

WestminsterResearch

<http://www.westminster.ac.uk/westminsterresearch>

**Using Cognitive Walkthrough and Hybrid Prototyping to Gather
User Requirements in Early Design Virtual Reality Prototypes
Doumanis, I. and Economou, D.**

This is an author's accepted manuscript of a paper published in Beck, D., Peña-Rios, A., Ogle, T., Economou, D., Mentzelopoulos, M., Morgado, L., Eckhardt, C., Pirker, J., Koitz-Hristov, R., Richter, J., Gütl, C. and Gardner, M. (eds.) 5th International Conference of Immersive Learning Research Network, London 23 - 27 Jun 2019.

The final authenticated publication is available at Springer via:

<https://dx.doi.org/10.1007/978-3-030-23089-0>

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of WestminsterResearch: (<http://westminsterresearch.wmin.ac.uk/>).

In case of abuse or copyright appearing without permission e-mail repository@westminster.ac.uk

Using Cognitive Walkthrough and hybrid prototyping to gather user requirements in early design Virtual Reality prototypes

Ioannis Doumanis¹, Daphne Economou²

¹ University of Central Lancashire, Preston, UK

² University of Westminster, London, UK

¹idoumanis@uclan.ac.uk; ²d.economou@westminster.ac.uk

Abstract. To evaluate Virtual Reality (VR) prototypes usability involves a variety of single-perspective or Hybrid methods. The latter has been suggested by literature as offering a more complete set of requirements highlighting both ‘in-world’ and user interface problems. This paper describes our experiences in using a single-perspective method for gathering user requirements in the REVERIE (Real and Virtual Engagement In Realistic Immersive Environment) project. The study reports results involving nine evaluators who reviewed two hybrid VR prototypes with educational context. It was found that this approach was effective in highlighting a plethora of usability problems covering all aspects of the two VR prototypes. The performance of our approach was similar to the literature. Although additional validation work is required, we can conclude that our approach may provide a viable option to evaluate early design VR prototypes when required (e.g., when the expertise needed to use a hybrid method is not available). Future work aims to compare the performance of our approach with two-stage and multiple stage hybrid methods.

Keywords: Usability methods; Cognitive Walkthrough; Virtual Worlds; User Interface; Interaction.

1 Introduction

With the advent of virtual reality platforms (e.g., Oculus VR1 and Steam VR2), it was not long before the medium was overrun with a plethora of applications. In the creation of such applications, designers typically collect and define user requirements by investigating the usability of VR prototypes [1]. This user-centered development process [2] uses either single-perspective or hybrid methods. A single-perspective method is an adapted Human-Computer Interaction (HCI) method [3] [4] to the requirements of the specific domain. A hybrid method applies more than one traditional HCI approach in the usability evaluation of VR prototypes (e.g., an extended cogni-

1 <https://www.oculus.com/>

2 <http://store.steampowered.com/steamvr>

tive walkthrough and virtual world heuristics [5]). As opposed to the former, hybrid methods can accommodate a greater range of usability problems by capturing domain-specific and user experience related issues (e.g., spatial navigation, orientation, UI, etc.). Thus, many researchers argue that using hybrid methods for usability evaluation may be more effective than using single-perspective methods [5].

Using a hybrid method may be well suited when experienced, and trained usability evaluators are available to review a VR prototype. However, in the absence of such expertise applying a hybrid method may be troublesome. In the study reported in this paper, we use a modified version of the cognitive walkthrough method [6]. This is an expert evaluation method used to examine the usability of a product. It requires one or more evaluators to walk through a series of tasks and ask a set of questions from the user perspective. We applied a modified method to two hybrid versions of the REVERIE prototypes [7] [8] (both REVERIES and the prototypes are described in section 3.1). Those prototypes immersed users in two virtual environments (EU parliament in Brussels and a Virtual Gallery filled with cultural artefacts from various historical eras) where they had to participate in various educational activities. As the software was still in the early-design stage, we augmented the prototypes with storyboards and videos which provided evaluators with a step-by-step illustration of the missing user tasks (we dubbed this a hybrid prototype).

We found that our approach, single-perspective method and hybrid prototyping (see Section 3.2), identified a plethora of usability problems covering all aspects of the VR prototypes. We translated the usability problems into a high-quality set of user requirements to guide the future design of the prototypes. Another important deliverable of the study was a new method for effectively prioritising requirements. As opposed to existing methods, it captures input from multiple stakeholders in the requirements prioritisation process.

Our analysis shows that the proposed approach was effective in eliciting requirements for the REVERIE project. Relevant literature suggests that our approach generates comparable results to hybrid methods in usability evaluation. The remaining of the paper is organized as follows: Section 2 presents a review of the related work in the area; Section 3 gives a detailed account of the two VR applications developed using the REVERIE framework and discusses the procedure followed during the cognitive walkthrough process; Section 4 presents and discusses the results of the study, and the paper ends in section 5 with the conclusions.

2 Related work

Sawyer et al., [5] suggest a two-stage hybrid method to evaluate the usability of VR prototypes. In the first stage, it uses an extended version of the Cognitive Walkthrough (CW) method [9] developed for 3D virtual environment systems. The goal of this stage is to identify usability problems related to ‘in-world’ interactions using a task-based approach. This method is composed of three cycles of interaction: task action; navigation; and system initiative. Within a given scenario, a user navigates around the VE to complete a given task. The system may interrupt task comple-

tion to provide guidance or help. The user may decline or accept the system initiative and resume navigation. In the second stage, the method uses a set of heuristics specifically developed for VEs [10] [11]. The set includes 16 usability heuristics and an associated usability checklist of 53 items that are grouped into three categories (i.e. Design and Aesthetics, Control and Navigation, and Errors and Help). The goal of this stage is to enhance the findings of the first stage by identifying usability problems in the user interface (UI). The researchers applied the method to a study designed to evaluate the usability of a VR application in the context of health and safety education. The cognitive walkthrough captured problems (3 problems) related to navigation. It also captured some problems (2 problems) related to task action. The system initiative did not occur within the selected scenario, and therefore it was not used. The heuristics found 36 problems mostly related to the design and aesthetics of the user interface (UI). The researchers conclude that using a hybrid method in usability evaluation may be more effective than using a single-perspective method.

This conclusion was further reiterated in the Alencar et al. study [12]. The researchers performed a usability evaluation in a technologically mature VR application (an oil platform visualisation) using a multiple-stage hybrid method consisting of several usability evaluation methods. The researchers applied heuristic evaluation [13], usage observation sessions [14], questionnaires and interviews [15] as well as the communicability evaluation method (CEM) [16] and compared the results. The combined methods identified 82 HCI issues with the VR prototype. The issues related to ‘in-world’ interactions and the user interface (UI) (e.g., speed of navigation and size of icons). The number of usability problems is significantly higher than the previous study which demonstrates the strength of hybrid methods in usability evaluation. However, the application of additional evaluation methods has several problems:

- it is an open question whether using a multiple-stage hybrid method is more effective than using a single-perspective or a two-stage hybrid method in early-design VR prototypes;
- it tends to increase the overall cost of the evaluation;
- some methods are complex to use even for HCI experts, for example to apply successfully the Communicability Evaluation Method (CEM) method requires evaluators to go through a list of complex steps [16].

For the evaluation of the hybrid REVERIE prototypes a simplified approach compared to the aforementioned studies was adopted. This approach consists of a modified cognitive walkthrough method. Although, utilising a fusion of methods might have extracted more usability problems our reviewers used it successfully to obtain useful results.

3 Material and methods

The evaluation of the hybrid REVERIE prototypes was conducted with evaluators using a modified version of the cognitive walkthrough method. The evaluators identified a range of usability problems with the two prototypes that led to the development

of a series of design recommendations. These design recommendations define both the “what” and “how” to meet the physical and cognitive needs of the two VR prototypes target audience. We have prioritised the requirements (the “what” part of the design recommendations) based on the MoSCoW prioritisation method [23]. The method includes three items indicating different prioritisation levels. The “must-have” item refers to the requirements which were considered as essential for the prototypes to become ready for user testing and were all expected to be met by the next software release. The “should-have” item refers to requirements which are beneficial or useful to have in the next release of the prototypes. The “could have” item refers to requirements which could be met in a future version of the prototypes.

3.1 The REVERIE VR prototypes

REVERIE’s educational scenarios integrate a wide range of technologies and features (e.g., social networking services; tools to create personalised lookalike avatars; navigation support services; spatial adaptation techniques, AI techniques for responding to a user’s emotional status) [7] [8] to create a realist and responsive learning experience for students and teachers online. In the first scenario, a group of students registered on the REVERIE social network are invited to a virtual educational trip to the EU Parliament in Brussels. The students can access an avatar authoring tool [17] which they can use to build custom avatars utilising their appearance (e.g., by mapping their face on the avatar). Once users are online, an Embodied Conversational Agent (ECA) invites them to an exploratory tour of the parliament VE. The participants’ semi-autonomous avatars can automatically follow the autonomous agent through the tour. The destination is automatically given to each of the participants’ avatars. The semi-autonomous avatars can also reflect each participant’s facial expressions using a standard webcam. The ECA constantly analyses the user’s attention and emotional status and responds accordingly much as a teacher would in a real world (e.g., try to get a student’s attention if it was lost). The agent can demonstrate a range of pre-scripted behaviours (e.g., clapping, waving, happy and angry expressions, etc.) in response to the user’s status. After the tour is over, the autonomous agent walks to the side of the parliament for the online debate session to start [18]. In the virtual debate session, each student presents a topic of their choice to their fellow students. Teachers can further engage and enthuse students by streaming video clips from TrueTube³ in the virtual world. Finally, after the completion of each presentation students can vote for their preferred presentations and capture screenshots to share on their favourite social media channels. The second scenario maintains all these realistic and responsive functionalities, but immerses users in a different virtual world. Users enter a Virtual Gallery environment filled with 3D models of historical artefacts from various historical eras. There is no ECA in this scenario and users can start an educational activity as soon as they enter the world. In groups, they can observe and discuss the 3D models in a naturalistic way much as they would do in a real-world gallery.

³ <https://www.truetube.co.uk/>

3.2 Hybrid Prototyping

At the time of running the study, the REVERIE prototypes were still on an early beta stage. To enable evaluators to review the prototypes, we augmented them with storyboards and videos to simulate the missing tasks. We call this approach hybrid prototyping (i.e., software prototype augmented with storyboards and videos). For example, the storyboard in Fig. 1, shows the required steps students have to take to capture a screenshot in the first VR prototype and share it on Facebook.

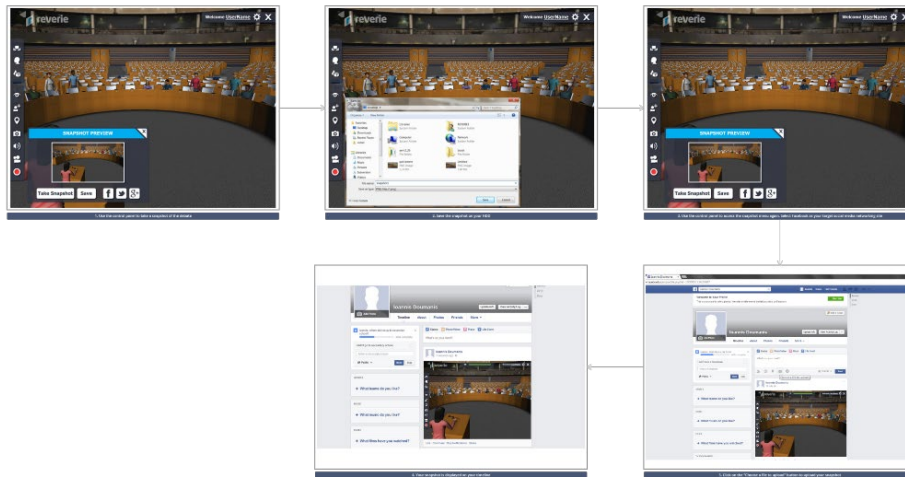


Fig. 1. One of the storyboards used to simulate the missing tasks in the EU Parliament scenario.

The video prototype was used to demonstrate the behaviour of the ECA. A series of videos using Living Actor Presenter⁴ have been created, featuring an ECA following the same script the autonomous guide agent would use in the VE. The videos were then assembled into an interactive video application using Articulate Storyline [19] and were displayed on the lab's main TV. The experimenter played the videos as required by the relevant tasks. A particularly challenging behaviour of the autonomous agent was its attention-grabbing capabilities. A video where the Living Actor agent displayed a similar to REVERIE agent attention-grabbing behaviour was included in the video application. The video was played as required by the experimenter when he thought that one or more of the evaluators were not paying attention to the guided tour.

3.3 The evaluators

In total, nine evaluators reviewed the prototypes for both educational scenarios. Three of the evaluators participated in a pilot review of the prototypes to validate the design of the study. Those reviewers completed the same tasks as the rest of the users but

⁴ <https://www.livingactor.com/Presenter/>

spent more time in the laboratory. They provided valuable feedback on the process and identified a range of bugs with the REVERIE prototypes that were logged and corrected prior to the main study. The remaining six evaluators were divided into two groups of three and had a variety of technical and media backgrounds. None of the evaluators had a HCI or cognitive science background. Finally, the evaluators had no previous experience using VR prototypes.

Table 1. The group of evaluators who reviewed the two educational scenarios

| Evaluators | Profile | Role | REVERIE Prototype |
|-------------|------------------------------|---------|-------------------------------------|
| Evaluator 1 | Media Producer | Teacher | Social networking & Virtual Gallery |
| Evaluator 2 | Media Producer | Student | Social networking & Virtual Gallery |
| Evaluator 3 | Media Producer | Student | Social networking & Virtual Gallery |
| Evaluator 4 | PSHE Teacher | Teacher | EU Parliament |
| Evaluator 5 | Office Assistant | Student | EU Parliament |
| Evaluator 6 | Marketing/Research assistant | Student | EU Parliament |

3.4 The modified Cognitive Walkthrough method

The modified cognitive walkthrough method [6] starts with an analysis of the required tasks, where the experimenter specifies a sequence of actions required by the user to complete the task and the system response(s) to those actions. The evaluators' walk through the steps, asking themselves the four questions below. Evaluators were required to answer the questions for each step of the assigned tasks. Answers to the questions have a binary (Yes/No) format, but evaluators are also required to comment on their preferred answer. Finally, the method required evaluators to indicate on a scale (0% to 100%) the likelihood users will have problems doing the right thing according to the requirements of each of the following question:

1. *Will the user realistically trying to do this action?*
This question finds problems with interfaces that make unrealistic assumptions about the level of knowledge or experience that users have).
2. *Is the control or the action visible?*
This question identifies problems with hidden controls (e.g., buried too deep within the navigation system) and controls that are not standard and unintuitive).
3. *Is there a strong link between the control and the action?*
This question highlights problems with ambiguous or jargon terms, or with other controls that look like a better choice. It also finds problems with actions that are physically difficult to execute.
4. *Is feedback appropriate?*
This question helps you find problems when feedback is missing, or easy to miss, or too brief, poorly worded, inappropriate or ambiguous.

We adapted the method by:

- providing additional text explanations under each question to guide evaluators on the kind of input expected (see above);
- providing evaluators with personas representing different users of the VR prototypes;
- integrating tasks into use cases reflecting the requirements of each educational scenario.

We designed the personas based on the initial user requirements gathered during the early stages of the REVERIE project [20]. We gathered quantitative data from 277 users using an online survey with questions about various aspects of the REVERIE system (e.g., avatar types, rendering style, the social network supported etc.). We also collected qualitative data from potential users from two informal usability inspections. The first inspection took place at the Education Innovation Conference & Exhibition in Manchester, UK, in February 2014 [21]. We asked teachers and students to review videos showing the REVERIE prototypes in action and to provide feedback on the camera. The second inspection took place internally with two of the REVERIE partners. We invited various evaluators (e.g., teachers and IT specialists) to use a preliminary version of the VR prototypes and to provide feedback about their usability and usefulness in education.

3.5 The evaluation sessions

In each group, two evaluators reviewed the tasks from a student perspective, while one expert from a teacher perspective. We provided evaluators a standard cognitive walkthrough form to use. The form listed the tasks evaluators had to review and for each task the tools they had to use to review the tasks (e.g., software prototype, storyboards or Internet browser). At the beginning of each session, we provided training on the use of the CW method. The training session lasted 10 minutes (instructions and Q&A) and it was deemed necessary as no evaluator had prior experience in evaluating VR prototypes. In total evaluators analysed 36 tasks grouped into four categories:

1. *user authentication and social networking tasks (11 tasks);*

The tasks included in this category, 4 tasks for teachers and 7 for students are related to the way users authenticate their credentials on the system as well as its social networking functionalities.

2. *REVERIE Avatar Authoring Tool (RAAT) (6 tasks);*

The tasks included in this category refer to REVERIE's integrated tool (RAAT) [17] for customising avatars, such as modifying the avatar's body features and mapping the user's face on an avatar.

3. *EU parliament scenario (20 tasks);*

This category includes 9 tasks for teachers and 11 tasks for students, and it refers to what users (teachers and students) can do in the virtual parliament scenario.

4. *Virtual 3D gallery (4 tasks);*

This category includes 2 tasks for teachers and 2 for students and it refers to what users (teachers and students) can do in the Virtual Gallery scenario. It includes

tasks such as exploring the Virtual Gallery to find a given object. Other tasks include rating the performance of a presenter using the system’s voting features.

We asked the first group of evaluators to review the first two tasks using the virtual 3D gallery scenario. The second group of evaluators reviewed the EU Parliament scenario. The set-up of the study was the same for both groups. Each expert conducted the walkthrough of the VR prototypes individually. This was done to ensure an independent and unbiased evaluation from each evaluator for the prototypes.

4 Results and discussion

After the walkthrough was completed, evaluators were asked to participate in a debriefing session to have their findings aggregated. The session was moderated by an external group moderator. We identified 47 usability problems with the VR prototypes. Most problems refer to the virtual parliament rather than the Virtual Gallery. This was to be expected as the Virtual Gallery scenario is much simpler to use. The user requirements were grouped into six macro-topics as appear in **Table 2** below.

Table 2. User requirements classification

| Topic | Requirements | Explanation |
|---|--------------|--|
| User authentication services and Social Networking integration | 14 | Requirements in this section refer to how users log-in to the VR prototypes and its social networking functionalities |
| Design of the Graphical User Interface (GUI) | 12 | Requirements in this section refer to the UI design of the VR prototypes & the RAAT tool |
| Media content | 3 | Requirements in this section refer to media content (external video links and 3D graphical assets) used in by the VR prototypes |
| Avatars | 1 | Requirements in this section refer to the appearance of the avatars and the way they interact in the virtual environment |
| Characteristics and functionality of REVERIE’s virtual environments | 7 | Requirements in this section refer to the appearance and usability of the VR prototypes |
| User-user and user-agent virtual interaction features | 10 | Requirements in this section refer to how the VR prototypes enable and facilitate interactions among users and with the autonomous agent |

44% of the generated requirements were considered as essential, while 28% as useful improvements and 28% as future improvements to the VR prototypes. In addition,

none of the problems discovered were discharged, or considered to have a cosmetic nature.

Table 3. Distribution of usability problems identified by each method by a number of problems

| Type of Problems | Our approach | Two-stage hybrid | Multiple Stage Hybrid |
|------------------|--------------|------------------|-----------------------|
| User Interface | 26 | 36 | 38 |
| In-World | 21 | 5 | 44 |

Although referring to different systems, it is possible to draw some conclusions about the performance of our approach compared to the literature (see Table 3). Specifically, the performance of our approach is comparable to a two-stage hybrid method [5], but not the multiple stage hybrid method [12]. It also captures a similar type of usability problems covering both ‘in-world’ interactions and in the user interface of the two VR prototypes. Future work aims to validate these findings by comparing the performance of the three methods using the REVERIE prototypes.

4.1 Likelihood of usability problems

Evaluators rated on a scale (0% – 100%) the likelihood a user would have a problem conducting an action in every step of the process. Below we present the average scores of the four questions of the modified cognitive walkthrough method (see Section 3.5) per task for the second group of evaluators. These were the evaluators who reviewed the EU parliament scenario. Table 4, shows the average scores of the teachers, while Table 5 the average scores of the students. Students had two more tasks to complete with the assistance of their teachers (see task 10 and task 11 in Table 5).

Table 4. Average scores assigned to each task by the teacher of the second group

| No. | Tasks | Average Scores | Std. Deviation |
|-----|--|----------------|----------------|
| 1 | Login to REVERIE using your TrueTube credentials | 12.5% | 25 |
| 2 | Select the first educational scenario | 12.5% | 25 |
| 3 | Select one of REVERIE’s standard avatars | 12.5% | 14.4 |
| 4 | Adjust the camera viewpoint to your preferred angle | 50% | 57.7 |
| 5 | Explore the 3D environment | 62.5% | 43.3 |
| 6 | Participate in the guided tour of the parliament | 37.5% | 14.4 |
| 7 | Start a debate on the topic “Multicultural London” with students | 25% | 0 |
| 8 | Debate on the topic “Multicultural London” with students | 43.75% | 37.5 |
| 9 | Ask students to rate the debate | 6.25% | 12.5 |

It is evident that the teacher thought that users would most likely have problems with the majority of the tasks in the virtual EU Parliament. However, he scored some

tasks lower than others which shows that he considered the importance of addressing specific usability problems more urgently than others. A particularly concerning task was number five (“Explore the 3D environment”). The teacher thought that there is a 62.5% likelihood that users will have problems with this task. Examples of usability problems teachers identified in this task were:

- the difficulty to accurately navigate the avatar in the environment using the navigation support tool;
- the difficulty to recognise the keyboard shortcut key (“M”) for activating the on-screen menu.

Table 5. Average scores assigned to each task by the students of the second group

| No. | Tasks | Student One | Std. Deviation | Student Two | Std. Deviation |
|-----|---|-------------|----------------|-------------|----------------|
| 1 | Login to REVERIE using your True-Tube credentials | 44% | 12.5 | 31% | 24 |
| 2 | Select the first educational scenario | 63% | 14.4 | 25% | 35.3 |
| 3 | Select one of REVERIE’s standard avatars | 56% | 12.5 | 31% | 37.5 |
| 4 | Adjust the camera viewpoint to your preferred angle | 44% | 37.5 | 69% | 47.3 |
| 5 | Explore the 3D environment | 62% | 47.8 | 62% | 48 |
| 6 | Participate in the guided tour of the parliament | 50% | 20.4 | 31% | 12.5 |
| 7 | Test the autonomous agent’s attention-grabbing features | 12.5% | 25 | 31% | 12.5 |
| 8 | Take a seat in the front row of the parliament | 62% | 25 | 25% | 20.4 |
| 9 | Debate on the topic “Multicultural London.” | 37% | 14.4 | 44% | 24 |
| 10 | Rate the debate on “Multicultural London.” | 44% | 12.5 | 12.5% | 25 |
| 11 | Share a snapshot of the 3D world on Facebook | 37.5% | 25 | 50% | 41 |

As opposed to the teacher, students scored all tasks higher, which show that they considered the usability problems found in all tasks as equally important. Students agreed with the teacher on task five (“Explore the 3D environment”). They thought that there is a 62% likelihood that users will have problems with this task. Examples of usability problems students identified in this task were:

- the fact that the users cannot view 360° around their avatar (e.g., behind you or left/right);
- it is standard in games to use WASD keys instead of a map to navigate in the environment;

- there is a need for system support (on screen information) on how to find the navigation system.

Students disagreed with the teacher in the first three tasks. They thought that there is a 42% probability that users will have problems with these tasks. Examples of problems students identified with these tasks were:

- no system response upon successful login to the system;
- the difference between “Avatar Library” and “Avatar Authoring Tool” is not clear;
- there is no description of what each scenario (entertainment and education) is about.

The teacher also highlighted several problems with these tasks. However, they thought that the likelihood of users having problems with these tasks is low (12.5%). Nevertheless, fixing navigation and UI problems in the VR prototype were given a priority in the next design iteration of the prototypes. Finally, we measured the agreement between students and teacher scores (only for the same tasks) by computing the intra-class correlation coefficient (ICC). The ICC score was 0.367 with 95% CI (-.461, 0.823) indicating poor agreement. This shows that the groups did not assess the likelihood of users having problems with each task consistently. A review of the data reveals that this is due to the number of problems each expert identified for each task. The use of personas also had an impact on the type and number of usability problems evaluators identified. Although we expect evaluators to identify different usability problems, poor disagreement reveals that they may not had the same level of understanding in the use of the method. Finally, the likelihood scores can significantly inform the process of requirements prioritisation. We recommend a method consisting of the following steps:

1. assign a weight for the importance of teacher and students likelihood scores, provided that both of the REVERIE prototypes were designed to be teacher-driven experiences, this weight should be 60/40;
2. recalculate the likelihood scores based on the assigned weights;
3. convert the likelihood score to a custom nine-point scale (0 = not important, 8 = extremely important) inspired by the planning poker agile method [22];
4. assign as the score to each requirement the average of the group.

For example, for Task 1 (see Table 4) the teacher will be assigned a score of 7%. Each student will be assigned 18% and 12% respectively. This gives an average score for the group of 2 on the 9-point scale. Any requirements matching the particular task should be assigned a score of 3 indicating moderate importance. This score can be further adjusted by the project partners to account for time and budget constraints. In the REVERIE project, we considered feedback only from the project partners and prioritised requirements according to the MoSCoW prioritisation [23] method. Our proposed method is better as it takes into consideration input from multiple stakeholders.

5 Conclusion

The main goal of this study was to test the effectiveness of the modified cognitive walkthrough method and hybrid VR prototypes approach in eliciting requirements for the design of VR prototypes of the REVERIE project. The approach was found to be highly useful predominantly due to its ability to capture a high-quality set of requirements in a cost-effective manner. The modified cognitive walkthrough captured several usability problems covering all aspects of the VR prototypes. The identified problems were into six clusters covering both ‘in-world’ interactions and UI (e.g., the design of the UI, navigation in the VE). Despite early design, the hybrid prototypes enabled evaluators to review the usability of the VR prototypes holistically. A comparison of the performance of our approach with the literature shows that it is slightly better than the two-stage hybrid method, but worse than the multiple-stage hybrid method. However, additional work is needed to compare the performance of the three methods using the REVERIE prototypes. We therefore conclude that our approach may provide a viable alternative to use in the evaluation of early-design VR prototypes when it is required (e.g., when the expertise needed to use a hybrid or a multiple-stage method is not available).

The first avenue for future work is to compare the performance of our approach to the two-stage hybrid and multiple stage hybrid methods using the REVERIE prototypes. We hope to validate that the performance of our approach is better or comparable to two-stage hybrid methods and to strengthen the conclusion above. Then, we plan to review the training we provide to evaluators on the use of the method. An instructional video at the beginning of each session holds the potential to significantly strengthen the evaluation consistency among evaluators. We would also like to explore increasing the participating stakeholders (e.g., developers) to the evaluation process to realise a more pluralistic walkthrough [24]. This is particularly important for R&D prototypes like REVERIE where the focus is on technological innovation, as it may teach technical people (e.g., developers, managers) to be more open to user experience requirements. Finally, we would like to apply our proposed requirement prioritisation method to real-world projects and gather feedback on its usefulness from stakeholders.

6 Acknowledgements

The research that led to this paper was supported in part by the European Commission under the Contract FP7-ICT-287723 REVERIE.

7 REFERENCES

1. Bellwood, P., Neuhaus, P., Juhra, C.: Adapting usability testing techniques to gather user requirements: an illustrative proposal, vol. 164 (2011).

2. Standardization, I. O. f.: Ergonomics of human-system interaction, Part 210: Human-centred design for interactive systems, ed, ISO 9241-210.
3. Bowman, D. A., Gabbard, J. L., Hix, D.: A survey of usability evaluation in virtual environments: classification and comparison of methods, *Presence: Teleoper. Virtual Environ.*, vol. 11, pp. 404-424 (2002).
4. Kaur, D.: A Usability Evaluation Method for Virtual Reality User Interfaces, (2018).
5. Sawyerr, W. Brown, E., Hobbs, M.: Using a hybrid method to evaluate the usability of a 3d virtual world user interface. In: 2nd International Conference on Human Computer Interaction & Learning Technologies (ICHCILT 2013) Proceedings, Abu Dhabi (2013), <https://goo.gl/TTSbzC>, last accessed 2019/05/14.
6. Bremin, S.: Rapid evaluation of TV interaction devices using a Cognitive Walkthrough method (2011).
7. Fechteler, P., Hilsmann, A., Eisert, P., Broeck, S. V., Stevens, C., Wall, J.: A framework for realistic 3D tele-immersion. In: 6th International Conference on Computer Vision / Computer Graphics Collaboration Techniques and Applications Proceedings, Berlin, Germany (2013).
8. Wall, J., Izquierdo, E., Argyriou, L., Monaghan, D. S., Connor, N. E., O. Poulakos, S.: REVERIE: Natural human interaction in virtual immersive environments. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 2165-2167 (2014).
9. Sutcliffe, A., Kaur, D. K.: Evaluating the usability of virtual reality user interfaces, vol. 19 (2000).
10. Rusu, C., Muñoz, R., Roncagliolo, S., Rudloff, S., Rusu, V., Figueroa, A.: Usability Heuristics for Virtual Worlds (2011).
11. Muñoz, R., Barcelos, T., Chalegre, V.: Defining and Validating Virtual Worlds Usability Heuristics. In: 30th International Conference of the Chilean Computer Science Society, pp. 171-178 (2011).
12. Alencar, M. F. C., Raposo, A. B., Barbosa, S. D. J.: Composition of HCI evaluation methods for hybrid virtual environments. In: 2011 ACM Symposium on Applied Computing, TaiChung, Taiwan (2011).
13. Gabbard J., Hix D.: A Taxonomy of Usability Characteristics in Virtual Environments, PhD thesis, (1999).
14. Anders Ericsson, K., Herber, A. S.: Protocol Analysis : Verbal Reports as Data, MIT Press (2018).
15. Goodman, E., Kuniavsky, M., Moed, A.: Observing the User Experience: A Practitioner's Guide to User Research (Second Edition), IEEE Transactions on Professional Communication, vol. 56, pp. 260-261 (2013).
16. deSouza, C. S.: The Semiotic Engineering of Human-Computer Interaction (Acting with Technology), The MIT Press (2005).
17. Apostolakis K. C., Daras, P.: RAAT - The reverie avatar authoring tool. In: 18th International Conference on Digital Signal Processing (DSP), pp. 1-6 (2013).

18. Kuijk, F., Apostolakis, K. C., Daras, P., Ravenet, B., Wei, H., Monaghan, D. S.: Autonomous agents and avatars in REVERIE's virtual environment. In: Proceedings of the 20th International Conference on 3D Web Technology, Heraklion, Crete, Greece (2015).
19. Articulate 360, ed: Articulate Global Inc. (2017).
20. Vet, J. d., Alfonso, D.: D2.1 Initial version of User Requirements Specification, European Commission (2012).
21. CTVC, REVERIE Informal Usability Inspection. In: Education Innovation Conference and Exhibition (2014).
22. Calefato F., Lanubile, F.: A Planning Poker Tool for Supporting Collaborative Estimation in Distributed Agile Development (2011).
23. Clegg, D., Barker, R.: Case Method Fast-Track: A Rad Approach: Addison-Wesley Longman Publishing Co., Inc. (1994).
24. Hollingsed, T., Novick, D. G.: Usability inspection methods after 15 years of research and practice. In: the 25th annual ACM international conference on Design of communication Proceedings, El Paso, Texas, USA, 2007.