Interactive Learning Environments
Vol. 25, No. 07, 2017, 1–28

RESEARCH ARTICLE

Virtual Worlds vs Books and Videos in History Education

Kiran Ijaz[a*], Anton Bogdanovych[b] and Tomas Trescak[b]
[a] *University of Sydney, NSW, Australia*
; [b] *School of Computing, Engineering and Mathematics, Western Sydney University, NSW, Australia*
(*Received 16 November 2015; final version received 27 February 2016*)

In this paper we investigate an application of virtual reality and artificial intelligence as a technological combination that has a potential to improve the learning experience and engage with the modern generation of students, "digital natives", who grew up with Google and Facebook and do not know the world before the Internet. We have our own personal experience in educating modern students and can certainly relate to some of the postulates and strongly believe that new ways of engaging the students must be developed. To address this need, we have created a virtual reality replica of one of humanity's first cities, the city of Uruk and populated this city with artificial intelligence controlled 3D avatars, which re-enact everyday life of ancient Sumerians in the period around 3000 B.C. Our hypothesis is that by immersing students into this environment and allowing them to learn by browsing through it and interacting with its virtual citizens can be more engaging and motivating than simply reading the corresponding history text or watching an educational video. To confirm this assumption, we have designed a study with three groups of students. One group was given a historical text about Uruk and everyday life of its citizens (created by our subject matter experts), the second group was shown a documentary video on Uruk and the third group was immersed into virtual Uruk and engaged into interactions with its virtual inhabitants. The outcomes of the study suggest that not only did people in the third group provide much more positive qualitative feedback about the learning experience, but they also showed a better comprehension of the study material by performing (on average) 20% better than the first two groups on the mini-exam that was conducted as a part of this study.

Keywords: virtual learning environments;agent-based learning environments; digital natives; artificial intelligence

1. Introduction

Traditional learning approaches often suffer from an inability to entertain the students and, consecutively, an inability to motivate them to learn. Reading textbooks, for example, sometimes fails to engage modern students and they often regard the traditional learning setup as boring and ineffective (Prensky 2007). On the other hand, young students spend a large amount of their time watching movies and playing video games and do not consider these activities being boring. We advocate that rather than eliminating videos and games from the classroom - it is more effective to use this technology as a facility to improve learning motivation and satisfaction.

*Corresponding author. Email: kiran.ijaz@sydney.edu.au

Further in this paper we show how to build educational systems based on the game-based learning paradigm (Prensky 2007). In particular, we discuss the technological details of creating educational virtual reality environments (virtual worlds) specifically targeted at improving students' academic performance and motivation. One of the key elements of success of such virtual worlds is the use of artificial intelligence controlled avatars (virtual agents) capable of engaging with students and providing them with interactive mentoring. Next, we provide the rationale for using virtual worlds in education, discuss the usefulness of virtual agents and describe a technological solution that combines a number of contemporary artificial intelligence techniques that simplify the creation of virtual agents. Through a case study (the city of Uruk 3000 B.C.) we illustrate an example of a virtual reality environment that utilises virtual agents for teaching ancient history. In order to measure the effectiveness of such virtual environments we have conducted a user study, in which learning in a virtual environment has been compared with learning from a text document or from an educational video. The study provides important evidence in favour of using virtual agents and interactive virtual worlds as a technological combination capable of improving students' motivation and academic performance.

2. Background

Virtual Worlds is a technology which resembles video games and has similar potential to provide learning with entertaining experiences. Virtual Worlds inherit qualities from games like interactivity, immersion, rich graphics to engage and motivate students in learning tasks. They also offer opportunities to collaborate irrespective of the physical location and time of the day, to support gestures, emotions and non-verbal communication. Virtual Worlds are not designed for one specific purpose and can be applied to any context as opposed to video games. General purpose Virtual Worlds give more flexibility to design the learning environment while providing a similar degree of entertainment and immersive experience to students.Virtual Worlds can be differentiated from video games as they do not often provide a clear script or narrative. In Virtual Worlds, learners themselves build knowledge through investigating, analysing, interpreting, problem solving and evaluating in an immersive environment, where they may not have pre-scripted instructions (Dawley and Dede 2014). Unlike computer games, Virtual Worlds allow users to create their personal world, interact with this virtual space and other users in it, rather than just interacting with a preprogrammed environment.

2.1. *Virtual Worlds in Education*

Several studies confirm the importance of using Virtual Worlds in education and highlight the beneficial nature of this technology in regards to improving learning outcomes. A detailed taxonomy in this area has been discussed by (Duncan et al. 2012). Enhancing learning through games, collaboration or interaction in virtual worlds has increasingly been a topic of many international consortiums and projects. The GAPS[1] consortium studies how computer games can help children and young students learn and teaches them to think like professionals in various jobs by providing them set of simulated tools (Chesler and Shaffer 2013, Steinkuehler and Sasha Barab 2012).

Important favourable findings were collected in Harvard within the River City project (Gibson 2007), suggesting that the next generation of games and simulations has potential to dramatically improve students' motivation and educational outcomes, as well as

[1] `http://edgaps.org/gaps/` (last visited 04/2014)

generating new insights into the nature of learning. In particular, the River City project has showed that the use of the Virtual Worlds technology had the most dramatic influence on engaging low performing students. In a science class, where many low performing students have previously indicated that they gave up on learning science the use of virtual world turned them into high achievers. The study also indicates that students who used the Virtual World grew equally regardless of their academic history, while the performance of the students who learned in a traditional classroom was often negatively influenced by poor academic history.

Further practical evidence was provided with a case study conducted by the Loyalist college in training future customs officers on the virtual reconstruction of the US-Canadian border (Loyalist College 2009). This project simulated customs and immigration training to provide real life border interactions among visitors and customs staff. The study reports that use of a simulated training program showed significant improvements with 56% course completion success in 2007, to 95% at the end of 2008.

Despite high potential for this technology as an educational tool there are also downsides. In traditional classrooms, teachers set the agenda and structure of the learning task towards attainment of learning outcomes. But in Virtual Worlds students are often left to themselves without a teacher and a structured learning approach. Virtual Worlds as a successful learning medium have been discussed previously (Clarke and Dede 2005, Hudson and Degast-Kennedy 2009) whereas a well designed structured learning approach is an area which needs further investigation (Martens and Maciuszek 2013). Some efforts have been made to provide structured learning in virtual learning environments in Sloodle (Daniel et al. 2009) and Lams (Cram and Eade 2010) projects. These learning environments are based on using complementary external resources such as text documents, web forums, online polls, notice boards, etc. Quest Atlantis (Barab et al. 2007) provides non player characters(NPC) with static narrative to achieve fixed learning goals. Having fixed objectives in such learning environments also leads to difficulties of assessing whether certain goals have been met (Benvenuti et al. 2010).

This free exploration with unlimited choice and endless information can make learning difficult, students are likely to get distracted from the learning goal due to the possibility of engagement in various other activities. For instance, a student may have been busy exploring a new shopping mall meanwhile his peers have learnt the given task and moved on. Other students may need assistance to choose between several options to complete a task and may lose interest in learning if such assistance is not provided.

Consequently to take full advantage of Virtual Worlds as suitable learning environments, we require a structured learning approach similar to a traditional classroom to guide students. In such learning environments, the role of teacher certainly becomes important. One of the several roles of a teacher is to guide students through the learning activities to attain desired outcomes. A well-structured learning approach can ensure learning task completion and focus learning activities while maintaining students' motivation in virtual worlds.

2.2. *Virtual Agents*

One possible way of providing structured learning in virtual worlds is through implementing a classroom setting similar to traditional learning. But traditional classroom setups in virtual worlds do not fully utilise the features provided by these immersive environments. Virtual Worlds encourage "learning by doing", parallel processing for various tasks at hand and independent investigation, which are rarely fully supported in traditional classroom settings (Benvenuti et al. 2010). Nevertheless, implementing classroom settings in virtual learning environments will have the advantage of better communication and coordination among students over the traditional classroom environment.

One of the benefits of using Virtual Worlds, in particular for students with poor academic history - is the possibility for everyone to study at their own pace. This is difficult to achieve in traditional classrooms as there is normally one teacher who has to deliver the material at the pace suitable for an average student. So, to fully enjoy the benefits of Virtual Worlds, support "learning by doing" and allow every student to study at their own pace, a traditional classroom approach might not be suitable. Providing a personal tutor for every student might be a possibility, but this is a very costly option.

An alternative cost effective approach to providing a personalised, engaging and focussed learning setup in Virtual Worlds is to use virtual agents[1]. Virtual agents can take the role of tutors inhabiting the study environment and provide reliable support in the virtual learning environment irrespective of time of the day or time differences in the case of international classes. According to (Woolf et al. 2010, Arroyo et al. 2011) virtual agents in learning environments can not only present information and structure learning, but can also improve communication between different students, which has proven to be a beneficial factor in terms of motivating and engaging students and results in better learning outcomes (Gibson 2007). Virtual agents do not have to fully replace the teacher, but can become an additional factor of engagement for students and help to structure their learning, while still allowing students to study at their own pace.

At present it is difficult to supply virtual agents with similar communication capabilities as human teachers. Successful communication is a complex concept as outlined in Littlejohn (2001) and a life long learning process. Virtual agents in learning environments must be able to communicate with students to share knowledge and build a bridge of understanding. Many existing agent-based systems have ad-hoc and developer specific communication mechanisms. Although these agents can achieve impressive tasks, participating agents are often based on context specific assumptions and can not accommodate tasks outside their capabilities Dignum and Greaves (2000). Human communication can be defined in terms of two categories: verbal and non-verbal communication. Both types of communication categories play vital roles to exchange knowledge. Currently, there is no technology available for an agent that fully supports fluent communication with humans using both verbal and non-verbal means effectively, but some progress has been made. Examples of virtual agents capable of dynamic interaction with human users include Johnson and Rickel (1997), Solutions (2001), Wallace (2004b), Corraze (1988), Lester et al. (2013). Agent communication procedures are still far from perfect and lack the features we aim to attain in this paper. Most of the verbal communication for virtual agents is pre-scripted and based on Natural Language Processing Wallace (2004b), Johnson and Rickel (1997). These verbal communication systems also depend on the virtual agent's designer to continuously extending their knowledge base either manually or automatically. Despite the limitations virtual agents have high potential and even in their present limited form can facilitate student learning.

2.3. *Virtual Worlds for History Simulation and Cultural Heritage*

Using 3D virtual environments to reconstruct lost sites of high historical and cultural significance has become very popular during the last decade. The primary focus of the majority of such approaches is on reconstructing ancient buildings, objects and even entire cities that are partially or completely destroyed at present. "Rome Reborn" (Bernard et al. 2013) recreates a historically accurate reconstruction of a large part of ancient Rome. The reconstruction includes virtual models of nearly 7000 buildings and the terrain of the city in the time of Constantine the Great in A.D. 320 (Bernard et al. 2013).

[1]Embodied graphical characters controlled by a computer program

Visitors to virtual heritage sites, similar to "Rome Reborn", are normally able to browse through digital models of the heritage objects and inspect them from different angles and proximity. They can also select a particular heritage object (i.e. Roman Colosseum) and closely explore its architectural details. While such an approach allows general audiences to examine the architectural details of the heritage site and the corresponding objects, it is certainly limited in regards to learning ancient history in such environments. Just having simulated buildings is not enough to understand how this site has been enacted in the past and how ancient people lived their daily lives there.

What is missing for converting a 3D simulation into a learning environment - is the use of avatars. Populating a historical site with virtual agents that behave similar to the ancient citizens that used to occupy given site could have a much higher potential in regards to preserving historical data and allowing students to learn from it. Through the use of Artificial Intelligence (A.I.) the virtual agents can absorb the relevant historical knowledge and become the knowledge carriers and tutors.

This potential, however, has not been properly developed. While there is a lot of technological progress in regards to systems that employ individual pedagogical (Rus et al. 2009, Nash and Shaffer 2013) or motivational (Baylor 2011) agents, the vast majority of large scale historical simulations instead of using advanced virtual agents normally focus on the so-called "virtual crowds" (Thalmann 2013). Such crowds normally consist of a large number of avatars dressed as local citizens of the reconstructed site. Instead of being capable of advanced action and multimodal interactions with users, individual virtual agents in these crowds often have very limited functionality. The state of the art in combining crowd simulation and 3D heritage can be observed on the example outlined in (Maiim et al. 2007) where a virtual City of Pompeii is populated with a large number of avatars that walk around the city avoiding collisions. In this work the avatars are simply moving around and are not involved into historically authentic interactions. So their presence is not much different from the presence of moving objects and they only extend the atmosphere of the culture simulation.

The few attempts that employ rich individual models of virtual agents for capturing some of the cultural attributes are mostly focused on building *culturally adaptive agents*. Such agents try to adjust their personality to better fit with the culture of the user. Some recent works are concerned with developing agents that strongly adhere to the given culture. An approach to encode some of the global cultural dimensions into the agent behavior is presented in (Aylett, Vannini, Andre, Paiva, Enz, and Hall 2009). Illustrating some attributes of a given culture through the use of rituals is outlined in (Mascarenhas et al. 2009). While being successful in highlighting a range of some cultural dimensions, the aforementioned works do not rely on a coherent formal model of culture and do not provide a general formal mechanism for culture preservation.

The use of virtual agents and virtual worlds in history and culture education is scarce. One of the most prominent projects in this field is Virtual Singapura project (Jacobson et al. 2009) that was designed to implement inquiry based learning in virtual environments where secondary school students have to solve an historical epidemic spread in 19th-Century Singapore (Kennedy-Clark and Thompson 2011). A framework for storytelling multi-agent architecture has been developed to facilitate less technical educators to apply teaching and learning in virtual learning environments and allowing it to be applied to various educational topics (from medicine to science). While being immersed into a historical setting of Singapore the work of (Kennedy-Clark and Thompson 2011) is not really focused on learning history of the place, but represents a quest related to epidemic spread that students have to solve. Moreover, the availability of virtual agents in virtual Singapura is rather limited, they are mainly static, predominantly offer communication with users and do not engage into complex interactions with the surrounding environment, its artefacts and other avatars.

Virtual guides are another popular direction in deploying avatars in historical simulations. These agents help users to navigate the virtual world that either preserves a given culture in a traditional form of a museum (Oberlander et al. 2008) or recreates the actual site as a virtual environment (Palace Museum and IBM 2009). The later case, the "Forbidden City" project, better utilises Virtual Worlds technology in regards to the heritage preservation. The project aims at simulating the culture of ancient China and replicating the one square-kilometre palace grounds called The Virtual Forbidden City. Similar to many other historical simulations a significant effort has been put into a realistic recreation of the architecture of the city, while a much smaller effort has been placed on the development of virtual agents. The agents in the Forbidden City are supplied with very limited "intelligence". Their actions are highly scripted and their ability to interact with the users is limited to scripted monologues. The number of available agents is also quite low and the majority of those act purely as guides rather than as virtual inhabitants of the city living their daily lives there.

3. The City of Uruk 3000 B.C.: Technological Solution

Existing work is quite limited in regards to building historical simulations that can be successfully employed for the purpose of history education. In such simulations it is important to have a large number of virtual agents playing different social roles, actively using surrounding objects, interacting with other agents as well as communicating with students. Modern video games are a good illustration in regards to the potential of having such simulations, but the cost of developing video games is enormous. For example, the estimated cost of developing Crysis 3, one of the popular modern video games, is $66 Million (Gauder 2013). It is hard to imagine such a level of spending when it comes to historical simulations for educational purposes, so populating a historical environment with virtual agents needs to be automated.

Aiming to achieve cost saving, some researchers do not model their societies at the level of individual agents, but employ "virtual crowds" (Thalmann 2013). While such crowds essentially consist of a large number of virtual agents, designing a crowd normally comes down to designing a few individuals and then replicating them a desired number of times with slight modifications so that the crowd appears to be diverse (Thalmann 2013). In regards to agent behaviour, classical crowds simulation techniques predominantly rely on automated algorithms for large scale obstacle and collision avoidance and individual agent behaviour is rarely complex enough to illustrate various aspects of daily life of the reconstructed society. Further in the paper we show how through simulation of physiological needs and motivations together with personality traits we can enrich classical crowd simulation with much more sophisticated simulations of human behaviour. Furthermore, employing genetic methods for inheriting personality traits and appearance characteristics together with connecting virtual agents with formalisations of social roles and social norms allows for a similar level of complexity in crowd based simulations as seen in commercial video games.

Next we emphasise the key technological aspect of our approach. It is being explained on the example of a case study (the city of Uruk 3000 B.C.), which aims at building a historical simulation of one of humanity's first cities.

3.1. *Approach*

Our general approach to building virtual heritage environments is as follows. We use the Second Life ®[1] Virtual World for recreating significant heritage objects from the results of archaeological excavations and available written sources. Participants are embodied as avatars and can freely move within the virtual world, interact with other participants and sometimes change the virtual world itself (very often in a dynamic manner using in-world building facilities).

The 3D virtual world is used for both preserving the culture and for teaching the resulting culture to the visitors. It is accessible by two types of participants: “visitors” and “experts”. Visitors participate in the environment with the aim to learn about the given culture through immersing in the virtual world and interacting with its virtual inhabitants. Experts are a key element in culture preservation. Through embodied interactions with other experts in the virtual world they share their knowledge and refine the appearance of the heritage environment, validate the correctness of the reconstructed buildings and artefacts, as well as help to refine the behaviour of virtual agents. As the result of the joint work of historians, archaeologists, designers and programmers the resulting heritage site is recreated in the virtual world and populated with virtual agents that look and behave similar to how the actual people that used to live in the given area looked like and behaved. Through the Virtual Institutions technology (Bogdanovych 2007) the agents can engage into complex interactions with other agents and humans, while adhering to the social norms of the reconstructed culture.

3.2. *Building the Virtual World*

Objects and artefacts in this Uruk environment were recreated by designers based on the photographs and measurements, while the buildings were designed on the basis of drawings and sketches produced by the archaeologists using data from archaeological excavations. The modelling of the buildings, animals and artefacts was conducted by our project partners in the Federation of American Scientists. The process of modelling the virtual heritage site followed the following three phases:

(1) The site is mapped by archaeologists using the existing site plans.
(2) A virtual three-dimensional model of the buildings is constructed using the given measurements.
(3) Each building model is positioned in the virtual world according to the plan of the site.

To illustrate how the resulting 3D models of the buildings were obtained, Figure 1. outlines these phases on the example of constructing the Ziggurat in Uruk.

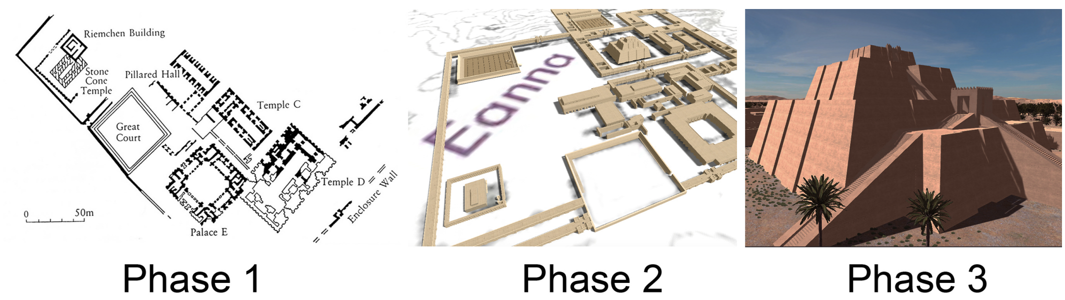

Figure 1. Three phases of building the virtual environment.

[1] `http://secondlife.com` (last visited 04/2014)

Figure 2. shows a screenshot of the resulting city. The city is located on the bank of the Euphrates river and is surrounded by a large wall for protection. The process of building the city is fairly standard, so we do not present further details, but mainly focus on the details of how it was populated.

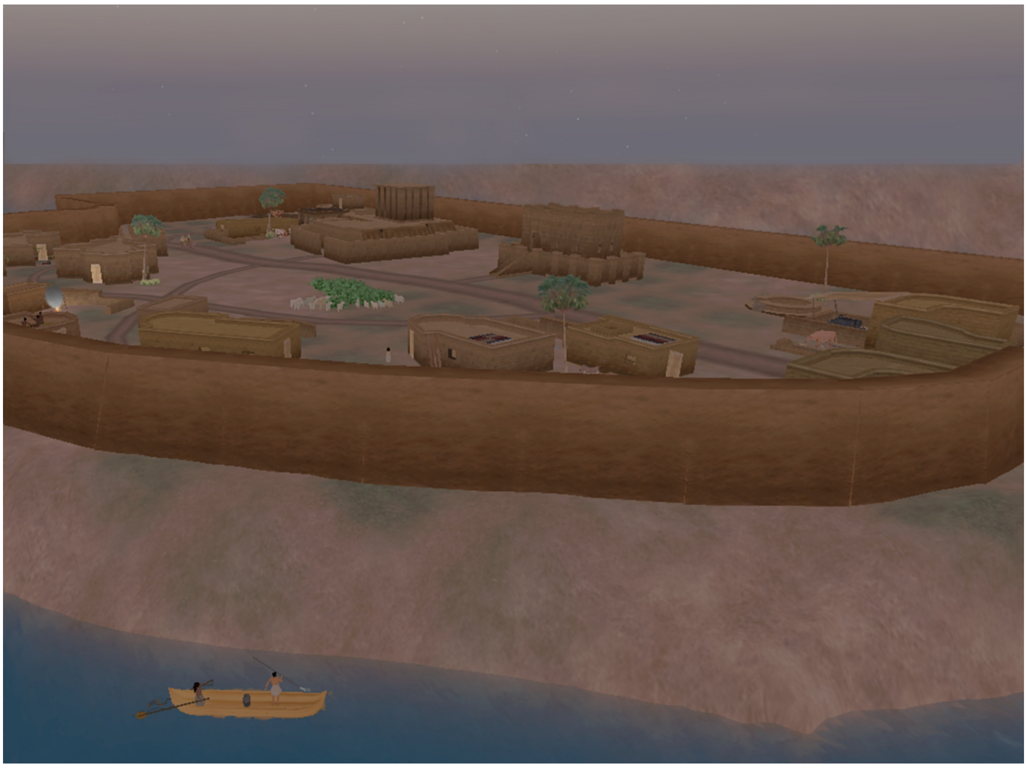

Figure 2. The City of Uruk, populated with buildings, objects and agents in the virtual world.

The main principle behind populating the environment with virtual agents is as follows. First we must formalise the social norms that govern and restrict agent behaviour and identify social roles for all participants. The social norms (or institution) will determine which agents with which roles are allowed to participate in which activities. For each activity the institution will define the protocol specifying how it must be enacted by which roles.

Once the institution has been defined, next we need to create the base population, which means manually designing at least two avatars for each social role. Additionally, for each avatar in the base population we need to specify its personality.

Finally, the base population is being used to automatically generate a large agent society of the desired size. The social role distribution will determine how many agents of each role we need. The appearance and personalities of the generated agents will be inherited from the "parents" - agents for the corresponding role from the base population. To introduce some variety into the generated crowd the appearance and personality features will undergo mutation (randomisation of some of the attributes).

Once the population of the city is generated - each agent will be driven by its simulated physiological needs (hunger, fatigue, etc.). The planner module will be sequentially matching each of the need and the personality type of the given agent with the sequence of actions that leads to satisfying the most pressing need. The underlying institution will help the planner to restrict the search space to only look at actions relevant for the particular role and will make sure that the resulting plan is historically accurate and adheres to the social norms of the reconstructed society.

The agents in the Uruk simulation represent a slice of Uruk society among which are fishermen families, priest, king and a number of workers (i.e. pot maker, spear maker). The agents can sense changes in the environment state, which result in them updating their beliefs accordingly. The current implementation features fishermen families where men's daily routines include sleeping, eating, fishing and chatting. The females do house work, sleep, eat bring water from the well and go to markets. The king agent walks around his palace and invites students to ask him about his ruling strategies. The priest agent conducts a prayer in the temple, accepts gifts and explores the city. Other agents represent various workers: pot makers, spear makers, etc. Those workers produce goods, exchange goods with one another, attend the prayer and in their spare time explore the city, provide information to students and simulate social interactions with other agents. Figure 3. shows some selected agents in Uruk.

Figure 3. Selected Uruk Agents: Wife, Priest, Fishermen, King.

Our agent architecture is based on the BDI (Belief Desire Intention) model and functions similarly to the Jack Intelligent Agents platform (Howden et al. 2001). To meet our requirements, we have created our own agent library (VIAgents) for developing agents that follow our formal model. In this section, we present the overview of key *modules* of the VI agent model[1].

3.2.1. Visualisation Module

This module facilitates creation of new avatars through specification of visual features, e.g. height, head shape, foot size. It stores the avatar definition in the form of visual parameters. Using randomisation, we are able to generate crowd of agents with unique appearance, although sometimes with unexpected results (anomalies) like unrealistically large noses or eyes. To deal with such anomalies, we use our genetics module as discussed later.

3.2.2. Physiology Module

Giving agents ability to sense physiological needs (hunger, thirst, fatigue and comfort) solves two important problems in crowd simulation. First, it allows agents to automatically generate goals based on current physiological state, e.g. they rest when they are tired, drink when they are thirsty, eat when they are hungry. Second, it solves the problem with the uniformity of performed actions, when agents execute actions at different times, increasing the believability of the simulation. This is achieved by assigning different *decay multipliers* for physiological needs for each agent. As a result, agents' physiological needs decay at various rates, reaching threshold values at different intervals.

3.2.3. Personality Module

Often, agents have various means to fulfil their goals. For example, when an agent gets hungry it can obtain food by trading on the market, going to fish, stealing or begging. One of the options is to select the possible action randomly. In our model, we are more

[1] A video illustrating the use of these techniques is available at: `http://www.youtube.com/watch?v=dU8ljJFNop4` (last visited 04/2014)

consistent by selecting current action depending on its compatibility with the agent's personality and current emotional state. Furthermore, performing an action affects the emotional state of the agent and affects its future actions, which increases the believability of the simulation (Loyall 1997). This module implements the well known OCEAN model (John and Srivastava 1999) for agent personality and OCC model (Bartneck 2002) for an agent's emotional state.

3.2.4. Genetic Module

Since the three above modules (visualisation, physiology, personality) use value parameters to represent an individual, we can randomise these parameters to generate agents with unique appearance and behaviour. But, as mentioned earlier, this can lead to producing unwanted results, anomalies. Moreover, if we desire to generate members of the same ethnicity with distinctive features (e.g. asian eyes), it is not possible, since randomisation does not respect such features.

To gain control over generation of of the unique crowd, we employ procedures from genetic reproduction. In our approach (Trescak et al. 2012), first, we define the base population that represents all the ethnic groups that are required to appear in the simulation. For each such ethnic group, we design at least two specimens, where each of them maintains their distinctive features, and we randomise the others. We apply the same to the definition of agent personalities and physiological decay factors. Second, we encode all current values into a genetic structure called chromosome. Using approaches from genetic programming we combine two parent chromosomes using a *crossover* operator to produce the new child's chromosome, used to recreate this individual. As a result, children carry visual features from their ancestors, respecting the appearance of the ethnic group they belong to. Also, in case greater variety is needed, when combining chromosomes we can mutate a number of genes to further increase novelty in generated avatars.

Figure 4. PotMaker Planning Example: AddWater → MakeClay → Work → Trade.

3.2.5. Sensing, Planning, Learning

Agents, using the Second Life API are capable of sensing changes in the environment, e.g. object position update. These changes in the virtual world might result in agents updating their goals. In order to achieve their goals they can use either static or dynamic planning. Static plans are instructions prescribed by an agent programmer and can't be changed at runtime. In the case of dynamic planning an agent can sense its current state in the environment and can react to environment changes and re-evaluate its current plan. Rather than having a complete recipe provided for every situation the agent can encounter - the agent is simply given the list of possible atomic actions and has to find a way of combining those to reach its goals. As mentioned previously in this section, if more than one plan to fulfil the goal exists, agents decide based on their personality and current emotional state. Figure 4. shows the example of dynamic planning, where potmaker role, from the list of atomic actions (Work, AddWater, WaterAvailable, MakeClay, Trade) constructs a plan to obtain food by creating a pot and trading it for food on a market: AddWater→MakeClay→Work→Trade→HaveFood. This search is enabled by the fact that every action an agent can perform is annotated with pre-conditions and post-conditions, so the agent can start with its desired state (like "Trade") and then search for an action that has this desired action as a post-condition. This process continues until the agents finds a sequence of actions that lead from its current state to the desired state. This search is conducted using the classical "backwards state-space search" method (Russel and Norvig 2003).

Furthermore, agents can be trained to respond to certain situations in a desired manner. They can learn at multiple levels of abstraction as described in (Bogdanovych et al. 2008). The Virtual Institution structures the learning process and provides the context for learning. Through imitation the agents can learn new plans for various goals. Such plans are represented as recursive-arc graphs (similar to recursive decision trees) with probabilities being assigned to the arcs of the graph as the training continues. We have also created a method for training the agents to perform different verbal behaviour in various situations. The method of modifying the AIML rules via learning is described in (Ijaz et al. 2011).

3.2.6. Interacting with the Environment

In order to convincingly simulate the daily life of the ancient people it was important to enable agents to use objects in the environment (i.e. take a spear, jump on a boat and go fishing). Intelligent object use is one of the aspects of A.I. that hasn't received appropriate attention. We implemented a fully-fledged solution to object use for the heritage context. We developed a designated library that provides a set of classes allowing agents to identify an object in the virtual world, attach it to the default attachment point, play a certain animation (i.e. rowing) associated with a given object, wear an object that represents a clothing item, detach the clothing item, drop an object to the ground and detach the object and hide it in the avatar's inventory.

Figure 5. illustrates the degree of complexity in object use that we faced. Here two agents perform the fishing task in a collaborative fashion and are capable of using the fishing basket, spear, boat and rowing paddles. The interactions with these objects include taking the objects, carrying them around, dropping objects, putting one object on top of another, moving while inside an object, etc.

Agents automatically search for interactive objects and navigate to them using the artificial potential fields algorithm (Ge and Cui 2002) for obstacle avoidance. This method was implemented from scratch taking into account the dynamic and unpredictable nature of Second Life.

Figure 5. Wide use of Objects in the Environment: Tools, Clothes, Boats, Fire, etc.

3.2.7. Interacting with Humans

All our agents can chat with human visitors on topics, related to the heritage environment through the instant message and chat mechanisms provided by Second Life. In order to participate in a conversation our agents employ the ALICE chat engine (Wallace 2004a) based on the AIML language (Wallace 2004b). Each agent uses a number of AIML files that represent what can be seen as a common sense database. Additionally to this database every agent is supplied by personalised AIML files that reflect on its personality and the data relevant for its role within the virtual society.

While the use of Alice and AIML is a common technique, what is normally missing in traditional approaches is the consistent integration of an agent's environment into verbal communication with users. We rely on the model of environment-, self- and interaction-awareness from (Ijaz et al. 2011) which is integrated using the method presented in (Ijaz et al. 2011). This allows our agents to be able to converse with users about the objects located in the environment, tasks they have to achieve, current state, expected next states, objects being used, goals, etc.

For the sake of enhanced believability, our agents are supplied with a programming solution dealing with idle gaze behaviour. When the agents are moving around their gaze is not fixed, but is continuously changing. The agent would shift its gaze between objects and avatars depending on the level of its interest in those. It will predominantly follow the movements of avatars approaching the agent at close proximity. Without a proper gaze model that agents would be always looking ahead while walking and their head would not move. In contrast, with our gaze model the agents continuously look around and turn their head in search for interesting objects, when such an object is found - the agent keeps the gaze focus on it until a more interesting object comes along.

4. Experimental Evaluation

To test the learning effectiveness of using virtual worlds and virtual agents in history education, we have conducted a user study. We aimed at comparing the novel way of teaching history in a virtual environment with more traditional approaches, in order to see what are the pros and cons of using interactive virtual worlds in history education. At this current stage of the study it was not our objective to test the effectiveness of supervised forms of learning, like comparing a lecture with a virtual world populated with pro-active virtual tutors. This is a rather complicated topic and such a study can be quite subjective as individual lecturers differ greatly in their ability to engage with the audience. Therefore, in this work we have instead focused on individual forms of learning, where a student learns in an un-supervised manner and with no support from a teacher. The most common approach currently being used in schools and universities for such individual learning is to rely on written sources (e.g. history books). But there is also a growing trend to include educational videos as a part of the learning curriculum. Therefore, in our study we compared three forms of learning: learning from a text document, learning from an educational video and learning from an interactive virtual environment.

The essence of the study was to select three very similar groups of participants and present each group with completely identical information, but use different forms of presenting this information. The topic of the study was history education and each of the groups was learning about the same set of historical facts (about the city of Uruk in ancient Mesopotamia) from either a text document, an educational video or an interactive virtual world. In this way we could maintain the same content and focus the study on testing whether the way in which the information is presented may affect learning outcomes.

For this study we have selected 60 students and separated them into three groups, each with 20 participants. First group of participants, which we call the "Traditional Group", was advised to read a history text describing the facts about our case study: City of Uruk (3000 B.C.) and its inhabitants. It is important to note that this text was widely illustrated similar to an average history book. Second group "Video Group" watched a video describing the city of Uruk. In this video the same text that was given to the Traditional Group was read by an actor and illustrated with the corresponding video footage. The footage was edited using relevant documentary videos and some of the missing elements were illustrated using the content from the virtual Uruk simulation. Finally, participants in the third group, named the "Virtual Group", were asked to visit the virtual city of Uruk to have an interactive learning experience there. The virtual Uruk, if fully explored, would present the same information as in the previous two groups, but this information will be pro-actively obtained by walking around, interacting with objects and having conversations with avatars. With the help of our subject matter experts we made sure that all three groups would be working with identical study content.

4.1. *Research Hypotheses*

Based on our research objective of evaluating the learning of history in 3D Virtual Worlds in comparison to video and text based learning, we had the following hypotheses:

- **H1**: Students from the virtual group will be more effective in terms of their learning outcomes in comparison to the students from the traditional and video groups.
- **H2**: Studying in the virtual world will be more appealing to the test subjects than learning from a text document or an educational video.

Next we will show how these hypotheses that were initially produced purely on the basis of our intuition were tested through a user study and what were the results of this study. But first we have to describe some of the specifics of using the virtual Uruk platform for individual learning and will discuss some limitations that were identified along the way.

4.2. *Specifics of the Virtual Group*

The participants in the Traditional Group and the Video Group received their study material in quite a structured form. Both the text document and the video had a well defined structure that was defined by the subject matter experts. In the case of the text document, they were given clear guidelines that the document is what they have to study from and that everything in the document will become the basis for the mini-exam conducted later on. Same holds for the educational video. All study participants were well familiar with these forms of study material and it was perfectly clear to them how to make sure that they can process all the necessary information that has been given to them.

The case of the Virtual Group was more complicated. The setup of the study with the virtual group is such that each study subject was given an avatar to represent him/her as a visitor in the city. With this avatar, each student was able to immerse into the virtual world and interact with objects and avatars in the virtual city. The information that a study subject had to absorb was scattered around a large area in the virtual space and was presented in various forms. Study subjects had to learn from moving around and looking at buildings, objects and people walking past them and performing daily chores. To successfully complete the study, all participants were instructed to interact with the system and virtual humans living in the city. The learning task in the virtual city of Uruk was based on two given activities:

- Visiting and exploring the virtual city: Participants could learn about the city by navigating, observing and exploring the various places in the virtual city. To learn more about major buildings and artefacts, these places have *'More Info'* pointers to give further information. These information pointers contain the same facts as were given to the traditional group in the Uruk text document.

- Conversing with Virtual Humans: Participants could also converse with virtual humans present in the city. These virtual humans were aware of their daily routines, tasks and interactions with other participants and were able to report on their actions and the surrounding environment.

So the ways of accessing the required information in the Virtual Group were quite diverse. Some information could have only been extracted from talking to particular agents. Other types of information could only be acquired from reading the text that was associated with "More Info" objects located in various places in the virtual city. To have a fair comparison with the other three groups it was important for Virtual Group participants to be exposed to all of this information and not to miss anything during their virtual Uruk tour. We had an intuitive feeling that leaving people to navigate the virtual world completely unsupervised may cause significant navigation problems for novel users and could result in some participants not having observed all the available information. An obvious choice that could help to ensure that all of this information has been presented to the study subject is to provide them with additional assistance. But in our study we aimed at comparing un-supervised learning methods. In this light, some forms of assistance could bring bias and invalidate the study outcomes. In order to come up with the right strategy of presenting the students with all the required information and at the same time avoiding bias we have conducted an additional qualitative pilot study.

4.3. Pilot Study for the Virtual Group

Given the aforementioned problems with the Virtual Group, we were concerned that some form of supervision in this group will be hard to avoid. To test this assumption we decided to conduct a qualitative study and investigate whether there is a significant difference in having participants left unsupervised in the virtual city or to have them supervised in some way. In order for the study results to be useful and for this supervision to have no bias over the learning outcomes the supervision could only be related to the technical problems with navigating the virtual world and with making sure that all participants are presented with the whole range of available learning material (e.g. explore the entire city rather than only a small portion). We strictly followed the principle that no supervision in relation to the study material itself was given to participants.

Our initial idea was to test the difference between participants who are left completely unsupervised and those who received help with navigation, but then we realised that help with navigation can be achieved in two different ways, so we had to test the following three methods of learning:

(1) The "unsupervised" method, where participants receive instructions on the objectives of the study, how to navigate the environment and what are the key forms of acquiring information in the virtual world. Then they were left to investigate the environment by themselves and were not provided with any assistance.
(2) The "supervised and controlled" method is where participants were given a brief explanation on how to navigate the virtual world, but also had an assistant sitting next to them and providing navigating help when participant required it or showed signs of confusion. If the participant was indicating that he/she is finished with exploration, but the environment was not fully examined - the assistant was

advising the participant to explore the missed areas and helped to navigate there.

(3) The "supervised" method is where participants are not navigating the virtual world directly, but have an assistant sitting next to them and operating the controls. They navigate by issuing verbal instructions (e.g "turn left", "approach this building").

To see how each of the aforementioned methods affects learning outcomes we have conducted a pilot study with 12 additional participants, who were evenly separated into three groups (four people in each pilot group). All study participants had no prior knowledge of ancient Mesopotamia and all were undergraduate students. The assistant was present in all three group (including the unsupervised group) and was sitting next to each participant. In the case of the unsupervised group the assistant was not allowed to provide any help and was simply observing the study subject. All studies were conducted with one study subject at a time. After exploring the city each of the test subjects took part in a mini-exam (the same exam that was later used in the main study). Apart from the exam an informal interview has been conducted with each of the participants. The key focus of the interview was on the navigation in the virtual world and whether it had any influence over their learning.

4.3.1. Pilot Study Results

To analyse which of the three methods of learning is the most suitable one for the main study we used the results of the mini-exam as well as the impressions of the assistant.

One of the key outcomes of the study with the **unsupervised** group (as indicated by the assistant) was that all participants did not properly explore the city. In all four cases either some important areas were left unexplored or some agents were not interacted with. The assistant has also witnessed that three of the four participants experienced frequent need for technical assistance (e.g navigation, re-login) during the sessions and this has also been confirmed by them during the interview. Although feedback also suggests that all four of them found the study interesting and motivating. Due to no supervision during the study, some students faced technical issues and did not fully achieve the learning objectives and scored low on the exam (and average of 14 out of 40 points possible).

Next was the study with the **supervised and controlled** group. Here the assistant was using the navigation controls following verbal instructions from the study subject. The exam results showed that the learning outcomes were better than in the unsupervised session (an average of 27 out of 40 points possible). But the interview feedback indicated a high degree of dependance on researcher of all study subjects, which was undesired in this study. The second major issue with this method was a demonstrated decrease in interactivity and immersion of participants during the session. This could also badly influence the aim of providing the interactive experience proposed by this work. The subjects in this group seemed least engaged in the experience and were keen to finish it quickly.

The third study was with the **supervised** group. Unlike the second group the participants navigated the environment by themselves. During the first few minutes all participants required a lot of technical help, but later they were able to master the controls and only asked occasional questions. In all four cases participants needed to be directed to the areas that were left unexplored and to talk with the some of the virtual agents that they missed. The exam results in this group were quite encouraging with the highest average score achieved across all the three groups (30 points). The interview has showed that all participants liked the study a lot and were quite engaged and motivated to learn.

Figure 6. shows the comparison between the three methods of learning in our pilot study and the average marks achieved by each group. Based on the aggregation of feedback and exam results the supervised method was selected for the Virtual Group in the main study as the most productive and the one that is most aligned with our aims.

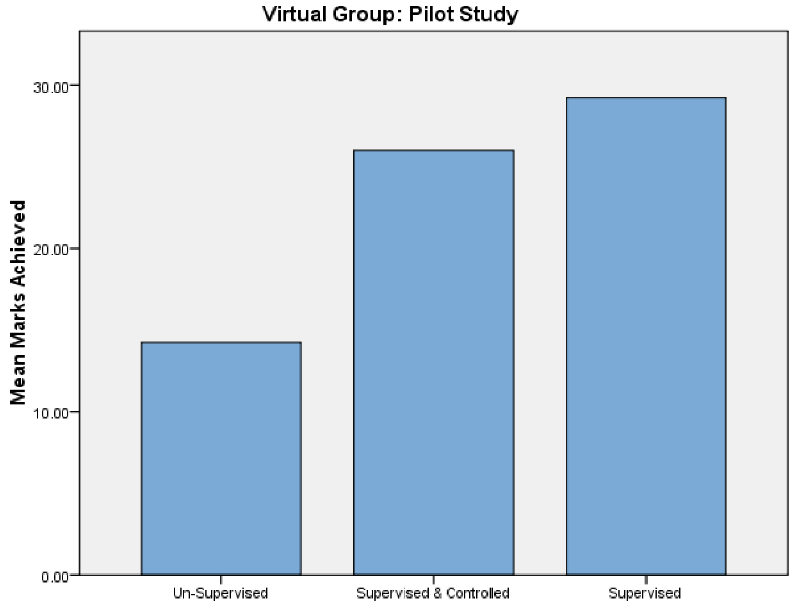

Figure 6. Pilot Study: Comparison of three Cases

The assistant's presence during the virtual city interaction and the interviews with test subjects helped us to identify technical problems that needed to be fixed, allowed us to spot poorly formulated questions and update them, as well as to modify the scenarios according to user feedback and simplify city navigation with additional signs placed in the virtual world.

4.4. The Main Study

After conducting the pilot study and selecting an appropriate method of testing the subjects in the Virtual Group the main study was launched. The objective of this study was to collect both quantitative and qualitative data about the usefulness of using virtual worlds and believable conversational agents in teaching ancient history. Therefore, we developed a questionnaire that contained two parts: the pre-test part collected the users' demographics and their background knowledge about the city of Uruk. Second part was a post-test questionnaire aiming to test the knowledge gained about the city of Uruk in the three study groups. The pre-test questionnaire consisted of six multiple choice questions about the participant's demographical information and one pre-coded (open ended) question to test previous knowledge of the historical city. The post-test part of the questionnaire aimed at examining the knowledge gained in these user groups (an open-ended questionnaire for participants' feedback). Only people with no prior knowledge about Uruk and ancient Mesopotamia were selected to continue after the pre-test. Other pre-test responses were used to make sure that the groups are comparable. The corresponding demographical factors helped to decide in which of the three groups to place each participant.

4.4.1. Research Instruments

All study sessions in all groups were individual. Before the start of each session a participant was given instructions about the aims of the study and how to interact with the given learning resource. The Traditional Group was provided with an illustrated text document describing the city of Uruk. The document was produced by our subject matter experts and contained information about people, buildings, climate of the city, historical significance, food, agriculture, trade and inventions made in that era. Similar to history books in schools, the document was widely illustrated. The resulting document comprised the same facts as those that were used in the documentary video and in the virtual city of Uruk.

The Video Group was requested to watch a 13 minute documentary video. The narrative of this video was directly taken from the text document used by the Traditional Group. The video footage that was used to illustrate the narrative was taken from existing documentaries about Uruk and ancient Mesopotamia and some of the missing parts were taken from video recordings in the virtual city.

Apart from the general set of instructions the subjects in the Virtual Group were given an extensive briefing about how to navigate around the virtual city. It was explicitly stated to the participants that they must explore the entire city, inspect all objects labelled with a "More Info" sign and talk with all agents who will initiate a conversation with them. The participants were then told that they can request technical assistance from the researcher present in the room. If those instructions are followed, then the knowledge Virtual Group participants would acquire is the same as in the Traditional and Video groups.

The post-test questionnaire (which we also refer to as the "mini-exam") was identical for all participants in all three groups. Its objective was to measure a student's knowledge of historical concepts presented in three different ways. The post-test questions fall under four major topics: a) climate and buildings; b) people/food/animals; c) agriculture and trade; and d) Uruk inventions. The scenarios for our three groups and the content were designed with the help of subject matter experts, who have also verified the study instruments. The post-test feedback and the informal student observations were aimed to gather all possible details that the researcher could collect regarding the student's engagement, motivation and enjoyment during the study.

4.4.2. Participants

The study was conducted in different stages and all groups were not selected and tested simultaneously. To ensure that this does not result in a bias that could possibly affect study outcomes, the study was organised as follows. Each group was formed based on the identical set of requirements (age and gender distribution, no prior knowledge of ancient Mesopotamia were the key criteria) and all participants in each group were selected at the same time and tested during a short period of time. But there was a significant time gap between selecting the study sample for other groups. Given that we have carefully maintained the similarities in the age/gender distribution in each study group, ensured that all participants are undergraduate university students and made sure that all of the selected participants had no prior knowledge of ancient Mesopotamia - we believe that there is no bias caused by not having the studies conducted simultaneously. In order to enforce the sample similarities in each group we have initially selected larger samples and then discarded participants that did not fit the desired distributions.

After the initial pre-test screening, we have selected 20 undergraduate university students with no previous knowledge of Uruk and ancient Mesopotamia to be tested for each study group (60 students in total). The first group worked with a text document (Traditional Group), the second group watched a video (Video Group), while the third group interacted with a virtual world (Virtual Group).

4.4.3. Study Procedure

All the participants were briefed about the study at the start of their respective sessions and then interacted with the corresponding media. No time limit has been enforced on the interaction with the selected media. The subjects in the Traditional Group were not limited in regards to how much time they can spend on reading the test and it was allowed to re-read it as many times as needed. The subjects in the virtual group were also allowed to spend as much time as needed in the virtual environment. As for the subjects in the virtual group, the length of the video shown to the participants was 13 minutes 9 seconds, which was artificially limiting the time of interaction, but all participants were given an opportunity to watch the video multiple times if required.

Immediately after each participant has finished learning about Uruk he/she was given a questionnaire to test the acquired knowledge. We aimed to have a similar setup to formal examination at universities, so we have strictly enforced the time for completing the questionnaire. The time required to fill out the post-test questionnaire was limited to 15 minutes and additionally 5 minutes were given to provide the post-test feedback. Those limitations were identical across the three groups.

4.5. Study Findings

The main objective of the study was to test our two research hypotheses (H1 and H2). To test the H1 we analysed the performance of each of the three study groups in terms of their scores on the mini-exam. We were also interested in seeing whether particular ways of presenting information in a virtual world could result performance variations, so in addition to comparing the Traditional, Video and Virtual groups for performance differences in general, we also wanted to see which ways of presenting information are associated with greater differences. Virtual Worlds participants were able to learn from simply observing objects and avatars in the virtual worlds. Apart from this they could learn from chat interactions with virtual agents as well as from reading text descriptions. Such text descriptions were represented by short text notes associated with some significant objects and labelled with a “More Info” sign (see Fig. 7). We were interested to see whether there is a significant difference in the effect of these different forms of learning. We tested how the information presented through text conversations with agents is different from that presented in the written document or a documentary video. We have also tested how presenting a piece of text inside a virtual world is different from presenting the same piece of text in a large text document. Our initial expectation was to see no significant difference in relation to learning performance in the two latter cases.

Figure 7. Text Note to Describe a School in Virtual Uruk

Text Note to Describe a School in Virtual Uruk Testing H2 was more difficult than testing H1, as engagement is harder to quantify. Therefore, we have predominantly relied on analysing informal feedback given by the participants in each group to simply understand whether there are some remarks about engagement being made and then analyse these remarks. A detailed description of our findings is further elaborated below.

General Comparison: Traditional Group vs Video Group vs Virtual Group
There was a visible difference in performance gained between the Traditional, Video and Virtual group students. The metric to measure any performance variations between three groups was based on marks achieved in the post-test questionnaire. This performance comparison is illustrated in Table 1. The maximum score that could have been obtained in each study is 40 points. A one-way between subjects ANOVA was conducted to compare the students' performance achieved in Traditional, Video and Virtual groups. All the relevant assumptions were tested and no violations were observed. In this study, marks achieved in post-test questionnaire was the dependent variable. There was a high variation in marks achieved in post test at the ($p < 0.05$) level for three conditions [$F(2, 57) = 17.47$, $p = 0.00$]. Descriptive analysis indicates that the mean score for the Virtual Group ($M = 26.11$, $SD = 5.82$) was quite different to that of the Video Group ($M = 16.25$, $SD = 6.31$) and the Traditional Group ($M = 17.65$, $SD = 4.90$). However, the Video Group did not significantly differ from the Traditional group. Surprisingly to us, the results suggest that visual aids like video does not seem to have any significant impact on learning with respect to textual and pictorial representation of the same facts. However, between these two groups the lowest performing student from the Traditional Group has shown slightly better marks achieved (Min = 10) in comparison to the Video Group (Min = 7.5). The highest achieving student in the Traditional group gained marks (Max = 28.75) which are lower than in the Video group (Max = 32.75). These differences, however can be due to individual differences and are not representative.

Table 1. Student's Performance

Comparison: Traditional vs Video vs Virtual Group					
	Students #	Mean	Std.Deviation	Min-Marks	Max-Marks
Traditional Group	20	17.65	4.9	10	28.75
Video Group	20	16.25	6.3	7.50	32.75
Virtual Group	20	26.11	5.8	17.50	38.00

The three groups were compared with post hoc using the Tukey HSD test which provides further in depth analysis on all pair wise contrasts. Table 2 presents all significantly different groups with ($p < 0.05$). The Virtual group significantly differs with the Traditional group ($MD = 8.46$, $p = 0.00$) and the Video group ($MD = 9.86$, $p = 0.00$). Nonetheless, between the Traditional Group and the Video Group no significant difference was observed with ($MD = 1.40$, $p = 0.72$).

Table 2. Virtual vs Video vs Text Group

Multiple Comparisons			
Group(I)	Group(J)	Mean Difference(I-J)	Significance
Virtual	Traditional	8.46	.000
Virtual	Video	9.86	.000
Traditional	Video	1.40	.719

The same trend in these group performances can be observed in Figure 8. Both Traditional and Video groups performed significantly lower than the Virtual group. Whereas, among Traditional and Video groups, not much difference has been observed.

Group formation considered same set of requirements and each group's participants were recruited at the same time for a short testing period. In the pre-test questionnaire,

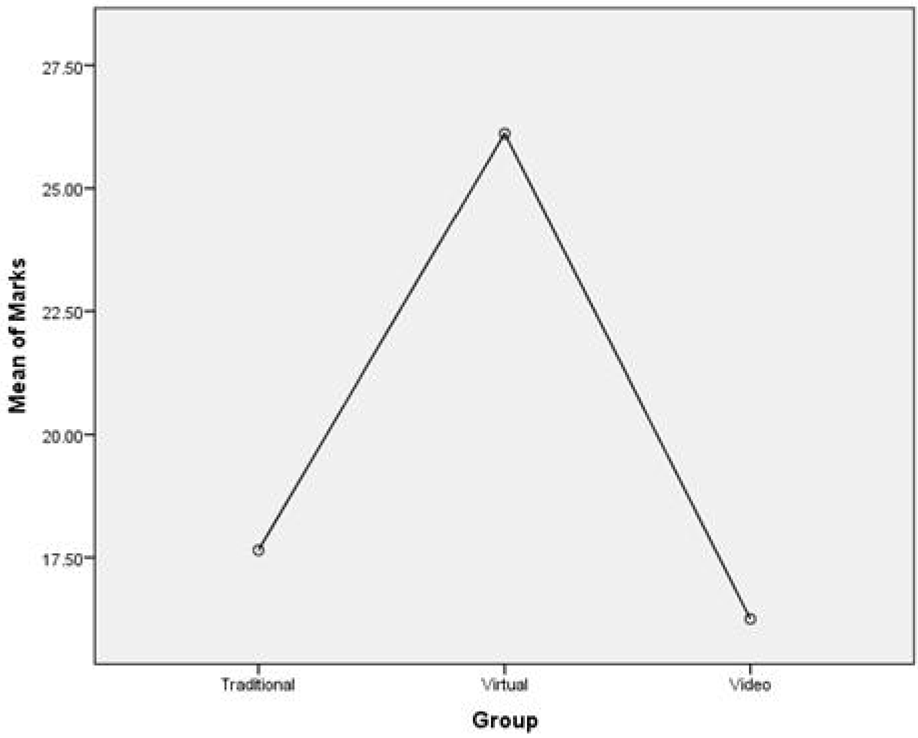

Figure 8. Comparing Students' Performance

students were requested to provide their demographics. We collected participants' information about gender, age, English language ability, game playing experience, qualification, field of study and any previous knowledge of the history of Ancient Mesopotamia. Our main interest was to examine any impact of these factors on student's post-test performance. Further we present the results of the main effect analysis, but only focus on critical factors. Our results revealed that female students (N = 30, M = 22.4, SE= 1.4) performed better than male students (N = 30, M =18, SE= 0.99) with $F(1,58) = 7.4$ and $p = 0.009$ across the three study groups. However, there was no significant difference for performance achieved among the different age groups of participants $F(4, 55) = 0.59$ and $p = 0.674$. Similarly, all other factors showed no significant effect on students' performance. Results suggest that students' age, English background, game playing frequency and qualification do not significantly impact his performance in post- test of this study. We were initially surprised about the lack of impact of the game playing experience. But we can explain this with the fact that all participants were able to request additional help with navigation, so in this way we managed to downplay the game playing experience factor.

Virtual Group - Agents' Conversation vs Reading Text:
During their visit to the virtual city of Uruk, participants had to explore and learn about the different places, buildings, artefacts and daily routines of virtual humans (agents) living in the city. There were three possible ways one could acquire information a) move around and observe; b) chat with virtual agents; c) inspect some significant locations by clicking on the "More Info" button (which results a text note being displayed in front of different places as shown in Figure 7). These notes contained relevant fragments of text from the document given to the Traditional group. Secondly, participants had to learn about the people living in the city through chat conversations with the corresponding virtual agents. The agents were able to explain their daily routines as well as describe their surroundings.

In the post-test questionnaire, we had specific questions about the facts learned by reading the "More Info" notes and another group of questions was targeting facts learned through conversations with virtual agents. A comparison of scores achieved for these two separate sets of questions indicate that learning from conversing with virtual humans was more effective than learning from text descriptions. As shown in Figure 9., most Virtual group participants achieved higher marks for the facts learned in conversations with virtual humans in comparison to "More Info". Admittedly, the minimum score gained as a result of conversations with virtual humans was 8.33% where the minimum performance for text reading was 22.92 %. But, comparison of maximum performance shows that two participants scored 100% for virtual human conversations. Whereas, by reading the text in "More Info" note, only one participant attained a high score and that was only 79.17%. Overall results suggest that on average participants scored 65.42% when they conversed with virtual humans(agents) where average score of learning facts from "More Info" was 48.23%.

While on average interacting with virtual agents resulted better learning, two students in this group gained lower marks than all students reading text from the "More Info" section. As reported by the researcher observing those two students, the most likely reason for it is the fact that these students had quite a lengthy interaction with some of the agents and their conversations often deviated from the main topic of study (e.g. they asked about personal life of these agents rather than talking about rituals they performed or objects they used). Thus, the students were overwhelmed with unrelated information and did not properly remember the important facts.

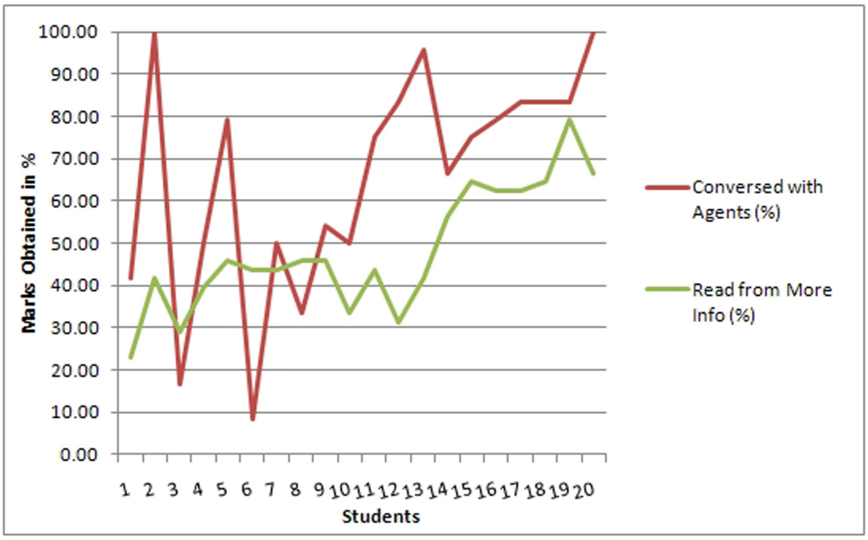

Figure 9. Virtual Human Conversations vs Text Reading in Virtual Worlds

Text Reading in Virtual World vs Traditional Method:

Apart from comparing learning from text descriptions and learning from interactive chats with virtual agents we were interested to test whether learning from a text description in a virtual world is different to reading the same text description as a part of a larger document. The virtual city "More Info" notes had the same text associated with the relevant artefacts as what was outlined in the text document for the Traditional Group. To test the effectiveness of text reading in both groups, we compared the post-test questionnaire for the Traditional and Virtual groups. This comparison has revealed that participants who read text in the virtual city showed better performance. The average score achieved by the traditional group for text based questions was 29.27 %, whereas, virtual group participants showed the better performance by attaining an average score of 48.23 %. Note that in our comparison we were dealing with a set of questions associated with the same text in both groups. The student with the minimum score in the Virtual group attained 22.92% in comparison to the low performing student with 4.17% in the Traditional group. Likewise, the highest score gained among the Virtual group participants was 79.17 % compared to the highest score 60.42% in the Traditional group. A comparison of both groups in regards to text reading is shown in Figure 10. Here we see the marks obtained by each of the students in every group with a clear indication that the Virtual Group has performed better than the Traditional Group.

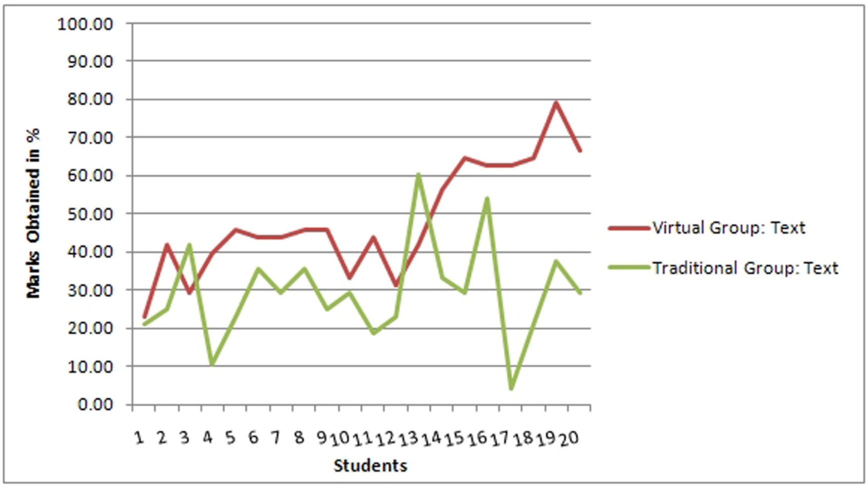

Figure 10. Text Reading in Traditional vs Virtual Group

4.5.1. Testing interest and engagement

One way of testing engagement was to actually ask the users about their motivation to learn and interest in the study in the post-study questionary. As the part of their overall feedback all participants were asked to briefly describe their study experience in the open-ended field of the questionary. As we expected after formulating H2, the reported engagement in the Virtual Group was significantly higher than that in the other two groups. In their open-ended responses 60% of participants have positively commented about their learning experience and indicated positive engagement. In contrast, only 10% of participants in the Traditional Group and 40% in the Video Group made comments that could be associated with positive interest and engagement. Most other comments in the Video Group and the Virtual Group were neutral. No single participant made negative remarks about the study in those groups. But, in the Traditional Group the situation was quite different. Some 20% of participants expressed their negative engagement with the study labelling it as boring and lengthy.

Another measure of engagement is the time spent by participants with the learning materials. It can be assumed that participants who spend more time learning are more motivated to learn and are more engaged in the learning process. Our findings indicated that the average time spent for text reading by traditional group participants was approximately 11 minutes, while in the video group this time is 14 minutes (most participants only watched the video once). Unlike to two other groups, virtual group participants spent considerably more time 37 minutes.

Finally, another way of assessing engagement was to observe the users and their non-verbal and verbal expressions during the study. We had the opportunity to observe this as the researcher was present in each study session. The researcher reports increased enthusiasm and excitement among the Virtual Group participants and rather neutral attitude towards the study by the other two groups. In the Virtual Group sessions, most students seemed to have particularly liked their conversations with virtual humans living in the city. They followed them to observe and inquire about their activities and spend far more time then needed interacting with them. In contrast, Traditional and Video group participants initially took interest in acquiring information about the city of Uruk. However, (35 - 40%) participants in these groups reported that it was hard to maintain interest in the later stages of the study.

4.6. *Discussion of Study Results*

Traditional, Video and Virtual groups were compared for performance differences in learning historical facts about the city of Uruk. Activities devised for three study groups were: traditional text reading to learn historical facts about Uruk from a text document; watching a documentary video to gain the same historical information; and learning historical facts by exploration in the Virtual World, talking with its virtual inhabitants and reading text notes associated with significant objects in the virtual environment. The mini-exam conducted based on post-test questionnaire depicts clear differences in students' performance among the three study groups. Better academic performance has been achieved by the Virtual group over the Traditional text reading and Video groups. Overall, the results of the study, confirm our hypothesis H1 and suggest that on average the students in the Virtual group have acquired approximately 21% higher marks in comparison to the Traditional text reading group and 25% higher marks than the Video group, who learned about Uruk from watching a documentary video .

The results of the study are consistent with our hypothesis (H1) that the studying in the virtual city would results better academic performance. It was initially surprising to see that study subjects who watched the documentary video about Uruk performed a little (4%) worse than the text reading group on the exam. We can explain this result by the fact that reading represents an active form of learning, while watching a video is a rather passive process. Information presented through text and images is also quicker to absorb, which is consistent with our study results. The text that was used as the narrative for the video and as the story-line for the design of the virtual world was processed by our study participants in 11 minutes on average, while the video was processed in 14 minutes and the virtual world was explored in 37 minutes. But despite high efficiency the subjects in the text reading group reported much lower engagement. Some of them have labelled the study as excessively lengthy and boring. In contrast, the subjects who watched the documentary video found the study satisfactory and no single user has reported it to be boring or excessively long. Even more interesting are the results of the virtual group, where test subjects who interacted with the virtual city spent over three times the amount of time of the text reading group, but instead of complaining that the study is long and boring, many of them have demonstrated a high degree of excitement and provided extremely positive feedback. Most of them spent significantly more time than it was required in the virtual city. The most prominent attraction that consumed most of the time were the virtual agents, who allowed students to converse with them and learn about the city through chat conversations with those agents. As the key result of this study, we may assume that learning from an illustrated text document is the most time-efficient form of learning, but is also the least engaging one. Learning in a game-like environment of the virtual world is the least time-efficient, but the most engaging and the most effective form of learning. The subjects were happy to sacrifice their personal time on learning and, as the result, showed much better academic performance on the exam. This high performance, of course, can be due to the fact that they spent much more time on the task, but we believe that the performance is much more strongly correlated with high engagement than the time of study. In support of this assumption, the students who watched the documentary video spent more time on it than the text reading group, but this extra time did not help to improve their performance.

We would like to stress that results of this study do not advocate discarding text books or video documentaries in favour of virtual worlds. What we suggest is to use virtual worlds as additional motivation for students to learn from traditional sources of knowledge. We believe that text books and videos should still remain the primary learning resources and must be used in combination with modern technology.

4.7. *Limitations of the Study*

There is a number of limitations of this study that the reader must be familiar with. One of the key limitations is that it was difficult for us to design the virtual city in such a way that after a quick introduction the users would be able to self-sufficiently interact with it. In our pilot study we tried to develop such a prototype and let the users interact with it in a completely unsupervised manner. But all virtual environments have their own navigation style that requires some time to get accustomed to. Most of test subjects in the pilot study had issues with navigating the virtual world, which had a negative impact on their learning. Furthermore, at this stage it was difficult for us to design the virtual world in such way that subjects who had no issues with navigation were able to effectively observe all relevant information and be exposed to the same amount of knowledge as those who watched a documentary video or read the text document. To address these problems we have decided to introduce a researcher who was able to help the test subjects with navigation and was instructing them to continue learning if some learning critical areas of the city remained unexplored or some agents were not approached for conversation. We believe that our study design helped to eliminate the bias of this additional form of assistance, but this would require further verification in the future.

Another issue that was not a part of this study is whether our results could be associated with the Novelty Effect (Houston-Price and Nakai 2004). The reason why most test subjects had high motivation and engagement with the virtual world could very well be the fact that they have experienced this new form of learning for the first time and like the new idea. It may happen that when they are exposed to such forms of learning on a regular basis the novelty will wear off and the engagement will drop. This, however was very difficult for us to test as such testing would require developing many prototypes for different study topics and disciplines, for which we don't have sufficient resources.

We also have to admit that increased motivation and engagement (which we associate with better academic performance) could be not directly linked with the virtual world itself, but with interactivity that it offers. Text reading and watching videos are non-interactive forms of learning, but maybe organising documents and videos in some interactive way could also result increased academic performance.

Another limitation of this work is that our assessment of learning was conducted immediately after the engaging with the study material, so it can be argued that what has been assessed is the short term impact on learning, as longer term learning effects were not examined in our study.

Finally, we have to explicitly state that it was not the intention of this study to compare learning in a proper classroom environment versus learning in a virtual world. Instead, the idea behind this study was to compare different study materials (e.g. books, videos and interactive virtual worlds), which a single student can use to *independently* study history. Our findings suggest that virtual agents add interactivity and may result greater student engagement, but such (or even greater) engagement is also achievable in traditional classroom scenarios with humans. At present no virtual agent is capable of being as engaging as a teacher or a fellow student can be. Therefore, our findings should not be interpreted as evidence in favour of using virtual worlds and virtual agents to replace traditional classroom instruction. In our study the participants had to read the text or view the video without processing it in any other way. Of course, texts and videos are rarely used in this way in a traditional classroom setup, where students are normally exposed to the combination of various study sources and not only read a text document, but also listen to the presentation of the lecturer, watch videos and discuss the material with fellow students. This mixture of techniques is likely to result greater engagement and better understanding of the study material. As an extension of our work, it would have been useful to compare learning with virtual agents in a virtual world versus learning in a traditional classroom. However, in a traditional classroom there is a large variety of variables that can have a significant impact on the outcome (e.g. how video and text are combined, which engagement practices are used, the personality and professionalism of the teacher, and fellow students, etc.). Therefore, it will be difficult to conduct and objective study that can properly capture all this variety and compare it with interactive teaching practices of artificial agents.

4.8. *Conclusions and Future Work*

Our future work avenues are directly related to the aforementioned limitations of this work. In the future we plan to develop a self-contained video game around Uruk (based on Unity rather than Second Life) and test this game as a part of the study curriculum in primary schools in Australia. Having a self-contained video game would mean that students will require no additional supervision apart from some initial briefing and can use this game to study the history of ancient Sumerians independently. In this way any possible bias related to receiving additional supervision can be completely eliminated. Through the school engagement program we have generated a lot of interest in using our game in the school curriculum within greater western Sydney. Having the game tested with much larger numbers of participants and those who are directly interested in studying the topic and be formally marked on their knowledge by a qualified history teacher will certainly help us to produce interesting results. Furthermore, being able to test it in such a classroom scenario would also help us to investigate the long-term learning impact of using virtual agents and virtual worlds in history education. We also plan to conduct additional studies to determine whether the virtual world and its immersive and highly visual nature are the key elements resulting increased learning engagement and academic performance or whether interactivity is the key factor and similar benefits can also be obtained with interactive text/video resources. Apart from working with subject matter experts, when developing the Uruk game we will also actively engage with professional educators to make sure that our study is educationally sound.

5. Acknowledgement

We would like to thank the Federation of American Scientists (especially Alex Cohen) and Tata Interactive Systems for developing the initial 3D models of Uruk that were used in our simulations and for connecting us with subject matter experts. We are very grateful to Alice Petty and Dan Bracewell for helping us to design the scenarios and produce the text for the Traditional group. We would also like to acknowledge the help of John Debenham and Simeon Simoff with various aspects of this research, including productive discussions of the technical details and advising on the study design.

References

Arroyo, I., B. Woolf, D. Cooper, W. Burleson, and K. Muldner (2011). The impact of animated pedagogical agents on girls' and boys' emotions, attitudes, behaviors and learning. In *Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference on*, pp. 506–510.

Aylett, R., N. Vannini, E. Andre, A. Paiva, S. Enz, and L. Hall (2009). But that was in another country: agents and intercultural empathy. In *AAMAS '09: Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems*, Richland, SC, pp. 329–336.

Barab, S., T. Dodge, H. Tuzun, K. Job-Sluder, C. Jackson, A. Arici, L. Job-Sluder, C. J. R, J. Gilbertson, and C. Heiselt (2007). The quest atlantis project: A socially-responsive play space for learning. *The educational design and use of simulation computer games*, 159–186.

Bartneck, C. (2002). Integrating the OCC Model of Emotions in Embodied Characters.

Baylor, A. (2011). The design of motivational agents and avatars. *Educational Technology Research and Development 59*(2), 291–300.

Benvenuti, L., E.-J. Hennipman, E.-J. R. Oppelaar, G. C. Veer, R. B. Cruijsberg, and G. Bakker (2010). Experiencing and learning with 3d virtual worlds. In J. M. Spector, D. Ifenthaler, P. Isaias, Kinshuk, and D. Sampson (Eds.), *Learning and Instruction in the Digital Age*, pp. 191–204. Springer US.

Bernard, F., A. Dean, K. Dylla, and G. Guidi (2013). Rome Reborn Project. http://romereborn.frischerconsulting.com/.

Bogdanovych, A. (2007). *Virtual Institutions.* Ph. D. thesis, UTS, Sydney, Australia.

Bogdanovych, A., S. Simoff, and M. Esteva (2008). Virtual institutions: Normative environments facilitating imitation learning in virtual agents. In *proceedings of the 8-th International Conference on Intelligent Virtual Agents (IVA 2008), Lecture Notes for Computer Science*, Berlin/Heidelberg, pp. 456–464. Springer.

Chesler, N., A. G. D. C. B. E. and D. Shaffer (2013). Design of a professional practice simulator for educating and motivating first-year engineering students. *Advances in Engineering Education 3*, 1–29.

Clarke, J. and C. Dede (2005). Making learning meaningful: An exploratory study of using multi-user environments (muves) in middle school science. In *Proceedings of the American Educational Research Association Conference*, Montreal, Canada, pp. 1–7.

Corraze, J. (1988). *Les communications non-verbales, 4e d. rev. et corr edition.* Paris: Presses universitaires de France.

Cram, A., L. K. and J. Eade (2010). Using LAMS to structure and support learning activities in virtual worlds. In *Proceedings of the 5th International LAMS Conference*, pp. 75–82.

Daniel, L., J. Kemp, E. Edgar, C. Surridge, and P. Bloomfield (2009). *Games-Based Learning Advancements for Multi-Sensory Human Computer Interfaces: Techniques and Effective Practices*, Chapter Multi-User Virtual Environments for Learning Meet Learning Management, pp. 34–50. IGI Global.

Dawley, L. and C. Dede (2014). Situated learning in virtual worlds and immersive simulations. In J. M. Spector, M. D. Merrill, J. Elen, and M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology*, pp. 723–734. Springer New York.

Dignum, F. and M. Greaves (Eds.) (2000). *Issues in Agent Communication*, London, UK. Springer-Verlag.

Duncan, I., A. Miller, and S. Jiang (2012). A taxonomy of virtual worlds usage in education. *British Journal of Educational Technology 43*(6), 949–964.

Gauder, J. (2013). Crysis 3 cost $66 million to make, can next gen sustain such budgets? GameChup Video Games News at http://www.gamechup.com/crysis-3-cost-66-million-to-make-can-next-gen-sustain-such-budgets/.

Ge, S. S. and Y. J. Cui (2002). Dynamic motion planning for mobile robots using potential field method. *Auton. Robots 13*(3), 207–222.

Gibson, D. (2007). *Games and Simulations in Online Learning: Research and Development Frameworks*. Arlington, VA, USA: Information Resources Press.

Houston-Price, C. and S. Nakai (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development 13*(4), 341–348.

Howden, N., R. Ronnquist, A. Hodgson, and A. Lucas (2001). JACK intelligent agents - summary of an agent infrastructure. In *Proceedings of the 5th ACM International Conference on Autonomous Agents*, pp. 1–6.

Hudson, K. and K. Degast-Kennedy (2009). Canadian border simulation at loyalist college. *Journal of Virtual Worlds Research* (2).

Ijaz, K., A. Bogdanovych, and S. Simoff (2011). Enhancing the believability of embodied conversational agents through environment-, self- and interaction-awareness. In M. Reynolds (Ed.), *Australasian Computer Science Conference (ACSC 2011)*, Volume 113 of *CRPIT*, Perth, Australia, pp. 107–116. ACS.

Jacobson, M. J., B. Kim, C. Miao, Z. Shen, and M. Chavez (2009). *Designs for learning environments of the future: International learning sciences theory and research perspectives*, Chapter Design perspectives for learning in virtual worlds., pp. 111–142. Springer.

John, O. P. and S. Srivastava (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research 2*, 102–138.

Johnson, W. L. and J. Rickel (1997). Steve: an animated pedagogical agent for procedural training in virtual environments. *SIGART Bull. 8*(1-4), 16–21.

Kennedy-Clark, S. and K. Thompson (2011). What do students learn when collaboratively using a computer game in the study of historical disease epidemics, and why? *6*, 513–537+.

Lester, J. C., E. Y. Ha, S. Y. Lee, B. W. Mott, J. P. Rowe, and J. L. Sabourin (2013). Serious games get smart: Intelligent game-based learning environments. *AI Magazine 34*(4).

Littlejohn, S. W. (2001, August). *Theories of Human Communication (with InfoTrac) (Wadsworth Series in Speech Communication)*. Wadsworth Publishing.

Loyalist College (2009). Virtual World Simulation Training Prepares Real Guards on the US-Canadian Border. Available online at http://secondlifegrid.net.s3.amazonaws.com/docs/Second_Life_Case_Loyalist_EN.pdf.

Loyall, A. B. (1997). *Believable Agents: Building Interactive Personalities*. Ph. D. thesis, Department of Computer Science, Carnegie Mellon University, Pittsburgh.

Maiim, J., S. Haegler, B. Yersin, P. Mueller, Thalmann, Daniel,, and L. Van Gool (2007). Populating Ancient Pompeii with Crowds of Virtual Romans. In *8th International Symposium on Virtual Reality, Archeology and Cultural Heritage - VAST*, pp. 109–116.

Martens, A. and D. Maciuszek (2013). *Serious Games and Virtual Worlds in Education, Professional Development, and Healthcare*, Chapter Balancing Instruction and Construction in Virtual World Learning, pp. 15–40. IGI Global.

Mascarenhas, S., J. Dias, N. Afonso, S. Enz, and A. Paiva (2009). Using rituals to express cultural differences in synthetic characters. In *AAMAS '09: Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems*, Richland, SC, pp. 305–312.

Nash, P. and D. Shaffer (2013). Epistemic trajectories: Mentoring in a game design practicum. *Instructional Science 41*, 745–771.

Oberlander, J., I. Androutsopoulos, A. Isard, and G. Karakatsiotis (2008). Building an adaptive museum gallery in second life. In *Proceedings of Museums and the Web*, pp. 749–753.

Palace Museum and IBM (2009). The Forbidden City: Beyond Space and Time. https://myvirtualreality.wordpress.com/2009/07/06/the-forbidden-city-beyond-space-and-time/.

Prensky, M. (2007). *Digital Game-Based Learning*. Paragon House.

Rus, V., M. C. Lintean, and R. Azevedo (2009). Automatic detection of student mental models during prior knowledge activation in metatutor. In T. Barnes, M. C. Desmarais, C. Romero, and S. Ventura (Eds.), *EDM*, pp. 161–170. www.educationaldatamining.org.

Russel, S. and P. Norvig (2003). *Artificial Intelligence a Modern Approach*. Upper Saddle River, New Jersey, USA: Prentice Hall. Pearson Education International.

Solutions, A. (2001). Chatbots Directory. http://www.chatbots.org/developer/artificial_solutions/P10. visited 05 September 2009.

Steinkuehler, Constance, K. S. and e. Sasha Barab (2012). Cambridge University Press.

Thalmann, Daniel, M. S. R. (2013). *Crowd Simulation*. Springer.

Trescak, T., A. Bogdanovych, S. Simoff, and I. Rodriguez (2012). Generating diverse ethnic groups with genetic algorithms. In *Proceedings of the 18th ACM symposium on Virtual reality software and technology*, VRST '12, New York, NY, USA, pp. 1–8. ACM.

Wallace, R. (2004a). Be your own botmaster, 2nd edition. Technical report, Alice A.I Foundation Inc.

Wallace, R. (2004b). The elements of AIML style. ALICE AI Foundation.

Woolf, B., I. Arroyo, K. Muldner, W. Burleson, D. Cooper, R. Dolan, and R. Christopherson (2010). The effect of motivational learning companions on low-achieving students and students with learning disabilities. In *International Conference on Intelligent Tutoring Systems*, Pittsburgh, pp. 327–337.