

# A comparative study between automated and human evaluation of sensory-motor skills in Interactive 3D Virtual Environments involving application of intramuscular injection

Alexandre M. dos Anjos  
 Laboratory for Interactive Technologies (Interlab)  
 Escola Politécnica  
 São Paulo, Brasil  
 dinteralexandre@usp.br

Leonardo Fernandes Cherubini  
 Federal Institute of Science Education and Technology  
 Mato Grosso, Brasil  
 cherubini18@hotmail.com

Romero Tori  
 Laboratory for Interactive Technologies (Interlab)  
 Escola Politécnica  
 São Paulo, Brasil  
 romero.tori@gmail.com

Fátima L. S. Nunes  
 Laboratory for Informatics Applications in Health (LApIS)  
 School of Arts, and Humanities Science (EACH/USP)  
 São Paulo, Brasil  
 fatima.nunes@usp.br

**Abstract**— The evolution of computational techniques in the field of virtual reality has enabled new ways of discussing and using information technology in educational contexts. Given this scenario, new challenges have emerged and, among them, the need to verify the efficiency of teaching techniques in interactive three-dimensional environments must be highlighted. This work aims to show the comparison between the experimental results of the evaluation of a theoretical evaluation model carried out by human experts and the results of an automated theoretical model of evaluation and sensorimotor skills acquisition in an Interactive Virtual Environment. The experiment involved procedures such as needle positioning, insertion, and withdrawal operations in injections in the gluteal region. Results indicate that the evaluation of sensory-motor skills is feasible in interactive 3-D virtual environments, showing a higher level of agreement with the human evaluation within the stage of discriminatory evaluation, followed by a relative agreement in evaluating processes of skill degrees distributed on the basis of an average point found in the experts' evaluation and that of an automated method.

**Keywords**—*Virtual Reality; Evaluation of Sensory-motor Skills Acquisition; Medical Training; Three-dimensional Virtual Environments.*

## I. INTRODUCTION

The introduction of technological changes observed in the global scene has made the process of incorporating information technologies used in a perspective of educational processes apparent over the last years. In this context, the field of acquisition and development of Sensory-Motor Skills (SMS) must be highlighted considering that teaching and learning evaluation techniques, once simply performed by

movement reproduction processes monitored by human visual perception, can now rely on new information technology-mediated possibilities.

New demands for interactivity, efficiency and, particularly in the medical area, the reduction in the use of live beings or corpses as training or teaching objects have been met by technology in the field of Virtual Reality (VR). The application of new teaching techniques is presented at the same time that reflections on the implementation of alternatives for training contexts become necessary, particularly in the field of Sensory-Motor Skills. Alternatives need to be emphasized when we observe that the motivation for educators to use teaching-learning virtual strategies is noteworthy, especially due to the fact that they offer solutions that minimize the need for high-cost experiences such as those that require physical material and resources in real and in-person educational environments [1].

## II. WORK RELATED

The use of VR applications in the evaluation of Sensory-Motor Skills has been promoting the development of solutions for different problems in various areas, especially in Healthcare, a field that will be explored in this work. Some areas such as kinematics, statistics, mathematics, and computing are highlighted in this work because of the similarity in their evaluation criteria. In order to enhance the understanding of factors that are part of the SMS acquisition evaluation methods present in this work, related studies are included in this section.

Panait et al. [30] present a technique training program for laparoscopic surgical practice. Virtual Reality simulators were used to perform tasks carried out by means of a laparoscopy haptic device with and without force feedback. This study consists in comparing tasks that apply techniques which follow practice standards for laparoscopic surgeries where techniques are applied to evaluate training in the field of Fundamentals of Laparoscopic Surgery (FLS).

This study consisted in developing activities at three different levels. The experiment measured results for each level aiming to analyze gains and losses in the use of the device with and without force feedback. In order to evaluate results, the study used the evaluation method from the FLS field together with mathematical scoring system based on completion time. As a result of these evaluation methods, the system showed that the devices of force feedback enabled faster results in the completion of more difficult tasks. However, the tactile simulation did not show an improvement in the performance in simpler systems. Based on the results obtained from this study, the authors recommend the use of software and force feedback devices in systems that offer a higher level of difficulty.

Kolesnikov et al. [31] developed a haptic simulator that was responsible for assisting in SMS acquisition in Odontology. This simulator is able to generate the model and position trajectories responsible for guiding the learner. Based on this resource, the expert professor can create a standard procedure to serve as basis for students. After that, the student can follow the correct trajectory set by the professor. These guidance procedures include data on positioning, force and the recording of the entire trajectory. Finally, some experimental results from the mathematical scoring system were presented, which were obtained from the evaluation questionnaire on the efficiency of the proposed method to guide learners in the execution of dental tasks represented by tactile reproduction.

Schmidt e Wrisberg [4] present a SMS evaluation approach that involves three specific dimensions: the person, the environment and the task. SMS evaluation dimensions taken into consideration, the current investigation aims to specifically focus on evaluating tasks from the perspective of the analysis of human movement [5], where skills may be subdivided into gross and fine. Schmidt e Wrisberg [4] furthermore define that the analysis of the execution of movements must take into account two important aspects: the way the tasks are organized (discrete, serial or continuous) and the level of environmental predictability observed during the execution of the movement.

In order to create a taxonomy for SMS evaluation, Kraiger, Ford and Salas [6] state that motor learning has traditionally been analyzed by means of conventional methods of human observation. Hence, there is need for alternative methods for automated evaluation from contributions in the field of VR in the form of training simulations in 3D Interactive Virtual Environments (3D IVEs), in order to minimize the occasional variation of human evaluation. The results of a previous investigation [7, 8] emphasize that experiences found in

literature are limited to a single type of procedure or method that involves evaluation parameters applied to specific tasks in a certain training domain. This previous investigation also showed that discussions about a SMS evaluation model that can be extended and applied to a certain virtual training domain in a more far-reaching way are still inchoate. SMS evaluation parameters have been found grouped in six specific categories: errors, time, trajectory recording, feedback, dexterity/success, and user experience/usability. Methods and techniques for SMS evaluation grouped according to evaluation approaches can be highlighted, categorized and identified within the fields of kinematics and robotics [9,10]; statistics [11, 12, 13, 14, 15, 16, 17]; mathematics [16], [18], [19]; information technology [11, 20] and other areas of human knowledge [21]. Based on the mapping of parameters and on an exploratory study, a conceptual model was built aiming to test the feasibility of an evaluation model for virtual tasks in the field of medical training. Thus, this article aims to discuss experimental outcomes based on the use of a theoretical model for the evaluation of Sensory-Motor Skills acquisition, implemented in a 3D IVE and compare these results to the evaluation of human experts.

### III. MATERIALS AND METHODS

Based on parameters and methods cited in literature for specific contexts, a theoretical model was defined aiming to evaluate skill acquisition in any context, providing for a phase for task discrimination and another phase for task classification. The theoretical model generated an architectural model, which, in turn led to a semi-automated computer system. The models and the system above mentioned are described further on in this article.

#### A. Theoretical Model of Evaluation

The devising of a theoretical model to evaluate SMS acquisition employed contributions from the field of motor learning based on aspects that reflect task organization at specific levels of environmental predictability [4]. Fig. 1 shows the model, considering two distinct blocks: a) calibration and b) evaluation.

The first block (Calibration) focuses on the settings of parameters that are part of the evaluation process. In the second block (Evaluation) is divided into four modules. The **Definition of values for each parameter** module shows information on the movement execution of individuals where the values are set for each evaluation parameter. The **Discrimination** module focuses on labelling virtual tasks as skilled and unskilled according to evaluation parameters and rules defined by experts. The **Classification** module classifies the tasks that were labelled as skilled and unskilled in the previous module, taking into account degrees of skill based on the observation of virtual tasks execution. The last module (**Results**) provides the evaluation metrics and results processed by the "Discrimination" and "Classification" modules. Both these modules also generate the results for skilled and unskilled tasks at different skill degrees.

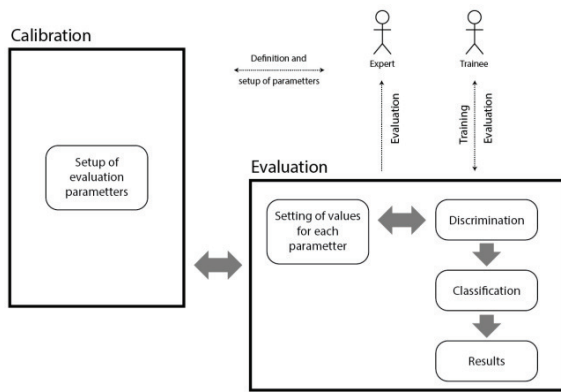


Fig. 1. Theoretical model of SMS evaluation for 3D IVE

Fig. 2 shows classification and discrimination models.

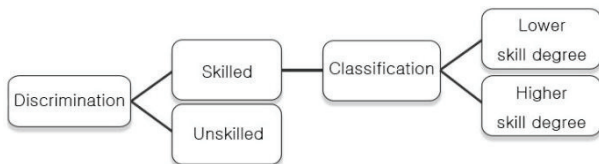


Fig. 2. Discrimination and Classification Processes

Unskilled executions may be identified when one or more evaluation parameters reveal an unacceptable condition in the task execution or in the task evaluation. Furthermore, identifying unskilled executions is especially important in risky situations when a certain task must be interrupted immediately not to offer risk to the procedure or to the executant himself (for example when piercing more than one injection site on a patient).

On the other hand, skilled executions will be recognized in this work when certain task objectives are met based on pre-set evaluation criteria.

**B. Semi-Automated Method**

In order to bring about the implementation of the theoretical model, a proposal for a system architecture is shown in Fig. 3, considering the components and modules of a semi-automated evaluation method for SMS acquisition in 3D IVEs. The term *semi-automated* is justified considering that the method created can use both parameters that are processed by a 3D IVE in an automated manner and parameters based on the experts' judgment.

In the architectural model, information on interactions, configuration parameters, method adjusting, training sessions and evaluation is processed by a computer system and stored in a database. In addition, an interaction interface was proposed with the haptic device complementing the mouse functionalities.

Similar to the theoretical model, the method architecture is divided in two components: i) calibration component and ii) evaluation component.

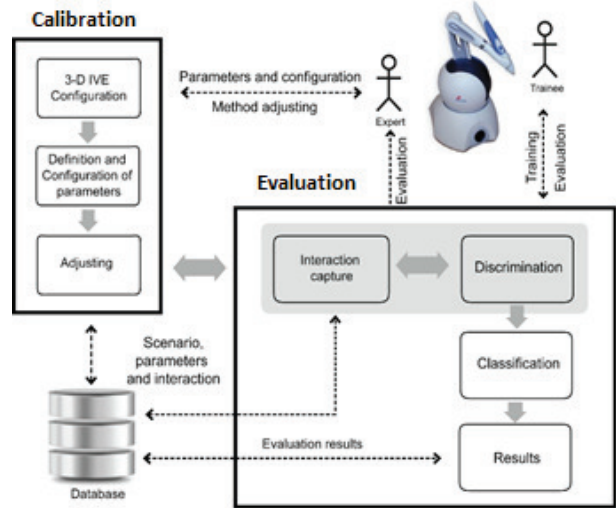


Fig. 3. Architecture of the semi-automated evaluation method of a 3D IVE

The **Calibration** block considers functionalities implemented in three modules: **3D IVE Configuration**, **Definition and Configuration of parameters** and **Adjusting**. The **Evaluation** block considers functionalities implemented in four modules: **Interaction capture**, **Discrimination**, **Classification** and **Results**.

While parameters are being configured, they can be defined as discriminating or classifying parameters. Discriminating parameters can be configured for the processing of tasks in the modules of interaction capture and discrimination, where tasks are discriminated as skilled and unskilled.

Different degrees of skill can be assessed in the architecture during the various stages of execution of a virtual task. Skill degrees are only assessed for tasks considered skilled during the classification process. In order to classify the different skill degrees evaluation parameters must be defined as classification module.

In addition, the model allows the use of data obtained in an automated manner (interaction capture module) or in a hybrid manner, combining automated evaluation parameters with judgments obtained from human experts.

Results from the processing module are made available in the last module (Results), providing a visualization of the tasks discriminated as skilled or unskilled and classified according to a higher or lower skill-degree.

*C. Evaluation performed by human experts and by the semi-automated method*

The model defined in Section II-A aims to offer conceptual definitions for the development of an evaluation method for

virtual tasks within the range of sensory-motor skills, and the semi-automated method (Section II-B) provides the visualization of the main functionalities of the architecture of an evaluation system for SMS acquisition. Evaluation procedures used by human experts and by a semi-automated method are described in the following sections so that they may be compared later on. Both procedures – human and automated – used, as an experimental basis, a 3D IVE to train the task of applying an injection in the gluteal region. The task that was to be evaluated was divided in three stages: needle positioning, needle insertion and needle withdrawal. Based on the consensus among experts, discriminating and classifying parameters were defined as shown in Table I.

TABLE I. EVALUATION PARAMETERS

Stages	Parameters	Type of evaluation
Positioning	PP1 – Do not hit an hematoma	Discriminating
	PP2 - Do not apply injection other than in upper outer quadrant of buttock	Discriminating
	PP3 – Rectilinear movements	Classifying
Insertion	IP1 – Insert needle up to its base	Discriminating
	IP2 - Insertion at a 90-degree angle (15 degrees tolerance)	Discriminating
	IP3 – Steadiness	Classifying
	IP4- Smoothness	Classifying
Withdrawal	WP1 - Withdraw needle at a 90-degree angle (15 degrees tolerance)	Discriminating
	WP2 – Steadiness	Classifying

<sup>(1)</sup>PPn – Positioning Parameter n; <sup>(2)</sup>IPn – Insertion Parameter n; <sup>(3)</sup>WPn – Withdrawal Parameter n.

### 1) Human expert's evaluation

Human experts were offered the chance to choose the method they would use to evaluate trainees in the performance of tasks in the 3D IVE. Thus, the tasks performed in the 3D IVE were assessed by experts, on the base of visual observation of the procedures and note taking of discriminatory values. Values for skilled and unskilled execution were judged as discriminatory parameters, and zero to ten grades were given to classifying parameters.

Three experts evaluated the tasks that were performed skillfully or unskillfully. In order to reach a final definition, the highest percentage in the judgments was taken into account, i.e., the individual was considered skilled when at least two of the experts considered him so. As for the classifying parameters, the average grade among the three experts was considered.

### 2) Evaluation by semi-automated method

The evaluation method is set off with the process of calibrating the 3D IVE based on the configuration of the scenario and of the objects that compose the simulated environment. After that is done, evaluation parameters are

configured based on human experts' experience. In order to set the discriminatory parameters, the system was configured so that it would verify if the task was performed correctly or not. As for the classifying parameters, techniques for the processing of metrics, which classify skills were linked, as shown in Table II.

TABLE II. TECHNIQUES FOR CAPTURING EVALUATION METRICS

Parameters	Technique for metrics processing
PP1 (D)	Marking the impact region Analysis of the impact site within a marked region
PP2 (D)	Marking the impact region Analysis of the impact site within a marked region
PP3 (C)	Analysis of the rectilinear movement or the most economical movement
IP1 (D)	Impact Detection
IP2 (D)	Angle analysis
IP3 (C)	Steadiness analysis
IP4 (C)	Smoothness analysis
WP1 (D)	Angle analysis
WP2 (C)	Steadiness analysis

<sup>(1)</sup>D - Discriminating; C- Classifying; <sup>(2)</sup>PPn – Positioning Parameter n; <sup>(3)</sup>IPn – Insertion Parameter n; <sup>(4)</sup>WPn – Withdrawal Parameter n.

In order to process those metrics, the points of each trajectory were captured during the virtual task execution every 100 milliseconds and impacts between objects were detected based on the regions marked by the experts during the configuration of the virtual environment. The information was automatically registered in a database. These data allow the analysis of the virtual task execution by individuals and the identification of patterns that resemble the “desired” pattern established by experts.

In the context of 3D IVEs, a virtual task can be analyzed by a simulated moving object and by its trajectory  $T$ , which is understood as a curve composed of a set of points  $p_t = (x_t, y_t, z_t)$ , where  $t = 1, 2, \dots, n$  that represent the three-dimensional coordinates at time  $t$ , as shown in Equation (1).

$$T = \langle p_1, p_2, \dots, p_n \rangle \quad (1)$$

From trajectories and pre-set stages, generic parameters can be obtained, such as movement of an object, angles of an object, steadiness in the movement of an object and smoothness in the movement of an object. Other parameters can be set depending on the domain of the task.

From the processing of metrics in order to analyze the rectilinear movement of a trajectory, and taking an initial and a final point as reference, a plane that represents the shortest or most economical way can be defined. The use of this plane

to verify the orientation of a rectilinear movement is justified by the need to measure distances in relation to an ideal plane, the tracing of which starts from a fixed-origin reference (withdrawal point of the instrument from the haptic device) and reaches a variable final reference (site where the instrument collides with the virtual organ).

A plane is represented by its general equation, defined in Equation (2) where a, b and c represent the points that determine a normal vector, d represents the perpendicular distance from the point to the origin on the plane and x, y, and z represent the coordinates of a rectilinear trajectory, as shown in Fig. 4.

$$ax_i + by_i + cz_i + d = 0 \quad (2)$$

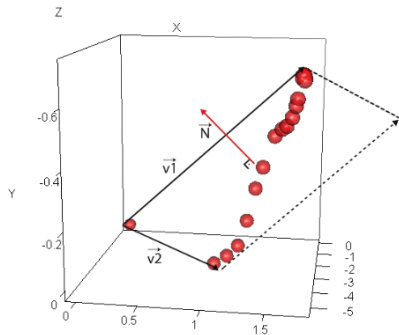


Fig. 4. Trajectory in relation to a plane

In order to check the distance of the trajectory performed by an individual to the plane, the distance from the trajectory points  $p_i = (x_i, y_i, z_i)$  to the plane must be calculated according to Equation (3) and terms originating from Equation (2).

$$D = \sum_{i=1}^n \frac{|ax_i + by_i + cz_i + d|}{\sqrt{a^2 + b^2 + c^2}} \quad (3)$$

By following this sequence, the best and worst cases of rectilinear movement can be analyzed, considering as best movements those represented by the shortest distance  $D$  and, inversely, the worst cases are represented by the longest distance. In a three-dimensional space, for each point  $p_i = (x_i, y_i, z_i)$  in space, a vector  $\vec{u}$  is associated and denoted by  $\vec{v} = (x_i, y_i, z_i)$  considering the initial point in its origin and final point in  $p_i$ .

The analysis of the positioning angle of objects can be conducted by applying a technique to verify the direction of a moving object in a certain trajectory. This metric considers vector  $\vec{u}$  and a normal vector  $\vec{N}$  of a plane that represents an impact surface (Fig. 5).

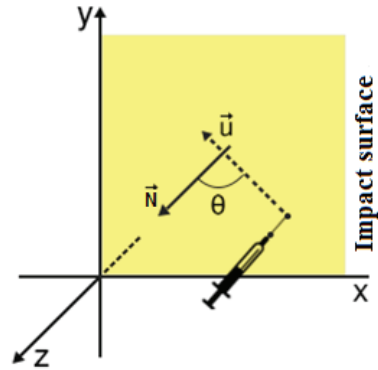


Fig. 5. Angle between a trajectory and a plane.

Thus, the positioning angle  $\theta$  is obtained by Equation (4), which represents the arc cos (acos) between vectors  $\vec{u}$  and  $\vec{N}$ .

$$\theta = \text{acos} \left( \frac{|\vec{u} \times \vec{N}|}{|\vec{u}| \times |\vec{N}|} \right) \quad (4)$$

It was also necessary to establish a technique to represent the trajectory's steadiness during the manipulation of objects in a virtual task. This technique was implemented to allow the analysis of the oscillation of the instrument as it was conducted along a particular trajectory.

Steadiness represents how much the individual's hand oscillates during insertion. During this procedure a  $\theta_T$  angle was calculated for each point  $p_i = (x_i, y_i, z_i)$  according to Equation (4), and then a value for steadiness was obtained as shown in Equation (5) where  $\bar{X}_\theta$  and  $S_\theta$  represent the standard mean and deviation, respectively, of the obtained angles  $\theta_T$ .

$$F = 1 - \frac{S_\theta}{\bar{X}_\theta} \quad (5)$$

In order to determine the smoothness parameter, instantaneous *Jerk* [10] is calculated by Equation (6), where  $\Delta a$  represents a change in acceleration and  $\Delta t$  represents a variation in time. Instantaneous *Jerk* represents the mean change rate of acceleration. The metric for smoothness was obtained from calculating the mean *Jerk* normalized by limit velocity [14].

$$J = \frac{\Delta a}{\Delta t} \quad (6)$$

So as to analyze discrimination agreement and disagreement between the experts' evaluation and the semi-automated method, a *kappa* [21] coefficient was used provided

by Equation (7) where  $p_o$  is the proportion of agreement observed and  $p_e$  is the expected proportion of agreement.

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (7)$$

In order to verify the skill degree found in the classification of the tasks performed by skilled individuals, classifying parameters that show values within the zero to ten range must be considered. Fisher's linear discriminant<sup>1</sup> [22], represented by Equation (8) is calculated for each evaluation where  $X_i$  are variables that represent the population of interest,  $\vec{L} = (a_1, a_2, \dots, a_n)$  is the discriminating vector and  $Z$  is the score attributed to each individual in the population. Based on Fisher's linear discriminant, it was possible to obtain skill degrees where the values are proportional to the skill degrees identified in the performed task.

$$Z = a_1 X_1 + a_2 X_2 + \dots + a_n X_n \quad (8)$$

Since the scores are obtained from discriminating functions, Kendall's rank correlation coefficient [23] is applied. This is a non-parametric statistical technique, that varies from zero ("lack of agreement") to one ("total agreement") represented by Equation (9) where  $k$  is the number of classification methods,  $n$  is the number of individuals and  $S$  is the sum of the obtained scores (Equation 10).

$$W = \frac{S}{\frac{1}{12} k^2 (n^3 - n)} \quad (9)$$

$$S = \sum_{i=1}^n \left( R_i - \frac{\sum_{i=1}^n R_i}{n} \right)^2 \quad (10)$$

#### D. Implementation of a semi-automated method

The possibility of capturing and processing evaluation metrics as was described in the previous section, enabled the planning of a semi-automated method. The method was implemented using Java programming language, Java 3D library [24] supported by a NetBeans [25] IDE [26] – SQLite database management system [27] connected via Java Database Connectivity (JDBC) library [28]. Open Haptics 3.1 library, provided by the equipment manufacturer, was used to establish communication with the haptic device using QuickHaptics library [29]. In addition, in order to make communication with the haptic device effective, considering that it uses C++ programming language [30], a communication layer had to be implemented using JNI (Java Native Interface) [31]. The equipment used for the execution of the virtual tasks included: Dell Optiplex 760 microcomputer, Intel(R), Core™ 2 Duo Processors. 3,00GHz, 2GB memory, 32bit Operating System, Windows 7 Professional, Service Pack 1 and a Phantom Omni – SensAble haptic device (Fig. 6).

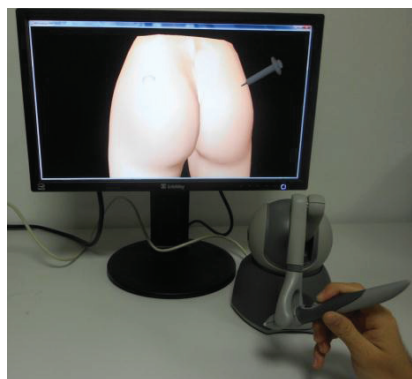


Fig. 6. 3D IVE and haptic device

<sup>1</sup> Fisher's linear discriminant finds the linear combinations of observed variables that best separate the subgroups of indicated individuals. The

variables associated to each individual's score are obtained by metrics that are automatically processed or judged according to human observation.

In order to present new functionalities in the generated application, functional requirements of the automated model can be observed in the diagram of use cases (Fig. 8), which projects the automated evaluation model for SMS acquisition. Fig. 7 shows the virtual object that simulates a human gluteal region and regions marked to discriminate skilled executions (impact within the boundary points – for example: P1, P2, P3 and P4) from unskilled executions (impact out of the boundary points) for PP1 and PP2 parameters of the positioning stage as described in Table I, section II-B.

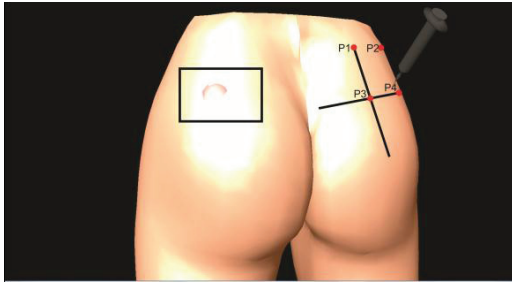


Fig. 7. Illustration of marked regions in virtual anatomical model

On the left side of Fig. 7, a hematoma can be seen in the middle of the marked region, which turns this impact site eliminatory for the application of injections.

E. Experiment

In order to verify the effectiveness of the semi-automated evaluation method, an experiment was developed aiming to analyze the quality of the evaluation of SMS in the method implemented in a 3D IVE, and then compare the results with the evaluations resulting from human experts. The task in the experiment consisted in applying an intramuscular injection in the gluteal region of a human being. To obtain parameters and calibrate the method, first a study case was carried out, involving three professors of the nursing program of a university. With this purpose in mind, a semi-structured questionnaire was prepared which allowed the collection of initial data, followed by a collective discussion that, by consensus, led to a single evaluation instrument. The experts are college professors, aged 37 to 49, with 15 – 28 years of experience.

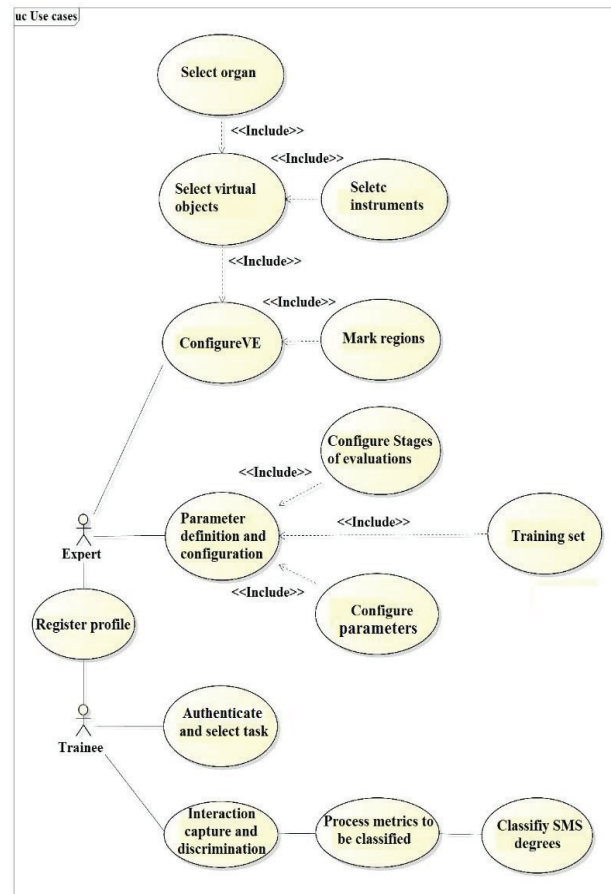


Fig. 8. Use case diagrams of semi-automated method

In a second stage, based on the discussion with the experts, the 3D IVE was adjusted where all the details for the virtual scenario, methods and necessary parameters for the task execution were configured. In a third stage the experiment objectives were shown collectively to students of a Nursing course. Forty-three students were selected, and divided into two groups. Group I included 20 volunteers in their fifth semester in Nursing, aged 20-30. These volunteers were considered experienced because they had already participated in an educational process to develop specific skills, which would be part of the task set for the experiment. Group II consisted of 23 volunteers, seven of which were masculine and 16 were feminine. These volunteers were considered beginners, as they were students who had just been admitted for the Nursing course at the institution and had no previous experience with giving injections. After a process of getting to know the 3D IVE, in a second session of the experiment, volunteers were asked to execute the virtual task. The 3D environment including trajectories, evaluation parameters and screen capture for each training session.

IV. RESULTS

Out of the 43 individuals (groups I and II) who were invited to participate in the experiment, only 38 individuals were interested in participating: 17 from group I and 21 from group II. In order to compare the evaluation results of the semi-automated method and those of the experts' evaluation, the following paragraphs will describe the results of both evaluation processes and then, data will be analyzed and discussed.

The semi-automated method allowed results to be obtained at two different points in time. First, skilled task executions were discriminated. Then, classifying skill degrees were attributed to the subset of skilled tasks. When individuals from group I (experienced) were assessed, 17 of them executed the task in the 3D IVE where the semi-automated method considered nine executions as skilled and eight unskilled. In the same way, out of the individuals from group II (inexperienced), 21 of them executed the task in the 3D IVE and the semi-automated method assessed six executions as skilled and 15 unskilled. The human experts' discriminating evaluation also judged the same 17 individuals from group I, where nine of the evaluated tasks were considered skilled and eight were considered unskilled. Out of all the 21 individuals that comprised group II and attended the experiment, five executions were assessed by the human experts as skilled and 16 as unskilled.

In the agreement analysis of the evaluations performed by the method and by the experts, there was total agreement (100%) in all the assessments related to the tasks executed by volunteers from group I. This relationship can be observed in Table III where the vertical analysis shows the data for the evaluation performed by the semi-automated method, and the horizontal analysis shows the experts' evaluation. The row and column intersection shows the convergence or consensus of results, while non-intersected cells show divergence.

TABLE III. EVALUATION BY SEMI-AUTOMATED METHOD VERSUS EXPERTS – EXPERIENCED STUDENTS (GROUP I).

Experts' evaluation	Evaluation by semi-automated method		
	Skilled	Unskilled	Total
Skilled	8	0	8
Unskilled	0	9	9
Total	8	9	17

As far as Group II is concerned, five judgments did not converge, which represents about 24% divergence in the comparison of the evaluation performed by the semi-automated method and that of the experts'. This relationship can be observed in Table IV where the vertical analysis shows data from the semi-automated method and the experts' evaluation can be seen horizontally. The row and column intersection for skilled and unskilled performances show convergence and the area outside the intersection represents divergence.

TABLE IV. EVALUATION BY SEMI-AUTOMATED METHOD VERSUS EXPERTS – INEXPERIENCED STUDENTS (GROUP II).

Experts' evaluation	Evaluation by semi-automated method		
	Skilled	Unskilled	Total
Skilled	3	3	6
Unskilled	2	13	15
Total	5	16	21

Although there is divergence in some cases in Group II, a substantial agreement was observed in the process of discrimination ( $\kappa=0,72$ ) when analyzed by the kappa coefficient (equation 6). Therefore, by consensus between both evaluations, twelve individuals were considered skilled, out of which nine belonged to Group I and three to Group II. In order to perform the process of classification by the semi-automated method, only a set of 12 tasks, which were discriminated by consensus between the human experts and the semi-automated method, were considered. Individuals were classified according to the evaluation of the 3D IVE based on Equation (8); a discriminant function (Equation 11) was generated based on the set of data captured by the semi-automated method in the 3D IVE.

$$Dx = 0.18d + 0.64Fe + 0.04Se + 1.86Fs - 20.08 \quad (11)$$

The best evaluated task was performed by individual 10 in Group I, with a score of 6.01. The worst case was that of individual 31 in Group II who obtained a -4.12 score.

The scores for the tasks executed by the individuals are shown in Fig 9 where, based on the evaluations of the semi-automated method, 10 task executions were considered above the skill's midpoint, eight of which belonged to Group I (Individuals 10, 13, 3, 17, 1, 12, 20 e 7) and two were part of Group II (Individuals 38 and 28). Two tasks were considered below the skill's midpoint, one from Group I (Individual 9) and one from Group II (Individual 31). To compare the classification process of the experts in a similar way to that performed by the semi-automated method in the 3D IVE, the average grade of the three professors was first calculated and then, the results were classified based on the midpoint found (7.38).

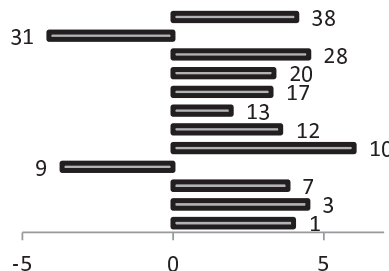


Fig. 9. Skill levels categorized in classes – Evaluation by the semi-automated method in a 3D IVE.



Individual executions were then grouped according to their evaluations: above and below the midpoint. The task that was best evaluated by the experts was performed by individual 1 in Group I, who obtained a 6.01 score. The worst case was that of individual 12 – Group I, who scored 5.33. Thus, similar to the evaluation by the semi-automated method in the 3D IVE, the results for the skill classification obtained from the experts' average can be observed above and below the skill's midpoint (7.38) for the evaluated group. According to the experts' evaluation, six executions were classified below the midpoint and six above the midpoint as can be seen in Fig. 10.

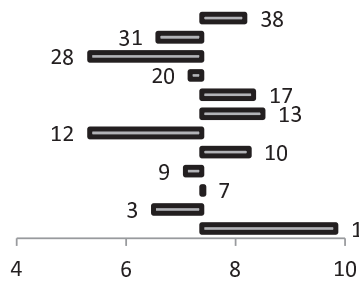


Fig. 10. Skill levels categorized in classes – Evaluation by human experts based on the average of attributed grades

Kendall's method was used to analyze the agreement in the task evaluations performed by the semi-automated method and the human experts, as defined in equation 9 where a 0.63 coefficient was obtained. The result obtained after applying the method shows there is a 63% agreement in the comparison of results between human experts and the semi-automated method.

## V. DISCUSSION

An agreement analysis by *kappa* (equation 7) was used to compare the process of discriminating tasks into skilled and unskilled by human experts and the semi-automated method, showing a substantial level of agreement in the discriminating process ( $\kappa=0,72$ ). The 28% divergence shown in the comparison between the evaluation by experts and by the semi-automated method is justified by five execution cases – individuals 22, 24, 26, 36 and 37 in Group II. The analysis of divergence showed two situations: some cases judged as skilled by the experts while the semi-automated method suggested the exclusion of those same skilled tasks (individuals 24, 36 and 37).

The metrics that represented those differences can be justified by the following discriminating parameters: positioning parameter PP2 (see Table I – section II-C) which analyses the impact within the quadrant where the injection should be applied, discriminating parameter IP2 for insertion angle which analyses the movement during the insertion of the instrument within the angles pre-set by experts and WP1 which represents free angles observed when the instrument is withdrawn from the virtual anatomical model. The analysis of the impact within or without the set quadrant made it possible

to verify the accuracy of the semi-automated method by processing techniques to calculate the distance between a given point and the plane (Equation 3) and the analysis of impacts achieved inside the marked quadrants in the simulated environment. This was done on the basis of data recorded in a database implemented in the semi-automated method which stored data on the trajectories, coordinates of the demarcated regions and videos showing movements executed in the 3D IVE. The results of this analysis showed that individuals 22 and 26 actually hit a region outside the quadrant, while human experts failed to detect the limits considered to judge according to evaluation criteria previously set and configured in the calibration block of the semi-automated method. On the other hand, human experts assessed inaccurately individuals 24, 36 and 37 when they hit the region defined by the pre-set criteria judging them unskilled, as shown in Fig. 11 A and B.

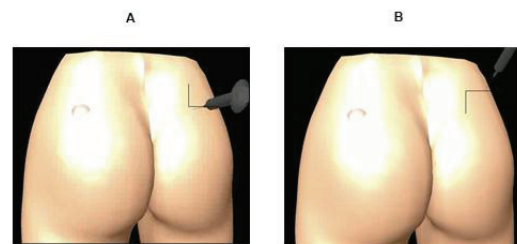


Fig. 11. Giving injections in regions bordering the demarcated area

The experts' perception of a movement made in the correct angle during the insertion and withdrawal of the instrument into and from the virtual anatomical model within pre-set degrees of freedom was another factor responsible for the differences in the discriminating process. These metrics were processed in the automated model based on angle analysis techniques (Equation 4), and then compared with the assessment of human experts. The analysis of the metrics relating to the insertion parameter IP2 also showed that human experts judged that individuals 24 and 37 performed their movement out of the degrees of freedom allowed, while the semi-automated method's judgment differed and demonstrated that those individuals executed movements within the evaluation criteria pre-set and configured in the calibration block. In addition, during the withdrawal of the instrument, experts judged the withdrawal angle of individual 24 as unskilled, while the semi-automated method considered the individual's execution as skilled. This comparison was possible when data on the execution of trajectories (equation 1) were checked, and videos that recorded the execution of the movement were analyzed.

In order to compare the classification process of the semi-automated method and that of the experts Kendall's method (Equation 9) was applied and the obtained result was a 64% level of agreement between both forms of evaluation. These results are highlighted in Figures 9 and 10 where eight cases show consensus between the two evaluations. This consensus can be verified on the basis of the following results: the executions of individuals 9 and 31 were judged below the

midpoint both by the experts and the semi-automated method, and the executions of individuals 1,7,10,13,17 and 38 were classified above the midpoint by both. The cases of classification divergence (36%) between experts and the semi-automated method are represented by the executions of individuals 3, 12, 20 and 28.

The consensus among the three experts who participated in the experiment on their evaluation was checked in order to analyze the just mentioned divergence. Results indicated considerable differences among them. The average standard deviation was 1.85, 0.25 maximum and 4.44 minimum points representing divergence of opinion or judgment among the experts. Given this analysis, it can be deduced that, although the evaluation criteria were clearly set, some degree of subjectivity present in human judgment can affect the results of the comparison between the two processes of SMS classification.

In order to verify a human opinion on the semi-automated method, the experts, who had pre-configured the 3D IVE, were invited to analyze the experiment results and contribute to the improvement of the semi-automated method. One of the factors pointed out by the experts was the need to adjust the visualization in the 3D IVE. They suggested displaying the task execution on more than one screen considering different angles. This adjustment would allow different viewing perspectives while monitoring the execution of the movement: front, side and needle at the time of impact.

These changes were developed and implemented in the architecture of the 3D IVE as shown in Fig. 12.

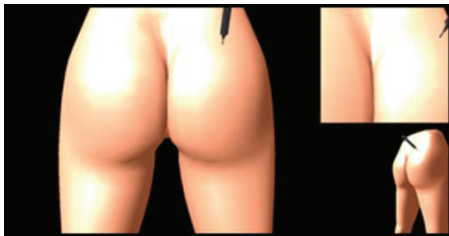


Fig. 12 Application including new viewing perspectives

The adjustment in the simulated scenario of the semi-automated method allowed the visualization of the task from different perspectives including details that would not be assessable from only one viewpoint (Fig. 12): visual details of the instrument's tip when it hits the virtual anatomical model – top right window; front view of the application – window on the left and side view of the application – bottom right screen.

In addition, the experts highlighted, as an advantage in the use of the semi-automated method, the possibility of evaluating SMS acquisition tasks after the virtual tasks are executed. This contribution may aggregate value, in the teaching practice or in SMS acquisition as well as it may allow trainees or evaluators to analyze learning goals by checking parameters that affected their performance during the execution of a virtual task.

Considering that the parameters for acquisition evaluation are configured and may be re-configured in a specific module, the flexibility of the automated model is verified when it allows experts to reconfigure evaluation criteria, automate new parameters or enter them manually if these cannot be automated. This conception adds the possibility of adjusting the semi-automated method to new evaluation strategies, according to the needs of the human experts.

Another factor to be highlighted is the ease of visualizing and representing evaluation results processed by the semi-automated method, considering that the data captured by trajectory pattern defined by Equation 1 can be easily organized in graphs in the three-dimensional (3D) space. Hence, not only is it possible to view evaluation results in 3D format, but they can also be tracked and analyzed in order to establish the association of three-dimensional coordinates with visual information frames stored in the database of the semi-automated method. In addition, due to the contributions coming from the field of virtual reality, it is possible to emphasize the ease of simulating experiences with low cost when compared to the use of human subjects and the advantages of carrying out teaching experiences and SMS acquisition activities repeatedly without the natural wear of physical simulators.

Finally, the results described in the previous section prove that automated evaluation techniques can result in more accurate metrics to support the evaluation of SMS acquisition, highlighting the analysis of trajectories, smoothness, steadiness, impact accuracy and region demarcation, in simulated scenarios implemented in a 3D IVE.

## VI. CONCLUSION AND FUTURE WORK

The main objective of this work was to compare results of a semi-automated method for the evaluation of SMS acquisition with the evaluation of human experts. In order to achieve these results, an experiment was conducted and an application was generated. Automatic methods were defined to discriminate and classify SMS and the conclusion was that it is possible to assess human skills in a 3D IVE on the basis of criteria pre-set by human experts.

In general, as a result of the comparison of the evaluations performed by the semi-automated method and the human experts, there was a higher level of agreement between both forms of evaluation in the process of discriminating SMS. On the other hand, as far as the processes of classifying SMS are concerned, the proposed method provided the chance to verify skill degrees based on a midpoint of SMS evaluation. In spite of this, there was a 63% agreement between the automated evaluation and the assessment by experts. In a process of post-experiment analysis, the results showed that factors such as computational accuracy and mathematics of the semi-automated method can justify the differences when the method is compared with the subjectivity of human judgment.

Finally, advantages in the use of the semi-automated method and contributions from the human experts to improve the method in a 3D IVE were described. Suggestions for future

work involve improving the semi-automated method and the application of the theoretical model in other domains of evaluation in the field of medical training.

## VII. ACKNOWLEDGMENTS

The authors would like to thank CNPq (National Council for Scientific and Technological Development) – process numbers: 559931/2010-7 e 401745/2013-9, Fapesp (Research Support Foundation of the State of São Paulo) – process number 2010/15691-0 and INCT-MACC Science and Technology National Institute – Scientific Computing-Assisted Medicine for their financial support.

## REFERENCES

- 1 Tori R, Nunes FLS, Gomes VHP, Tokunaga DM (2009) VIDA: atlas anatômico 3D interativo para treinamento a distância. In: XXIX Congresso da Sociedade Brasileira de Computação, *Anais ...* (WIE 2009 XV Workshop sobre informática na escola, Bento Gonçalves), Porto Alegre: SBC, 2009. p. 1–10. Available em: <<http://www.lbd.dcc.ufmg.br/>>. Acesso em: 20 Jul 2010.
- 2 Panait, L. et al. The role of haptic feedback in laparoscopic simulation training. *Journal of Surgical Research*, v. 156, n. 2, p. 312-316, 2009.
- 3 Kolesnikov, M.; zefran, M.; steinberg, A.D.; bashook, P.G. PerioSim: Haptic virtual reality simulator for sensorimotor skill acquisition in dentistry. Robotics and Automation, 2009. In: ICRA '09. IEEE International Conference on, vol., no., pp.689-694, 12-17 May 2009. Disponível em: URL:<<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5152751&isnumber=5152175>>. Acesso em: 09 mar. 2010.
- 4 Schmidt RA, Wrisberg CA (2010) *Aprendizagem e performance motora: uma abordagem da aprendizagem baseada na situação*. Trad. Ricardo D. de Souza Petersen. 4. ed. Porto Alegre: Artmed, 2010, 416p.
- 5 Teixeira LA (2000) Revisão de livro. *Aprendizagem motora: conceitos e aplicações*. (Magill R. Ano: 2000, 5. ed., publicado nos EUA em 1998), *Motriz*, 6 (1): 35-36.
- 6 Kraiger K, Ford JK, Salas E (1993) Application of cognitive skill-based and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, 78: 311-328.
- 7 Kitchenham B (2004) Procedures for performing systematic reviews. *Joint Technical Report*. TR/SE-0401. Keele University Technical Report, Australia. ISSN 1353-7776.
- 8 Anjos AM, Nunes FLS, Tori R (2012) Avaliação de habilidades sensório-motoras em ambientes de RV para treinamento médico: uma revisão sistemática. *Journal of Health Informatics*, 4: 28-34.
- 9 ChenW, Xiong C, Huang X, Sun R, Xiong Y (2010) Kinematic analysis and dexterity evaluation of upper extremity in activities of daily living. *Gait and Posture*, 32 (4): 475-481. DOI 10.1016/j.gaitpost.2010.07.005.
- 10 Tanaka Y, Ishii T, Imamura N (2008) Modeling and evaluation of human motor skills in a virtual tennis task. In: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'08. *Proceedings ...* - Personalized healthcare through technology, art. n. 4650133: 4190-4193.
- 11 Chalmers N, Seaborn G, Jung JY, Glasgow JI, Scott SH (2009) Recombination of common sensory-motor impairment evaluation techniques using a committee of classifiers. In: Annual Conference of the IEEE Engineering in Medicine and Biology Society. *Conference Proceedings... 2009*, Minneapolis, MN, pp 857-860. DOI [10.1109/IEMBS.2009.5332555](https://doi.org/10.1109/IEMBS.2009.5332555)
- 12 King RC, Guang-Zhong Y (2009) Development of a wireless sensor glove for surgical skills assessment. *Information Technology in Biomedicine, IEEE Transactions* 13 (5): 673-679.
- 13 Vuillermot S, Pescatore A, Holper L, Kiper DC, Eng K (2009) An extended drawing test for the assessment of arm and hand function with a performance invariant for healthy subjects. *Journal of Neuroscience Methods*, 177 (2): 452-460.
- 14 Colombo R, Pisano FM, Mazzone A, Delconte C, Carrozza MC, Dario P, Minuco G (2008) Assessing mechanisms of recovery during robot-aided neurorehabilitation of the upper limb. *Neurorehabilitation and Neural Repair*, 22 (1): 50-63.
- 15 Yoshikawao T, Koeda M, Sugihashi M (2007) A quantitative evaluation method of handedness using haptic virtual reality technology. In: *Proceedings IEEE Workshop on Robot and Human Interactive Communication*, art. n. 4415098, pp 298-303.
- 16 Stylopoulos N, Vosburgh KG (2007) Assessing technical skill in surgery and endoscopy: a set of metrics and an algorithm (C-PASS) to assess skills in surgical and endoscopic procedures. *Surgical Innovation*, 14 (2): 113-121.
- 17 Choi KS, Soo S, Chung FL (2009) A virtual training simulator for learning cataract surgery with phacoemulsification. *Comput. Biol. Med.*, 39 (11): 1020-1031.
- 18 Stefanidis D, Scerbo Junior MW, Scott DJ (2007) Redefining simulator proficiency using automaticity theory. *American Journal of Surgery*, 193 (4): 502-506.
- 19 Baldas V, Tang L, Saleh G, Koutsouris D (2010) A real-time automatic instrument tracking system on cataract surgery videos for dexterity assessment. In: 10th IEEE Conference on *Information Technology and Applications in Biomedicine*, 2010, 3-5 Nov, 2010, pp 1-4.
- 20 Rissanen M, Kuroda Y, Nakao M, Kume N, Kuroda T, Yoshihara H (2006) Annotated surgical manipulation for simulator-based surgical skill-transfer using SiRE – simulation record editor. In: Third Conference on Biomedical Simulation (ISBMS'06), *Proceedings ...* Berlin, Heidelberg, pp 122-131.
- 21 Raphael G, Bhneman A, Tan V, Poiman N, Berka C (2011) *Interactive Neuro-Educational Technologies (I-NET)*: development of a novel platform for neurogaming. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6780 LNAI, pp 452-461.
- 22 Johnson, R. A., Wichern, D. W. Applied multivariate statistical analysis. 4th ed. Upper Saddle River, New Jersey: Prentice-Hall, 1999, 815 p.
- 23 Siegel S, Castellan N (1988) *Nonparametric statistics for the behavioral sciences*. 2<sup>th</sup>. ed. New York: McGraw-Hill, 1988. pp. 281-285
- 24 Java 3D, available em <http://www.java3d.org/>. Acessado em 02 Janeiro de 2014.
- 25 NetBeans, available em <https://netbeans.org/>. Acessado em 09 Janeiro de 2014.
- 26 A Beginner's Guide to Integrated Development Environments, available em <http://mashable.com/2010/10/06/ide-guide/>. Acessado em 09 Janeiro de 2014.
- 27 SQLite, available em <http://www.sqlite.org/>. Acessado em 02 Dezembro de 2013.
- 28 Java SE Documentation, available em <http://docs.oracle.com/javase/7/docs/technotes/guides/jdbc/>. Acessado em 02 Dezembro de 2013.
- 29 SensAble technologies. OpenHaptics Toolkit version 3.0 Programmer's Guide. Woburn, 2008. pp 13-23.
- 30 Cplusplus, available em <http://www.cplusplus.com/>. Acessado em 02 Dezembro de 2013
- 31 Java SE Documentation, available em <http://docs.oracle.com/javase/6/docs/technotes/guides/jni/>. Acessado em 02 Dezembro de 2011