Development and preliminary research on the measure properties of a perceptual and motor demands assessment protocol for virtual reality systems

Fábio Fabregas Cairolli, Giordano Marcio Gatihno Bonuzzi, Gisele Carla dos Santos Palma, Marcos Antônio Arlindo Soares, José Eduardo Pompeu, Christina Danielli Coelho de Morais Faria, Camila Torriani-Pasin

ABSTRACT

The virtual reality has been frequently required in rehabilitation settings. However, it still lacks specificity, making it necessary to establish specific criteria to classify the most relevant aspects of electronic games to allow interventions based on virtual reality. Therefore, the purpose of this study is to develop a “Perceptual and motor demands assessment protocol for virtual reality systems” and to investigate its content validity and intra and inter observer agreement. The protocol was created through a literature review including classical studies as well as a review of recent articles about motor behavior, physical training, cognitive neuroscience and virtual reality. The previous versions were presented in study group meetings and congresses, and modified accordingly to suggestions of experts. Three examiners used the final version to analyze twice the total of 20 videos of individuals in a virtual environment and answered a questionnaire about its content validity. Most of the obtained values were classified as “good” (concordance from 80% to 89%) or “excellent” (concordance from 90% to 100%) by the three examiners and the protocol’s content validity was adequate. The protocol is valid, applicable and practical for analyzing different requirements of electronic games in a virtual environment.

Keywords: rehabilitation, virtual reality, assessment, measure properties

INTRODUCTION

Virtual Reality (VR) has received special attention since 1990 in rehabilitation researchers, due to its recent spread and increased usage in different ages and health conditions (Keshner, 2004). The VR term is defined as the interface between a human user and a computer which allows navigation and interaction in a three-dimensional environment through multisensory devices for feedback and participation (Kirner & Tori, 2004).

VR can also be seen as a therapeutic condition, which promotes a wide range of usage possibilities due to its features. Descriptions of its usage are present in literature by electronic gaming or more sophisticated interfaces. The use of electronic games as a rehabilitation resource can be seen as a specific format of VR, in which the task is enjoyable and the environment is motivating and fun. However, the exact definition of “playing” is controversial (Kirner & Kirner, 2011) and, probably for this reason, the definition of VR is commonly used as a synonym for electronic gaming (e.g., Deutsch et al., 2011; Laver, George, Thomas, Deutsch, & Crotty, 2015).

The VR environment can be classified as immersive, interactive (Cho et al., 2002; Kirner & Kirner, 2011), surrounding, realistic, safe, and consistent in the info provided to users (Rizzo, Schultheis, Kerns, & Mateer, 2004; Standen & Brown, 2005). It offers varied and quantitative...
feedback (Riener & Harders, 2012), which can favor the development of complex tasks presented in increasing difficult levels (Stewart, Whitford, McLaughlin, Rizzo, & Winstein, 2006). Such immersion and adjust of these levels make the virtual environment motivating and interesting (Levin, Weiss, & Keshner, 2015), while the previous features allow to analyze and optimize the environment itself, for research and individuals’ evaluation and treatment (Rizzo et al., 2004). The virtual environment can also be changed in all ways to adjust its features to achieve specific goals. It is considered safe for people with neurological disabilities to exercise, since doing so in the real world is often more difficult and dangerous (Standen & Brown, 2005). According to Levin, Weiss, and Keshner (2015), VR is more motivating than conventional rehabilitation, which potentiates the effects of proposed interventions, keeping the individual engaged in an interactive environment.

There are already registers of these technologies being used for neuromotor recovery and effective rehabilitation of some health conditions, such as Down Syndrome (Lin & Wuang, 2012), stroke (Laver et al., 2015; Pompeu, Alonso, Masson, Pompeu, & Torriani-Pasin, 2014), spinal cord injury (Carlozzi, Gade, Rizzo, & Tulsky, 2013), Parkinson’s Disease (PD) (Pompeu et al., 2014), psychological disorders as autism, Attention deficit hyperactivity disorder (Wuang, Chiang, Su, & Wang, 2011), anxiety and phobias (Parsons & Rizzo, 2008), and also obesity and eating disorders (Riva, 2011).

Research is also being done in order to show the development of software, interfaces, instruments and other aids for VR users’ needs (Bell & Weinstein, 2011; Broeren, Claesson, Goude, Rydmark, & Sunnerhagen, 2008). However, VR usage in rehabilitation is still empirical and lacks specificity for neurological patients’ needs. Thus, more research is needed in order to establish this resource’s use (Monteiro, 2011) and, specially, studies about the proper choice of electronic games to be used accordingly to the goals established for each individual’s rehabilitation.

Therefore, in order to make VR-based interventions for rehabilitation, specific criteria for classifying the most relevant features provided to the user of some electronic games in VR-based systems need to be established. After analyzing these features, it is possible to list the main features and demands of each game, enabling a more precise and adequate use to address each population’s needs. In this context, Deutsch et al. (2011) investigated thoroughly two games from the console Nintendo Wii, Wii Sports and Wii Fit, aiming to validate the system use itself. The work described the activities of each game, the feedback provided by the system in each case and if (and in some cases how) it challenged the user’s balance, coordination, endurance, strength and upper extremity control. However, the validation model used in this study has some limitations to the number of features examined, including only the motor domain. Besides, only one system was used, which makes difficult a wide usage of this classification system for other VR systems.

Due to the diversity of available VR systems, ranging from the originally produced for domestic use (Morrow, Docan, Burdea, & Merians, 2006; Yavuzer, Senel, Atay, & Stam, 2008) (usually known as consoles) to the ones developed specifically for therapeutic goals, developing a protocol to classify VR system demands can aid in the choice of an appropriate VR-based task to each user’s needs.

The purpose of this study is to develop a “Perceptual and motor demands assessment protocol for virtual reality systems” and to investigate its content validity and intra and inter observer agreement by using the referred protocol in two different consoles.

**METHODS**

This study was submitted and approved by the School of Physical Education and Sports of the University of São Paulo Ethics Committee (registration number: 318.659/2013). The planning, development and content validity investigation procedures followed the guidelines proposed by Benson and Clark (1982) and Davis (1996) and are described below. Recent studies related to the development and measure
properties’ investigation of other instruments from the rehabilitation field also followed those guidelines (Natalio et al., 2011; Natalio, Faria, Teixeira-Salmela, & Michaelsen, 2014).

**Instrument**

The first step on the development of this assessment protocol was a broad review of classic references and recent articles related to motor behavior (Gallahue & Ozmun, 2002; Schmidt & Lee, 2011), exercise and physical/sports training (Platonov, 2008; Weineck, 1999), cognitive neurosciences (Gazzaniga & Reuter-Lorenz, 2010) and VR (Keshner, 2004; Saposnik & Levin, 2011). This way, the first items to be included were identified. Also, there was an important role the use of the model of motor skills classification by Fleishman (Fleishman, 1964) and Gentile’s taxonomy of tasks (Gentile, 2000) on the content of the protocol. The first classified a wide range of skills in two main groups (perceptual-motor and physical proficiency abilities), while the later produced a classification based on environmental demands (static or dynamic) and the function of the action (locomotion, manipulation and stabilization).

After reviewing the referred literature, two independent examiners from the areas of Sport Science and Physiotherapy selected and prepared the relevant items to compose the protocol with a third examiner helping in case of divergence and voting on the inclusion or exclusion of the items. Preliminary versions of the protocol were presented in the study group meetings in order to get to the first pilot version. This version was then presented in the Brazilian Congress of Motor Behavior (Bonuzzi, Palma, Torriani-Pasin, Soares, & Antunes, 2012) and, after specialists’ opinions, modified accordingly. Then it was presented in two other congresses (Palma, Cairoli, Bonuzzi, Soares, & Torriani-Pasin, 2014; Soares et al., 2014), going through the same process and being adjusted again, until the first version was established; then, its content validity and reliability could be tested. The final version of the “Perceptual and motor demands assessment protocol for virtual reality systems” (Figure 1) is composed by 25 items, divided in three domains: 1- Task features (seven items), 2- Environment features (four items), 3- Biological features, divided in 3a-Biological motor features (four items), 3b- Biological conditional features (seven items) and 3c- Biological perceptual and cognitive features (three items). The items had two or three answer options, in which only one could be chosen. The domains are described below:

1- Task features: presence or absence of dual task, augmented feedback (type, when it is provided, how it is used to control movement) and intent in stability, locomotion and manipulation to the task’s goal.

2- Environment features: Ecological validity, Avatar representation on the virtual environment, Avatar’s appearance, Environment stability.

3- Biological features:
   a) Motor features: presence or absence of multi-limb coordination, bimanual coordination, bimanual asymmetry, muscle involvement;
   b) Conditional features: flexibility, muscle endurance, cardiovascular endurance, speed frequency, agility, dynamics balance and static balance relate to the task goal.
   c) Perceptual and cognitive features: presence and absence of reaction time, timing, Motor synchronization/rhythm.

Detailed definitions of each item were presented as part of the protocol in supplementary document (attachment 1)

**Sample and Procedures**

In order to investigate the assessment protocol's content validity, a semi-structured questionnaire was developed and divided in three parts. The first part had questions about the examiner’s previous experience and formation (both academic and professional). The second part had questions about each domain of the protocol (Task, Environment and Biological features) and the items’ relevance, items’ clarity, if the answer options were adequate and if some item should be included or excluded. The third part asked for the time spent to fill in the protocol during the videos’ analysis (described below) and comments about whether using the protocol was practical or not.

Additionally, five young college students were
recorded with a SONY DCR-PJ5 camcorder (after signing a consent) performing different activities in a VR environment. Two different consoles were used in this phase of the study (Nintendo Wii and XBOX 360 with motion sensor Kinect). The selected games were five from the “Kinect Adventures!” (20,000 Leaks, Rally Ball, Reflex Ridge, River Rush, Space Pop) and five from the game “Wii Sports” (Baseball, Bowling, Boxing, Golf, Tennis). A total of 20 videos were recorded, with two different individuals for each activity.

Three examiners with previous academic experience in Physiotherapy and Sport Science, with two of them also having clinical/professional experience, participated in the following part of the study after signing a consent. A more detailed description of their experience can be seen in Table 1.

<table>
<thead>
<tr>
<th>Task features</th>
<th>Present (✓)</th>
<th>Absent (☐)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of feedback</td>
<td>Knowledge of performance (☐)</td>
<td>Knowledge of results (☐)</td>
</tr>
<tr>
<td>Time of feedback</td>
<td>Concurrent (☐)</td>
<td>Delayed (☐)</td>
</tr>
<tr>
<td>Feedback processing and movement control</td>
<td>Online control (☐)</td>
<td>Offline control (☐)</td>
</tr>
<tr>
<td>Stability tasks</td>
<td>Related to task goal (☐)</td>
<td>Partially required (☐)</td>
</tr>
<tr>
<td>Location tasks</td>
<td>Related to task goal (☐)</td>
<td>Partially required (☐)</td>
</tr>
<tr>
<td>Manipulative tasks</td>
<td>Related to task goal (☐)</td>
<td>Partially required (☐)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Environment features</th>
<th>Ecological validity</th>
<th>High (☐)</th>
<th>Medium (☐)</th>
<th>Low (☐)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avatar representation on the virtual environment</td>
<td>First person (☐)</td>
<td>Third person (☐)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avatar appearance</td>
<td>Customized (☐)</td>
<td>Previously defined (☐)</td>
<td>Player's representation (☐)</td>
<td></td>
</tr>
<tr>
<td>Environment stability</td>
<td>Unstable environment (☐)</td>
<td>Stable environment (☐)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Biological features</th>
<th>Motor features</th>
<th>Present (✓)</th>
<th>Absent (☐)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexibility</td>
<td>Related to task goal (☐)</td>
<td>Partially required (☐)</td>
<td>Not required (☐)</td>
</tr>
<tr>
<td>Muscle endurance</td>
<td>Related to task goal (☐)</td>
<td>Partially required (☐)</td>
<td>Not required (☐)</td>
</tr>
<tr>
<td>Cardiovascular endurance</td>
<td>Related to task goal (☐)</td>
<td>Partially required (☐)</td>
<td>Not required (☐)</td>
</tr>
<tr>
<td>Speed frequency</td>
<td>Related to task goal (☐)</td>
<td>Partially required (☐)</td>
<td>Not required (☐)</td>
</tr>
<tr>
<td>Agility</td>
<td>Related to task goal (☐)</td>
<td>Partially required (☐)</td>
<td>Not required (☐)</td>
</tr>
<tr>
<td>Dynamic balance</td>
<td>Related to task goal (☐)</td>
<td>Partially required (☐)</td>
<td>Not required (☐)</td>
</tr>
<tr>
<td>Static balance</td>
<td>Related to task goal (☐)</td>
<td>Partially required (☐)</td>
<td>Not required (☐)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sensory and cognitive features</th>
<th>Reaction time</th>
<th>Present (✓)</th>
<th>Absent (☐)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timing</td>
<td>Present (✓)</td>
<td>Absent (☐)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Perceptual and motor demands assessment protocol for virtual reality systems

<table>
<thead>
<tr>
<th>Table 1: Examiner’s Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant</td>
</tr>
<tr>
<td>Examiners 1</td>
</tr>
<tr>
<td>Examiners 2</td>
</tr>
<tr>
<td>Examiners 3</td>
</tr>
</tbody>
</table>

Legend: VR – Virtual Reality

The examiners evaluated independently the assessment protocol’s first version. Each one of them got a copy of the assessment protocol with a text explaining its development phase and the
recorded videos. They were oriented to independently read everything first, watch the videos, classify them using the protocol and, finally, answer the questionnaire for the content validity evaluation. All study’s phase were conducted in Portuguese.

For the reliability test, inter-observer agreement rates were used, considering the first analysis of the 20 videos by the three examiners. After 15 days, the examiners analyzed the same videos again in a random order, also independently, with the same version of the protocol. This second data was used to test the intra-observer agreement reliability.

**Statistical Analyses**

For statistical analysis of the content validity, a descriptive analysis was made considering the simple rates of the examiners’ answers on the questions about relevance and clarity of items and domains/subdomains in the questionnaire.

For inferential statistical analysis of reliability, which was done separately for each console (Nintendo Wii and XBOX 360), the observer agreement test (Thomas, Nelson, & Silverman, 2012) was used for each item of the protocol. The obtained results were then classified according to Giannichi’s parameters (Giannichi, 1984): 0-0.59 is classified as “weak”, 0.60-0.79 is “regular”, 0.8-0.89 is “good” and 0.9-1 is “excellent”. It was set as reliable the concordance higher than 0.7 to determine the test’s reliability, according to Kiss (1987) and Tabachnick and Fidell (2007).

**RESULTS**

Concerning the content validity, all the items in the three domains of the assessment protocol were considered relevant to its context of evaluation, seen as well-explained and the answer options were appropriate on all items by all the three examiners. None of them requested the inclusion of a new item or the exclusion of an existing one in any of the three domains. Finally, the examiners spent an average of nine minutes to fill in the protocol.

![Table 2](image-url)
Intra and inter-observer agreement rates obtained for each of the consoles are presented in tables 2 and 3.

As it can be seen, most of the intra-observer agreement rates can be classified as “excellent”, and most of the inter-observer agreement rates are classified as “good”. However, there are some values classified in lower standards, especially on the Biological features from XBOX 360’s analysis: Bimanual asymmetry, Flexibility, Muscle endurance, Cardiovascular endurance, Agility, Timing. Inter-observer agreement rates for these items were of 0.6, while high intra-observer agreement rates were obtained. The mean value for Biological features in XBOX 360’s analysis was the lowest (0.785), being the only one classified as “regular”, despite still being in the established validity standards.

The lowest rates found in Nintendo Wii’s analysis were on the items “Ecological Validity” (0.4 in inter-observer agreement and 0.6 in intra-observer agreement from the second examiner) and “Motor synchronization/ rhythm” (0.6 inter-observer agreement).

Table 3
Intra and inter-observer agreement rates obtained in XBOX 360’s environments.

<table>
<thead>
<tr>
<th>Items</th>
<th>Task features</th>
<th>Dual Task</th>
<th>Type of feedback</th>
<th>Time of feedback</th>
<th>Feedback processing and</th>
<th>Stability tasks</th>
<th>Locomotion tasks</th>
<th>Manipulative tasks</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Examiner 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Examiner 2</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Examiner 3</td>
<td>0.971</td>
<td>0.942</td>
<td>1</td>
<td>0.8</td>
<td>0.914</td>
<td>1</td>
<td>1</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Environmental features</th>
<th>Ecological validity</th>
<th>Avatar representation on</th>
<th>Avatar appearance</th>
<th>Environment stability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examiner 1</td>
<td>0.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Examiner 2</td>
<td>0.8</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Examiner 3</td>
<td>0.85</td>
<td>1</td>
<td>1</td>
<td>0.85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Biological features</th>
<th>Multi-limb coordination</th>
<th>Bimanual coordination</th>
<th>Bimanual asymmetry</th>
<th>Muscle involvement</th>
<th>Flexibility</th>
<th>Muscle endurance</th>
<th>Cardiovascular endurance</th>
<th>Speed frequency</th>
<th>Dynamic balance</th>
<th>Static balance</th>
<th>Reaction time</th>
<th>Timing</th>
<th>Motor synchronization/ rhythm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examiner 1</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>0.8</td>
<td>0.8</td>
<td>1</td>
</tr>
<tr>
<td>Examiner 2</td>
<td>1</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>1</td>
<td>0.8</td>
<td>0.8</td>
<td>1</td>
</tr>
<tr>
<td>Examiner 3</td>
<td>0.871</td>
<td>0.914</td>
<td>0.985</td>
<td>0.785</td>
<td>0.85</td>
<td>0.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

| TOTAL                     | 0.896                   | 0.936                 | 0.992              | 0.832             | 0.85        | 0.8              | 1                      | 1              | 1              | 1             | 1             | 1      | 1                           |

DISCUSSION

Results showed that the proposed version of the assessment protocol is adequate and its items have an acceptable content validity, according to the three consulted specialists.

However, it can be seen that the reliability results varied in both systems: despite the great unity concerning the intra-observer agreement rates, inter-observer rates had their lowest values on Wii’s analysis of Environment features, while XBOX 360’s lowest values in this case were found on Biological features.

Such results may have appeared due to peculiarities from both systems, such as the fairly “childish” look from Nintendo’s games, interfering on the ecological validity’s evaluation. For its greater precision and consequently being able to show a greater spectrum of corporal demands, XBOX 360’s motion sensing input device, Kinect, might also justify these findings.
In these cases, the protocol itself can be seen as a tool capable of bringing up these peculiarities from different systems.

It is also possible to speculate that the description of an item was precise, but the answer options given were limited to describe the possibilities for that item. For example, the Biological features had three answer options in most of the items (“Mainly required”, “Partially required”, “Not required”). Still, every item from this part can be seen as partially linked to success in any task, which may have contributed for lower inter-observer rates.

However, such division aims to evidence the relative importance of each capacity and at least try to avoid “present” and “absent”-type answers, since each examiner may perceive the same action in different ways. The use of different methods or a previous training, as suggested by an examiner, could be an option. However, it would demand more time before the beginning of its usage by an individual and make the protocol less practical to be used.

With the obtained results, it is expected that the protocol will be used to a selection of VR-based games which is more specific to a population’s needs. For example, it can be used to help treating people with PD (which usually have impaired balance, worse performance in dual task, etc) (Pompeu et al., 2014), by choosing a game that stimulates these abilities, such as XBOX 360’s “20,000 leaks” or Nintendo Wii’s “Tennis”. Thus, using the assessment protocol allows the alignment of game features and user needs in order to find the best resource to treat a patient.

Such descriptions can also help raising the specificity of systems to be created, working as a guide to game production to be used for game industry workers who are less used to the possibilities/needs within rehabilitation and movement areas.

We can point as a study limitation, the time of experience of the examiners who assessed the protocol. However, we have to consider that the advent of VR technology is recent and the studies using this technology, as well as, the development of these VR tools can be considered relatively new, especially in development countries. Therefore, we considered at least 4 years of experience in the research and in the clinical practice with VR. As an expertise parameter, it does not corroborate with the literature which considers at least 10 years of experience for this classification (Kiss, 1987; Thomas et al., 2012).

Lastly, future studies are necessary to test the use of the protocol in other types of systems, in specific populations, as well as a second, most detailed analysis of its content validity.

**CONCLUSION**

The “Perceptual and motor demands assessment protocol for virtual reality systems” presented valid content, practicality and adequate intra and inter-observer reliability for both tested consoles.

**Acknowledgments:**
Nothing to declare

**Conflict of interests:**
Nothing to declare

**Funding:**
This research was performed with support from CAPES, organization of the Brazilian government dedicated for developing human resources.

**REFERENCES**


