



# UQE-3D: a usability evaluation method of 3D user interfaces for the elderly

Manoela Rogofski Brum Postal  [ Universidade de Passo Fundo | [manoela.brum@hotmail.com](mailto:manoela.brum@hotmail.com) ]

Luciane Daroit  [ Universidade de Passo Fundo | [ludaroit@upf.br](mailto:ludaroit@upf.br) ]

Rafael Rieder  [ Universidade de Passo Fundo | [rieder@upf.br](mailto:rieder@upf.br) ]

**Abstract** Health professionals could use 3D user interfaces to support elderly rehabilitation, offering fun and engagement during physical and cognitive activities. The evaluation of these immersive applications needs instruments designed for the specific context. This study presents and validates a new usability evaluation method for 3D user interfaces for the elderly. We developed the UQE-3D questionnaire from previous studies keeping in mind the 3D aspects and technological language suitable for the target public. To apply it, we used a protocol considering a demographic questionnaire, the Mini-Mental State Exam, the 15-item Geriatric Depression Scale, the SUS Scale, the UQE-3D, and a structured interview. We also executed an experiment considering a heterogeneous group with 30 subjects (60+ years), where seven participants were institutionalized elderly. UQE-3D presented good results showing a mean score of 82.60 (range 1-100). UQE-3D and SUS scores did not show a statistically significant difference, highlighting the UQE-3D to be sufficient and effective in identifying the usability issues seen in the assessment of 3D user interfaces for the elderly. The method has the potential to evaluate and ensure the quality of 3D context-specific applications, considering appropriate terminology and contributing to the development of technologies fitting for the elderly.

**Keywords:** 3D user interface, elderly, evaluation, UQE-3D, usability

## 1 Introduction

According to Osagie et al. (2017), usability evaluation provides support for solution acceptance and adoption, reduce costs, and influence on design, improving return on investment and impacting end-user satisfaction. Solano et al. (2016) highlight that usability is a fundamental quality characteristic for the success of interactive systems, like games or immersive applications.

According to Cockton (2013), methods and metrics contribute to determine the usability extension, measuring the robustness, the goals and the reliance - when the usability evaluation points about the utility of a system or a device. Because of this, it is necessary to use methods or protocols including reliably evaluation.

Considering Nielsen (1994) and Nielsen (1996), the main usability characteristics to evaluate are the easiness and the efficiency during the task performance, the easiness to reuse resources, the reestablishment of the services after system faults, and the satisfaction experienced by the participant during the use of the system.

Three-dimensional user interfaces (3DUI), as Virtual Reality (VR) and Augmented Reality (AR) applications, require user evaluations considering the spatial context to improve the usability (Kharoub et al., 2019). Traditional interfaces (non-3D) can use different evaluation approaches, from informal studies with users to formal experiments based on heuristics evaluations. Regarding 3DUI, it is difficult to evaluate the usability without the aid of real users interacting with tasks in immersive or mixed environments, because it is hard to measure how much an interface is being intuitive and easy to use without the user interaction (LaViola Jr et al., 2017).

VR serious games for elderly are examples of applications

recently used in clinical intervention of rehabilitation (Tierl et al., 2018; Aminov et al., 2018). A serious game follows the same entertainment principles of interactive digital game, but the aim is to transmit also an educational content or user training. In the elderly case, they can stimulate the practice of beneficial activities to the human body and increase the interest of the patient for the treatment, because the traditional intervention usually is slow and painful (Miller et al., 2017; Khaled et al., 2018). However, these objectives are easier to reach if the interface has been appropriate to its public and has a good usability.

Efforts have been applied to project and to develop games destined to the older people because these tools have shown their efficiency to improve the reaction time, the visual perception, the cognitive abilities, the self-confidence and, consequently, the life quality and welfare of this category of users (Simor, 2016). Fua et al. (2013) relate improvements in memory and attention of the elderly with the use of a serious game developed to this aim. Authors also mention that computer games are capable of improving the reaction time and motor abilities, as coordination and dexterity, and that these improvements can have a direct impact on the daily of the elderly.

Applications for seniors require, like any other system, an evaluation method to test the quality of the interaction, to demystify the lack of access, practice or fear of this public (Cota et al., 2015). Besides, it is necessary to evaluate the interface considering its usability to ensure if the solution is proper to the elderly profile. Usability interface issues can cause problems in the use learning, in the efficient use itself or the user's satisfaction degree.

Research in this context justifies it because the elderly population have increased around the world. This public also

search for games and 3DUI solutions to entertainment, rehabilitation or training - independent of being serious games or games intended to their age group. According to Hall and Marston (2015), due to the increase of life expectancy, games to the elderly have shown it as valuable tools for health promotion and education. It also promotes social interaction with people of the same age group and allows for a closer connection with younger (Osmanovic and Pecchioni, 2016).

To work with a specific public such as elderly, it is important to use tools that fit in specific classification contexts (Brace, 2018; Krosnick, 2018), like 3DUI, and intends to contribute in quality to this age group. In this case, developing a instrument to a specific context of elderly people is relevant because it allows evaluate if the interface is appropriate (or no) to elderly profile. An appropriate evaluation method to this profile can help in the development, for example, of better VR systems to a public that, gradually, get closer technology through entertainment with its descendants, serious games to rehabilitation (Trombetta et al., 2017) and interactive applications that stimulate the physical exercise practice (Konstantinidis et al., 2014; Báez et al., 2016). In face of this, the 3DUI usability evaluation is fundamental to ensure a good experience for elderly people, minimizing issues inherent to older age limitations.

With this in mind, we performed a literature systematic review to identify evaluation methods and instruments used during interventions with seniors involving virtual environments (Postal and Rieder, 2019b). The idea was to ground to posterior development of a methodology to evaluate the 3DUI, directed exclusively to experiments with seniors. Our research were generic, without any focus in device, interface aspects, or interaction technique.

Our systematic review did not point to a tendency concerning interface usability evaluation performed in experiments with older users. We have not found tools, techniques, or methods that consider the specific concept classification pointed by Brace (2018), Krosnick (2018) and LaViola Jr et al. (2017). The selected studies did not mention about adaptations of the tools used for the specific context with the elderly samples. About the 3DUI, we also have not found a specific method to evaluate virtual environments, interface aspects, or interaction techniques considering the elderly.

In view thereof, this study presents and validates the UQE-3D questionnaire, a new context-specific method focused on the elderly to evaluate 3DUI usability. We studied well-defined evaluation techniques and some methods used by selected studies in our systematic review. UQE-3D proposes items using a technological language more appropriate for the elderly to evaluate 3DUI applications designed for them with more reliability. With this in mind, we used a VR-based serious game for elderly rehabilitation in our case study with a convenience sample of seniors and suggesting an application protocol to use the UQE-3D. Our proposal showed itself enabled us to identify subject opinions and sensations about the evaluated application, keeping in mind the language of the target public.

This document is organized as follow: Section 2 presents our questionnaire and the methodology used to define it; Section 3 reports the results to validate the UQE-3D, considering an experiment involving elderly subjects to evaluate a VR

exergame; Section 4 shows discussions about the proposed method and its validation; Section 5 presents conclusions and future work.

## 2 Materials and Methods

Considering the relevance to use specific tools to evaluate 3DUI (LaViola Jr et al., 2017), the popularity of VR/AR for healthcare solutions for older adults (Huygelier et al., 2019), and the results obtained in previously systematic review (Postal and Rieder, 2019b), we decided to propose the UQE-3D questionnaire, a new context-specific method to evaluate the usability of 3DUI for the elderly. Our intention is to verify if this new questionnaire, using a technological language more appropriate of the target group, can report issues inherent to the usability of VR/AR applications designed for the elderly. Therefore, the creation of a 3DUI context-specific tool for the elderly can guarantee, for example, that an evaluation method can be clearly understood by the subjects, minimizing errors and research bias because of semantic misunderstanding.

As a basis to create our questionnaire, we used the VR/AR concepts listed by Burdea and Coiffet (2003), the 3DUI evaluation aspects cataloged by LaViola Jr et al. (2017), the question and questionnaire design by Brace (2018) and Krosnick (2018), and Nielsen's heuristics for user-interface design (Nielsen, 1994, 1996). We also used results of pilot studies in previous experiments proposed and tested by Simor (2016) and Postal and Rieder (2019a). These authors describe their methods considering a protocol in three stages: Pre-Test, Test, and Post-Test (Table 1). We also suggested this protocol format to apply our method. Table 2 presents the main concerns related by the authors during their experiments. Figure 1 shows the conceptual framework to illustrate how our questionnaire was thought and designed. The UQE-3D is highlighted in Section 2.3.

According to the protocol presented in Table 1, our propose considers three stages: Pre-Test, Test, and Post-Test, presented in next subsections. The time duration of each procedure considered pilot studies (Simor, 2016; Postal and Rieder, 2019a) before the experiment. We recommended the use of two sessions, 20 minutes, offering a good experience and reducing subjects' boring and tiring. If necessary, the time can be reallocate between stages, considering the context of use of the protocol. Since the application of our method focus on a public that might be having their first contact with the immersive technology, these details can collaborate to getting accurate results. First section applies pre-test questionnaires and training steps, and second session executes user experiment (test) and post-test stages.

**Table 1.** Proposal for a protocol in which our method is applied.

Day	Stage	Procedure	Duration
1	Pre-Test	Questionnaires	~15 minutes
		Training	5 minutes
2	Test	Interaction	1 minute
		Rest	2 minutes
	Post-Test	Questionnaires	~5 minutes
		Interview	~12 minutes

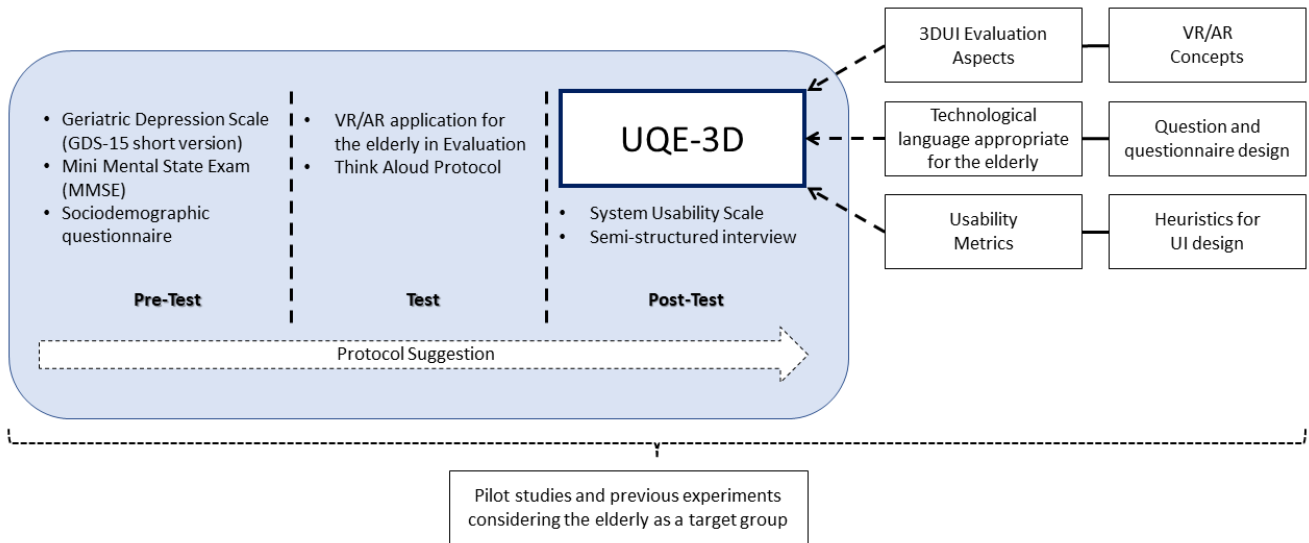


Figure 1. The conceptual framework for designing the UQE-3D questionnaire.

Table 2. Weakness listed by Simor (2016) and Postal and Rieder (2019a), and some solutions implemented in our proposed application protocol.

Weakness to be solved	Implemented solutions in our application protocol
Wait time to participate of the test	Personal experiment, scheduling a day and an hour with each participant
Time spent to apply the application protocol	Divide the execution of application protocol in two days (training and experiment), distributing the time equally
Interaction time	Minimize the interaction time to be enough, but not exhaustive
Number of questionnaires involved in pre-test	We cannot minimize the questionnaires' number
Lack of a training phase	We created a moment to familiarize the user with the 3DUI interaction process
Elderly difficulties about specific terms	Create questions considering an expert review assessment, and insert an interview in the post-test stage

### 2.1 Pre-Test

This stage considers the first contact between subject and observer. First, the observer explains all experiment steps and the research goal, and requires the subject to read and sign the Informed Consent Form. Moreover, pre-test uses instruments to select and categorize the participants, verifying previous knowledge about technologies and some physical and cognitive limitations. This step is relevant to posterior application of our method because can determinate the participants background, what can further a better analyze of the results. We apply the following instruments:

- Geriatric Depression Scale (GDS-15, short version) (Sheikh and Yesavage, 1986);
- Mini Mental State Exam (MMSE) (Brucki et al., 2003);
- Sociodemographic Questionnaire (Postal and Rieder, 2019a).

The GDS-15 form is a superficial evaluation of the participant to verify depression degree. According to Sheikh and Yesavage (1986), a depressive subject tends to present not reliable data, because there is a possibility that his psychological state interferes with the test result. This test shows the participants that present depression severe degrees, no recommended to continue the experiment. Questions verify if the participant is satisfied with himself/herself and with his/her life, answering just "yes" or "no". For this experiment, we adopted the cut-off score of 5 for healthy subjects, and score of 10 for subjects with minor depression cases (Almeida and Almeida, 1999).

MMSE is a score test to evaluate the user cognitive function ( 7 minutes to examine), easy to apply, and does not require a specific material. We used it and GDS-15 to define which participants are prepared to the next steps of the experiment. The questions involve issues of spatial orientation, temporal orientation, immediate memory, evocation memory, math, naming language, repeating, understanding, reading, and copy of a draw. The cut-off score of 25 to elderly with higher education, 18 to elderly with primary or secondary education, and 13 to pre-primary or illiterate (Lourenço and Veras, 2006).

The scores applied to the MMSE and GDS-15 were defined according to literature (Sheikh and Yesavage, 1986; Brucki et al., 2003).

The sociodemographic questionnaire aims to collect general information of the sample. We considered questions about age, level of education, motor limitation, and technological familiarity (especially with devices or resources used during the experiment). This questionnaire is adaptable according to the evaluation focus.

The Informed Consent Form is a statement that the study involves research. Our document also informs that the participant can leave the experiment any time, and the collected data will be anonymous and exclusive to the study.

After filling questionnaires, the participants starts a training phase, interacting with 3DUI resources (scene, tasks and equipment) during 5 minutes. They receive an explanation

about the VR/AR system operation and can interact freely. This procedure is relevant for the participants in next stage ("test") (Postal and Rieder, 2019a), avoiding novelty and contributing to a more critical evaluation. Moreover, the training helps to prevent breaks in presence during an immersion experience, and to allow user clarifies doubts about devices and interaction process tasks. We inserted in our application protocol this phase considering previous pilot studies (Simor, 2016; Postal and Rieder, 2019a).

## 2.2 Test

In this stage, the participants interact with the system in evaluation, performing a specific task, and using the Think Aloud protocol (Nielsen, 1994). In this way, the observer will be able to identify the participants' difficulties and impressions more easily, taking notes during the interaction process.

We recommend does not exceed two minutes for each interaction task, especially if the user needs to repeat natural movements. Considering older users, a long exposure time can tire the participant and risks to discourage their sincerity and participation in the research, or cause discontentment with the task, interfering in the evaluation process.

## 2.3 Post-test

During the post-test, we apply our questionnaire. In this step, each participant transmits his/her impressions and opinions about the interface and the interactions through two questionnaires and one semi-structured interview. The questionnaires are:

- System Usability Scale (SUS) (Brooke, 1996);
- 3DUI Usability Evaluation Questionnaire for Elderly (UQE-3D: Table 5).

SUS questionnaire was included in our application protocol because it was the most used in the selected studies, providing a reliable tool for measuring usability (although not suitable for specific contexts, like 3DUI applications and elderly samples). In the next session, we will discuss the validity of keeping it on the application protocol or not, due to the specificity level of our approach.

UQE-3D is the questionnaire we have developed aiming to evaluate 3DUI considering the elderly as a target public. This tool was thought to be applied directly in experiments involving subjects with 60+ years (Brace, 2018; Krosnick, 2018), revised by four professionals: two healthcare researchers (gerontologists), and two computer science researchers (usability engineers). This revision aimed to elect points to improve the questionnaire so that it can catch up better the target group, like the adequacy of terms to the specific context of the target group and the verification of the statements' senses. Table 3 resumes the main appointments of the revision performed by the professionals according to their areas of expertise.

To attend the professional's appointments about the questionnaire's first version, we perform the changes proposed by them. Table 4 presents the final version, which is adapted to our test experiment (which one we used a VR exergame with natural interactions and immersive visualization). To provide

**Table 3.** Issues listed by the experts to be altered in the questionnaire.

Items	Computer Science Experts
all	The descriptive form of the answers can confuse the users
all	Attend to the items senses (positive and negative)
all	Standardize the expressions used
1	Evaluate just one aspect (nice or comfortable)
2	Change the expression "real world"
6	Specify the interaction mode
8	Change the term "stare"
9-12	Reinforce that these items are relating to the image
11	Change the expression "game interaction elements"
12	Change the expression "aural elements"
12	Focus the item of listening or assimilate
14	Change the term "sounded"
	<b>Healthcare Experts</b>
all	Change the descriptive answers to a scale pattern
	Add a question: Do you consider easy to locate yourself into the game? (the answers will show the easiness or hardness degree)
	Add questions about the rest and task time
4	Change the expression "geographically-oriented"
6	Relocate this item nearest to the item about fun
8	Give a sense (positive or negative) to the item
9	Evaluate just one aspect
9-11	Summarize in two items
11	Evaluate very much aspects of the same item
12	This item are not relate to the image

a way for observers to adapt our questionnaire to evaluate different types of applications and interactions (game, virtual environment, simulation, etc.), we have also produced a table highlighting the purpose of each item (Table 5). However there will not be significant changes in the items if the interface do not be a game, once that the words to be changed will be the ones that refers to the interface, the kind of the interaction (moves) or the devices used.

A part of UQE-3D aims to identify the following interface aspects: comfort, welfare, immersion, presence, and perception (intuitive visual and aural elements). These aspects are from 3DUI evaluation metrics considered by LaViola Jr et al. (2017), and defined as VR/AR basic principles by Burdea and Coiffet (2003).

According to LaViola Jr et al. (2017), 3DUI should be intuitive, offering ways for good sensory feedback, besides should not be intrusive, providing the sense of comfort and welfare during the use (minimizing cybersickness). The authors highlight that presence and immersion are essential concepts to evaluate 3DUI properly.

Burdea and Coiffet (2003) define three VR primordial concepts: Immersion, Interaction, and Imagination. Huang et al. (2010) explains these concepts stating that immersion divides itself in mental and sensory (the first congruent with was defined by LaViola Jr et al. (2017), being a consequence of the second, which happens with the interaction with the interface through sensory stimulus). The interaction promotes the immersion sense, once that proffers the system reciprocity in contact with the user. Imagination relates itself with involving because it stimulates the abstraction and the capacity of the human mind to perceive and to realize creatively, according to the stimulus received.

Another part of the UQE-3D considers usability metrics defined by Nielsen (1994) and Nielsen (1996): easy to learn, efficient to use, easy to remember, error tolerant, and pleasant to use. We also add the relationship between scenario

**Table 4.** UQE-3D items and respective relationships considering Nielsen (1994), Nielsen (1996), Burdea and Coiffet (2003), and LaViola Jr et al. (2017) principles.

#	Item	Investigated Aspects and Components
1	I felt comfortable with the equipment during the game.	Comfort
2	I felt like a member of the game while playing.	Immersion and presence
3	The game caused uneasiness such as nausea, headache, etc.	Welfare
4	I was spatially aware of my location during the game.	Immersion, presence, and visual perception
5	I could easily have a sense of direction during the game.	Visual perception and ease of use the application
6	I executed easily the game tasks.	Ease of use the application and ease to perform the task
7	I got tired of making movements during the game.	Welfare and task execution time
8	It was good to wear a visualization helmet during the game.	Visual perception, immersion, and visual elements
9	The images helped me understand the game.	Relationship between scene and task, visual elements and visual perception
10	The images helped me understand how to play.	Clarity about the task to be performed, relationship between scene and task, visual perception, and visual elements
11	The sounds produce during the game helped me understand how to play.	Clarity about the task to be performed and sound elements
12	I think that playing the game was fun.	Immersion and welfare
13	I enjoyed playing the game with gestures.	Immersion and welfare
14	I think that the game was appropriate for my age.	Welfare and presence (indirectly, because interfere in affect)
15	I think that I had enough time to play.	Task execution time
16	I think that I had enough time to rest after the game.	Rest time interval

**Table 5.** General purpose of each item of the UQE-3D questionnaire.

#	UQE-3D item purposes
1	Equipment Comfort
2	Feel as though in a virtual environment (immersion & presence)
3	Cybersickness
4	Spatial orientation
5	Difficulty to establish the spatial orientation
6	Facility to perform the tasks
7	Analysis of the required efforts during the interaction process
8	Impact of the VR devices
9	Interface intuitiveness about the whole interaction process
10	Interface intuitiveness about the task (visual feedback)
11	Interface intuitiveness about the task (aural feedback)
12	Fun and boredom about the task
13	Fun and boredom about the interaction process
14	Relation between the application and the participant age
15	Task time measurement
16	Rest time measurement

and task because the scenario can help to understand the task (Nielsen, 1996), and the task execution time and rest time interval to identify if the task duration was satisfactory and/or enough for the participants. The questionnaire uses 5-point Likert Scale.

UQE-3D statements can be adaptable to the experiment context. For example, item 7 considers evaluating user movements. As the validation experiment developed used a serious game, we have demonstrated the questionnaire in the "game" context. We evaluated a game that uses arms' movements, but we could adapt this item considering an interaction process using legs', hands' or any body movement. We can also extend adjusts considering the type of 3DUI application, like a VR simulator or a VR training system. Moreover, our questionnaire observed the use of appropriated language to better attends the misunderstanding technological demands of the target group (Brace, 2018; Krosnick, 2018). Therefore, the UQE-3D combine usability and VR/AR characteristics mentioned, and it can be suit for each context of the application.

Figure 2 represents our suggestion about relationships between questionnaire items and 3DUI aspects/ usability metrics based on the literature previously cited. The dashed arrows represent factors that are not directly related, but we

comprehend that they may potentially interfere with the assessment of the item (for instance, error-forgiving situations may or may not result in breaks in immersion). We used as reference the questionnaire developed by Simor (2016). One of the substantial differences between SUS and UQE-3D questionnaires is the specificity of our proposal to evaluate VR/AR applications, considering 3DUI metrics like immersion and presence (Figure 3). For this reason, we will discuss the need to keep (or not) the SUS on our application protocol.

After completing the questionnaires, the participant can comment about the test openly, allowing additional information to collect by the observer.

We also adopted the use of a semi-structured interview to validate the questionnaire responses. We formulated similar items to get responses suggestive to the same aspects and components investigated with the UQE-3D questionnaire.

Table 6 shows the questions of the interview and their relationships with the UQE-3D. Some interview questions have no direct textual relationship with the UQE-3D items; however, based on participants' responses, we can understand their experience and associate with the questionnaire statements. For example, if the item 4 response of the interview was affirmative, then can mean the user felt immersed in the game because he had a sense of presence and a sense of direction in the virtual environment (item 4 and 5 of the UQE-3D). If there is incoherence about the questionnaire responses, it is possible to perform a subjective analysis of the user behavior during his interaction with the application, considering the MMSE and GDS-15 scores and sociodemographic responses, to identify which possible factors influenced the discrepancy.

The interviews are very significant to the evaluation experiment because of the specific context to which it is applied. Generally, it is easier to get information about elderly experiences using interviews, especially being applications supporting VR devices and used by older people. The interview helps the observer to identify semantic misunderstandings about the technical terms and to explain them, because, sometimes, they are not aspects of the participants' daily life. It is a way to ensure the best understanding of the user sen-

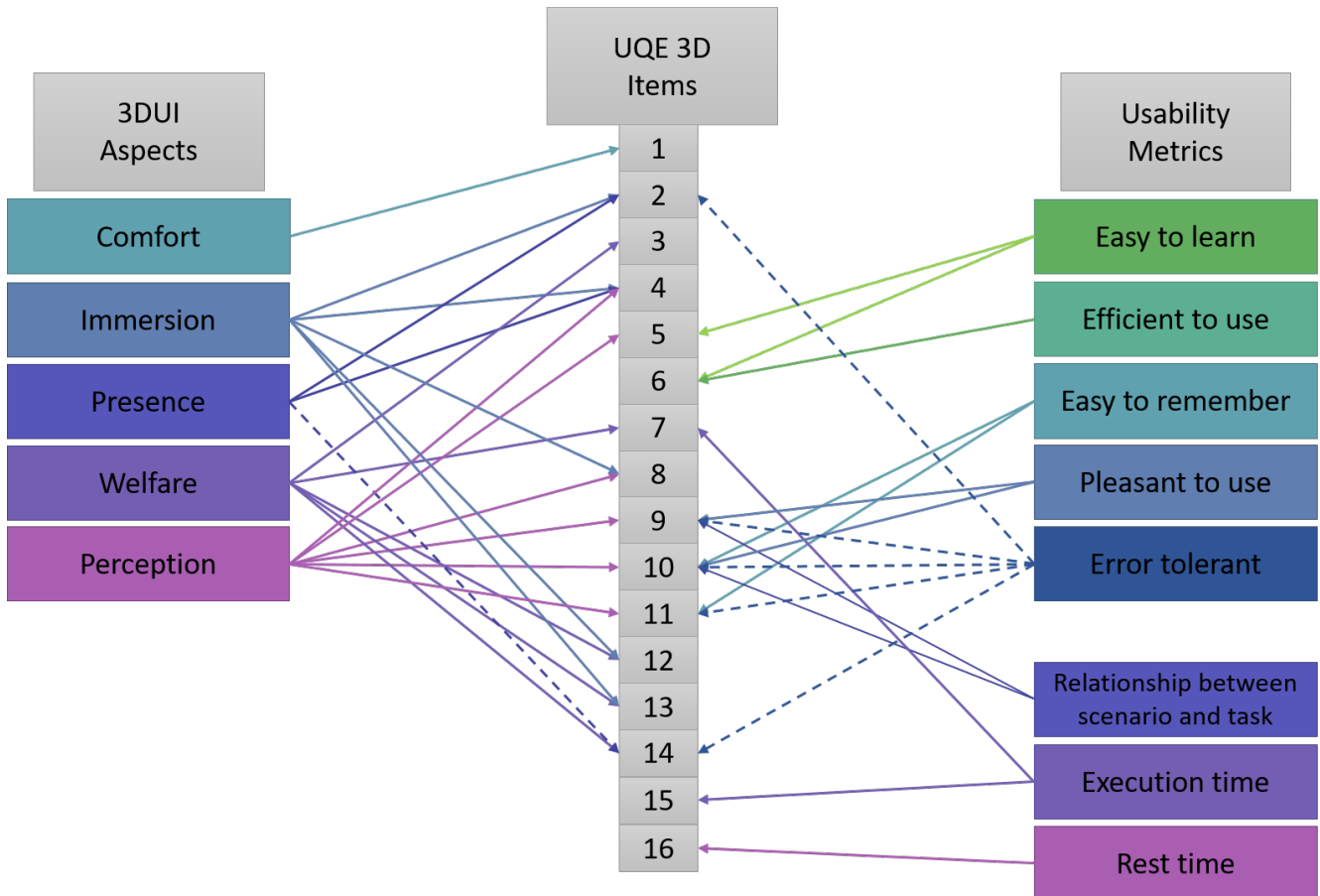


Figure 2. Relationships between the UQE-3D items and 3DUI aspects/ usability metrics.

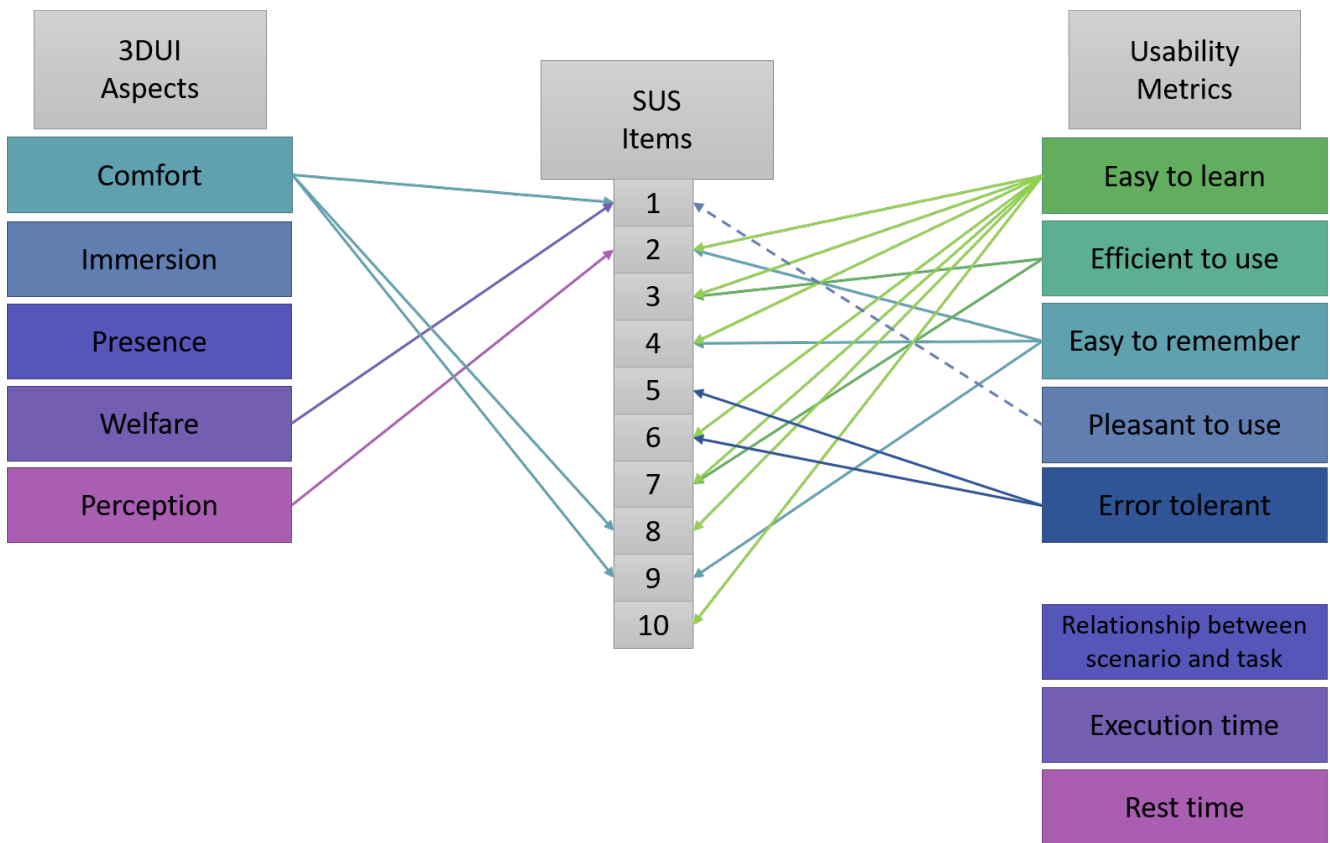


Figure 3. Relationships between the SUS items and 3DUI aspects/ usability metrics.

**Table 6.** Semi-structured interview questions and their relationships with the UQE-3D.

#	Interview question	Connections with UQE-3D items
1	Have you liked using your arms to interact with the game?	7 and 13
2	Did you like to use the visualization headset? Why?	2 and 8
3	If you had a game like this, would you play it?	12
4	Would you be able to play the game by yourself, with no one to teach you?	4, 5, 9, 10, and 11
5	Were you able to understand what you were supposed to do in the game? What made it difficult?	2, 9, 10, and 11
6	Was there any time when you had questions about what you were doing while playing?	9, 10 and 11
7	What did you think that was tiring in the game?	7
8	Is the game adequate to your age range? What points can improve?	14
9	Does the game stimulate the practice of beneficial activities to yourself? Physically or mentally?	–
10	Does the game contribute to increasing your concern in computational tools to education, welfare, and health?	–
11	Does the game allow the increase of your connection with other people that use technology, like youngsters or other seniors?	–

sations during the interaction process.

The high point of our questionnaire (UQE-3D) is what it intends to capture from the experiment's subjects. In our experiment we have used the interview don't just as an way to capture other impressions from the subjects, but also to validate our items. Alternatively, the interview can be used as an application way to the proposed method.

## 3 Results

In order to validate our method, we apply it in an experiment involving elderly and VR exergame. We detail the experiment and results in next subsections.

### 3.1 Experiment

For a better didactic development, this subsection refers to the application protocol in an experiment to test and evaluates the UQE-3D. The results of this evaluation and the discussion of the results will be in the next subsection. This study was approved by the Ethics Committee of the University of Passo Fundo, Brazil, under number 79809317.2.0000.5342.

#### 3.1.1 Participants

Our evaluation considered convenience sampling, assuming the population available for recruitment at the study sites. We obtained 30 volunteer participants with 60 years or more ( $71.40 \pm 10.29$ ). We composed two groups with elderly of different contexts to verify the applicability of our method, organized as follow:

- Group A, Active Seniors: elderly that perform activities daily, participate in social programs for your age group, and no present depressive symptoms, considering scores MMSE > 18 and GDS-15 < 5;
- Group B, Non-Active Seniors: institutionalized elderly (nursing home residents), presenting mild cognitive impairment or low level of depressive symptoms, considering scores MMSE > 18 and GDS-15 < 10.

The inclusion criteria, common to both Groups, considers participants literate (primary education at least) and no motor impairment.

Group A was composed of 23 participants ( $67.47 \pm 6.45$ ), four men and 19 women; and Group B was composed of seven participants ( $84.28 \pm 10.32$ ), two men and five women.

#### 3.1.2 Task

To validate our approach, we executed an evaluation using one of the game levels of Motion Rehab AVE 3D, a program registered with the Brazilian Institute of Industrial Property under the number BR 51 2016 001373-7 (Trombetta et al., 2017). This software aims to assist health professionals in elderly motor and cognitive rehabilitation activities, with VR support (immersive visualization and spatial interaction). Figure 4 shows one of the scenes in the first-person view. We configured UQE-3D questionnaire considering the game application context, the interaction task, and devices involved.

The experiment task consists of using the arms to touch with virtual hands the beach volleyball balls (context objects) thrown towards the avatar. Randomly, pencils (distractor objects) may arise replacing the balls, demanding that the player does not execute the touch movement (he can let down his arms or dodge his body). Figure 5 illustrates the body movements of the seniors during the interaction process. The task time is 30 seconds. The participant is free to repeat the task within the time set by the protocol (one minute).

#### 3.1.3 Devices

We used the Oculus Rift DK 1 Model to support 3D immersive visualization. The motion-sensing input device used to map gestures and body movements was the Microsoft Kinect One motion sensor.

### 3.2 Validation

Here we conducted the UQE-3D validation. So we verified if the proposed questionnaire achieved its evaluation purpose, if it was proper for the target public, and if it had internal consistency reliability.

After the internal consistency evaluation, we calculated a score to UQE-3D based in SUS score formula, and we defined a confidence interval to apply to the UQE-3D a measure of quality based on the subject scores. We also executed a paired T-test and an F-Test to compare SUS and UQE-3D between groups.

Finally, we used the frequency distribution to relate questionnaire and interview responses. We performed this comparison to confirm the validity of the questionnaire responses





Figure 4. First-person view in Motion Rehab AVE 3D during an interaction task.



Figure 5. Seniors in the evaluation experiment using the game interface. They have given permission for us to use their photos through the consent form.

and to analyze the adequacy of the instrument for the target public.

### 3.2.1 Factor Analysis

To evaluate the internal consistency of our questionnaire, we used the Exploratory Factor Analysis technique. To estimate the reliability, we applied the Cronbach's Alpha. We defined two hypotheses:

- H0: The data is not suitable for Factor Analysis;
- The data is suitable for Factor Analysis.

Considering these techniques and a convenience sample of 30 subjects to analyze the UQE-3D, we obtained a value of 0.669 for KMO test and 0.000 for Bartlett's Test. Both the

tests verify the data adjustments' degree to Factor Analysis. In the same way, the Cronbach's Alpha evaluates the questionnaire's internal consistency verifying if the items have coherency between itself. The UQE-3D presented a score of 0.868 to Cronbach's Alpha.

The barely acceptable values to the KMO test should be greater than 0.5 to indicate suitability, while Bartlett's Test should be lower than 0.1 to deny the null hypothesis. According to Indrayan and Malhotra, the acceptable Cronbach's Alpha scores should be 0.8 or above to suggest a high internal consistency (Indrayan and Malhotra, 2017).

From the extraction method, five components (factors) presented eigenvalues greater than one. These factors explain  $\approx 78.332\%$  of the total variability. Considering the rotation method to analyze the factor loadings of each variable about



the five main factors extracted, we can categorize our questionnaire as follows:

- Factor 1: 3DUI usability (Items 1, 2, 4, 5, 6, 10, 11, 15, 16), variance 41.417%;
- Factor 2: Fun and pleasure (Items 12, 13, 14), variance 12.807%;
- Factor 3: Confidence and uneasiness (Items 3, 9), variance 9.577%;
- Factor 4: Device functionality (Item 8), variance 8.277%;
- Factor 5: Tiredness (Item 7), variance 6.254%.

Table 7 presents the UQE-3D factor matrix highlighting the factor loadings by component. Factor 1 considers most of the usability assessment items of 3DUI in our questionnaire, presenting high coefficients (greater than 0.7) for seven of the nine variables. Factor 2 has a good connection with most items related to immersion and entertainment. Factor 3 is related to concerns about understanding and feeling good during the interaction process. Factors 4 and 5 are independent, reporting situations listed to equipment importance and user fatigue, respectively. Based on these five factors, it is possible to infer what an elderly expects from a good VR/AR application for their age, considering the proposed 3DUI usability questionnaire.

**Table 7.** Factor loadings of the UQE-3D items. Highest factor loading is in boldface.

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
1	<b>0.543</b>	-0.418	-0.164	0.203	-0.090
2	<b>0.845</b>	-0.049	-0.159	-0.102	-0.060
3	-0.384	0.358	<b>0.483</b>	0.482	-0.125
4	<b>0.810</b>	-0.224	-0.149	0.039	0.136
5	<b>0.821</b>	-0.118	0.041	0.245	0.065
6	<b>0.844</b>	-0.110	-0.020	-0.004	-0.075
7	-0.037	0.164	-0.250	0.415	<b>0.834</b>
8	0.330	-0.163	0.226	<b>0.799</b>	-0.185
9	0.542	0.001	<b>0.708</b>	-0.184	0.314
10	<b>0.877</b>	-0.048	0.003	-0.002	-0.010
11	<b>0.451</b>	0.355	-0.527	-0.023	0.029
12	0.313	<b>0.796</b>	-0.031	-0.117	0.095
13	0.615	<b>0.651</b>	-0.129	0.137	-0.187
14	0.547	<b>0.611</b>	0.232	-0.045	-0.149
15	<b>0.715</b>	-0.146	0.492	-0.315	0.191
16	<b>0.874</b>	-0.184	-0.125	-0.004	-0.116

According to the results, we can note good internal consistency for the UQE-3D, with fit data for analysis. Moreover, the commonality analysis approach returned all values greater than 0.5, proving satisfactory so that no exclude any of the questionnaire items.

### 3.2.2 Confidence Interval

We also computed a confidence interval (CI) for the UQE-3D questionnaire based on the experiment results. In doing so, we used the same method define by Brooke to define the SUS score (Brooke, 1996), considered an intuitive way to calculate (Lewis, 2018).

Firstly, we simulated a system evaluation using the questionnaire to determine a multiplicative factor value considering the best results for each item, in a 5-point Likert Scale.

This way, for each item, we subtracted one from the subject responses to positive statements ( $x-1$ , when  $x$  is 5 in the best-expected result); and we subtracted from 5 to negative statements ( $5-x$ , when  $x$  is 1 in the best-expected result). This scales all values from 0 to 4. After, we added all these values to get the maximum possible score. The multiplicative factor (SUS questionnaire is 2.5) is the result of 100 divided by maximum score (100/maximum score). Figure 6 shows the calculation scheme.

As a result, we multiply the subject score (sum of values from each item) by the UQE-3D multiplicative factor (1.5625) to obtain the final score (between 1 and 100).

We also aimed to apply a CI to the questionnaire as a quality measure, considering our sample. We used  $\alpha = 0.05$  and the data sample, Group A - Active Seniors, Group B - Institutionalized Seniors, and Group General (A + B). Table 11 shows the results from statistical analysis and respective CIs by group.

As expected, Table 8 shows a significant difference between the mean scores from Group A (Active Seniors) and Group B (Institutionalized Seniors). For this reason, a generic CI, considering the two groups (Group General) can be inappropriate to individuals with characteristics similar to the subjects from Group B of this experiment. Therefore, we recommend the use of the CI specifically for each user group (Active Seniors and Non-Active Seniors), when the sample characteristics suit the same of this experiment. If the sample characteristic is different, it is possible to develop a new CI using the same parameters discussed in this paper.

**Table 8.** Statistical analysis applied to define the Confidence Interval (CI).

Stats	Group A	Group B	Group General
$\alpha$	0.05	0.05	0.05
SD	9.810082845	9.059677311	14.45357399
Sample	23	7	30
Mean	88.52	63.17	82.60
CI	[84.28; 92.76]%	[54.79; 71.55]%	[77.81; 88.00]%

### 3.2.3 T-Test of Student and F-Test of Fisher

We applied the Student's paired T-test to compare if there is a significant difference between SUS and UQE-3D for groups A, B, and General, and to validate the UQE-3D statistically. We used the mean scores from each questionnaire, once that the scores are calculated similarly.

Table 9 shows the comparisons between groups, presenting a no significant difference between SUS and UQE-3D, regardless of group. This result exposes that the UQE-3D can be equivalent to SUS evaluating usability issues. However, UQE-3D considers the assessment of 3DUI concepts (e.g. presence and immersion) and the specific context application (elderly).

We also applied Fisher's F-Test (Hahs-Vaughn and Lomax, 2013) only the general group, because of the difference in mean and standard deviation (SD) values presented between SUS and UQE-3D. This test checks whether the variances of the two samples are equal, assuming that they do not deviate from normality. According to this test, and considering the sample size, the SD between SUS and UQE-3D

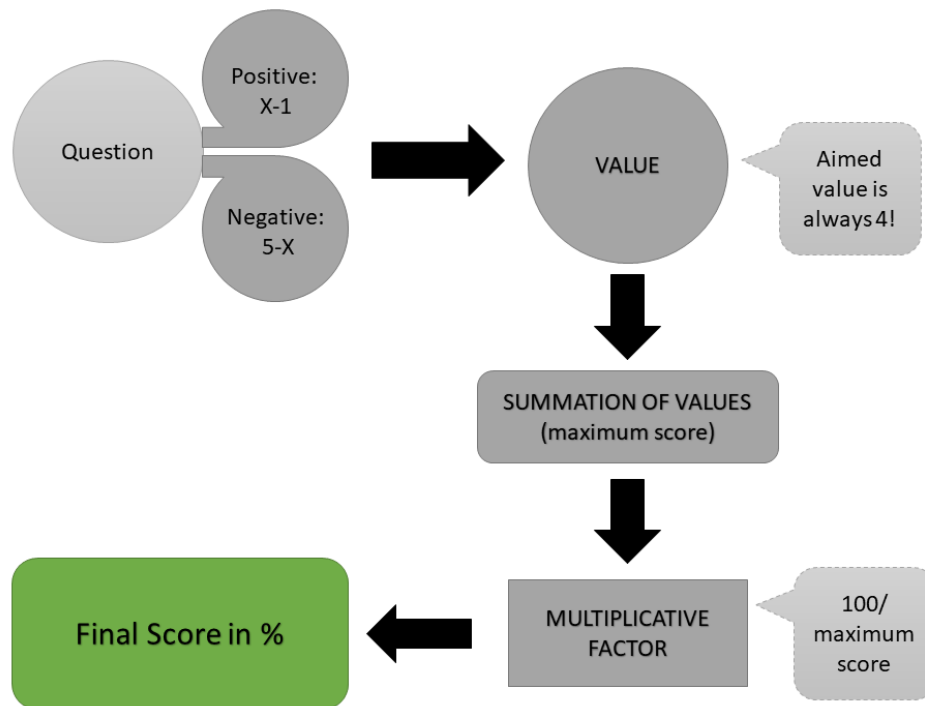


Figure 6. Explanation scheme to SUS calculation replicated on UQE-3D.

can be considered equals, showing no statistically significant difference.

Table 9. Student’s t-test results presented by group.

Group A			
Questionnaire	Sample	Mean	SD
SUS	23	88.0435	8.08184
UQE-3D	23	88.5200	9.80923
p = 0.8580 > 0.05		No significant difference	
Group B			
Questionnaire	Sample	Mean	SD
SUS	7	71.4286	10.78745
UQE-3D	7	63.1700	9.05828
p = 0.1470 > 0.05		No significant difference	
Group General			
Questionnaire	Sample	Mean	SD
SUS	30	84.1667	11.16748
UQE-3D	30	82.6050	14.45316
p = 0.6410 > 0.05		No significant difference	

The high p-value of the active elderly group (Group A) may be justified by the ease to understand the interaction process during the experiment, and by no present difficulties to interpret the questionnaires and relate it into the virtual environment experience.

## 4 Discussion

Statistical analysis showed a difference between Group A (Active Seniors) and Group B (Institutionalized Seniors). With this in mind, we decided to perform the distribution and frequency analysis of the UQE-3D responses separately, by group, allowing us to establish a relation between them, besides the interview responses relation, summarized in Table 10.

Table 11 and Table 12 show the distribution and frequency of the UQE-3D answers considering Group A and Group B, respectively. We noted that all members of Group A felt comfortable using the equipment and did not get tired throughout the interaction process. On the other hand, 42.9% of Group B participants felt comfortable, and 42.9% somewhat. Besides, the majority of Group B (85.7%, six individuals) did not feel tired during the interaction. These results elucidate institutionalized seniors have a different lifestyle and no frequent access to new technologies.

Regarding immersion, 82.6% of Group A participants felt like a game member while playing (providing a sense of presence), and 87% could easily have a sense of direction during the game. In Group B, just one person felt connected to the game while playing, and 57.1% presented spatial awareness during the game (according to the UQE-3D answers). This difference may be associated with an abstraction difficulty found in Group B individuals, which presented lower MMSE and GDS-15 scores. This situation reinforces the relevance of evaluating elderly groups with different lifestyles separately.

During the interview, 71% of Group B participants mentioned they felt more connected in the virtual environment wearing an HMD. For item 8 of the UQE-3D, all subjects responded as having good experience in wearing a visualization helmet. So, these subjects cannot comprehend the aim of items 2, 4, and 5 of the UQE-3D (which may also be justified by the lower MMSE and GDS-15 scores), generating a difference between the interview and questionnaire.

Still on immersion, in the interview, 91.2% of Group A participants manifested that they enjoyed interacting using gestures and display devices, feeling a sense of presence during the interaction process with scene and game elements. UQE-3D confirmed these responses to Group A, considering a mean of 87% best-expected responses to items 8 and

**Table 10.** Summary of the interview best-expected responses collected.

Item	1	2	3	4	5	6	7	8	9	10	11
Group A	91.2%	91.2%	86.9%	56.6%	100%	8.6%	0%	82.6%	91.2%	82.6%	82.6%
Group B	85.0%	71.0%	57.0%	0%	100%	42.0%	0%	42.0%	100%	0%	57.0%

**Table 11.** Responses’ distribution and frequency marked by Group A using UQE-3D.

Item	Responses (Likert Scale)					Without response
	1	2	3	4	5	
1				43.5%	52.2%	4.3%
2			17.4%	26.1%	56.5%	
3	95.7%		43.3%			
4			13%	26.1%	60.9%	
5			13%	17.4%	69.6%	
6			8.7%	4.3%	87%	
7	91.3%	4.3%	4.3%			
8			13%	34.8%	52.2%	
9			8.7%	17.4%	69.6%	4.3%
10			8.7%	17.4%	73.9%	
11	4.3%	8.7%	4.3%	26.1%	56.5%	
12			13%	47.8%	39.1%	
13			4.3%	34.8%	60.9%	
14		4.3%	8.7%	34.8%	52.2%	
15		4.3%	4.3%	26.1%	60.9%	4.3%
16			4.3%	17.4%	78.3%	

**Table 12.** Responses’ distribution and frequency marked by Group B using UQE-3D.

Item	Responses (Likert Scale)					Without response
	1	2	3	4	5	
1		14.3%	42.9%	42.9%		
2	28.6%	14.3%	42.9%	14.3%		
3	57.1%	28.6%	14.3%			
4	14.3%		28.6%	57.1%		
5		14.3%	28.6%	57.1%		
6		14.3%	28.6%	57.1%		
7	71.4%	14.3%	14.3%			
8				85.7%	14.3%	
9			42.9%	57.1%		
10		14.3%	42.9%	42.9%		
11	14.3%	14.3%	14.3%	57.1%		
12				85.7%	14.3%	
13				100%		
14	14.3%	42.9%	28.6%	14.3%		
15		42.9%	28.6%	28.6%		
16		28.6%	28.6%	42.9%		

12.

In Group A, 86.9% of the subjects declared during the interview that they would like to play the game frequently (56.5% without requiring help to play); in Group B, this number was 57%. Regarding the game intuitiveness, all participants of both groups mentioned that ease understood the interaction task. In the same way, considering the UQE-3D means, 87% of Group A and 85.7% of Group B responded positively about the fun user experience and the ease to use of the application (items 5, 6, 9, 10, 11, and 12).

In Group A, 82.6%, and Group B, 14.3% of the sample considered the game adequate for their age range and thought that the game could contribute to increasing their technology interests. Participants also highlighted this experience as a factor in expanding connections with people that use technology frequently, not only seniors but also younger. Five participants declared that they could talk with the younger about this, increasing their interaction and connection; three

participants manifested to indicate the experience for other seniors; two participants commented that the device experience made them feel more modern and digital as a person, and because of this, nearest to the today’s youth.

Moreover, 91.2% of Group A and 100% of Group B members believed that the game is beneficial to health and promotes the practice of physical (because it requires a routine and precision of body movements) and cognitive activities (because it stimulates concentration and quick thinking). Only two participants of Group A did not feel challenged by the game, probably because the game considers rehabilitation tasks. However, we informed all the participants about this feature before the test.

None of the participants reported tiredness or motion sickness during or after the experiment. Only three subjects of Group B (42%) declared in the interview felt uncertainty about the correct procedure to perform the task while playing. Items 9, 10, and 11 of the UQE-3D explored how visual (images) and aural (sounds) feedback help the user to understand the interaction process, resulting in a mean of 87% best-expected responses to Group A, and 52.3% to Group B. For this reason, we noted that the visual and aural feedback is not enough to prevent feelings like uncertainty or insecurity during the interaction process with Group B elderly. In the interview, some seniors of this group manifested they heard the sounds produced by the game, justifying a negative score of item 11. Considering items 9 and 10, we can assume that the results of Group B have some relation with the participants’ characteristics (low scores of MMSE and GDS-15) because these subjects have reduced contact with new technologies.

The sociodemographic questionnaire, applied in Pre-Test, pointed out that none of the participants knew about VR technologies used, and only two elderly said that they played computer games frequently. Nonetheless, the experiment showed that the elderly play easily and intuitively, considering the 3DUI usability aspects.

The proposed method was able to evaluate the VR game interface satisfactorily in our case study, identifying good usability for the elderly. It allowed the task execution easily and intuitively, providing immersion in the virtual experience and promoting fun and welfare to the participants.

Besides, the experiment identified a difference in perception of the virtual environment considering institutionalized seniors. For this reason, we suggest defining a specific CI to evaluate a 3DUI when the sample has similar characteristics to this experiment. Evaluations using a heterogeneous group are an opportunity to increment our method.

Also, the experiment pointed out that interviews are a valuable ally to validate the application of a questionnaire with the elderly because some people share more impressions by talking than using a questionnaire. As a result, the interview could also be a UQE-3D’s application way.

We could affirm that this study validated our method be-

cause it assisted in the 3DUI assessment with the elderly. We also highlight the UQE-3D questionnaire can be applied regardless of our proposed methodology, becoming a new option for evaluation processes. Considering the application of all steps of the application protocol, we also emphasize that it is relevant to a better understanding of the results, the relationship between components during an evaluation process with the elderly, crossing values of different context-specific instruments (sociodemographic, MMSE, GDS-15, UQE-3D, and interview). As checked by the experiment, without the information obtained by the other steps of the application protocol, we would have less complete knowledge of the user experience and usability issues experienced by the seniors of Group B.

However, new experiments are necessary to confirm what this experiment showed (considering other elderly groups) and to execute a strengthened validation with 3DUI experts to identify some biases to be corrected.

The SUS questionnaire does not demonstrate results beyond the identified by UQE-3D since 3DUI metrics, like immersion or presence, is not observed. The mean statistic similarity of both suggests that the UQE-3D is enough to identify the usability issues aimed at a 3DUI evaluation with the elderly.

Considering the statistical analysis and identifying nuances of evaluation with the elderly, our questionnaire produced satisfying results to 3DUI usability evaluation with the elderly. Besides, the UQE-3D had presented a statistically similar score with the SUS questionnaire, which might mean that the UQE-3D is equivalent to SUS when analyzing the usability of an interface, however, designed for interventions with seniors in 3DUI evaluations.

We applied the UQE-3D questionnaire in an experiment involving only the elderly. However, we noted that other audiences could use this questionnaire in the 3DUI evaluation process (even designed to facilitate the elderly understanding). Therefore, UQE-3D could help groups unfamiliar with new technologies. It is noteworthy that this questionnaire focused primarily on the elderly and, changing the specificity of the public, it is necessary to pay attention to the adaptation of the language used in the questionnaire (Brace, 2018; Krosnick, 2018; Huygelier et al., 2019; Lee et al., 2019).

One limitation of the application protocol is the total execution time. As mentioned, very long times can generate a disinterest in older participants. The solution found was the division of the application protocol performance in two days. However, future efforts can identify other ways to reduce the time, and new experiments can investigate step exclusion or its execution in the domestic environment. For instance, the participants could complete questionnaires and bring them to the observer on the experiment day.

Regarding the method validation, one limitation was the sample size. The sample was enough for the CI definition to verify that the method can identify what it is supposed to do and to certify that it can reach different categories of seniors and capture their perceptions about the interface. It would be interesting to direct efforts to engage more participants, which could confirm the validation of the method and identify new points for improvement.

Finally, the main advantage of using the UQE-3D in an

evaluation process with the elderly is that it was tailor-made for this audience, not yet accustomed to interacting in 3DUI. The language used for the questionnaire items considered terms and sentences to facilitate comprehension of the evaluation aspects. We noted the participants had missed the meaning of the questions from the SUS questionnaire. But during the use of the UQE-3D, we observed they seemed more comfortable about the questions' understanding. For example, it was difficult for them to understand questions like "I found the system unnecessarily complex" and "I found the system very cumbersome to use". On the other hand, they provide more reliable answers about interface problems for related items through the UQE-3D questionnaire, such as "I easily executed the game tasks" and "I felt comfortable with the equipment during the game."

Another advantage is that this method can assess usability aspects presented only in 3DUI, such as immersion and presence. It can help VR/AR researchers to improve the quality of the immersive applications and user experience, driving the development of technologies more adapted to this audience.

## 5 Conclusions

Our work presented a proposal of the usability evaluation method of 3DUI for the elderly, a context-specific questionnaire, the UQE-3D (3DUI Usability Evaluation Questionnaire for Elderly).

Our application protocol considers three stages: Pre-test, Test, and Post-test, applying in two sessions, on different days. The first day, Pre-test stage involves the use of the MMSE, GDS-15, and sociodemographic questionnaires, and a training section using devices and 3DUI application in evaluation. The second day, subjects experiment (Test stage) and evaluate (Post-test stage) the 3DUI application, considering the same device configuration and the use of UQE-3D questionnaire and a semi-structured interview.

To define our method, we performed a literature systematic review to identify usability evaluation instruments usually used in interventions with the elderly. This review pointed out 28 related work and demonstrated a usage pattern of questionnaires. We considered these results and conducted, previously, two pilot studies to adjust our purpose and concept the UQE-3D. Before validation presented by this work, we also had a UQE-3D revision by two healthcare researchers and two computer science researchers.

We applied the method considering a sample of 30 seniors of both the genres, and we analyzed the results to validate it. We performed statistic tests to verify the UQE-3D internal consistency, and to elaborate UQE-3D score calculation and CI. We used these resources to analyze the questionnaire results and to check the similarity with the SUS questionnaire. Besides, we compared the UQE-3D and the interview responses, considering two groups (active and institutionalized seniors), with the purpose of results validation of the questionnaire.

The main contribution of our work is the UQE-3D questionnaire, validated in a context-specific game, develop to evaluate 3DUI applications focused on the elderly. The ques-

tionnaire is suitable to apply in the 3DUI development, user-centered design, to identify if the interface is attending to the listed requirements, as also at the end of the development process to 3DUI validation. Among other relevant contributions, we could identify that interviews are important to determine variations in the comprehension of the interaction process not captured by questionnaires considering elderly groups. They were also useful for clearly identifying user perceptions of the assessment with questionnaires, or even to questionnaire's application way. Therefore, the UQE-3D has shown itself effective and efficient to identify the 3DUI usability to this public.

Our method also presents a contribution to the VR/AR area, offering a new tool to evaluate and to secure the quality of 3DUI for the elderly. Immersive solutions are still not widely used by this specific public. In this context, the UQE-3D could help to improve interfaces and stimulate the development of VR/AR technologies to the elderly. Furthermore, the UQE-3D proved appropriate to apply in 3DUI experiments, once it identifies the user's perceptions during the interaction process.

Moreover, evaluations with other public can use the UQE-3D. Even designed to facilitate the understanding of the elderly, we noted that your conception is adequate to groups unfamiliar with new technologies too.

As future work, we propose a new validation considering a heterogeneous elderly sample (and superior of 30 subjects), to confirm the results of this study and to spread the method's use. We also recommend the validation considering other VR/AR applications. Other suggestion is to apply our questionnaire considering different age groups with participants not familiarized with the VR/AR technology, investigating its adaptation to other contexts.

In addition, it is worth exploring the use of the UQE-3D questionnaire, independent of the proposed application protocol, to evaluate the extent of its effectiveness in the 3DUI evaluation. We suggest investigating ways to apply our questionnaire combined with automated approaches, as the user's eye or body tracking, to study the relationship between presence, immersion, and interactivity. This kind of approach can also guarantee that there were no interference during the interaction process, providing a correct usability evaluation.

## Disclosure statement

The authors report no conflicts of interest.

## Funding

This work was supported in part by Brazilian National Council for Scientific and Technological Development (CNPq), Grant Number 420226/2018-4.

## Ethics approval

This study was approved by the Ethics Committee of the University of Passo Fundo, Brazil, under number 79809317.2.0000.5342.

## Consent to participate

Informed Consent Form was signed by the study participants and approved by the Ethics Committee of the University of Passo Fundo, Brazil, under number 79809317.2.0000.5342.

## ORCID iD author statement

Manoela Rogofski Brum Postal: Conceptualization, Methodology, Resources, Validation, Visualization, Data Curation, Investigation, Writing - Original Draft.

Luciane Daroit: Formal analysis, Data Curation, Investigation, Writing - Original Draft.

Rafael Rieder: Supervision, Project administration, Methodology, Software, Resources, Visualization, Formal analysis, Writing - Review & Editing.

## References

- Almeida, O. P. and Almeida, S. A. (1999). Reliability of the brazilian version of the geriatric depression scale (GDS) short form. *Arquivos de Neuro-psiquiatria*, 57(2B):421–426.
- Aminov, A., Rogers, J. M., Middleton, S., Caeyenberghs, K., and Wilson, P. H. (2018). What do randomized controlled trials say about virtual rehabilitation in stroke? a systematic literature review and meta-analysis of upper-limb and cognitive outcomes. *Journal of neuroengineering and rehabilitation*, 15(1):1–24.
- Báez, M., Ibarra, F., Far, I. K., Ferron, M., and Casati, F. (2016). Online group-exercises for older adults of different physical abilities. In *2016 International Conference on Collaboration Technologies and Systems (CTS)*, pages 524–533. IEEE.
- Brace, I. (2018). *Questionnaire design: How to plan, structure and write survey material for effective market research*. Kogan Page Publishers.
- Brooke, J. (1996). SUS—a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.
- Brucki, S., Nitri, R., Caramelli, P., Bertolucci, P., and Okamoto, I. H. (2003). Suggestions for utilization of the mini-mental state examination in Brazil. *Arquivos de neuro-psiquiatria*, 61(3B):777–781.
- Burdea, G. C. and Coiffet, P. (2003). *Virtual reality technology*. John Wiley & Sons.
- Cockton, G. (2013). *The Encyclopedia of Human-Computer Interaction*, chapter Usability Evaluation, pages 817–940. Interaction Design Foundation, 2 edition.
- Cota, T. T., Ishitani, L., and Vieira Jr, N. (2015). Mobile game design for the elderly: A study with focus on the motivation to play. *Computers in Human Behavior*, 51:96–105.
- Fua, K. C., Gupta, S., Pautler, D., and Farber, I. (2013). Designing serious games for elders. In *FDG*, pages 291–297.
- Hahs-Vaughn, D. L. and Lomax, R. G. (2013). *Statistical Concepts-A Second Course: A Second Course*. Routledge.



- Hall, A. K. and Marston, H. (2015). Games for health in the home: Gaming and older adults in the digital age of health-care. *Handbook of Smart Homes, Health Care, and Well-being*. Charn, Switzerland: Springer International Publishing.
- Huang, H.-M., Rauch, U., and Liaw, S.-S. (2010). Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. *Computers & Education*, 55(3):1171–1182.
- Huygelier, H., Schraepen, B., van Ee, R., Abeele, V. V., and Gillebert, C. R. (2019). Acceptance of immersive head-mounted virtual reality in older adults. *Scientific reports*, 9(1):1–12.
- Indrayan, A. and Malhotra, R. K. (2017). *Medical biostatistics*. CRC Press.
- Khaled, A. B. H., Khalfallah, A., and Bouhlel, M. S. (2018). Fall prevention exergame using occupational therapy based on kinect. In *International conference on the Sciences of Electronics, Technologies of Information and Telecommunications*, pages 479–493. Springer.
- Kharoub, H., Lataifeh, M., and Ahmed, N. (2019). 3D user interface design and usability for immersive VR. *Applied Sciences*, 9(22):4861.
- Konstantinidis, E. I., Billis, A. S., Mouzakidis, C. A., Zilidou, V. I., Antoniou, P. E., and Bamidis, P. D. (2014). Design, implementation, and wide pilot deployment of Fit-ForAll: an easy to use exergaming platform improving physical fitness and life quality of senior citizens. *IEEE journal of biomedical and health informatics*, 20(1):189–200.
- Krosnick, J. A. (2018). Questionnaire design. In *The Palgrave handbook of survey research*, pages 439–455. Springer.
- LaViola Jr, J. J., Kruijff, E., McMahan, R. P., Bowman, D., and Poupyrev, I. P. (2017). *3D user interfaces: theory and practice*. Addison-Wesley Professional.
- Lee, C. C., Czaja, S. J., Moxley, J. H., Sharit, J., Boot, W. R., Charness, N., and Rogers, W. A. (2019). Attitudes toward computers across adulthood from 1994 to 2013. *The Gerontologist*, 59(1):22–33.
- Lewis, J. R. (2018). The System Usability Scale: Past, present, and future. *International Journal of Human-Computer Interaction*, 34(7):577–590.
- Lourenço, R. A. and Veras, R. P. (2006). Mini-Mental State Examination: psychometric characteristics in elderly outpatients. *Revista de saúde pública*, 40:712–719.
- Miller, K. K., Porter, R. E., DeBaun-Sprague, E., Van Puymbroeck, M., and Schmid, A. A. (2017). Exercise after stroke: patient adherence and beliefs after discharge from rehabilitation. *Topics in stroke rehabilitation*, 24(2):142–148.
- Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 152–158.
- Nielsen, J. (1996). Usability metrics: Tracking interface improvements. *IEEE software*, 13(6):1–2.
- Osagie, E., Waqar, M., Adebayo, S., Stasiewicz, A., Porwol, L., and Ojo, A. (2017). Usability evaluation of an open data platform. In *Proceedings of the 18th Annual International Conference on Digital Government Research*, pages 495–504.
- Osmanovic, S. and Pecchioni, L. (2016). Beyond entertainment: Motivations and outcomes of video game playing by older adults and their younger family members. *Games and Culture*, 11(1-2):130–149.
- Postal, M. R. B. and Rieder, R. (2019a). A usability evaluation experiment of 3D user interfaces for elderly. *Revista Brasileira de Computação Aplicada*, 11(3):28–38.
- Postal, M. R. B. and Rieder, R. (2019b). User interface evaluation methods for elderly: A systematic review. In *2019 21st Symposium on Virtual and Augmented Reality (SVR)*, pages 84–91. IEEE.
- Sheikh, J. I. and Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. *Clinical Gerontologist: The Journal of Aging and Mental Health*.
- Simor, F. W. (2016). Usability evaluation method for serious games with gestural interaction for seniors. Master's thesis, University of Passo Fundo, Passo Fundo, RS, Brazil.
- Solano, A., Collazos, C. A., Rusu, C., and Fardoun, H. M. (2016). Combinations of methods for collaborative evaluation of the usability of interactive software systems. *Advances in Human-Computer Interaction*, 2016.
- Tieri, G., Morone, G., Paolucci, S., and Iosa, M. (2018). Virtual reality in cognitive and motor rehabilitation: facts, fiction and fallacies. *Expert review of medical devices*, 15(2):107–117.
- Trombetta, M., Henrique, P. P. B., Brum, M. R., Colussi, E. L., De Marchi, A. C. B., and Rieder, R. (2017). Motion Rehab AVE 3D: A VR-based exergame for post-stroke rehabilitation. *Computer methods and programs in biomedicine*, 151:15–20.