; (ii) a registration component which overlays a virtual layer to the real scene in a position determined by the tracking component; and (iii) a visualisation component, that consists in a head-mounted display (HMD) where AR content is overlaid on the real world in the wearer’s field of view (FOV), directly. [8], [9]

In this work, the potential of OST-AR for training insertion of needles for central venous catheterization (CVC) is explored. This procedure consists in the insertion of a catheter into the superior vena cava, through the internal jugular or the subclavian veins. [10] It allows to measure hemodynamic variables, to deliver medications and nutritional support and provides a portal for intravascular volume replacement, being performed several times in the daily routine of the hospital. [11], [12] This procedure requires precise placement of the injection needle, being associated with potentially life-threatening adverse events, such as arterial puncture and pneumothorax, that are both hazardous to patients and expensive to treat. [12]–[14] Therefore it is very important that medical professionals which commonly perform
CVC (including internal and emergency medicine doctors, general surgeons and mainly, anaesthesiologists [14], [15]) are as much prepared as possible before starting to perform this procedure in patients. For that, efficient and effective methods for training CVC are needed. Some studies already shown that a simulation-based training increase residents’ skills in CVC insertion and decrease related complications. [14]–[17] However, the research regarding the potential of AR as a complement to the CVC training is limited.

The main research questions here addressed are: "Does OST-AR assist the training of needle insertion in CVC?" and “How well does OST-AR needle insertion CVC training compares to conventional training methods?”. To answer these questions, it was developed an OST-AR application that assists the train of CVC. Our hypothesis is that this system can complement the conventional training system, reducing the dependence of the instructor without affect the quality of training.

2 Related work

VR and AR applications have been developed to train or educate medical professionals, as a navigation tool during surgical procedures, to enhance visualisation at the operating room and as a therapeutic tool in the treatment of patients. [3] In this section, some examples of the applications described in literature developed for training, mainly for procedures involving insertion of needles will be presented.

O. Grottke et al. (2009) [18] and Saad Ali et al. (2018) [19] developed VR simulators to train needle-based interventions. These simulators allow repetitive execution of procedures and provide precise metrics for the evaluation of the trainee performance, having great potential in medical training. However, they also have limitations: the very high costs and, mainly, the unrealistic simulation of the visual and haptic sensations between virtual objects and instruments. [4], [20]

AR simulation combines the advantages of physical and VR simulators within a unique environment, avoiding some of the drawbacks mentioned. [21] In 2011, Caitlin T. Yeo et al. proposed a configuration for Perk Tutor (an open-source training platform for image-guided needle insertions) to augment the effectiveness of computed tomography (CT)-guided facet joint injections. In the proposed method, the tracking of the needle is done using an electromagnetic tracker sensor and a laser overlay system is used to guide the procedure in the FOV of the user. The results of the user tests shown that AR image overlay can assist medical trainees in learning the correct placement of a needle. [13] A similar work was developed by E. Moult et al. (2013) for ultrasound (US)-guided facet joint injections. The tool developed allow the training in a synthetic phantom of the vertebral column and surrounding tissues, with the visualisation in an external monitor of the position of the tools with respect to the live US and the vertebral column. The user tests performed to compare this method to conventional training revealed a higher success rate for the group that used AR. One limitation of this approach relatively to the first is the requirement for the trainee look away from the phantom to view the image on an external display. [22] However, both systems have limitations in common regarding the realism of the simulation: neither the phantoms used neither the setup of the tests are much realistic.

In fact, today it is common to perform procedures that involve insertion of needles recurring to medical image for guidance. Nonetheless, medical doctors have to be prepared to execute them even without guidance. Yuichiro Abe et al. (2013) developed an AR guidance technique to visualise the needle insertion point on the skin and the 3D trajectory path using an HMD for a percutaneous vertebroplasty procedure. Lauryn R. Rochlen et al. (2017) developed and tested a prototype that consists in an AR trainer on needle insertion for CVC. Wearing an HMD, the trainee is capable of visualise the projected internal anatomy, revealing the target point for needle insertion. The system provides feedback to the user according to its performance: if the needle is inserted correctly into the vein, it turned red to simulate blood return; if the carotid artery was inadvertently punctured, the anatomy is revealed with an expanding hematoma. [23] Tianyu Song et al. (2018) developed a HMD-based AR prototype to help with planning and to provide guidance in endodontic therapy. The proposed system contains AR instruction overlay, radiographic image, voice control and audio feedback. During interaction, both the tool and the tooth are tracked using markers and there are some visual indicators that change colour and shape based on the depth, distance and orientation between them. [24]

The examples presented in this section demonstrate the potential that AR has in the medical field, particularly in training of procedures that include insertion of needles. However, 3D registering, i.e., 3D alignment between real and virtual objects, is still a complex problem. [25] Moreover, there are some limitations common to all applications using AR regarding the HMDs available: limited FOV and some discomfort and unnatural feeling described by the users. The majority of the systems described here present these limitations even using complex and expensive AR hardware and recur to external systems to resolve the registration problem. Regarding the graphical interfaces, the systems described here focus only on one of two types of AR content to guide needle insertion: projection of internal structures or geometrical information about the position and orientation of the needle. There was not found any system that combined these two types of additional information. Besides that, the user studies presented in literature have several limitations mainly in the design process.

3 User studies

Design a new solution is not a straightforward pathway and cannot be solved with technological or scientific approaches alone. It is necessary to understand the needs of the users and translate that needs into solutions. [26] The framework that guided this study is schematised in Figure 1.

3.1 Observation sessions

3.1.1 Methodology

First, two informal observation sessions in workshops for training CVC were conducted: the first with about 20 trainees
3.2 Co-design workshop

3.2.1 Methodology

Co-design workshops were conducted to understand what are the key design requirements and to collect feedback about the initial prototype. Two physicians (attending specialists of anaesthesiology with 2 and 27 years of experience) with experience as instructors participated in the workshops. Video and audio of the sessions were recorded with the consent of all the participants. Participants were asked to talk about the current experience as medical trainers. Then, it was shown to the participants the initial prototype developed and they were invited to try out the system and to give their feedback. To stimulate their suggestions some previously prepared questions such as "What you currently enjoy/not enjoy about the way that training of CVC is done?" and "What are the features of this propose that you like/do not like?" were raised.

In order to analyse the collected data, a structured method was used - thematic analysis. This method consists in establish themes, i.e., "short words or phrases that assign a summative, salient and/or evocative attribute to a portion of written or visual data" (Tomitsch et al., 2018, p.122), to summarise the data. [26] The resultant set of themes was, then, used to iterate the pre-prototype of PIÑATA.

3.1.2 Results and Discussion

The main results obtained from the co-design were some specific suggestions to improve the design of PIÑATA and some of the requirements that should be met. From the thematic analysis process, it was defined a set of themes:

- **Familiarity**
  - This was one of the identified requirements: a training tool should resemble the conventional training. Quoting one of the participants referring to the way to handle the needle, "As we do not do the procedure in that way, the train should follow the way it is done in practice". From this, we conclude that the design of PIÑATA should be simple, promoting the habitual training workflow.

- **Stability**
  - While experimenting the pre-prototype of PIÑATA, the professionals refer that sometimes the virtual elements disappeared or changed position. This was already an expected limitation related to the shape, size and design of the image targets. Therefore, this was one of the improvements applied in PIÑATA after the co-design.

- **Virtual elements of the design**
  - When asked about the virtual information displayed, the professionals agreed that it is useful. They also suggest "it could make sense to project the carotid and jugular to have an additional reference". The main implication of these aspects for the design of PIÑATA is to obtain the 3D model of the vessels to be projected in the simulator.

- **Complement the instructor**
  - This theme expresses some of the advantages of PIÑATA relatively to the conventional way to train CVC. According to the participants, "This is useful for the trainees become more autonomous, and even allow them to train more frequently". With the additional information provided by PIÑATA, the trainees will be capable to train without relying constantly on the trainer.

- **Debriefing of training**
  - This was also an advantage of PIÑATA pointed by the professionals: "This allows them to understand if they are performing well". Today, there is not a formal way to evaluate the simulation. As this tool allow to measure the angles and depth of the needle over time, it is possible to define performance measures that allow to evaluate the
trainee. The colour code defined also helps the trainee to understand and correct the gesture in real time.

These results allowed to iterate the pre-prototype of PIÑATA and to obtain the final prototype detailed in the next section.

4 PIÑATA

We developed an OST-AR application with the objective of complement the conventional medical training of the insertion of needles for the CVC procedure. From now on, the developed system will be designated PIÑATA, that stands for Pinpoint Insertion of intravenous Needles via Augmented reality Training Assistance. PIÑATA consists of four essential components illustrated in Figure 2.

The application was developed in Unity3D (version 2018.3.14f1), a game engine with built-in support for AR applications, particularly for Vuforia Engine (version 8.1.7) and for Aryzon software development kit (SDK) (version 2.1).

4.1 Real world component

This component consists in the real objects to which the user interacts directly, the medical instruments - a syringe and a Seldinger needle - and the physical simulator. The physical simulator of the upper torso and neck used in this work was the CentraLineMan® System. This simple plastic mannequin features clinically relevant landmarks and anatomy, including accurate internal and external landmarks that are palpable. [27]

In order to register the virtual objects to this simulator and to project the vascular anatomy in AR, it was necessary to create the 3D model of both the external surface of the simulator and of the simulated vessels. To obtain these 3D models, CT images of the simulator were acquired. The 3D reconstruction from the medical images was made following the pipeline described by N. S. Ribeiro et al. (2009). [28] The first step consists in the automatic segmentation of the 3D image dataset to identify the tissues and their boundaries, using ITK-SNAP [29], generating a very close approximation of the anatomical structures to be segmented. The models created present a non-smooth aspect and has an excess of vertices and faces, that are irrelevant and turn the mesh too heavy. To deal with these details, the mesh was imported into Paraview [30] and a sequence of smoothing and decimation filters was applied. After this, the result was imported into MeshLab [31], where the interior of the mesh was discharged to reduce the size of the resultant mesh. The last step was to export the result in the 3D model format (.obj), being ready to be imported and used in the Unity environment.

4.2 Registration and display component

The AR system used in PIÑATA was the Aryzon headset, a sub-$30 headset. This OST-AR system makes use of the power of a smartphone and of an optical system to create a layer of digital information. With the camera and sensor data of the phone, it is possible to perceive the real world and to track image markers. In the smartphone display it is created a stereo image in order to give the perception of 3D depth. Before arriving to the viewer’s eyes, this image passes through a mirror, lightweight stereoscopic lenses and a combiner glass. Aryzon has an open-source SDK for Unity and support for Vuforia Engine. [32]

4.3 Tracking component

The tracking component of this system was accomplished using Vuforia (version 8.1.7). [33] It was used marker-based tracking using image targets: the detection is based on the comparison of the features extracted from the target image and from the live camera image in real time. In the tool developed, it is necessary to track two objects in the real world: the simulator and the needle. The targets produced are represented in Figure 3.

The design of the targets considered the ideal characteristics related to the physical properties and the intrinsic attributes of the images. For both objects, multi-image targets are used that allow the tracking of the entire target when any of its child targets (faces) is detected. To track the simulator, it is used a multi-image target composed by two 50x72 cm rectangular posters assembled as demonstrated in Figure 3(a) to augment the chances for the target to be detected. To perform the tracking of the needle, it is used a polyhedron adapted from a cube, as demonstrated in Figure 3(b), where the child target facing the user has an inclination in order to its normal become better aligned with the camera viewing direction. The width of about
4 cm was defined considering a trade-off between a minimal size that allows an acceptable tracking and a maximal size that does not perturb too much the needle handling. Another very important factor to optimise the tracking is the image target’s features. All the image targets designed in this work have a common aspect: quick response (QR) code based patterns, fulfilling the recommended characteristics to optimise target detection and tracking stability: (i) being rich in detail; (ii) having good contrast; (iii) do not having repetitive patterns.

### 4.4 Virtual component

This component consists in all the elements that are registered to the real world in order to guide the insertion of the needle in the correct way. In PIÑATA, three different spaces are augmented: the space of the simulator’s surface, the space inside the simulator and the space of the needle.

The 3D model of the vessels is projected in the space inside the simulator. This model includes part of the superior vena cava, the subclavian artery and vein, the internal jugular vein and the carotid artery, as represented in Figure 4.

![Figure 4: AR projection of the vascular anatomy in the simulator. Veins represented in blue and arteries in red.](image)

In the surface of the simulator, it is represented the point of insertion using four arrows pointing inwards and a placeholder for the needle using a conical shape suggesting the right orientation in which the needle should be inserted. These visual aids were implemented for the two different cannulation routes: for the internal jugular vein, the arrows were placed pointing to the superior apex of the triangle formed by the two heads of the sternocleidomastoid muscle and the clavicle and the placeholder makes an angle of 45° with the coronal plane pointing toward the ipsilateral nipple, as represented in Figure 5(a); for the subclavian vein, the point of insertion is just superior to the clavicle at the midclavicular line and the placeholder is placed in the coronal plane pointing toward the sternal notch, as illustrated in Figure 5(b). [10]

![Figure 5: AR representation of the point of insertion and placeholder of the needle in the simulator for training CVC through the (a) internal jugular vein and (b) subclavian vein. The landmarks used to implement the AR elements to each access site are represented.](image)

The interaction is achieved through the direct manipulation of the needle which results in some visual aids to the user. It is used a colour code to guide the user during training: green encodes the correct orientation of the needle and red means that the needle is not oriented in the correct way. The interaction of the needle with the simulator also results in the modification of the appearance of the conical placeholder: as the needle enters in the placeholder area, it starts to become transparent to give the idea of progression in the execution of the task.

### 5 Performance study

#### 5.1 Methodology

A user study was performed in order to assess if PIÑATA can complement the conventional training of the CVC procedure. The independent variables are defined based on the training paradigm (conventional or PIÑATA) and on the access site (jugular or subclavian), which gives a total of 4 independent variables. The dependent variables considered are task completion time and number of errors (objective measures) and participant preferences (subjective measures).

#### 5.1.1 Participants

A total of 18 invited participants performed the user study (6 male and 12 female), 9 (50%) novices - medical residents - and 9 (50%) experts - attending specialists of anaesthesiology with experience between 3 and 32 years. Participants were selected to match the user profile of both trainees and instructors. From all the participants that had previous experience in
training sessions for the CVC (83.3%), only 4 participants (26.7%) had already used systems with VR/AR and almost all (86.7%) had already trained in dummies of the upper torso and neck. All participants agreed that the access site more frequently trained for the CVC was the internal jugular vein. They received no compensation for taking part in the evaluation.

5.1.2 Apparatus

The study took place in an office room of an hospital’s operating theatre. The setup of the study consists simply in the prototype placed over a table and the Aryzon headset using an Android smartphone (16nm Octa core 2.36GHzx4+1.7GHzx4 3GB ram with 2160x1080 pixels screen) to run the application. A portable computer was used to fill all the questionnaires.

5.1.3 Tasks

In order to receive as broad feedback as possible, each participant took part independently. Also an user study with only one group was considered, i.e, all the participants execute the training with conventional and PI˜NATA systems. Initially, users were asked to perform an habituation task without anatomical meaning so that they could get used to the system. Then participants were asked to complete a total of 4 different tasks (2 access sites x 2 paradigms of train per access site). A task was considered successful when the liquid that circulates in the simulated vein is aspirated to the syringe. For each task, the participant has at most 5 trials, after which the task was interrupted and considered as unsuccess. A trial is defined as an insertion of needle without going back to the skin. Task completion time was measured since the beginning of each task declared by the participant until the task is completed. The number of errors is the number of trials before the one where the participant completed the task. There were accounted two types of errors: remove the needle to the skin for redirection and arterial puncture.

5.1.4 Procedure

The tests followed a previously prepared protocol. At the beginning of the session, participants filled in an informed consent and a demographics questionnaire regarding their gender, level of expertise and previous experience. To each participant it was previously defined the order to perform the tasks using Latin squares’ permutations to obtain a as balance as possible number of participants per initial access site and training paradigm. This is an attempt to prevent biased results. After explain the steps of the study and the execution of the tasks using Latin squares’ permutations to obtain a as balance as possible number of participants per initial access site and training paradigm. This is an attempt to prevent biased results. After explain the steps of the study and the execution of the habituation task, the participants performed the tasks, as is illustrated in Figure 7. The time required for task completion was measured and the number of errors was registered.

After each set of two tasks (for the same access site), the participant was asked to complete a satisfaction questionnaire comparing some aspects of the two forms of training: (1) easiness to identify the anatomical landmarks; (2) easiness to use; (3) learnability; (4) usefulness; (5) recallability; (6) easiness to debriefing. The responses were given on a 6-level Likert scale from strongly disagree to strongly agree.

After performing all the tasks, each participant was asked to complete a general questionnaire that is divided in three sections: face and content validity, participants’ satisfaction (using the system usability scale (SUS) system) and perceived workload (using the NASA task load index (NASA-TLX) assessment tool). At the end, it was conducted a semi-structured interview to the participant, with previously prepared questions such as "In your opinion, PIÑATA complements the conventional training system for the insertion of needles in CVC?", "What were the difficulties that you felt while using PIÑATA?" and "Do you think that the adoption of this kind of technology for medical training is possible?". This allow to collect a more detailed and broad feedback. The interviews were documented by voice recording. A full session lasted for approximately 20 minutes including all parts.

5.1.5 Statistical analysis and interpretation of results

Statistical analysis was performed using the Mann-Whitney U test, the Wilcoxon signed-ranks test and the Spearman’s rank order correlation test. All statistical analyses were carried out using IBM SPSS Statistics for Windows, version 25 (IBM Corp., Armonk, N.Y., USA). [34] For all tests, a p-value of less than 0.05 was considered statistically significant.

To interpret the results obtained using the SUS questionnaire, it is used the SUS score, a single number that represents a measure of the overall usability of the system. SUS scores have a range of 0 to 100, that have a corresponding interpretation: 100 is a really good user experience with high acceptability; 68 is considered average; anything below 68 is considered as not acceptable for the users. [35]

To analyse the results for the NASA-TLX questionnaire, the unweighted scores between 0 and 100 are calculated simply by converting the answers in the 21-item scale. This score correlates with the workload of the system: 0-9 is considered low; 10-29 medium; 30-49 is considered somewhat high; 50-79 is high; and 80-100 means a very high workload. [36]

The subjective data obtained from the semi-structured interviews was analysed with the thematic analysis method.
5.2 Results and Discussion

The data recoiled during the user evaluation tests include performance measures using both PIÑATA and the conventional training method, the answers to the satisfaction questionnaires and the more detailed opinions of the professionals obtained during the semi-structured interviews.

5.2.1 Face and content validity

In the context of this work, assessing the face validity of PIÑATA consists in verifying if training with the prototype resembles the real working situation. To attest the content validity of PIÑATA, its educational content should be uniformly and positively evaluated by the professionals in the field.

The central tendencies of responses to each statement were summarised by using median (Mdn), with dispersion measured by interquartile range (IQR) in Table 1. The median of all the responses is higher than 3.5, meaning that the participants positively evaluate the realism and content of PIÑATA. A one-sample Wilcoxon signed-ranks test was used to determine the significance of the responses evaluating if the operators were significantly more likely to agree or disagree with each of the statements. For that, the null hypothesis (H0) considered was “the difference between the median of each Likert item and the median of the scale (3.5) is zero”. All the results are statistically significant (p < 0.05) attesting the face and content validity of PIÑATA. It is relevant to note that the item most strongly agreed (Mdn = 6.0) is related to the projection of the vessels that was one of the improvements made in the prototype after the suggestions of the instructors during the co-design sessions. By contrast, the item less agreed (Mdn = 4.0) is related to the usefulness of graphical elements to indicate the orientation of the needle that can be explained by the fact that that element may be distracting and hide part of the view of the insertion site.

5.2.2 Concurrent validity

To infer about the concurrent validity, it is important to prove that the performance measures (time and number of errors) in training using PIÑATA and using the conventional training method are correlated.

Task completion time

The comparison between the mean time to perform the insertion task using PIÑATA and using the conventional system is represented in Figure 8. The mean and standard deviation (SD) values for each training method and for each access site are in Table 2.

A Shapiro-Wilk test was used to assess if the sample follows a normal distribution. Since it does not follow a normal distribution and the same group of participants tested both the PIÑATA and conventional training systems, a Spearman’s rank order correlation test was performed. The null hypothesis (H0) defined was “There is no monotonic relationship between the task completion times for PIÑATA and for conventional training methods”. The correlation coefficients and p-values obtained are presented in Table 2.

From the descriptive statistics, the efficiency of PIÑATA seems to be comparable to the efficiency of the conventional training method since the task completion time of the two training methods is similar. It is observable that the mean time is slightly lower in the conventional method, being this explained for the fact that the participants are not used to train with the AR glasses leading to higher duration times to complete the tasks. Moreover, the Spearman’s Table 1: Median (Mdn) and interquartile range of the responses to the Likert items and p-value of the one-sample Wilcoxon signed-ranks test comparing their median with the median of the scale (3.5).

<table>
<thead>
<tr>
<th>Statement</th>
<th>Mdn (IQR)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The anatomy of the mannequin is realistic.</td>
<td>5.0 (1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2. The anatomy of the projected vessels is realistic.</td>
<td>4.0 (1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3. Allows to identify the anatomical landmarks.</td>
<td>5.0 (1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>4. The possibility to see the internal anatomy (vessels) is useful.</td>
<td>6.0 (1)</td>
<td>0.002</td>
</tr>
<tr>
<td>5. The indication of the value of needle’s angles is useful.</td>
<td>5.0 (1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>6. The indication of the value of needle’s depth is useful.</td>
<td>5.0 (0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>7. The graphical elements represented to indicate the insertion site are useful.</td>
<td>4.5 (1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>8. The graphical elements represented to indicate the orientation of the needle are useful.</td>
<td>4.0 (0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>9. It is useful for training the insertion of needles.</td>
<td>5.0 (1)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Figure 8: Comparison of time to task completion (in seconds) for the jugular and subclavian access sites between PIÑATA and the conventional training methods.

Table 2: Mean and standard deviation (SD) of the task completion time (in seconds) and correlation coefficient (rS) and p-value of the Spearman’s rank order correlation test comparing the two training methods.

<table>
<thead>
<tr>
<th>Access site</th>
<th>Training method</th>
<th>Mean (SD)</th>
<th>rS</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jugular</td>
<td>PIÑATA</td>
<td>32.76 (19.17)</td>
<td>0.872</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Conventional</td>
<td>32.22 (20.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subclavian</td>
<td>PIÑATA</td>
<td>31.63 (18.98)</td>
<td>0.730</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Conventional</td>
<td>27.63 (15.35)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Access site</th>
<th>Training method</th>
<th>Mean (SD)</th>
<th>rS</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jugular</td>
<td>PIÑATA</td>
<td>32.76 (19.17)</td>
<td>0.872</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Conventional</td>
<td>32.22 (20.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subclavian</td>
<td>PIÑATA</td>
<td>31.63 (18.98)</td>
<td>0.730</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Conventional</td>
<td>27.63 (15.35)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
correlation test shown a strong, positive correlation between the task completion times, which was statistically significant \((p < 0.05)\), supporting the concurrent validity of \(\text{PI} \text{NATA}\).

**Number of errors**

The comparison between the number of errors (considering the total number obtained by the sum of the number of the redirection and arterial puncture errors) to perform the insertion task using \(\text{PI} \text{NATA}\) and using the conventional system is represented in Figure 9.

As the normal distribution condition was not verified by the Shapiro-Wilk test, a Spearman’s rank order correlation test was performed. The null hypothesis \((H_0)\) in this case is defined as “There is no monotonic relationship between the number of errors using \(\text{PI} \text{NATA}\) and using conventional training methods”. The median and IQR values of the number of errors and the correlation coefficients and p-values obtained are presented in Table 3.

Table 3: Median (Mdn) and interquartile range (IQR) of the number of errors and correlation coefficient \((r_S)\) and p-value of the Spearman’s rank order correlation test comparing the two training methods.

<table>
<thead>
<tr>
<th>Access site</th>
<th>Training method</th>
<th>Mdn (IQR)</th>
<th>(r_S)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jugular</td>
<td>(\text{PI} \text{NATA})</td>
<td>0 (1)</td>
<td>0.421</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>Conventional</td>
<td>1 (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subclavian</td>
<td>(\text{PI} \text{NATA})</td>
<td>1 (1)</td>
<td>0.198</td>
<td>0.431</td>
</tr>
<tr>
<td></td>
<td>Conventional</td>
<td>1 (1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The median number of errors is similar in both methods, suggesting that the effectiveness of training with \(\text{PI} \text{NATA}\) is similar to the effectiveness of the conventional training method. The results of the test were not statistically significant \((p > 0.05)\), not allowing to conclude about the concurrent validity in the number of errors. A possible reason for this result is the fact that the number of errors is an ordinal variable and in the majority of the analysed cases took only the values 0 or 1, being the rest of the cases considered outliers as represented in Figure 9. As the Spearman’s rank order correlation test is not very sensitive to outliers [37], this can be a reason for the unexpected result.

The step regarding the concurrency in the validation process of the proposed training method was partially completed since the results only proved the correlation in one of the two considered performance measures, the task completion time.

### 5.2.3 Construct validity

To attribute construct validity to \(\text{PI} \text{NATA}\), one should prove that there is an inherent difference in the training outcomes using \(\text{PI} \text{NATA}\) between experts (attending specialist) and novices (medical residents) in the field.

**Task completion time**

The comparison between the mean time to the experts and the novices perform the insertion task using \(\text{PI} \text{NATA}\) is represented in Figure 10. The mean and SD values for each group of participants and for each access site are in Table 4.

As expected, observing Figure 10, the difference between the two groups is evident: the group of experts need much less time to complete the tasks than the group of novices. As these are independent groups, a Mann-Whitney U test was used. The null hypothesis \((H_0)\) considered was “The population distributions of the task completion times using \(\text{PI} \text{NATA}\) by experts and novices are identical”. The p-values and Z-scores obtained are presented in Table 4.

Table 4: Mean and standard deviation (SD) of the task completion time (in seconds) using \(\text{PI} \text{NATA}\) and p-value and Z-score of the Mann-Whitney U test comparing the two groups of participants.

<table>
<thead>
<tr>
<th>Access site</th>
<th>Level of expertise</th>
<th>Mean (SD)</th>
<th>p-value</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jugular</td>
<td>Attending specialist Medical resident</td>
<td>22.66 (9.02)</td>
<td>0.040</td>
<td>-2.075</td>
</tr>
<tr>
<td>Subclavian</td>
<td>Attending specialist Medical resident</td>
<td>23.86 (13.60)</td>
<td>0.094</td>
<td>-1.722</td>
</tr>
</tbody>
</table>

The test indicated that there are statistical evidence to reject \(H_0\) \((p < 0.05)\) in the case of the jugular access site, i.e., there are significant differences in the task completion time between medical residents and attending specialists. However, for the subclavian access site, the null hypothesis cannot be rejected \((p > 0.05)\) probably due to small size of the sample (9 for
each group) or to the high variance.

**Number of errors**

The comparison between the number of errors made by expert and novice group while performing the insertion task using PIÑATA is illustrated in Figure 11. The median and IQR values are also in Table 5.

![Figure 11: Comparison of number of errors using PIÑATA for the jugular and subclavian access sites between group of experts (attending specialists) and novices (medical residents).](image)

The difference expected between the two groups is visible in Figure 11: the group of experts made less errors during the execution of the tasks than the group of novices. As previously, a Mann-Whitney U test was used with the null hypothesis (H₀) “The population distributions of the number of errors using PIÑATA by experts and novices are identical”. The p-values and Z-scores obtained are presented in Table 5.

**Table 5: Median (Mdn) and interquartile range (IQR) of the number of errors using PIÑATA for the jugular and subclavian access sites.**

<table>
<thead>
<tr>
<th>Access site</th>
<th>Level of expertise</th>
<th>Mdn (IQR)</th>
<th>p-value</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jugular</td>
<td>Attending specialist</td>
<td>0 (0)</td>
<td>0.036</td>
<td>-2.217</td>
</tr>
<tr>
<td></td>
<td>Medical resident</td>
<td>1 (3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subclavian</td>
<td>Attending specialist</td>
<td>0 (1)</td>
<td>0.067</td>
<td>-2.021</td>
</tr>
<tr>
<td></td>
<td>Medical resident</td>
<td>1 (3)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As for the task completion time, the test indicated that there are statistical evidence to reject H₀ (p < 0.05) only in the case of the jugular access site. Therefore, there is only statistical evidence to prove the inherent difference of the number of errors between experts and novices, in the case of the jugular access site. One possible explanation for this is the fact that the participants of the tests (both experts and novices) do not perform the CVC using the subclavian access site very often, being their level of expertise in that specific procedure not significantly different.

**5.2.4 User satisfaction**

The participants were asked to fill a satisfaction questionnaire comparing some aspects of the two training methods for each access site. The median and IQR of the responses are summarised in Table 6. It is important to note that in the way the questionnaires are elaborated, positive answers favour PIÑATA in comparison to the conventional method. A one-sample Wilcoxon signed-ranks test was used to determine the significance of the responses using the null hypothesis (H₀), “the difference between the median of each Likert item and the central value of the scale (3.5) is zero”.

**Table 6: Median (Mdn) and interquartile range (IQR) of the responses using PIÑATA in comparison to the conventional method.**

<table>
<thead>
<tr>
<th>Statement</th>
<th>Jugular Mdn (IQR)</th>
<th>p-value</th>
<th>Subclavian Mdn (IQR)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Easiness to use.</td>
<td>4.0 (1)</td>
<td>0.033</td>
<td>4.0 (1)</td>
<td>0.072</td>
</tr>
<tr>
<td>2. Learnability.</td>
<td>4.0 (1)</td>
<td>0.012</td>
<td>4.0 (1)</td>
<td>0.062</td>
</tr>
<tr>
<td>3. Usefulness.</td>
<td>5.0 (1)</td>
<td>0.001</td>
<td>4.0 (1)</td>
<td>0.001</td>
</tr>
<tr>
<td>4. Recallability.</td>
<td>5.0 (1)</td>
<td>0.001</td>
<td>4.5 (1)</td>
<td>0.002</td>
</tr>
<tr>
<td>5. Easiness to debriefing.</td>
<td>5.0 (2)</td>
<td>0.001</td>
<td>5.0 (1)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

The median of all the responses is higher than 3.5, confirming that the participants positively evaluate the usability of PIÑATA, considering that it facilitates the training of CVC, being easy to use, to learn and to recall. However, not all the results are statistically significant, including the easiness of use and the learnability (p > 0.05). This can be related to the fact that wearing the AR glasses requires a certain time for habituation being more difficult to use and to learn in the beginning.

For further validation it was also used the SUS system - a proved valuable evaluation tool widely used for assessing the perceived usability of a system. [35] The SUS score was calculated to each participant. The mean score obtained was 75.69 (SD = 9.07). To evaluate if this is a statistical significant result, a one-sample Wilcoxon signed-ranks test was run to test the following null hypothesis (H₀): “the difference between the mean of the SUS score and the considered average score of the scale (68) is zero”. A p-value of 0.002 indicates that PIÑATA was significantly considered as a good user experience and that has high acceptability rates.

**5.2.5 Perceived workload**

To measure the perceived workload, it was used a validated assessment tool - NASA-TLX. A one-sample Wilcoxon signed-ranks test was used to determine the significance of the responses evaluating the following null hypothesis (H₀) “the difference between the mean score of each parameter and the value from which the workload is considered somewhat high (30) is zero”. The mean, SD and p-values of the test associated to each parameter are presented in Table 7.

It is noted that the mean score for almost all parameters lies in the range were the workload of the system is considered medium. There is statistical evidence to reject H₀ for all parameters except for the level of effort probably due to the increased effort of wearing the AR glasses. Overall, from this
Table 7: Mean and standard deviation of the responses to each parameter of the NASA-TLX scale and p-values of the one-sample Wilcoxon signed-ranks test comparing their mean with the value from which the workload is considered somewhat high (30).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean (SD)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental demand</td>
<td>21.11 (13.56)</td>
<td>0.019</td>
</tr>
<tr>
<td>Physical demand</td>
<td>18.06 (12.96)</td>
<td>0.003</td>
</tr>
<tr>
<td>Temporal demand</td>
<td>7.78 (7.12)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Overall performance</td>
<td>20.83 (15.53)</td>
<td>0.014</td>
</tr>
<tr>
<td>Effort</td>
<td>29.17 (15.36)</td>
<td>0.856</td>
</tr>
<tr>
<td>Frustration level</td>
<td>14.72 (13.88)</td>
<td>0.001</td>
</tr>
<tr>
<td>Overall</td>
<td>19.26 (10.34)</td>
<td>0.002</td>
</tr>
</tbody>
</table>

analysis, it seems that PIÑATA is not considered to have high demand levels, being a valuable tool for training of CVC.

5.2.6 Verbal user’s feedback

The professionals provide a more detailed feedback in the semi-structured interviews about the advantages and disadvantages of PIÑATA, their difficulties during the tasks and their suggestions for further improvement. After a thematic analysis, this feedback was summarised in the following themes:

- **Complement to conventional training**
  Indeed, almost all the participants (14) considered that PIÑATA can complement the conventional training system, supporting the hypothesis stated in the beginning of this work: "Anything that helps to better understand the correct way of doing the procedures complements the train of inexperienced doctors". Moreover, they agreed that the image overlay is reliable and that the feedback during task execution is useful.

- **Content useful**
  When asked about the aspects that they liked most, the majority of the professionals (10) highlighted the projection of the vessels inside the simulator and some (5) referred the information about the angles and depth of the needle. This corroborates the content validity of PIÑATA already supported by the results of the questionnaires.

- **Discomfort and unnatural feeling**
  Regarding the limitations, the most mentioned topics were related to the AR headset: limited FOV and sense of discomfort wearing it. Another difficulty described by the participants was some lack of stability of the virtual elements. This is due to the tracking system that requires a conditioned environment to work properly. All these factors combined with the fact that it is the first time that professionals were exposed to this tool contribute for some non naturally expressed using PIÑATA.

- **High acceptability**
  When asked if the adoption of this type of tool and technology in medical training is possible, almost all the professionals (13) agreed. They considered the simulation as the future of medical education. Moreover, they considered that this is particularly useful in an early stage of training, in medical schools.

  - **US integration**
    Integrate US in PIÑATA was one of the main suggestions made by the interviewed. Today, it is common to perform this type of procedures with US guidance. Then, they considered that a mode to use the echograph would be a good addition for training.

6 Conclusion and Future work

This study consisted in a comparison between the proposed system and the conventional training system. The results suggest that PIÑATA fulfils the criteria of validation of a new method for training medical procedures. Only predictive validity was not assessed in this work since it requires a randomised controlled trial comparing performance performing the procedure in real patients. [3] Therefore, the initial hypothesis was confirmed: PIÑATA complements the conventional training system, reducing the dependence of the instructor without affect the quality of training. More participants and runs may make this conclusion more evident.

This system leaves room for improvement and further investigation. One of its main limitations is the tracking system that is affected by external factors including the light conditions. Also, the HMD causes some discomfort and has a limited FOV. While these are important limitations, they are described in the state of the art of AR applications showing that it is necessary to keep improving this technology. In the future, another potentialities of this type of tool can be explored including a collaborative setup for training using handheld devices. Finally, the reasoning behind the new training method proposed here could be applied to more procedures, obtaining a common AR framework to train needle-based interventions, promoting a better medical education and, consequently, improving the patient care.

Acknowledgement: This document was written and made publically available as an institutional academic requirement and as a part of the evaluation of the MSc thesis in Biomedical Engineering of the author at Instituto Superior Técnico. The work described herein was performed at Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento (INESC-ID, Lisboa, Portugal), under the supervision of Prof. Dr. Daniel Simões Lopes, and at Hospital da Luz (Lisbon, Portugal), under the co-supervision of Cátia Isabel Andrade Botelho Costa, during the period February–October 2019.

References


