The impact of multimodal collaborative virtual environments on learning: A gamified online debate

Doumanis, Ioannis, Economou, Daphne, Sim, Gavin Robert and Porter, Stuart

Doumanis, Ioannis ORCID: 0000-0002-4898-7209, Economou, Daphne, Sim, Gavin Robert ORCID: 0000-0002-9713-9388 and Porter, Stuart (2018) The impact of multimodal collaborative virtual environments on learning: A gamified online debate. Computers and Education, 130 . pp. 121-138. ISSN 0360-1315

Elsevier Editorial System(tm) for Computers & Education
Manuscript Draft

Manuscript Number: CAE-D-18-01171R1

Title: The impact of multimodal Collaborative Virtual Environments on learning: A gamified online debate

Article Type: Research Paper

Keywords: online learning platforms; virtual environments; collaboration; immersion; multimodal interaction

Corresponding Author: Dr. Ioannis Doumanis,

Corresponding Author's Institution: University of Central Lancashire

First Author: Ioannis Doumanis

Order of Authors: Ioannis Doumanis; DAPHNE ECONOMOU; Gavin Robert Sim; Stuart Porter

Abstract: Online learning platforms are integrated systems designed to provide students and teachers with information, tools and resources to facilitate and enhance the delivery and management of learning. In recent years platform designers have introduced gamification and multimodal interaction as ways to make online courses more engaging and immersive. Current web-based platforms provide a limited degree of immersion in learning experiences that diminish learning impact. To improve immersion, it is necessary to stimulate some or all of human senses by engaging users in an environment that perceptually surrounds them and allows intuitive and rich interaction with other users and its content. Learning in these collaborative virtual environments (CVEs) can be aided by increasing motivation and engagement through gamification of the educational task. This rich interaction that combines multimodal stimulation and gamification of the learning experience has the potential to draw students into the learning experience and improve learning outcomes. This paper presents the results of an experimental study designed to evaluate the impact of multimodal real-time interaction on user experience and learning of gamified educational tasks completed in a CVE. Secondary school teachers and students between ages 11 and 18 participated in the study. The multimodal CVE is an accurate reconstruction of the European Parliament in Brussels, developed using the REVERIE (Real and Virtual Engagement In Realistic Immersive Environment) framework. In the study, we compared the impact of the VR Parliament to a non-multimodal control (an educational platform called Edu-Simulation) for the same educational tasks. Our experiment results show that the multimodal CVE improves student learning performance and aspects of subjective experience when compared to the non-multimodal control. More specifically it resulted in a more positive effect on the ability of the students to generate ideas compared to a non-multimodal control. It also facilitated some sense of presence for students in the VE in the form of emotional immersion. The paper concludes with a discussion of future work that focusses on combining the best features of both systems in a hybrid system to increase its educational impact and evaluate the prototype in real-world educational scenarios.

*Title Page -- anything identifying the author should be on this page.

The impact of multimodal Collaborative Virtual Environments on Learning: A gamified online debate

Ioannis Doumanis[1], Daphne Economou[2], Gavin Robert Sim[1], Stuart Porter[3],

[1]School of Physical Sciences and Computing; University of Central Lancashire

[2]Department of Computer Science; University of Westminster

[3]Consultant

Corresponding Author:

Ioannis Doumanis

School of Physical Sciences and Computing

University of Central Lancashire

UK

Running Title: The Impact of Collaborative Virtual Environments on Learning: A gamified online debate

Word Count: 11,949

The impact of multimodal Collaborative Virtual Environments on learning: A gamified online debate

Online learning platforms are integrated systems designed to provide students and teachers with information, tools and resources to facilitate and enhance the delivery and management of learning. In recent years platform designers have introduced gamification and multimodal interaction as ways to make online courses more engaging and immersive. Current web-based platforms provide a limited degree of immersion in learning experiences that diminish learning impact. To improve immersion, it is necessary to stimulate some or all of human senses by engaging users in an environment that perceptually surrounds them and allows intuitive and rich interaction with other users and its content. Learning in these collaborative virtual environments (CVEs) can be aided by increasing motivation and engagement through gamification of the educational task. This rich interaction that combines multimodal stimulation and gamification of the learning experience has the potential to draw students into the learning experience and improve learning outcomes. This paper presents the results of an experimental study designed to evaluate the impact of multimodal real-time interaction on user experience and learning of gamified educational tasks completed in a CVE. Secondary school teachers and students between ages 11 and 18 participated in the study. The multimodal CVE is an accurate reconstruction of the European Parliament in Brussels, developed using the REVERIE (Real and Virtual Engagement In Realistic Immersive Environment) framework. In the study, we compared the impact of the VR Parliament to a non-multimodal control (an educational platform called Edu-Simulation) for the same educational tasks. Our experiment results show that the multimodal CVE improves student learning performance and aspects of subjective experience when compared to the non-multimodal control. More specifically it resulted in a more positive effect on the ability of the students to generate ideas compared to a non-multimodal control. It also facilitated some sense of presence for students in the VE in the form of emotional immersion. The paper concludes with a discussion of future work that focusses on combining the best features of both systems in a hybrid system to increase its educational impact and evaluate the prototype in real-world educational scenarios.

Keywords: online learning platforms; virtual environments; collaboration; immersion; multimodal interaction; gamification; learning performance; subjective experience

1. INTRODUCTION

Technology is rapidly changing the way teachers create and pass knowledge to students and vice versa. The use of e-learning platforms (e.g., Moodle (Al-Ajlan & Zedan, 2008) and Blackboard (Coopman, 2009)) is gaining momentum in almost all levels of education. These platforms provide teachers with powerful tools to enhance and improve student learning both inside and outside the classroom. Research has suggested that e-learning is at least as effective as traditional "brick and mortar" classrooms (Russell, 1999). This "no significant difference" phenomenon means that the impact of digital tools on education is already significant. To move beyond this phenomenon technology has the potential to revolutionise education by complementing traditional education. We subscribe to the view that traditional education plus digital tools are greater than either on their own. The possible disruptions in education of the marriage of traditional and online learning are endless (e.g., new pedagogies and teaching cultures). However, there are several design and technical challenges that will need to be addressed before e-learning becomes a vital part of the curriculum in education. This paper focuses on the research of gamification and multimodal interaction in online learning environments. Gamification refers to the use of game mechanics and game design thinking to make online courses fun and engaging (Huotari & Hamari, 2012) (e.g., the Level Up[1] plugin for the Moodle platform). Although gamification can boost the learner's motivation (Egenfeldt-Nielsen, 2011) (de Freitas, 2006) to engage with online learning materials alone, it does not guarantee that the intended learning will occur. Most online courses follow a single-mode (unimodal) approach to instruction (e.g., read the

[1] https://moodle.org/plugins/block_xp

PowerPoint slides) with little opportunities for students to get involved. This limitation creates an environment that stifles the student immersion in the learning experience and their ability to create rapport with the teacher and other students. To address these limitations, it is necessary to consider a multimodal approach to instruction. A multimodal instruction involves some or all of the human senses (e.g., vision, hearing, touch) when interacting with the material and other users to fully immerse students in the learning experience.

Some entry-level multimodal e-learning platforms already exist in the market (e.g., ilearn (University, 2017) and FrogPlay (Inc, 2017)). These platforms use a wide range of technologies and tools (e.g., virtualisation, web conferencing, analytics, mini-games, gamification) to enable teachers to design instructional strategies involving mostly asynchronous (and less often synchronous) multimodal interactions. These instructional strategies have the potential to draw students into the learning experience (e.g., through collaborative problem solving) and improve student learning outcomes. Although these learning experiences achieve a degree of immersion, the mostly asynchronous communication with a faceless teacher prevents students from developing a deep feeling of immersion that diminishes learning effectiveness (Georgiou & Kyza, 2017). By immersion, we refer to three dimensions of immersion: *spatial immersion*, *emotional immersion* and *temporal immersion* (Ryan, 2001). *Spatial immersion* refers to the capability of the online educational experience to construct a setting or a virtual space where learning can occur. *Emotional Immersion* is about evoking emotional participation in the educational experience. *Temporal immersion* is about producing stimulating educational experiences which students have the desire to follow to completion. The use of multimodal Collaborative Virtual Environments (CVEs) in online learning experiences holds the potential to successfully address these dimensions of immersion. A multimodal CVE is a virtual environment involving representations of teachers, students and / or learning content. In such environments users interact via real-time multimodal interactions. This rich sensational interaction coupled with real-time responses (simulated or from other human users) produces a deep feeling of immersion (Burdea, Richard, & Coiffet, 1996). In turn, teachers can design better online courses aimed squarely at improving student learning outcomes. To explain how enhanced learning performances can be achieved in multimodal CVEs we introduce an adapted version of the pedagogical model P2 (Bronack et al., 2008). The model provides the theoretical framework for this paper. As the technology has only recently enabled multimodal CVEs (Education, 2017), few empirical studies have addressed their impact on online learning (Isabwe, Moxnes, Ristesund, & Woodgate, 2018; Zizza et al., 2017). However, studies are yet to examine the impact of immersion (spatial, emotional and temporal) on the student subjective experiences and learning performance.

We conducted a field experiment to measure the impact of immersion on students' learning performance and subjective experiences with gamified educational tasks in a multimodal CVE. We recruited secondary school teachers and students between ages 11 to 18 to participate in the study. Both gamified educational tasks immerse users in a virtual environment (VE) reassembling the European Parliament in Brussels where they had to participate in an online debate. Participants completed the educational activities once using the immersive virtual Parliament and another using a non-immersive control (an educational Web platform called Edu-Simulation

(Economou et al., 2015)). The virtual Parliament (VP) was built using the REVERIE (Real and Virtual Engagement In Realistic Immersive Environments) (Fechteler et al., 2013) framework. REVERIE is a framework designed to facilitate real-time multimodal interaction on the Web. It integrates a wide range of tools (e.g., realistic 3D environments, human puppeted avatars, Embodied Conversational Agents (ECAs), human body reconstruction as a replica, spatial audio adaptation techniques (Bai, Richard, & Daudet, 2015) and others) that can be used to develop multimodal CVEs aimed at affecting spatial and emotional immersion. This type of multimodal CVEs can also affect temporal immersion, but as REVERIE does not offer representations for learning content, any impact is limited. The results of the experiment showed that the virtual parliament improves student learning performance and user satisfaction compared to the non-immersive platform. Also, the results suggest that the two platforms (multimodal CVE and Web platform) complement each other and future developments should be directed towards merging their functionalities into a hybrid platform. The remaining of the paper is organized as follows: Section 2 explains the theory behind the type of learning supported by REVERIE VP; Section 3 gives a detailed account of the REVERIE VP prototype; Section 4 gives a detailed account of the user trials, research environment, and goals; Section 5 presents the results of the study; Section 6 discusses the lessons learned from the study; Section 7 presents a list of design recommendations developed based on the study; and the paper ends in Section 8 with the conclusions and future work.

2. THEORETICAL FRAMEWORK

The REVERIE VP educational scenario implements an adapted version of the presence pedagogy (or P2 model) (Bronack et al., 2008). P2 is a new pedagogy that prepares students as global citizens by providing opportunities for active learning, interactive experiences, access to subject matter experts, collaborative projects, peer and social exchanges, and a deepened understanding of global diversity and interconnectedness. A critical attribute of the P2 model is immersion, and how its pedagogical principles can help to effect and affect it. In this paper, we discuss immersion for each pedagogical principle in light of the following three dimensions:

- *Spatial immersion,* which occurs when the educational experience constructs a setting for a potential learning narrative action. It is about creating a space to which students and teachers can relate, and populating this space with individual objects so it constitutes a viable world and becomes a setting where learning narrative can occur.
- *Emotional immersion,* which refers to the capability of the learning experience to invoke emotional participation; feelings of happiness or sadness towards the experience and its participants (e.g., peers, teachers).
- *Temporal immersion,* which necessitates the accumulation of the learning narrative. Successful learning experiences persistently stimulate the student's desire to see the learning narrative progress until it reaches a satisfactory conclusion. A core element of this dimension of immersion is the skills of the instructional designer to produce learning activities which stimulate students' interest.

Below, we present the ten principles of the P2 model, along with a discussion on how the REVERIE VP educational scenario that has been created implements each

principle offering an environment and educational activities tailored to activate background knowledge and expertise in useful ways and foster learning.
#Principle 1: Ask questions and correct misconceptions

Question and answer (Q&A) sessions are an integral part of learning enabling facilitation of self-exploration of learning narrative which is key in affecting temporal immersion. REVERIE VP complies with this principle by offering an appropriate environment where teachers and students can engage in collaborative inquiry and build a mental model of the space (which affects spatial immersion) and serve as a catalyst in promoting learning. For example, the visuals of the VE (i.e., an accurate representation of the European Parliament in Brussels) and the use of spatial audio adaptation techniques (Bai et al., 2015) suggest a virtual space for questioning, pondering, and discussion.

#Principle 2: Stimulate background knowledge and expertise

Empowering students to impart their knowledge and personal experiences in educational practice is critical in affecting temporal immersion. Enabling students to share their current knowledge helps to identify gaps that need to be addressed. In REVERIE VP a group of educational tasks have been designed that encourage students to share what they know about a topic with their group peers and come up with an answer to a specific question. The gamification of the educational tasks (e.g., competition) creates a competitive environment which further stimulates students to support the group.

#Principle 3: Capitalise on the presence of others

The formation of a mental model in the learning environment of who is actively engaged in the learning activity and who is the knowledgeable source learner should relate to in order to receive support are important steps towards developing spatial and emotional immersion respectively. In REVERIE VP students and teachers are represented by a range of digital representations, that indicate explicitly the user's role (student/teacher) and status (active/inactive). Avatars act as virtual "bodies" via which users experience the mediated environment.

#Principle 4: Facilitating interactions and encouraging community

Offering support for community formation, interaction and collaboration between community members play an important role in affecting emotional immersion. In REVERIE VP, puppeted avatars (with a user-adapted look and feel) and spatial audio (Bai et al., 2015) are used to facilitate interactions between participants. Puppeted avatars map user facial expressions (by mapping those on the avatar's face) and user engagement (via the affective tool that indicates if the users are engaged/or disengaged with the activity). Such features add an emotional dimension to user interaction.

#Principle 5: Support distributed cognition

VEs can readily facilitate distributed cognition by providing spaces that encourage participants to: interact with each other and objects in the VE, creating in this way spatial immersion; and collaborate with each other enabling the formation of affective relations, which is prerequisite for emotional immersion. REVERIE VP offers the virtual environment to form groups, navigate and discuss the requirements of a task. The learning sessions can be recorded allowing asynchronous access to conversations of individuals and a community of students.

#Principle 6: Share tools and resources

Learning in a VE needs to be facilitated by tools and resources easily identifiable, accessible and easy-to-use. Those factors play a key role in effecting temporal immersion and support participants to progress with their learning. In REVERIE VP, a range of tools is provided (e.g., spatial audio and multimedia creation tools) for participants to exploit the power of collaborative and active learning. These tools are easily identifiable and available to all participants through a common GUI. Also, the REVERIE social network (Fechteler et al., 2013; Wall et al., 2014) provides a common knowledge base resulting in a shared cognitive base for activity throughout students and teachers.

#Principle 7: Encourage exploration and discovery

The P2 model assumes that to sustain presence, a VE needs to be rich with resources and encourage learners to take the time to discover and explore those resources. By engaging students in activities which require the utilisation of shared-in-world tools, resources and knowledgebase allow them to easily build a mental model of the VE. REVERIE VP supports exploration in the virtual space and discovery of educational resources (e.g., videos from the TrueTube[2] platform that they can select and stream to all participants in the VE), which are key elements in affecting spatial immersion.

#Principle 8: Providing and delineating context and goals

A learning environment should provide the context for educational activity/s that address students and teachers' goals, with personal meaning and relevance to students. This can act as a catalyst towards temporal immersion that motivates students to explore a learning narrative to its completion. REVERIE VP delineates context and meaning by providing the environment (e.g., virtual parliament, pathways, chairs, avatars, interface elements, flags of different countries) to facilitate an educational activity on the topic of multiculturalism and accessibility to required tools and resources for the successful completion of the activity.

#Principle 9: Foster reflective practice

Successful learning activities require students not only to perform but also to reflect upon the outcome of their learning. Fostering reflective practice draws participants deeper into the learning narrative and helps them develop affective relations with the community (e.g., other students). VEs can nurture reflective practice using a variety of techniques (e.g., gamification and public speaking) and multimodal/multimedia tools. The educational activity that has been designed in REVERIE VP requires students to reflect upon the topic of multiculturalism before

[2] https://www.truetube.co.uk/

presenting their views to their peers. The learning process is guided by an observant teacher who provides feedback during every step of the process. After completing their presentation, students award each other points (1 to 5) and provided feedback using spatial audio. A GUI element (participants' menu) displayed the total score achieved for each presentation next to the name of each student. The guided activity, the verbal feedback and scores enabled students to reflect on their views and understand the significance of their arguments.

#Principle 10: Utilise technology to achieve and disseminate results

The REVERIE framework is consisted of the five following components (Fechteler et al., 2013): (1) multimodal and multimedia signal acquisition; (2) interaction and autonomy; (3) composition and visualisation; (4) networking and immersive communication and (5) social networking. These components work together to facilitate communication and collaboration between teachers and students online. Each of these components implements a range of functionalities (e.g., fully puppeted avatars and autonomous Embodied Conversational Agents (ECAs) to affect and effect immersion (spatial, emotional and temporal) of participants in the educational experience.

3. THE VIRTUAL PARLIAMENT PROTOTYPE

We used the REVERIE framework (Fechteler et al., 2013) to implement an educational scenario which immersed participants in a guided tour of the virtual European Union (EU) parliament followed by an online debate session. In this role-playing scenario participants (in the role of teachers or students) interacted with each other and an Embodied Conversational Agent (ECA) to complete educational tasks designed to promote dialogic learning (Hajhosseiny, 2012).

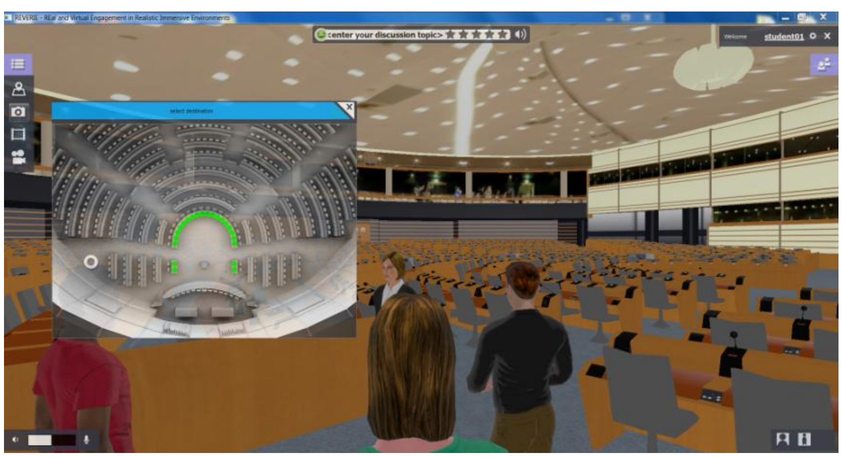

Figure 1: The REVERIE's Virtual Parliament educational scenario

The REVERIE VR environment features an immersive 3D representation of the European Union (EU) Parliament (see Figure 1); a participants' list button; the main menu button; and two bars (viewpoint and volume). Participants could explore the environment, communicate with other participants and the autonomous ECA (in the role of a tour guide) using avatars in a multimodal manner (e.g., using spatial audio and nonverbal communication). In addition, users could create multimedia content to share with other users on social media. Students could create personalised avatars using the RAAT tool (Apostolakis & Daras, 2013), while teachers were assigned with a default avatar. Each user sessions was moderated by the teacher assigned to each group (e.g., to permit students to speak, monitor and prevent cyberbullying).

4. EXPERIMENTAL APPARATUS

The study was designed to evaluate the **learning performance** of students after completing gamified educational tasks within a multimodal CVE (REVERIE VP). We were also interested in evaluating the impact of the multimodal CVE on the **users' subjective experience** of the system. The success of the REVERIE VP in imparting knowledge depends on the **usability** of the system and the **cognitive accessibility (CoA)** of the educational tasks. This is because to successfully acquire knowledge users should be enabled to: (a) complete the assigned tasks with completeness (effectiveness), little effort (efficiency) and satisfaction (Nielsen, 2012) and (b) cognitively process (access, interpret and respond to) the information conveyed by the virtual environment (simulated or by other human users) (Seeman & Cooper, 2015). Information in a multimodal CVE such as the REVERIE VP can be conveyed through multiple multimodal and multimedia means of communication and artefacts (e.g., ECA, other human users, etc.). The educational tasks required participants to take a guided tour of a virtual parliament and to participate in a debate on the topic of multiculturalism with other participants. The experimental evaluation activities for REVERIE's VP educational scenario included lab-based testing and contextual studies in schools.

4.1 Edu-Simulation (non-multimodal control)

The Edu-Simulation web platform (see Figure 2) is a prototypical online learning platform (similar to Moodle). It features a menu bar at the top of the page with the learning scenarios (or simulations) the user participates.

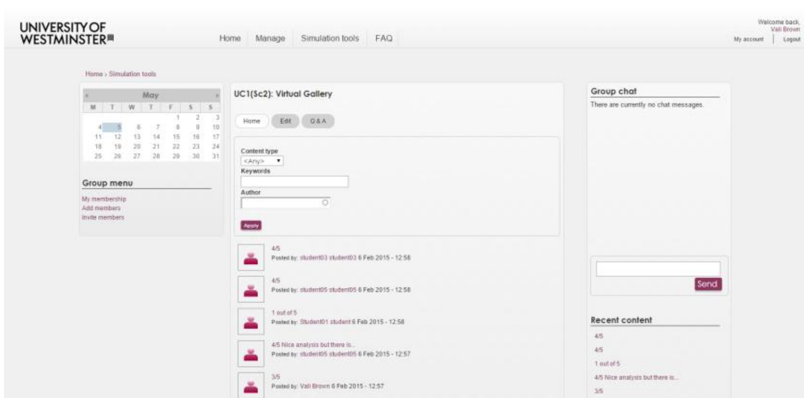

Figure 2: The Edu-Simulation Educational Platform

The web platform has all the necessary features (e.g., role-playing, voting among students) for users to complete the same tasks to REVERIE VP, but interaction is restricted to a conventional multimedia Web environment. As on REVERIE VP system, online learning sessions were moderated by the teacher assigned to each group (e.g., to permit students to speak, monitor and prevent cyberbullying, etc.).

4.2 Overview

The first study was conducted in the laboratory in a set-up simulating an actual classroom environment (see Figure 3). The same setting was replicated in two schools in the UK where teachers and students performed the same educational tasks as those in the laboratory. We manipulated the following variables:

- The type of systems (i.e., multimodal CVE vs multimedia Web)
- The type of educational content (multimodal vs multimedia)
- Type of educational activity (individual vs group)
- The order of systems (REVERIE VP vs Edu-Simulation vs vice versa) to study practice effects

Four REVERIE researchers were present in the lab to record each session and to provide the necessary support (technical and logistical) for the successful completion of each session.

Fig. 3. One of the evaluation sessions in the laboratory

Participants used a standard computer (with a keyboard and wireless mouse) to complete the educational tasks on both platforms. Also, they had access to a Bluetooth headset and a web camera when interacted with REVERIE VP. At the beginning of each session, participants were given a short training session to become familiar with the use of the prototypes. All user sessions were recorded on HD video. About the possible effects of each platform we had the following hypotheses:

H1: The amount of learning of a topic is higher in a multimodal CVEs (such as REVERIE VP) compared to a conventional multimedia Web environment. This is because the educational activity implements the presence pedagogy (P2) and hence, produces a more immersive experience in the multimodal CVE compared to the multimedia Web. The increased immersion renders the interaction smoother between users (and the VE), enhances the understanding of the content, thus supporting greater learning.

H2: The students' subjective satisfaction is higher with REVERIE VP regardless of the type of educational activity (individual vs group). This is because the game-like environment and educational tasks positively impact how they perceive the fun and enjoyment of the learning experience. The use of multimodal real-time interaction enhances this effect by enabling a more natural way of communicating with their peers and teachers.

4.3 Participants

In total, 48 participants have participated in this study. Six of the participants were used in a pilot study, to ensure that the main study will run problem free. Those six participants completed the same tasks as the others but spent overall more time in the lab to discuss improvements in the instruments of research and identify any bugs the prototypes might have. The remaining 42 participants (36 students and 6 teachers) were assigned at random to the study conditions. Each group included eighteen participants (including teachers); Twelve of the participants were females and twenty-four males; the age range of the first group was 11-14 while the second group 11-18. All participants were English–speakers (either native or as a second language) and had a variety of familiarity with video games and social networking sites (e.g., Facebook, Twitter, etc.).

In the following sections, we describe the data which we collected and the methods that we used to evaluate those two hypotheses. Section 4.3 covers the research instruments that had to be created, and section 4.5 explains the experimental conditions.

4.4 Measures and Methods:

We measured usability as effectiveness, efficiency and satisfaction (Nielsen, 2012). Effectiveness was measured as task completion and error rates, i.e., the number of tasks participants completed and the errors they made when attempted to complete a task. Another important measure of effectiveness was rating/voting, i.e., the total number of votes participants cast. Efficiency was measured as the time needed to perform a task. Satisfaction was evaluated using the Post-Study System Usability Questionnaire (PSSUQ) (Lewis, 2002). We measured Cognitive accessibility (CoA) as user satisfaction using a standardised questionnaire (Adams, 2007). The user experience includes all of the previous measures as well as objective and subjective measures of immersion. Finally, we measured learning performance as the fluency of argumentative behaviour of students (Frijters, ten Dam, & Rijlaarsdam, 2008).

Objective measures:

- **Completion rate:** The completion rate was recorded as a binary measure of task success (coded as 1) or task failure (coded as 0). The completion rate for the educational scenario is the number of users who completed the assigned task (group or individual) divided by the total number of users who attempted it.

- **User errors:** A user error was defined as the case in which a user: (a) did not choose the appropriate method to achieve their objective (e.g., the correct UI button to increase the volume or the gesture needed to start the interactive tour); (b) did choose the appropriate method to achieve their objective but did not use the method correctly (e.g., the keyboard shortcut properly navigate their avatars in REVERIE VP) (Norman, 2013).

- **Total time:** This is the average time (in seconds) of users completing tasks with both systems (REVERIE VP and Edu-Simulation).

- **Rating/Voting:** Voting was an integral part of the educational activities on both systems (REVERIE VP and Edu-Simulation). As students were able to vote on both platforms (REVERIE VP and Edu-Simulation) the total number of votes per group of users was taken as a measurement of task effectiveness.

- **Assessment of learning:** We examined the impact of each system on the dialogic learning of the topic of multiculturalism. In particular, we were interested in measuring the fluency of argumentative behaviour of students (Frijters et al., 2008) when completed educational tasks on each system. In the individual activity, we measured the number of arguments in support of the position taken by the student (for/against multiculturalism). In the group activity, we measured the number of arguments used by the representative of the group to support the provided solution. Using the Toulmin Model of Argument (Toulmin, 2003), we analysed the validity of the arguments used by students. Specifically, we distinguished three parts of a basic argument (Claim, Data and Conclusion) and counted how often they appeared in each student's response. The total score of each group quantifies the short-term post learning effect of each system. The score was also considered as evidence of the degree of students' temporal immersion in the learning narrative of each activity. To ensure reliability of measurements, we asked an external researcher to assess a sample of the data. We measured inter-rater reliability by computing the intra-class correlation coefficient (ICC). The ICC score was 0.821 with 95% CI (0.604, 0.914) indicating good inter-rater reliability.

- **Amount of immersion:** This is the degree of immersion elicited by each system. We thought that the sense of emotional immersion of each group of users would be reflected on the user's facial expressions. The more **attentive** and **emotionally engaged** a group of users was to the task, the more immersed it was thought to be in the experience (Paul Cairns, 2014). Using the Crowdsight Toolkit,[3] we analysed the videos of each group for attention and emotional engagement. The toolkit measures attention and emotional valence (positive, neutral, negative) (Kensinger, 2004). As emotional engagement, we computed the average of emotional valence for each group of users in the study. The developers say that the cross-validated accuracy of the application is 83.5% (Valenti, 2017) The technology behind the toolkit has also been used in several scientific publications (Shan, Guo, You, Lu, & Bie, 2017) (Machajdik et al., 2011) which ensures reliability of results. Both the reliability and validity of the immersion measurements can be increased by human coding. However, using human coding requires expertise in FACS (Facial Analysis Coding System) (Sayette, Cohn, Wertz, Perrott, & Parrott, 2001). As we did not have access to such expertise, we decided to rely on the computational approach.

Subjective Measures:

[3] http://sightcorp.com/

The responses to the individual items of the four electronic questionnaires: All, but the first questionnaire used a seven-point Likert scale (1=strongly disagree, 7 = strongly agree). Both students and teachers completed the same type of questionnaires (addressing the same aspects of the prototypes), but their length and complexity differed. Students had to answer shorter versions of the questionnaires with fewer and less complex questions for each aspect of the two prototypes.

1) The first questionnaire used mixed format (binary and open-ended) questions, and it was designed to assess the **users' spatial immersion** with the educational activities in REVERIE VP. Participants were asked to indicate (yes/no) whether they saw specific objects in the VE and if they could recall their names.
2) The second questionnaire addressed the **usability** of the systems. The usability questionnaire is based on the standardised Post-Study System Usability Questionnaire (PSSUQ) (Lewis, 2002) questionnaire.
3) The third questionnaire assessed the **cognitive accessibility (CoA)** (Adams, 2007) of the educational tasks completed with the systems. It addressed the effectiveness of each prototype in completing an educational scenario and numerous aspects of the user's satisfaction.
4) The fourth questionnaire examined qualities of the virtual representations (ECA and user avatars) used in REVERIE VP. This questionnaire was divided into two areas, the first addressing qualities of the users' avatars (e.g., fidelity, realism, interaction, etc.) and the second, addressing qualities of the tour guide ECA (e.g., quality of voice, gesturing, etc.).

We determined the internal reliability of the four questionnaires for both prototypes (see Table 1). Cronbach alpha's scores range from 0.648 to 0.926 indicating good internal consistency. The KR-20 score for spatial immersion also indicates good internal consistency. The student version of the CoA and usability questionnaires has lower internal consistency (< 0.70). This is to be expected because of the low number of questions (14 items) addressing multiple constructs. The validity of the CoA and usability questionnaires have been determined in prior studies (Adams, 2007; Fruhling & Lee, 2005). We assessed the content validity of the remaining two questionnaires by consulting relevant experts in the project consortium.

Table 1. Reliability of the four questionnaires

Questionnaire	Cronbach's alpha (REVERIE VP)	KR-20 (REVERIE VP)	Cronbach's alpha (Edu-Simulation)
Cognitive Accessibility (CoA)	0.809	N/A	0.926
Usability	0.648	N/A	0.897
Spatial Immersion	N/A	0.60	N/A
Virtual Representation	0.842	N/A	N/A

- The answers to the semi-structured group interview. The interview was conducted at the end of each session with each group of students. It was led by a REVERIE researcher, and it had a mixed structure. Each interview session combined standardised questions and session-specific questions that were asked based on participant observation made in the particular session.

Both types of questions were open-ended and provided participants with an opportunity to give their impressions about the platforms (REVERIE VP vs Edu-Simulation) and offer suggestions, about what we should improve in future versions. Each interview lasted 5-10 minutes.

4.5 Educational Tasks

Participants were administered in groups of six. Each group had to complete two educational activities, one individually and another as a group. Both activities required students to present their views about multiculturalism. Because of the complexity of the topic, we asked teachers to complete a pre-learning task with students to brainstorm ideas. In the individual activity students freely selected a topic of interest (e.g., multiculturalism and food). In the group activity, students had to consider the impact of multiculturalism on an aspect of the society (e.g., schools and communities) given by their teacher. To complete each activity groups had to use the interactive tools available on each system (e.g., spatial audio on REVERIE VP and group chat on Edu-Simulation). We gamified each task by adding relevant elements (e.g., exploration, rewards and competition).

The individual activity had the following format:

- Teachers and students participated in a guided tour of the virtual parliament given by a tour guide Embodied Conversational Agent (ECA) (applicable only to REVERIE VP)
- The teacher asked the first student to give a presentation.
- After the presentation, the teacher had a discussion with the student and asked the class to provide feedback (verbally and using points) about the presentation.

After students had finished presenting, the teacher announced the winner of the activity (i.e., the student with the highest total points).

The group activity had the following format:

- Students discussed the topic in separate groups.
- Each group voted for a representative whose task was to present the group's view to the class.
- After each group's presentation, the teacher awarded points for good aspects of the presentation (e.g., clarity of language) and deducted points for bad aspects of the presentation (e.g., racial comments).
- The group with the highest points won the group activity.

4.6 Experimental Design

We conducted the study with a mixed factorial design because apart from not being possible to expose each participant to every single variable, a mixed design might be favoured to avoid practice effects. The type of the system (multimodal CVE vs multimedia Web), type of content (multimodal vs multimedia) and type of educational task (individual vs group) was measured as within-subjects variables. The order of task (group vs individual vs vice versa) and order of systems (REVERIE VP vs Edu-Simulation or Edu-Simulation vs REVERIE VP) were measured as between-subjects variables to observe any practice effects. Participants were randomly assigned to four groups of nine participants and the eight study conditions.

Table 2. Experimental Design

Participants (N = 42)	**REVERIE VP (Multimodal CVE)**	**Edu-Simulation (Multimedia Web)**
1 – 9 Students (+ 1 teacher)	*Group Educational Activity*	*Individual Educational Activity*
10 – 18 Students (+ 2 teachers)	*Individual Educational Activity*	*Group Educational Activity*
	Edu-Simulation (Multimedia Web)	**REVERIE VP (Multimodal CVE)**
19 – 28 Students (+ 2 teachers)	*Individual Educational Activity*	*Group Educational Activity*
29 – 36 Students (+ 1 teacher)	*Group Educational Activity*	*Individual Educational Activity*
Dependent Variables	*Objective and subjective measures/ Learning Performance*	

Table 2, shows the eight study conditions. 1) REVERIE VP with group task vs Edu-Simulation with individual task or 2) REVERIE VP with individual task vs Edu-Simulation with group task or 3) Edu-Simulation with individual task vs REVERIE VP with group task or 4) Edu-Simulation with group task vs REVERIE VP with individual task.

5. RESULTS

Objective measures. In this section, we provide the results of the objective user assessment. This includes analysis of the data for each of the following measures.

Completion rate. Completion rate of users on REVERIE VP was 100% with both educational activities. However, the completion rate on Edu-Simulation was 94.4% with the individual task and 100% with the group task.

User Errors. An analysis of the videos for each user session revealed the following about user errors:

REVERIE VP:

(1) Most participants (teachers and students) could not find how to permit the RAAT tool to use the computers' camera. The two buttons (Deny/Allow) at the top of the browser were not immediately visible, which confused as to what they needed to do to continue.
(2) Out of the six teachers, three had problems identifying the correct gestures (i.e., head nod or head shake) to properly interact with the ECA at the beginning of the guided tour.
(3) Out of the six students, five had problems avoiding clashing their avatars in the virtual parliament.

Edu-Simulation:
(1) Out of the six students, three required additional help to find specific pages on Edu-Simulation as part of an educational activity.

Total Time. In Table 3, we present the total time needed to complete the educational tasks with REVERIE VP and Edu-Simulation. A two-factorial ANOVA revealed that the average time (in seconds) to complete the educational tasks (individual and group) differed as a function of the type of system ($F(1,6)=14.193$; $p < .05$). Specifically, participants spend significantly more time on REVERIE VP (mean REVERIE VP = 1872.8 sec.) than on Edu-Simulation (mean = Edu-Simulation 1254.8 sec.). No other ANOVA comparison reached a significance level.

Table 3. Time (seconds) as a function of type and order of systems

Type of System	**REVERIE VP vs. Edu-Simulation (N = 42)**	**Std. Deviation**	**Edu-Simulation vs. REVERIE VP (N = 42)**	**Std. Deviation**
REVERIE VP	1912.65	355.4	1833	329.5
Edu-Simulation	1194.15	173.7	1315.5	191.6

Rating/Voting. All students (36/36) voted for each other when completed the individual educational activity on REVERIE VP. However, on Edu-Simulation not all students voted for each other (20/36).

Assessment of learning. A two-factorial ANOVA revealed that students made significantly more claims (Table 4) when completed educational activities on REVERIE VP ($F(1, 34) = 4{,}451$; $p<0.05$) than on Edu-Simulation. No other comparisons reached statistical significance.

Table 4. Assessment of learning

Type of system	**Claim**	**Data**	**Conclusions**
REVERIE VP	35	31	13
Edu-Simulation	23	19	15

Amount of Immersion. The amount of users' immersion as a factor of emotional engagement (positive, negative and neutral valance) with each of the systems is shown below. It is clear that participants experienced a high proportion of negative and neutral valance emotions on both systems. A one-way ANOVA showed a significant main effect for system (F (1,36422) = 176.361; p<.001). Participants experienced significantly more negative valance emotions on Edu-Simulation (mean Edu-Simulation = 16.1%) than on REVERIE VP (mean REVERIE VP = 14.3%). The most disruptive emotional state experienced by participants was sadness. However, as the Crowdsight toolkit does not recognise facial expressions of boredom, it is unknown if sadness was indeed the most frequently negative emotion experienced by participants. Then, additional ANOVA comparisons showed a significant main effect of system for both Neutral Valance (F (1, 9104) = 22.797; p < .001) and Positive Valance (F (1, 18210) = 14.013; p <.001). The results for neutral valance show that participants experienced significantly more neutral emotions on REVERIE VP (mean REVERIE VP= 33.9%) than on Edu-Simulation (mean Edu-Simulation = 32%). Then, the results for positive valance suggest that participants experienced significantly more positive valance emotions on REVERIE VP (mean REVERIE VP = 9.85%) than on Edu-Simulation (mean Edu-Simulation = 9.15%).

Table 5. Amount of emotional engagement using the two systems

Type of System	**Positive Valance**	**Std. Dev.**	**Negative Valance**	**Std. Dev.**	**Neutral Valence**	**Std. Dev.**
REVERIE VP	9.85%	13.11	14.3%	12.0	33.9%	20.8
Edu-Simulation	9.15%	12.13	16.1%	13.0	32.0%	16.1

Regarding attention, were more attentive when completed tasks on Edu-Simulation (mean = 15.8 seconds) than REVERIE VP (mean = 12.01 seconds). A one-way ANOVA showed that the difference is statistically significant (F (1, 9103) = 80.819; p < .001).

Table 6. Amount of attention using the two prototypes

Type of System	**Attention (sec.)**	**Std. Dev.**
REVERIE VG	12.01	15.6
Edu-Simulation	15.8	23.6

Subjective Measures. This section provides the results of the subjective user assessment. This includes analysis of the user feedback gathered through the four questionnaires and the exit group interviews.

Spatial Immersion Questionnaire. The left column of Table 7 shows that teachers recognised 74% (29 out of 36) of the objects they encountered in REVERIE VP.

Table 7. Spatial immersion results

Objects(Y/N)	**Teachers (N = 6)**	**Students (N = 36)**
Recognised	29/7	161/55

Attached Meaning	30/6	162/54

Also, they could attach meaning to 83% of the objects they recognised (30 out of 36). Students (see right column of Table 7), recognised 80% of the objects in the VE. They could also recall 75% of the names of the objects they recognised.

Usability Questionnaire (Teachers Only). To identify the impact of the independent variables on the usability qualities, we performed a series of one-way ANOVAs. One of the ANOVAs revealed a significant main effect of type of system ($F (1, 94) = 4.657$; $p < .05$) on the Interface Quality (IQ). Teachers rated the interface quality of REVERIE VP significantly higher (mean IQ = 5.13) than Edu-Simulation (mean IQ = 4.50).

Table 8. Teachers' mean usability ratings for the two systems

Usability Qualities	**REVERIE VP**	**Std. Deviation**	**Edu-Simulation**	**Std. Deviation**
System Quality	5.33	1.2	4.86	1.5
Information Quality	4.53	1.6	4.19	1.9
Interface Quality	5.13	1.5	4.50	1.4

Cognitive Accessibility (CoA) Questionnaire (Teachers Only). The following table (Table 9) shows how teachers rated the cognitive accessibility (CoA) of the educational tasks. To investigate the potential impact of the independent variables on specific CoA measures, we conducted a series of two-factorial ANOVAs.

Table 9. Teachers mean CoA ratings for REVERIE VP and Edu-Simulation

CoA Qualities	**REVERIE VP**	**Std. Deviation**	**Edu-Simulation**	**Std. Deviation**
The organisation and implementation requirements of the educational task	5.27	1.05	4.63	1.31
Input Modalities	4.70	1.15	4.08	1.77
Feedback Mechanisms	4.64	1.42	4.70	1.66
Short Term Memory Requirements	4.13	1.96	3.93	1.73
Emotional Responses	5.67	1.03	4.94	1.26
Long Term Memory Requirements	4.63	0.98	4.63	1.31
Building a Mental Map	4.79	1.75	4.22	1.26
User responses	4.25	1.75	4.17	1.86
Complex user responses	4.50	1.53	4.50	1.69

The ANOVA comparisons showed an effect of task on the following questionnaire items for REVERIE VP:

- Item 1 (*"The educational activities in the Virtual Parliament were too simple*") ($F (1, 4) = 8.00$; $p < .05$)
- Item 18 *("I did not have to think hard about what I was doing in order to respond in the virtual parliament")* ($F (1, 4) = 49.00$; $p < .05$)

- Item 16 (*"I did not need help to properly navigate in the virtual parliament"*) ($F(1, 4) = 25.00; p < .05$)

Teachers thought that the group task (mean Group = 5.67) was simpler that the individual task (mean Individual = 4.33). They also felt that they had to think less in the group task (mean Group = 7.00) than in the individual task (mean Individual = 4.66). Finally, they felt they needed less help navigating the virtual parliament in the group task (mean Group = 4.66) than in the individual task (mean Individual = 3.0).

It also showed an effect of task on the following questionnaire item for Edu-Simulation:

- Item 19 *("The platform responded appropriately when I am confused or overloaded with information (e.g., by allowing me to create my own To-do list")* ($F(1, 4) = 16.00; p < .05$)

Teachers thought that Edu-Simulation responded more appropriately when they were confused or overloaded with information in the individual task (mean Individual = 3.3) than in the group task (mean Group = 2.0).

Students' version (Usability and Cognitive Accessibility). Between the usability qualities, a one-way ANOVA test did not reveal any significant differences in the way students rated the usability of the systems.

Table 10. Students mean usability ratings for REVERIE VP and Edu-Simulation

Usability Qualities	**REVERIE VP**	**Std. Deviation**	**Edu-Simulation**	**Std. Deviation**
System Quality	4.83	1.207	5.36	1.222
Information Quality	2.64	1.515	3.08	1.763
Interface Quality	4.81	1.283	5.08	1.228

Additional two-factorial ANOVA tests on each questionnaire item did not reveal any significant main effects for either type of task or order of systems. A one-way ANOVA test, however, did reveal a significant main effect of type of system on the following CoA qualities:

- Short-Term Memory Requirements ($F(1,72) = 14.006; p < .001$)
- Emotional responses ($F(1,72) = 4.900; p < .05$)

Students thought that the educational activities they completed on REVERIE VP had less demands (mean REVERIE VP = 4.31) from their working memory compared to Edu-Simulation (mean Edu-Simulation = 5.53). Also, they experienced significantly higher emotional responses when they completed tasks with REVERIE VP (mean REVERIE VP = 5.83) than with Edu-Simulation (mean Edu-Simulation = 5.17).

Table 11. Students mean CoA ratings for REVERIE VP and Edu-Simulation

CoA Qualities	**REVERIE**	**Std.**	**Edu-**	**Std.**

	VP	Deviation	Simulation	Deviation
The organisation and implementation requirements of the educational task	3.97	1.230	4.36	1.606
Input Modalities	4.19	1.369	3.67	1434
Feedback Mechanisms	3.33	1.639	3.25	1.381
Short Term Memory Requirements	4.31	1.489	5.53	1.183
Emotional Responses	5.83	1.108	5.17	1.298
Long Term Memory Requirements	4.38	1.614	5.06	1.264
Building a Mental Map	5.38	1.399	4.72	1.323
User responses	2.97	1.594	2.92	1.538
Complex user responses	3.36	1.676	3.56	1.557

We also found a main effect for the type of task on the following questionnaire items of Edu-Simulation:

- Item 1 *("The educational activities in Edu-Simulation was too simple")* (F (1, 34) = 14.268 p< .01)
- Item 2 *("The ways to communicate in Edu-Simulation (e.g., text and images) are enough to immerse me in the educational scenario. I wouldn't like to use a different environment (e.g., more game-like)")* (F (1, 34) = 4.857; p <.05)

Table 12. Summary of responses for questionnaire items with a significant type of task effects for Edu-Simulation (Students)

	Type of Task			
Item	**Group**	**Std. Deviation**	**Individual**	**Std. Deviation**
Item 1	3.5	1.6	5.2	1.0
Item 2	3.1	1.38	4.1	1.33

Virtual Representations Questionnaire (Teachers Only). We compared the teachers' ratings between VR qualities below (Table 13) using one-way ANOVA tests. We did not find any significant differences between the way teachers rated the qualities of REVERIE VP.

Table 13. Teachers mean ratings of the virtual representations used in REVERIE VP

Qualities	REVERIE VP	Std. Deviation
User Avatar	4.22	1.73
Communication between avatars	4.25	1.56
Feedback Mechanisms	4.33	1.39
Tour guide ECA	4.83	1.54

Virtual Representations Questionnaire (Students Only). However, a one-way ANOVA test revealed a statistically significant difference (F (1, 3) = 4.139; p < .05) between the way students rated the virtual representation qualities of REVERIE VP. A

Tukey's post-hoc test revealed that the mean score for the dimension "communication between avatars" (M= 4.34, SD =1.56) was significantly different than the dimension "user avatar" (M= 4.22, SD=1.73). There was no statistically significant difference between the remaining qualities.

Table 14. Students mean ratings of the virtual representation qualities of REVERIE VP

Qualities	**REVERIE VP**	**Std. Deviation**
User Avatar	3.92	1.8
Communication between avatars	4.34	1.5
Feedback Mechanisms	4.21	1.5
Tour guide ECA	4.13	1.6

Finally, additional two-way ANOVA comparisons showed a significant main effect of type of task for Item 29 *("The virtual guide helped me to understand more about the EU Parliament (areas, operation, etc.)")* (F (1, 34) = 7.026; p< .05).

Table 15. Summary of responses for Item 29 of the virtual representation questionnaire

Questionnaire	**Type of Task**			
Item	**Group**	**Std. Deviation**	**Individual**	**Std. Deviation**
Item 29	4.8	0.9	3.8	1.3

Post-Task Group Interviews. In the interview teachers and students had several comments about the systems. We used thematic analysis (Braun & Clarke, 2006) to analyse the data. Specifically, we clustered the participant comments in three categories related to learning performance, immersion and subjective experiences. Below, we present the three categories that provide further user insights.

Students' learning performance with REVERIE VP:

- Close simulation of reality where people can freely navigate and explore space.
- Natural speech-based communication which made easier for students to explain things and elaborate their thoughts that would take more time if they had to type.
- It aided recognition (students could identify their classmates or friends) and allowed expression of emotions with the change of the tone of voice.
- Students felt more comfortable presenting their ideas in-front of an audience disguised behind an avatar.

Students' emotional immersion with REVERIE VP:

- Being able to use their voice allowed them to recognise others and to change the tone of voice allowed the expression of emotions.
- The spatial audio, especially in an open space like the EU parliament felt like "in real life" and contributed to the feeling of immersion. However, the absence of

control over its use often resulted in noise which was distracting and disengaged students from the experience.

- Having their avatar and customising it to their liking allowed users to effectively express their personality and their emotional state (e.g. using funny accessories allowed expressing a jolly mood).
- Being able to facially puppet their avatars allowed them to effectively communicate with others (e.g., to know when it is their turn to speak).

Teachers' subjective perceptions:

- On REVERIE VP teachers raised the need for better control of the class. This includes guiding groups of students through tasks; follow a group discussion and individual user progress; coordinate turn taking; being able to talk to individual users; allowed tools for online note-taking and be able to give accurate feedback.
- The section of Edu-Simulation where students had to post their presentations (named Q&A) appeared disconnected, and it was difficult to follow the student activity.

There was a consensus among teachers that the educational scope of the systems is similar. They suggested combining the best features (e.g. 3D environment and access to information and documents) of the systems into a hybrid platform.

Students' subjective perceptions:

- Students found the lack of interaction with objects in REVERIE VP unsatisfactory. They thought that for students to be engaged and motivated, they should be able to fully interact with the virtual environment.
- Students mentioned that the asynchronous text communication on Edu-Simulation was distracting and limited their effectiveness when working on the educational tasks.

6. DISCUSSION

Objective Assessment. About the objective assessment results show the following about the variables measured.

- *Completion rate.* The high user completion rate cannot be solely attributed to the design of the systems. As both systems have an average Technological Readiness Level (TRL) (*Technology readiness levels (TRL)*, 2014), it was necessary to train our users at the beginning of each session. Hence, it can be said that both systems effectively enabled participants to complete the assigned tasks, but this can be partly attributed to the set-up of the experimental study.

- *User Errors*. In all experimental conditions, we encountered two type of errors: (1) interaction errors and (2) system design errors. Both type of errors made it difficult for participants to complete the assigned tasks. In Edu-Simulation, system design errors prevented two of the participants to fully complete the individual task.

- *Total Time:* A possible explanation for why students spent less time on Edu-Simulation than REVERIE VP could be due to different emotional reactions to the educational tasks. If students disliked the tasks on Edu-Simulation, it is possible that they spent overall less time than on REVERIE VP.

- *Rating/Voting.* The pattern of voting on both systems shows that it was difficult for students to vote using the available voting mechanism on Edu-Simulation. However, the voting mechanism on REVERIE VP was straightforward to use, which in turn, enabled them to vote for all of their peers.

- *Assessment of learning.* A possible explanation for the learning results could be found in the modalities used by each system. The modalities used by REVERIE VP (e.g., spatial audio and avatars) encouraged communication between students and teachers which resulted in a higher productivity of ideas (in the form of claims) compared to Edu-Simulation. This effect is independent of the type of educational task, which shows that participants were equally immersed (temporal immersion) in the educational activities. It is possible, therefore, that students might have retained information about multiculturalism after the study was well over. However, because of the prototypical nature of the systems, we did not measure any retention effects (either long-term or short-term) after the completion of the study. Our results support the main hypothesis of the study (see **H1**), at least referring to the ability of students to produce ideas on each system.

- *Amount of Immersion:* Neutral emotions represent the majority of emotions experienced by participants on both systems. It is not practical to consider these findings as evidence that something is amiss with REVERIE VP. Although, it is known that positive emotions facilitate learning, their presence is not required for students to learn (D'Mello, Lehman, Pekrun, & Graesser, 2014). Neutral emotions, therefore, may represent a state where students were neither distracted by negative activating emotions (e.g., anger or anxiety) or disengaged by negative deactivating emotions (e.g., boredom) (Pekrun, Goetz, Titz, & Perry, 2002) and did what they needed to do: learn about multiculturalism. On the other hand, students were more affected by negative emotions in Edu-Simulation. They perceived both educational tasks quite negatively on Edu-Simulation (mean Group = 15.55% and mean Individual = 16.67%) which can also explain the significant finding for total time. As students did not enjoy completing tasks on Edu-Simulation, they spent overall less time compared to REVERIE VP. Regarding attention, we believe that because participants had to read and produce large amounts of text on Edu-Simulation, they were more attentive to the educational tasks compared to REVERIE VP.

Subjective Assessment. The subjective assessment shows the following about each of the questionnaires:

- *Spatial Immersion Questionnaire.* The high recognition rates show that both teachers and students had a good understanding of the virtual space. It is possible that the increased attentiveness of participants to the educational tasks (see Table 6), facilitated some sense of spatial presence in the VE. As a result, participants were able to recognise and attach meaning to the majority of the objects.

- *Usability Questionnaire (Teacher Only*): The significant difference for interface quality can be attributed to problems teachers experienced with the user interface of Edu-Simulation. Teachers explained in the interview that the user interface of Edu-Simulation made it difficult to follow the activity of their students. The asynchronous chat communication and some sections of the platform which appeared disconnected prevented them from properly following what each student was doing in both educational tasks.

- *Cognitive Accessibility (CoA) Questionnaire (Teachers Only):* Beginning with the REVERIE VP questionnaire and Item 1 (*"The educational activities in the Virtual Parliament were too simple*"), teachers explained in the interview that the system does not provide the necessary tools to control their class (e.g., coordinate turn taking). Because the individual task required teachers to manage every single student, they may have thought that the group task was simpler than the individual task. The results for Item 18 *("I did not have to think hard about what I was doing in order to respond in the virtual parliament")* align with this finding. It is possible that teachers felt that micromanaging the students required more cognitive effort than giving them the freedom to discuss a question in a group. Finally, the results for Item 16 (*"I did not need help to properly navigate in the virtual parliament"*) show that teachers required a different degree of assistance to navigate the VE in different educational tasks. In the group task teachers did not have to navigate in the virtual space. They had to wait for their students to finish discussing a topic in different areas of the parliament. As a result, they may have thought that they needed less help navigating the virtual space in the group task than in the individual task. Continuing with the Edu-Simulation questionnaire, the results for Item 19 ("*The platform responded appropriately when I am confused or overloaded with information (e.g., by allowing me to create my own To-do list"))* suggest that teachers thought that the system responded differently when they were confused or overloaded with information in different educational tasks. In the group task teachers had to use extensively an asynchronous text chat to communicate with their students. The inability to communicate in real-time with their students might have created an impression that Edu-Simulation responded less appropriately in the group task than the individual task.

- *Students' version (Usability and Cognitive Accessibility):* Overall, students rated the usability of Edu-Simulation higher than REVERIE VP. One possible explanation is because of prior familiarity with e-learning platforms (e.g., Moodle). Then, students thought that the tasks completed on REVERIE VP had fewer demands from their working memory. It is possible that because REVERIE VP simulated a real-world scenario (e.g., a class debate), it was easier for students to understand the structure of the educational tasks compared to Edu-Simulation. Then, students experienced significantly higher emotional responses when completed tasks on REVERIE VP than on Edu-Simulation. These results corroborate with our findings for immersion and learning. Students may have thought that REVERIE VP is a more suitable environment to debate about multiculturalism. Also, students may have perceived the educational tasks on REVERIE VP as more fun and enjoyable than on Edu-Simulation. The absence of significant differences for type of task shows that the pre-learning activity teachers carried out in the laboratory was effective in distilling an understanding to students about the topic of multiculturalism. Then the significant results for the Edu-Simulation questionnaire follow the findings for teachers. The main effect of task on item 1 *("The educational activities in Edu-Simulation was too simple")* show that students found the group activity more difficult than the individual activity. As their teacher's students may have felt that the asynchronous chat communication in Edu-Simulation had a detrimental impact in their experience of the group task. The results for Item 2 *("The ways to communicate in Edu-Simulation (e.g., text and images) are enough to immerse me in the educational scenario. I wouldn't like to use a different environment (e.g., more game-like)")* show that students prefer Edu-Simulation over a more game-like environment for the individual task than the group task. If students found the group educational task more complex than the individual educational task, then it is to be expected that they would prefer a different environment to complete it than Edu-Simulation.

The above findings support the hypotheses **H2**. Also, these findings show that only the emotional participation (emotional immersion) of students in the educational activities was sufficient for successful learning to occur (see Table 4). Emotional immersion is a core element in the following principles of the presence pedagogy:

- #Principle 3: Capitalise on the presence of others

REVERIE VP uses multiple tools (e.g., different avatars for teachers and students) to ensure that students can benefit from the expertise of all participants in the learning experience. This facilitated the development of affective relations that led to high emotional responses about the learning experience.

- #Principle 4: Facilitating interactions and encouraging community

REVERIE VP uses multiple tools (e.g., puppeted avatars and spatial audio) to facilitate interaction and collaboration between group members (peers and teacher). This multimodal interaction facilitated group creation and stronger relations between group members that led to high emotional responses about the learning experience.

- #Principle 5: Support distributed cognition

REVERIE VP offers a safe and open virtual space where participants can interact. This open space further facilitated interaction between the group members that led to high emotional responses about the experience. For example, students were able to walk to different areas of the virtual environment to discuss the requirements of the group activity.

- *Virtual Representations Questionnaire (Teachers).* Teachers rated the virtual representations used in REVERIE VP slightly above average (with a higher score indicating a higher degree of satisfaction). It is interesting, however, that they gave the highest score for the tour guide ECA. Teachers perceived the visual qualities of the tour guide better than the other qualities. It is possible that teachers thought that the visual qualities of the tour guide were suitable for the assigned tasks.

- *Virtual Representations Questionnaire (Students).* The significant effect for visual qualities shows that students may have thought that some work has been done in the area of communication between avatars, but the overall fidelity of REVERIE VP is not "quite there yet" to create enhanced spatial immersion. As students explained in the interview the spatial audio contributed to the feeling of participation (e.g., in the discussions between the group members). However, the absence of control over its use often resulted in noise (e.g., because of students talking on top of each other) which disengaged students from the experience. As a result, it may have been easier for students to build a setting for the learning narrative of each educational task on Edu-Simulation. Finally, a possible explanation for the main effect on Item 29 *("The virtual guide helped me to understand more about the EU Parliament (areas, operation, etc.)")* is that students paid more attention to the tour in the group task and hence, found it more helpful than the individual task.

7. DESIGN RECOMMENDATIONS

Based on the findings of the empirical study reported above, we generated a list of design recommendations for optimising the user experience (including immersion) in REVERIE VP. To prioritise the recommendations, we used a custom nine-point scale (0 = not important, 8 = extremely important) inspired by the planning poker agile method (Calefato & Lanubile, 2011). These recommendations are highly actionable and situated to CVEs and hence, can be implemented in similar systems. About REVERIE VP, we deemed these recommendations are necessary for future versions of the system. In total we have identified ten important recommendations that we present in layman terms below:

1) Design for emotional immersion (priority = 8)

Successful dialogic learning scenarios are based on three pedagogical principles - capitalise on the presence of others, facilitate interactions and encourage community and support distributed cognition.

2) Merge the two systems into a hybrid platform (priority = 7)

Extend the architecture of the REVERIE framework and update the design of the virtual parliament to merge it with Edu-Simulation. Consider a hybrid system which combines the following features:

Features from REVERIE VP:

- 3D environment
- Virtual representation with an avatar
- Navigation in the world
- Speech with aided control

Features from Edu-simulation:

- Access to information and documents
- Grouping users
- Textual communication
- Public and private communication
- Dialogue log.

3) Allow teachers to control the ECA (priority = 6)

Teachers should be able to turn on/off the tour guide ECA as needed. When the ECA is not needed, do not include it in the VE.

4) Provide teachers with full control of their classes (priority = 8)

Teachers should have full control of their classes. Try, empowering them to:

- Control the navigation of students. An “auto” navigation option can automatically navigate students to an area of the VE selected by the teacher.
- See what their students see. A “surrogate avatar” option can enable teachers to temporarily take control of a student’s avatar in the VE. This way teachers can effectively moderate each session (e.g., advice students on how to use the system).
- Group students into teams as needed. The participants’ list menu could include a secondary tab where teachers can create and add students in teams.

5) Give control of avatar facial puppeting to users (priority = 6)

Offer users control of their avatar facial puppeting so they can adapt it to their needs. For example, users may decide that for a given discourse fully multimodal communication is not appropriate or even desirable.

6) Provide more interaction opportunities in the VE (priority = 5)

Provide users with more interaction opportunities in the VE. For example, users could select an object in the VE using their mouse to examine it closely.

7) Use system-assisted navigation when a collision is imminent (priority = 5)

Integrate a proximity alert to warn users that they are about to collide with another avatar or object. When the proximity alert activates, the automated navigation system takes over and navigates the user safely around a potential collision.

8) **Give control of the spatial audio to teachers (priority = 7)**

This is about giving control over the use of spatial audio to teachers and information about their students' conversational behaviour. Consider:

- Allowing to mute/unmute individual students as required
- Allowing to mute/unmute group of students as required
- Allowing to enable/disable groups of students to talk to each other
- Offering private audio communication with selected students
- Providing summary information about overlapping speech (e.g., students who frequently talk on top of each other)

9) **Provide visible alerts in the RAAT tool (priority = 4)**

Enable users to allow/deny permission for the RAAT tool to use the computer's camera and microphone through separate alerts displayed at the centre of the browser's window.

10) **Worlds within a world (priority = 8)**

Place at least one separate room within the main VE for teachers to use as needed. For example, teachers may wish to have a brainstorming session in the room with students before entering the main world for the main educational activity to start.

8. CONCLUSIONS AND FUTURE WORK

The study demonstrated the potential of REVERIE VP as an educational tool to enhance online dialogic learning experiences. The analysis of the data showed the following. First, REVERIE VP results in a more positive impact in generating ideas about multiculturalism, compared to Edu-Simulation. This effect is independent of the type of educational task (individual vs group) which shows that students were equally immersed (temporal immersion) in the learning narratives. Second, REVERIE VP facilitated some sense of presence in the virtual Parliament in the form of *emotional immersion*. However, it was not sufficient to enable participants to create a more viable space *(spatial immersion)* for the learning narratives compared to Edu-Simulation. Third, teachers perceived the CoA of the educational tasks similarly across the two systems. However, in specific CoA dimensions (e.g., complexity of the task) they rated the group task higher than the individual task on REVERIE VP. On the other hand, teachers perceived the group task as more cumbersome than the individual task on Edu-Simulation. In terms of usability, teachers perceived the quality of REVERIE VP interface better than Edu-Simulation. Finally, teachers suggested combining the best features of both systems into a single platform. The subjective experiences of students show they perceived the usability of the two systems similarly. It also shows that they did not perceive the educational tasks differently on REVERIE VP. This clearly shows the effectiveness of the pre-learning activity teacher conducted prior to each session with the systems. The students' experiences on REVERIE VP re-iterate the need for a hybrid platform. Given the positive impact on perception, there is the prospect of improved learning outcomes in real-world scenarios. However, the REVERIE framework does not offer virtual representations and the necessary tools to manage learning material. Using REVERIE VP to support real-world online learning scenarios requires integrating the REVERIE framework with an existing e-learning platform, such as Edu-Simulation. Finally, the students' subjective experiences on Edu-Simulation follow the teachers; they show a clear preference for the individual task over the group task.

The first avenue for future work is to implement the above recommendations and re-evaluate the impact of the system using alternative dialogic learning scenarios. A hybrid system alone can have a positive impact on all evaluation metrics, including learning performance. Participants in the current study indicated that a hybrid system could deliver significant improvements in a plethora of online learning scenarios compared to each system alone. Second, because REVERIE VP was not designed as a cloud application, it was difficult to evaluate the user's experience and learning performance in an actual class environment. Although the prototype was taken to schools in the UK, it was not used during class time. Asking schools to evaluate a hybrid system as part of their curriculum can provide deep insights into the real-world impact of multimodal CVE's. These insights refer to how a multimodal CVE can affect long-term and short-term knowledge retention as well as the student experience. Finally, the commercialisation of immersive displays (e.g., Oculus Rift) holds the potential to enhance the educational impact of REVERIE VP. With HMD users can immerse deeper in the educational scenario, which can have a significant impact on how they perceive the system and the educational tasks. As previous evaluations of such immersive set-ups are few, we believe that it is a suitable area where future development and evaluation work of the REVERIE project should be directed.

9. REFERENCES

Adams, R. (2007). Decision and stress: cognition and e-accessibility in the information workplace. *Univers. Access Inf. Soc., 5*(4), 363-379. doi:10.1007/s10209-006-0061-9

Al-Ajlan, A., & Zedan, H. (2008, 21-23 Oct. 2008). *Why Moodle.* Paper presented at the 2008 12th IEEE International Workshop on Future Trends of Distributed Computing Systems.

Apostolakis, K. C., & Daras, P. (2013, 1-3 July 2013). *RAAT - The reverie avatar authoring tool.* Paper presented at the 2013 18th International Conference on Digital Signal Processing (DSP).

Bai, H., Richard, G., & Daudet, L. (2015, 18-21 Oct. 2015). *Geometric-based reverberator using acoustic rendering networks.* Paper presented at the 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA).

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. doi:10.1191/1478088706qp063oa

Bronack, S., Sanders, R., Cheney, A., Riedl, R., Tashner, J., & Matzen, N. (2008). Presence pedagogy: Teaching and learning in a 3D virtual immersive world. *International journal of teaching and learning in higher education, 20*(1), 59-69.

Burdea, G., Richard, P., & Coiffet, P. (1996). Multimodal virtual reality: Input- output devices, system integration, and human factors. *International*

Journal of Human-Computer Interaction, 8(1), 5-24. doi:10.1080/10447319609526138

Calefato, F., & Lanubile, F. (2011). *A Planning Poker Tool for Supporting Collaborative Estimation in Distributed Agile Development*.

Coopman, S. J. (2009). A critical examination of Blackboard's e-learning environment. *First Monday, 14*(6).

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29*, 153-170. doi:http://dx.doi.org/10.1016/j.learninstruc.2012.05.003

de Freitas, S. I. (2006). Using games and simulations for supporting learning. *Learning, Media and Technology, 31*(4), 343-358. doi:10.1080/17439880601021967

Economou, D., Doumanis, I., Bouki, V., Pedersen, F., Mentzelopoulos, M., & Georgalas, N. (2015, 19-20 Nov. 2015). *Edu-simulation: A serious games platform designed to engage and motivate students*. Paper presented at the 2015 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL).

Education, I. (2017). Engage (Version 0.2). Retrieved from http://immersivevreducation.com/engage-education-platform/

Egenfeldt-Nielsen, S. (2011). *Beyond Edutainment Exploring the Educational Potential of Computer Games*: Lulu.com, 2011.

Fechteler, P., Hilsmann, A., Eisert, P., Broeck, S. V., Stevens, C., Wall, J., Zahariadis, T. (2013). *A framework for realistic 3D tele-immersion*. Paper presented at the Proceedings of the 6th International Conference on Computer Vision / Computer Graphics Collaboration Techniques and Applications, Berlin, Germany.

Frijters, S., ten Dam, G., & Rijlaarsdam, G. (2008). Effects of dialogic learning on value-loaded critical thinking. *Learning and Instruction, 18*(1), 66-82. doi:http://dx.doi.org/10.1016/j.learninstruc.2006.11.001

Fruhling, A. L., & Lee, S. M. (2005). *Assessing the Reliability, Validity and Adaptability of PSSUQ*. Paper presented at the AMCIS. http://dblp.uni-trier.de/db/conf/amcis/amcis2005.html#FruhlingL05

Georgiou, Y., & Kyza, E. A. (2017). *A design-based approach to augmented reality location-based activities: Investigating immersion in relation to student learning*. Paper presented at the Proceedings of the 16th World Conference on Mobile and Contextual Learning, Larnaca, Cyprus.

Hajhosseiny, M. (2012). The Effect of Dialogic Teaching on Students' Critical Thinking Disposition. *Procedia - Social and Behavioral Sciences, 69*, 1358-1368. doi:http://dx.doi.org/10.1016/j.sbspro.2012.12.073

Huotari, K., & Hamari, J. (2012). Defining gamification. 17. doi:10.1145/2393132.2393137

Inc, F. E. (2017). FrogPlay Game-based learning platform Retrieved from https://www.frogeducation.com/

Isabwe, G. M. N., Moxnes, M., Ristesund, M., & Woodgate, D. (2018). Children's Interactions Within a Virtual Reality Environment for Learning Chemistry. In T. Andre (Ed.), *Advances in Human Factors in Training, Education, and Learning Sciences: Proceedings of the AHFE 2017 International Conference on Human Factors in Training, Education, and Learning Sciences, July 17-21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA* (pp. 221-233). Cham: Springer International Publishing.

Kensinger, E. A. (2004). Remembering emotional experiences: The contribution of valence and arousal. *Reviews in the Neurosciences, 15*(4), 241-252.

Lewis, J. R. (2002). Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. *International Journal of Human–Computer Interaction, 14*(3-4), 463-488. doi:10.1080/10447318.2002.9669130

Machajdik, J., Stöttinger, J., Danelova, E., Pongratz, M., Kavicky, L., Valenti, R., & Hanbury, A. (2011, 13-15 Sept. 2011). *Providing feedback on emotional experiences and decision making.* Paper presented at the IEEE Africon '11.

Nielsen, J. (2012). Usability 101: Introduction to Usability. Retrieved from https://goo.gl/m3mDbF

Norman, D. (2013). *The design of everyday things: Revised and expanded edition*: Basic Books (AZ).

Paul Cairns, A. C., A. Imran Nordin. (2014). Immersion in Digital Games a Review of Gaming Experience Research *Handbook of Digital Games* (pp. 339-361): Wiley Blackwell.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist, 37*(2), 91-105.

Russell, T. L. (1999). *The no significant difference phenomenon: As reported in 355 research reports, summaries and papers*: North Carolina State University.

Ryan, M.-L. (2001). *Narrative as Virtual Reality: Immersion and Interactivity in Literature and Electronic Media*: Johns Hopkins University Press.

Sayette, M. A., Cohn, J. F., Wertz, J. M., Perrott, M. A., & Parrott, D. J. (2001). A Psychometric Evaluation of the Facial Action Coding System for Assessing Spontaneous Expression. *Journal of Nonverbal Behavior, 25*(3), 167-185. doi:10.1023/a:1010671109788

Seeman, L., & Cooper, M. (2015). Cognitive Accessibility User Research. Retrieved from https://www.w3.org/TR/coga-user-research/

Shan, K., Guo, J., You, W., Lu, D., & Bie, R. (2017, 7-9 June 2017). *Automatic facial expression recognition based on a deep convolutional-neural-network structure.* Paper presented at the 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA).

Technology readiness levels (TRL). (2014). HORIZON 2020 – WORK PROGRAMME.

Toulmin, S. E. (2003). *The uses of argument*: Cambridge university press.

University, A. (2017). iLearn [online]. Retrieved from https://moodle.bl.rdi.co.uk/login/index.php

Valenti, R. (2017). SDK Accuracy Calculation and Accuracy Levels. Retrieved from https://sightcorp.zendesk.com/hc/en-us/articles/200618248-SDK-Accuracy-Calculation-and-Accuracy-Levels

Wall, J., Izquierdo, E., Argyriou, L., Monaghan, D. S., Connor, N. E. O., Poulakos, S., Mekuria, R. (2014, 27-30 Oct. 2014). *REVERIE: Natural human interaction in virtual immersive environments.* Paper presented at the 2014 IEEE International Conference on Image Processing (ICIP).

Zizza, C., Starr, A., Hudson, D., Nuguri, S. S., Calyam, P., & He, Z. (2017). Towards a Social Virtual Reality Learning Environment in High Fidelity. *CoRR, abs/1707.05859.*

Highlights:

- A multimodal CVE fosters student learning performance in dialogic scenarios.
- Emotional immersion is essential for successful dialogic learning scenarios.
- A successful dialogic learning scenario is based on three pedagogical principles.
- Gamification of the task is sufficient for successful dialogic learning scenarios.
- Merging a multimodal CVE with an LMS is essential for real-world applications.