An Observation Framework for Recognising Learning Evidence in 3D Collaborative Virtual Environments

Samah Saad Felemban

A thesis submitted for the degree of PhD in Computer Science School of Computer Science and Electronic Engineering

University of Essex

July 2021

Abstract

Immersive environments such as 3D virtual spaces enable collaborative learning and facilitate better connections between students, virtually. Learners do acquire new knowledge or skills while practising collaborative activities in such spaces. However, recognising evidence of learning to assess students is a critical issue which must be considered when organising learning activities in virtual environments. Although there is extensive coverage in the empirical literature regarding assessing learning in real-world classrooms, there is a lack of research focused on identifying learning evidence and assessing students who are performing educational activities within virtual worlds. This thesis aims to fill this research gap, exploit the affordances of immersive environments, and investigate appropriate methods for identifying users' performance within these.

This research proposes a computational framework, and a number of virtual observation models, for classifying learning evidence in immersive environments – and then maps all these elements to an appropriate learning design. In order to implement the computational framework required, the research includes the construction of a proof-of-concept prototype. The prototype employs virtual observation components and applies fuzzy logic and multi-agents approaches in order to assess students' performance in real-time; this is from a number of different perspectives and based on multiple pedagogical frameworks.

The present study also goes on to evaluate the research framework proposed by putting together a large number of educational sessions which are then carried out in a virtual world. These evaluation sessions involve both student and expert participants collaborating together to validate the model used. Subsequently, the thesis discusses the findings from the experimental sessions and their broader significance for the research area. Overall, the results

strongly supported the effectiveness and usefulness of using the proposed virtual observation method when assessing collaborative students performing within immersive environments.

بسم الله الرحمن الرحيم

To my parents . . .

To my husband . . . To my son and daughters. . .

With warm love and full respect . . .

Acknowledgements

Firstly, I would like to express my gratitude to Allah for all the mercy and blessings He has showered me in the course of my life and on my PhD journey, giving me the perseverance, patience and strength to undertake this work.

I cannot find the words to express my gratitude to my parents. My simple thanks go to my father, Saad Felemban - who passed away years ago, though his soul is always with me. I say to him, you have always been my role-model in terms of seeking knowledge and working hard; you taught me that education could be my weapon in this life. My deepest thankfulness is owed to you for all your love and for all that you did for me. I wholeheartedly wish you were here to observe my achievement, of which I believe you would be very proud.

To my heavenly and beloved mother, Nawal Sendi. I cannot express my appreciation for your ceaseless love and support. Your faith in me and your prayers day and night have motivated me to put in the necessary hard work and have given me the light and the power to continue with this research. Without you, I could not have reached this stage. I am very grateful to Allah to have you in my life; you are everything to me.

Huge thanks go to my treasured siblings, Suha, Nesreen, Mohammed, Nermeen, Amenah, and Ahmed and their families for supporting me and encouraging me all throughout my life and in my PhD journey. Each one of you has stood beside me in different situations and provided me with ceaseless support and special love.

My deepest appreciation goes to Dr Michael Gardner whose guidance has supported me throughout this research. His expertise, motivation, patience and support were treasured and have allowed me to accomplish the goals of this project; they have also assisted me to overcome other issues beyond the scope of PhD study. Equally, my deepest gratitude to Prof Victor Callaghan whose valuable knowledge and advice has inspired me over the many different stages of the work and have made enormous contributions to this research.

My deepest gratitude to my sponsors Umm Al-Qura University and the Saudi Government for awarding me with this scholarship so that I could undertake my PhD research, and for their generous financial support.

My honest expressions of obligation go to my colleagues at the Immersive Education Lab and the staff members of the School of Computer Science and Electronic Engineering. I am particularly thankful to Dr Anasol Pena Rios for sharing her platform and for providing me with technical support and generous advice whenever these were needed. And special thanks go to Dr Enas Jambi whose care, love and friendship have aided and supported me a great deal throughout my study and also beyond this project. My thanks also to the board meeting members (Pro Hani Hagras, Prof Richard Bartle and Prof Dariush Mirshekar-Syahkal) for their valuable comments and feedback. Moreover, I would like to thank those researchers and students who have assisted me and especially those who gave of their valuable time in order to participate in the evaluation experiments.

To my friends who have continuously supported me, inspired me and become part of my family - thank you for your love and for all the great times that we have had together. My deepest gratitude to Arwa Basabreen, Huda Altaisan, Manal Alghannam, Nora Alkhamees and Wafaa Bihari. In addition, special thanks go to my friends in Saudi Arabia who have constantly encouraged me from a distance.

Last but by no means least, to my husband (Fawaz Mereani), I express my deepest thank you for sharing your life with me and encouraging me to carry on with my work. Thank you for the unconditional love, care and support you have provided me with all along the way. For staying by my side and sharing the long nights working at home and for all those valuable suggestions my appreciation is unbounded. Additionally, I cannot sufficiently express my thanks to my fantastic son, Ayman, and my gorgeous daughters, Toleen, Talia and Eleen. You started with me on this adventure when you were little kids and you have walked with me all the way, giving me the power to continue with your love. I love you all and I am thankful to have this wonderful family in my life. This achievement belongs to all of us.

Table of Contents

List of Figures

List of Tables

Glossary

Term	Description
VW	Virtual World
3D	Three-Dimensional
VLEs	Virtual Learning Environments
IEs	Immersive Environments
MAS	Multi-Agent System
FL	Fuzzy Logic
MF	Membership Function
MIVO	Mixed Intelligent Virtual Observation Framework
OLens	Observation Lenses Model
MixAgent	Mixed Agents Model
SA	Software Agents
NA	Natural Agent
TAM	Technology Acceptance Model
PEOU	Perceived Ease of Use
PU	Perceived Usefulness
IU	Intention to Use
AEQ	Assessment Experience Questionnaire
QTF	Quantity and Timing of The Feedback
QF	Quality of The Feedback
UF	Use of Feedback
PA	Students' Preferred Approach
COMM	Perception of Chat Communication
NA	Perception of Natural Agent Rating Procedure
E-PA	Experts' Preferred Approach
EXP	Experts' Assessment Experience
REF	Experts' Reflection on the Assessment Processes

Chapter 1

1. Introduction

"The secret of getting ahead is getting started."

— Mark Twain

In this, the era of the Internet, knowledge and learning are everything. Acquiring knowledge and accumulating intellectual capital have become key factors in the success of ventures and, indeed, of societies. The old saying, 'two heads are better than one' quite accurately represents the collaborative way of learning. With this, learners try to connect and collaborate within either physical or virtual environments in order to increase their knowledge and achieve better understanding. A technology which greatly enables collaborative learning and facilitates better connection between students, virtually, is the immersive environment such as a three-dimensional (3D) virtual world (VW). The use of such 3D virtual spaces is growing in popularity because such spaces facilitate the connection of geographically dispersed learners in real-time [1].

The features of such virtual 3D spaces which distinguish them from other kinds of virtual platforms include engagement, interactivity and exploration. Learner engagement and interactivity are enhanced when educational applications operate in virtual worlds. Virtual worlds (VWs) facilitate the exploration of ideas, places and situations which may be difficult or impossible to access in real life. Also, VWs have the ability to visually represent 3D objects which relate to a field of study in order to help explain complex phenomena to students. For example, Alrashidi [2] suggested an approach to the enabling of the currently unrecognised pedagogical gains offered by embedded computing activities - by providing learners with real-time feedback. 3D VWs can also enable students to conduct practical experiments that may be difficult to perform in the real world [3]. It has been shown that such VWs can make it considerably easier for participants to collaborate in real-time and share their ideas in a group setting [4-6]. Such collaboration is particularly beneficial in terms of supporting shared experience and students' knowledge development.

However, the use of collaborative learning is beset by a number of issues, including task allocation, communication, evaluation and assessment. According to Blumenfeld, working with peers is not just a matter of gathering people together; successful collaboration depends on who the members are, what the tasks are, and how the members are assessed individually and as a group [7]. Learners acquire new knowledge and skills in the course of taking part in learning activities, and the new knowledge and skills can often be of a kind which is difficult to assess using summative tests. Consequently, rather than looking only at the final product as evidence of learning, instructors should assess the whole process [8]. Schallert et al. [9] reported that learners can provide evidence of learning while they are taking part in online discussions and collaborative work. On the other hand, collecting data to trace students' behaviours as they interact within 3D virtual environments is more challenging than doing the same in face-to-face, traditional learning sessions [10].

Several issues can arise when assessing a group of learners who are interacting within an immersive environment. First, in immersive environments, observing users' behaviours dynamically and collecting evidence of learning are complex tasks. Second, numerous skills, including communication and negotiation skills, can be gained from collaborative activities, but it is difficult to automatically detect evidence of these as they emerge in these spaces.

This research has enhanced the learning affordances of immersive environments by presenting students' evaluations in a visual way in order to improve the accessibility of the kind of assessments which are possible vis-a-vis activities in such spaces. The research introduced a novel computational-architectural framework that specifically facilitates the observing and recording of collaborative learning activities which take place in VW environments; this was done in order to enhance the evaluation of the learning outcomes. In particular, this research created a virtual observation model that can map between virtual observations and the evaluations of students acquired from physical settings. It focused more on providing methods of identifying and classifying learning evidence and assessing group working than on mapping all these elements to an appropriate learning design. To achieve these goals, the virtual observation model provided a mechanism that mixes natural agents and software agents in order to support the recording and labelling of learning evidence gathered from virtual activities and so simulate teachers' observations. The research demonstrated the ways in which employing learning theories alongside technology can enhance the recognition of learning evidence within VWs. Such learning evidence can support the creation of lifelong learning portfolios, providing feedback for learners so that they can identify their weaknesses and strengths and can also enable the identification, to teachers, of lessons which have proved effective, based on participants' performance. The computational-architectural framework is made available in order to bridge the gap between the collecting of learning evidence by technology and the collecting of such evidence by human educationalists, and such a framework can solve some of the issues which emerge from the evaluation of differing interactions, activities, knowledge and skills in collaborative immersive environments.

1.1. Motivation

The motivations for this work have come from many different directions. First, there is the growth in the use of immersive environments and virtual reality in the education sector and the benefits of applying these powerful tools in education. These technologies deliver effective solutions for educational situations where the most relevant physical surroundings cannot be made available due to, among other things, their high cost, or the physical risks to the students which would be entailed [11, 12]. This motivation is becoming stronger on a daily basis because of the use of virtual technologies in training and teaching and their other applications in education and the workplace.

Secondly, there is the need to improve the measuring of students' learning in relation to 3D VWs and to think about assessing outcomes in novel ways in order to cope with today's advanced learning environments, generally. Accordingly, the aim of this study was to expand the affordances of 3D VWs as regards learning through applying existing educational frameworks in order to further explore the potential of observing and evaluating collaborative learning activities within 3D spaces. Gardner and Elliott [13] stated that 'learning within a technology creates a pedagogical shift that requires teachers to think about measuring outcomes in non-traditional ways'.

The third motivation was personal and was related to my current employer. Since the use of immersive learning environments in my country (Saudi Arabia) is limited, I was asked by my employer (my PhD sponsor) to conduct a research study which could contribute to the improvement of online learning and educational technologies (both in Saudi Arabia and elsewhere).

1.2. Research Issues

This research looks at the following issues and problems. The first problem it investigates is the recognition and analysis of the learning evidence generated by users in real-time. Where several learners are interacting in an immersive environment at the same time, tracking and analysing are made much more complex. This issue necessitates the creation of a suitable mechanism for collecting and managing such data, generated from such spaces.

The second problem is that it is hard to trace and capture the entirety of the learning outcomes which can emerge in 3D environments. Numerous skills can be gained from collaborative activities, but it is difficult to automatically detect evidence of these. Thus, a method which can identify the learning outcomes achieved could allow 3D VWs to provide definite advantages over conventional approaches.

Therefore, finding a technique which can dynamically recognise users' behaviour, collect learning evidence and analyse events to measure the quality and quantity of learning outcomes is necessary. Such a mechanism could help to identify and gather proof of learning in the course of collaborative activities within immersive worlds and correlate the evidence with learning objectives in order to assess the overall outcomes of the learning processes.

1.3. Research Hypotheses

This research proposed the following hypotheses:

It is possible to create a computational observation framework which can support the gathering of the learning evidence from collaborative distance students using immersive environments – and which is capable of being used as the basis for student assessment.

1. *Such an observation framework will be able to integrate between software and human agents - such that users will have positive attitudes towards their roles as human agents when performing distance-learning tasks in the virtual world (**H1**).*
2. *Systems based on this observation framework will provide collaborative distance learners with assessment feedback, and users will report positive experiences of such assessments (**H2**).*
3. *Systems based on this observation framework will provide assessments that are very similar to human-expert assessments; these system assessments will be produced using less effort overall (H3).*
4. *The assessment results and feedback from such an observation system will be preferred by users over and above that yielded from human experts (**H4**).*
5. *Users of the observation system will express their acceptance of it (**H5**).*

1.4. Research Contributions

The main contributions of this research are:

1. A synthesis of the existing learning theories into a novel virtual observation computational framework for observing and assessing collaborative learning within 3D environments (*MIVO Framework* - *Chapter 3*). This framework to consist of two innovative models:
    - The Observation Lenses model (*OLens*) – an observation model for identifying and structuring evidence of student learning within 3D virtual environments; this forms part of an e-learning assessment process (*OLens* - *Chapter 3),*

- *MixAgent* model - a computational model that enables the integration of software and human agents in order to demonstrate mechanisms for collecting learners' data. (*MixAgent - Chapter 3).*

2. A proof-of-concept prototype system that implements the *MIVO* framework with its observation models and which utilises fuzzy logic approaches for assessing students' performance from different perspectives, based on a pedagogical framework (*Observe Portal - Chapter 4 and Chapter 5*).
3. Empirical research findings derived from the evaluation of the prototype which demonstrate the effectiveness of the approaches and models used for collaborative learning activities in 3D spaces, and which allow for a comparison with equivalent, traditional approaches (*Chapter 6 and Chapter 7*).

The present study also delivered the following secondary (in terms of being less important) contributions, these were also accomplished as a result of conducting this research:

1- The design of assessment interfaces driven by learners' performance (to assess students taking part in collaborative tasks).

2- The creation of hierarchical fuzzy logic systems which emulate human reasoning as expressed in 3D virtual environments.

3- The encoding of a collaborative problem-solving skills taxonomy [14] and applying this in a 3D virtual environment to assess student collaborative skills.

4- The introduction of instruments for measuring the effectiveness of the natural agent tools; these instruments measure, specifically, the perception of the chat communication (COMM) subsystem and the perception of the natural agent rating procedure (NA).

5- The use of instruments for measuring the experts’ assessment experience (EXP) and the experts’ reflections on the effectiveness of the assessment methods (REF).

1.5. List of Publications

Some of the thesis contributions listed here have also been presented in the following publications:

- **Conference Publications:**

    1. S. Felemban, "Distributed pedagogical virtual machine (d-pvm)," presented at The Immersive Learning Research Network Conference (iLRN 2015), 2015.

    2. S. Felemban, M. Gardner, and V. Callaghan, "Virtual observation lenses for assessing online collaborative learning environments," presented at the Immersive Learning Research Network (iLRN 2016), Santa Barbra, USA, 2016.

    3. S. Felemban, M. Gardner, and V. Callaghan, "An Event Detection Approach for Identifying Learning Evidence in Collaborative Virtual Environments," in 2016 8th Computer Science and Electronic Engineering Conference (CEEC), 2016.

    4. S. Felemban, M. Gardner, V. Callaghan, and A. Pena-Rios, "Towards observing and assessing collaborative learning activities in immersive environments," presented at the Immersive Learning Research Network: Third International Conference, iLRN 2017 Proceedings, Coimbra, Portugal, 2017.

- **Journal Publications:**

5. S. Felemban, M. Gardner, and V. Callaghan, "Towards Recognising Learning Evidence in Collaborative Virtual Environments: A Mixed Agents Approach," Computers, vol. 6, 2017.

6. S. Felemban, M. Gardner, V. Callaghan, and A. Pena-Rios, "Mixed Agents Virtual Observation Lenses for Immersive Learning Environments," Journal of Universal Computer Science, vol. 24, pp. 171-191, 2018.

1.6. Thesis Outline

- Chapter 1 (this chapter) presents the introduction, the motivation behind this thesis, the contributions which are intended and the basic hypotheses informing this research - in addition to the list of publications that have arisen from this work.
- Chapter 2 starts by introducing various learning theories/frameworks and the methods which have been applied in the past to the collection of learning evidence and to the assessment of students' learning. It also provides a detailed literature review with regard to virtual learning environments and the numerous learning affordances of 3D worlds. In addition, this chapter discusses many techniques that have been used previously to evaluate students' progress in virtual worlds. Further, it provides a review of the fuzzy logic and multi-agent approaches, and the means by which these can be integrated into immersive environments so that learning can be assessed.
- Chapter 3 introduces the innovative computational framework (*MIVO*) used here; this includes the Virtual Observation Lenses (*OLens*) and the Mixed Agents model (*MixAgent*); these together simulate teachers' observation and assessment of

collaborative students in virtual worlds. The *OLens* model and its observation layers are then explained in detail by describing the lenses utilised and by providing supporting examples.

- Chapter 4 describes the experimental work associated with the creation of the proof-of-concept prototype, *Observe Portal*. It describes the practical work which has been undertaken in the learning environment to incorporate the research models within this, including the development of the 3D virtual world and of the evidence collection agents and also the implementation of the observation lenses. Additionally, it explains the construction of the assessment presentation unit and the design of the feedback dashboard interface.
- Chapter 5 continues by describing the experimental work which was undertaken; in particular, it provides a detailed narration of the development of the fuzzy logic systems implemented in the research prototype. The chapter describes the way in which the fuzzy logic method used combines all the data produced by the agents in order to make decisions about student performance and so assess their learning.
- Chapter 6 starts by introducing the approaches used in the research experiments to measure the effectiveness of the research models. Then it explains the collaborative learning activities employed to study the students' and the instructors' experiences, and also the activities employed in the experiments.
- Chapter 7 discusses the results of the evaluation and their analysis in detail. It describes the data analysis procedures employed and demonstrates the results yielded from the hands-on student learning activities and human expert assessments carried out in the virtual world.

- Chapter 8 discusses the outcomes and findings of the evaluation experiments, debating their wider importance in this research field.
- Chapter 9 summarises the achievements of the research and its contributions. This chapter also describes the future work leading on from this research and its challenges.

Chapter 2

2. Literature Review

"An investment in knowledge pays the best interest."

— Benjamin Franklin

The capabilities of computers and networking have led to the development of technologies that support learning and connect geographically dispersed learners via systems which provide enhanced learning experiences. This chapter provides a review of a number of important topics which are associated with this research. It begins by introducing some learning theories and frameworks which have been applied in education, and it presents the methods utilised to assess students and gather evidence of learning. Additionally, it gives a detailed view of virtual learning environments and the several learning affordances of three-dimensional spaces. It also discusses some computational techniques which have been used previously to evaluate students' progress within virtual worlds. In particular, it discusses agents and multi-agent approaches and how these can be employed within applications. Also, it looks at fuzzy logic reasoning methods and the ways in which these can be integrated with immersive environments in order to assist in assessing learning.

2.1. Learning

In general, learning is the process of acquiring new knowledge, skills, values, behaviours or visions and may involve combining various types of information. People may use differing methods and settings in order to learn new skills or acquire new knowledge. Schunk defined learning as "an enduring change in behaviour, or in the capacity to behave in a given fashion, which results from practice or other forms of experience (p. 2)" [15].

2.1.1. Learning Theories

Three main learning theories have been recognized by educators and researchers:

- **Behaviourism Theory:** This theory stresses that learning is achieved when an appropriate response to a specific environmental *stimulus* is demonstrated and that all behaviours are directly affected by external s*timuli*; such behaviours, this theory maintains, are not, necessarily, based on inner mental situations or awareness [16]. The theory focuses on the sequences involved with the making of a connection between a *stimulus* and a response — as this association is recognised, reinforced and then sustained.
- **Cognitivism Theory:** This emphasises the acquiring and storing of information in the inner mental structures and is more concerned, than behaviourism, with the mental activities of the learner (how the knowledge is obtained, stored, organised and reused by the brain) [16]. Learning is associated with isolated changes in states of information rather than, necessarily, with changes in the likelihood of particular responses (as with behaviourism). However, the environment nevertheless plays a significant role in both behaviourism and cognitivism.
- **Constructivism Learning Theory:** This is "the theory of knowledge acquisition obtained through interactions and building upon [learners'] own knowledge and which produces the highest type of learning [17]." This approach stresses learning-by-doing: learners can reach advanced levels in terms of gaining knowledge and understanding through taking part in activities [18]. According to Dalgarno [19], as far as constructivism is concerned, learning occurs when students discover a new domain which reveals the existence of a gap between their experiences and their current representations of knowledge. In addition, for learning fulfilment, learners

should be involved in an associated social context and actively interact and debate with others in order to construct knowledge [19]. According to the descriptions of this process, the educator's role is to monitor the learning processes and encourage students to explore principles with which they can construct information by working to solve problems [20].

2.1.2. Collaborative Learning

Collaborative learning is considered to be an approach based on constructivism. Students achieve learning through working with their peers, who support them to enhance their level of information and skills, resulting in the construction of new knowledge and experience.

Gokhale [21] conducted a comparative study between individual and collaborative learning; the study showed that students in groups significantly performed better in critical-thinking evaluation tasks than did individual students who had acquired the necessary knowledge on their own. Collaborative learning also enhances problem-solving strategies because learners come to view a given situation from the differing perspectives and understandings provided by other students. So, it is possible for learners to acquire both critical-thinking and external-knowledge skills from a system of co-operation and implement those skills in future logical operations [22]. A group's ability to perform tasks relevant to it successfully is called collective intelligence [23]. The factors which influence collective intelligence are as follows: (a) the group's composition (the members' intelligence, skills, and diversity); and (b) the group's means of interacting (structures, processes, and norms). Thus, when formulating frameworks for collaborative groups, we should consider these factors in order to achieve effective collaboration and optimal group performance.

In recent years, a popular movement has arisen which clearly illustrates the value of constructivism and collaborative learning: the ‘maker’ or ‘hacker’ space. A maker space can be viewed as a kind of informal learning space because maker space groups include less-skilled participants who learn from experts by working collaboratively with them [24]. The group acts as a team when developing a project and all tend to have an enthusiasm for creating and building new products. This collaborative way of creating helps makers to acquire new skills, resulting in them being available for more opportunities and challenges in the future. Figure 2-1 illustrates the world’s largest maker event (so far); this occurred in 2018, in Saudi Arabia, and was called (Hajj Hackathon) [25]. It facilitated participants meeting in groups and working together to create software and/or hardware to assist Muslim people to perform Hajj more easily. Such an event illustrates collaborative learning very well and the advantages of working with other people to enhance each individual’s knowledge and skills.

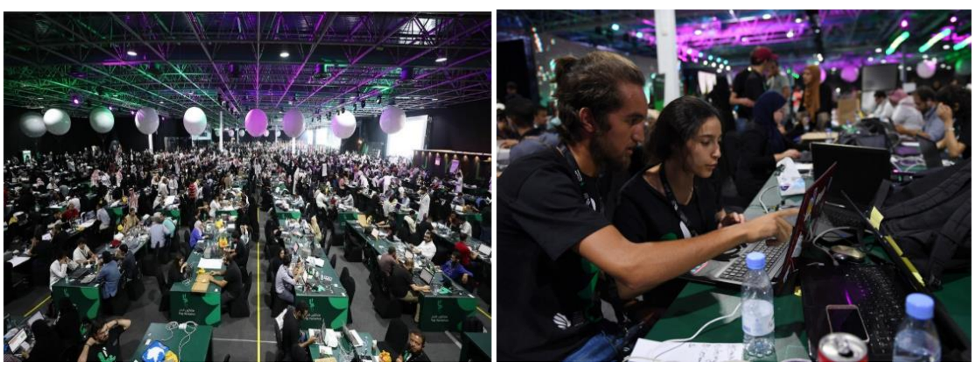

Figure 2-1: Hajj Hackathon - Collaborative Learning [25]

2.2.Assessment and Learning Evidence

Angelo defines assessment as “an ongoing process aimed at understanding and improving student learning. It involves making expectations explicit and public; setting appropriate criteria and high standards for learning quality; systematically gathering, analysing and interpreting evidence to determine how well performance matches those expectations and standards; and using the resulting information to document, explain and improve performance" [26]. Moreover, Suskie [27] said, "The more evidence you collect and consider, the greater confidence you will have in your conclusions about students learning (p.46)."

In school-situated formal learning, teachers evaluate their students by gathering evidence of their (the students') learning and analysing this evidence based on the specified learning goals and objectives in place. According to Suskie [28], there are two methods by which evidence is collected in education: direct and indirect. The direct method examines whether the student has acquired knowledge of certain subjects, or the use of a specific skill, or is able to perform a particular task and demonstrate work of a required quality. Examples of direct measures used to provide evidence of learning are class participation, presentations, research projects, quizzes, theses, and the resultant grades and scores. Indirect measurement methods merely infer that learning has taken place and focus on characteristics that are related to learning. These characteristics might be focused on individual learners, e.g., the number of hours a student spends on a project or class activity, or they may be measures which can reveal a whole class' or institution's learning result, such as course evaluations, the identification of skills and concepts covered in tests, the percentage of time a class spends in active learning, the course grade average, or the number of scholarships and awards earned by students [28].

In relation to group work, Webb[29] argued that it is important to identify, explicitly, the assessment purpose and to take into account the goal of the collaborative work and the processes of the collaboration. Thus, three purposes of assessment in collaborative work have been defined: the assessment of the individuals' levels of knowledge and skill after the learning process, the measurement of the productivity of the group, and the evaluation of the learners’ capabilities in terms of communicating and interacting with other group members.

On the other hand, Vygotsky assumed that evaluating learners taking part in group projects should be undertaken during the learning process, not just after finishing the learning sessions, because students usually acquire new knowledge while practising learning activities [30]. To evaluate collaborative groups effectively, it is critical to assess the product (the produced work) and the process (the students’ performance) [31-34]. Wells [8] agreed and stated that educators should evaluate the whole learning process when leading learning activities rather than just look at the final artefact as evidence of learning.

Evaluating the process of collaborative work can be achieved using many different measures, such as the ability to create a range of ideas, to listen respectfully, to communicate effectively and to resolve differences [34]. The following evaluation methods can often be useful in this regard:

- Group-evaluations: learners evaluate the dynamics of the whole group.
- Peer-evaluations: students assess the contributions of the other team members.
- Self-evaluations: learners evaluate their own contributions to the group.

However, no one assessment solution can be generalised so that it suits every scenario because different situations demand different methods and systems of assessment.

The 21st-Century Skills Assessment

In the last few years, educational research has found that a number of particular kinds of skills, called 21st-century skills, need to be developed by learners. Numerous authors have suggested various classifications for these skills - which are essential for facing the demands of modern life and jobs [35-37]. Although there is no one definition of these skills, Kyllonen [38] has developed a useful 21st-century skills taxonomy: cognitive skills (creativity, problem-solving, and critical thinking), interpersonal skills (social skills, teamwork, communication skills, cultural sensitivity, dealing with adversity) and intrapersonal skills (self-development - lifelong learning, time management, self-regulation, adaptability, self-management, executive functioning). Assessing such skills requires the setting of well-designed tasks which permit participants to interact and communicate with other learners and also training professionals. Additionally, providing students with feedback about their performance is important, so that they can recognize the levels of such skills that they exhibit.

Many studies have suggested that peer-evaluation and/or self-evaluation are useful approaches for the assessment of interpersonal skills. According to Kyllonen [38], assessing interpersonal and cognitive skills cannot be effectively accomplished by simple assessment approaches such as multiple-choice questions, essays or other tests. These skills require more advanced means of measurement, so Kyllonen suggested using self-rating or other rating approaches. Self-rating has been used in many educational studies to evaluate individual skills and assess personal experience. However, other studies have proposed that peer-rating assessments are more predictive and accurate than self-rating when evaluating 21st-century skills [39].

2.3.Observation

Observing students is another method which can be used to assess learning outcomes. “Observation involves teachers in observing students as they participate in planned activities. Teacher observation occurs on a continuous basis as a natural part of the learning and teaching process and can be used to gather a broad range of information about students’ exhibiting of learning outcomes"[40]. Therefore, observing learners can help educators to evaluate students by gathering evidence about their learning. The evidence can be saved and recorded so that it can subsequently be used to provide feedback for learners. Observations can take place in a number of different settings and using a variety of different methods. Observation may emphasise learners' performance in the course of a single activity or it may be applied to a group of activities. Applying observations in class requires the determination of when to observe, what to observe, and how often to observe. Moreover, tutors must plan how to record what they see and hear. Maxwell [40] summarised the types of evidence which can be gathered by educators as in Figure 2-2.

Focus on product — keep artefact(s)

Focus on activity — record process

- Direct record ('trace')
  - Audio-tape
  - Video-tape
  - Photographs
- Written record
  - Observation sheet
    - Checklist (high inference)
    - Description (low inference)
  - Logbook
    - Description of critical incidents
    - Description of key performance features

Figure 2-2: Types of Evidence [40]

2.3.1. Types of Observation

Teachers can observe learning activities (lab classes, seminars, lectures), teaching materials and documents presented, learning resources, interactions between participants, and also learning environments. According to [41], there are two types of observation: unstructured observation and structured observation.

- Unstructured observation or informal observation occurs when teachers observe students without determining what and whom to observe. Through this type of observation, educators may become aware of, for instance, which students work independently and which collaboratively and of the students who require more guidance.

- Structured observation is a formal and very systematic approach. It has a particular focus on students behaviour. It can include predetermined components of evidence such as checklists, frequency of performing particular actions, or specific events to be looked for in a specific situation. It is difficult for a teacher to assess problem-solving skills in a lab activity among a team without observing these in use.

Observation can help teachers gather information about individuals' and groups' behaviours and skills. According to [42], teachers can use observation for formative assessments of, for instance, particular behaviours, ways of thinking, writing skills, speaking skills, social skills, or athletic skills. In groups, students can also use observations and checklists to evaluate the progress of their peers and to monitor the entire group's development [42]. Some studies observe groups and gather information to assess the success of a collaboration based on task and time management, the success of collaborations and of individualised tasks. Others focus on group performance based on social interactions and communication. The third type assesses the process of constructing knowledge and skills within a group [43].

2.3.2. Observation Frames

Borich [44] introduced conceptual frameworks that follow educational standards in order to define the basic frames for observing. Because observing classrooms is very complex, he suggested that each teacher should select a specific frame or ‘lens’ to gain more insight into a specific classroom characteristic. The ‘lenses’ are identified as follows (Table *2-1*):

1. The learning climate
2. Classroom management
3. Lesson clarity
4. Instructional variety
5. Teacher's task orientation and content presentation
6. Students' engagement in the learning process
7. Student success
8. Students' higher thought processes and performance outcomes

1. The learning climate	• Degree to which students can express their feelings and opinions • Frequency with which student responses are used and extended • Amount of interaction and sharing among learners
2. Classroom management	• Use of preestablished classroom rules • Use of instructional routines • System of incentives and consequences
3. Lesson clarity	• Frequency of examples, illustrations, and demonstrations • Percentage of students who can follow directions given • Use of review and summary
4. Instructional variety	• Use of attention-gaining devices • Changes in voice inflection, body movement, and eye contact • Use of a mix of learning modalities (visual, oral)
5. Teacher's task orientation and content presentation	• Orderliness of transitions • Teacher's preorganization of administrative tasks • Cycles of review, testing, and feedback
6. Students' engagement in the learning process	• Use of exercises and activities to elicit student responses • Monitoring and checking during seatwork • Use of remedial or programmed materials for lower-performing
7. Student success	• Number of correct or partially correct answers • Number of right answers acknowledged or reinforced • Number of delayed corrections vs. immediate corrections
8. Students' higher thought processes and performance outcomes	• Use of teaming, pairing, or other cooperative activities that encourage student problem solving • Display of student products and projects • Opportunities for independent practice and application

Table 2-1: 'The Observable Signs Pertaining to the Eight Question Areas' [44]

Although observation is a great method that assists educators to evaluate learner skills and knowledge, it has some limitations. First, observation requires many resources and is time-consuming. Second, observing activities may affect participants' behaviours because they might be concerned about what the observer is evaluating. Third, students' thinking cannot easily be observed, so some educators utilise other methods, such as surveys or tests, to evaluate creative and critical thinking.

2.4. Learning Environments

The capabilities of computers and networking have led to the development of technologies which support learning and connect geographically dispersed learners with the purpose of enhancing learning experiences. In addition, the building of proficient and intelligent online learning systems has attracted several researchers [45-47]. Many educational technologies have been widely applied to connect scholars and educators in order to provide a variety of different types of activities and to access learning sessions remotely without requiring physical attendance. Organizations can more easily educate and train learners without reserving specific physical venues or hiring a large number of geographically dispersed tutors. The various educational environments which have been used to enhance educational activities are discussed in the following:

2.4.1. Virtual Learning Environments (VLEs)

Virtual Learning Environments (VLEs) are technology-based environments created to allow instructors and learners to share and access resources remotely without leaving their own localities. These environments were developed for managing the resources and activities required for successful computer-based learning [48]. A number of different types of VLE have been created. These environments can be categorized as either web-based applications such as *Blackboard*[1], *WebCT*[2], *Moodle*[3], 3D virtual worlds or a mixture of these different applications (*Sloodle*[4]). Several features [49] of VLEs have been categorized, as follows:

- They are developed as information spaces.
- They are designed as social spaces in which learning interactions can happen.
- In VLEs, the social spaces and the information spaces can be represented in several forms, from text to 3D worlds.
- Students are not only active but can also be actors who construct the virtual world.
- VLEs enhance school activities.
- The environments are constructed using many different technologies and a number of different educational approaches.
- Most VLEs overlap with and simulate real learning environments.

2.4.2. Immersive Environments (IEs)

Generally, immersion is achieved, in terms of the impressions of users, if users are allowed to interact with digital, realistic, and/or virtual environments that are able to deliver a sensation of presence [50, 51]. Well-designed educational immersive environments (IEs) can provide great experiences for learners. For example, the Immersive Education Group projects at the University of Essex [4, 52-57] exemplify different IEs that support learning activities. These projects have shown the advantages of using such environments to enhance learning. Conversely, the use of IEs in learning should be integrated fully with the

[1] https://www.blackboard.com
[2] https://www.elearninglearning.com/webct/
[3] https://moodle.org/
[4] https://www.sloodle.org/

educational setting; the implementors should not just be concerned with the technology [58]. The environment should be conceptualised beyond the technology to form an effective learning system. IEs have attracted a great deal of attention in education and there is interest in all the different types available, including mixed reality (MR), virtual reality (VR) and multi-user virtual environments (MUVEs).

Mixed Reality (MR)

MR technology allows learners to interact with objects in new ways; it merges physical and virtual worlds in order to create new visualised worlds wherein users can interact with both real and virtual objects. MR environments allow learners to travel beyond virtual worlds, and allow them to seem to become embedded within the associated physical world [13].

In 1994, Milgram and Kishio introduced an important model for MR technologies; this is shown in Figure 2-3. They defined a mixed reality environment as "one in which real world and virtual world objects are presented together within a single display, that is, anywhere between the extremes of the virtuality continuum" [59].

Figure 2-3: Reality–Virtuality Continuum[59]

This model defined four stages of the "*Reality–Virtuality Continuum",* spanning the space between the physical and virtual worlds. The real/physical environment includes only physical objects which exist in the real world. Augmented reality (AR) merges the physical environment with 3D objects which are generated by computers and are able to interact in real-time; AR can produce feelings of immersion in the user through the overlay of the computer-generated objects onto the real environment, and facilitates interaction between these two environments [60]. Augmented virtuality (AV) projects physical objects into the virtual environment in order to enhance virtual spaces using real data [61]. Finally, the virtual environment (VE) provides a simulation of real world experiences by generating virtual objects with the aim of producing virtual spaces wherein to practise activities that are a simulation of activities in the physical environment. Examples of mixed reality environments are shown in *Figure 2-4*.

a) Using AR in a biology class b) Using AV in a chemistry class

Figure 2-4: Mixed Reality Environments in Education [62]

Educational researchers [63-65] have recognized the importance of combining virtual and physical environments to provide innovative opportunities for learning [58]. Via the development of an interactive MR environment, students are enabled to use 2D or 3D objects and practise activities that are problematic in terms of being executed in other spaces [66]. In addition, MR may help learners to experience the exploration of phenomena that do not exist in the physical world. Moreover, it assists students to better understand complex concepts and visualize relationships between the elements of such concepts [67-70].

Virtual Reality (VR)

Virtual reality (VR) is an enhanced form of virtual environments for which users generally use VR equipment (Figure *2-5*), log-in to a computer-generated environment, and obtain the impression of 'being there' – a feeling of presence [51], providing users with a greater sense of reality in relation to the virtual space [71, 72]. The key purpose of VR is to simulate an imaginary or real environment [73] and the principle unique features of VR are that the environment will respond to a player's body movements and actions in real-time, the level of immersion possible, and the use of devices for unusual human-machine interaction such as haptic devices and data gloves [74, 75].

Figure 2-5: Virtual Reality (VR) [76]

Multi-User Virtual Environments

Multi-user virtual environments (MUVEs), also called 3D virtual worlds (3D-VWs), are a type of virtual environment, as mentioned by Milgram in [59], where people can meet simultaneously with many others (Figure *2-6*). Such environments provide users with avatars in order to enhance the visual interaction between users and virtual objects. Examples of such spaces are *Active World*[5], *Open Wonderland*[6], *Open Sim*[7] *and Second Life*[8]. In general, it is intended that people's behaviour in virtual environments remains what it might have been expected to be in the physical world [77]. Such 3D-VWs do not have levels or scores in the way that virtual games do, and usually, these environments operate in real-time, so users can

[5] https://www.activeworlds.com/
[6] http://openwonderland.org/
[7] http://opensimulator.org/wiki/Main_Page
[8] https://secondlife.com/

collaborate and communicate in a natural way [1, 78]. In this research, we applied a 3D-VW as our immersive environment of choice to implement our proof-of-concept prototype.

Figure 2-6: Multi-User Virtual Environments/3D Virtual Worlds [79]

2.5.Learning Affordances of 3D Environments

Educational researchers have recognised the importance of 3D environments as innovative opportunities for learning. Numerous organizations worldwide have implemented virtual worlds for distance learning, created virtual classes and developed virtual campuses. Educators and students have stated that such environments provide them with a sense of ‘being there’, as if they are attending physical classes but with virtual people [80-82]. Many studies address presence and immersion as important aspects of learning in 3D environments. Presence is the feeling of being in a real place and being immersed in the measurable characteristics and technical capabilities of the technology that leads to this feeling of presence [3].

Others have also reported that using MUVEs can stimulate users' enjoyment, engagement and motivation [82-84]. Moreover, MUVEs can help students to better understand complex concepts and visualise relationships between elements. Other advantages of 3D virtual spaces were reported in [85]: they enhance teamwork and facilitate decision-making within groups and increase shared understanding between learners. Coffman and Klinger [86] argued that learning with 3D virtual worlds is valuable when applying a constructivist approach. This approach emphasises learning-by-doing: learners are enabled to reach advanced levels of understanding through collaboration and discussion [18]. Therefore, these spaces can connect students and create collaborative activities to help them discover meaningful connections between academic content and the real world.

Dalgarno [3] documented a framework that illustrates the affordances of applying 3D virtual learning environments. He summarized the affordances in the following points:

1- They facilitate learning activities which enable the development of knowledge representations for the field which is to be explored;

2- They support the exploration of phenomena which do not exist in the physical world;

3- They improve learning engagement and motivation;

4- They support the transference of skills and knowledge into real situations through learning contextualisation;

5- They facilitate collaborative learning in a more effective way than do 2D environments.

However, there is a need for more research into 3D-VWs so that their potential for learning can be enhanced and to increase the affordances of using such spaces [87]. Many issues need to be resolved before the learning benefits of virtual worlds can be fully realised; the determining of students’ performance and skills, particularly in relation to collaborative learning activities, is one of these issues. Identifying learning evidence is problematic because of the large number of variables involved and the complex relationships between these variables and students' performance in these spaces. The next section highlights the issues, frameworks and approaches associated with evaluating multiple students in such spaces.

Identifying Learners' Performance

One way that the level of students' learning can be estimated is from text-based data collected from online discussions [9]. On the other hand, in simulation spaces, learning evidence can also be found to be embedded in users' patterns of action or in data generated from decisions that students have made [27]. An essential step towards achieving effective assessment is to create a record of everything which has taken place in the course of a learning course or session [28]. However, collecting learning evidence from simulation and 3D environments is more challenging than it is in traditional face-to-face learning sessions or from the use of tests [10]. Identifying evidence of learning is very straightforward when using, for instance, the multiple-choice test format, but becomes more problematic in 3D virtual environments; this is because of the large number of observational variables involved and the complex relationships between these variables and students' performance [88]. The relationships between the basis of the decisions and the actions that students take when solving problems in such spaces is not so clear as the relationships between learning and the answers to a test.

Although the high level of technology involved in 3D-VWs can mitigate this situation and assist in the recording of the data which is generated, understanding and analysing the composite data that results require more complex processes. Mostly, in 3D-VWs or learning games, the students' performance is extracted from the log files which are auto-generated while they are playing [89]. Learners can be assessed by analysing the data logs and tracking the students' pathways and the decisions they made when accomplishing tasks [90]. However, the log files save all learner responses to the given educational problems, including the students’ faults and mistakes [91]; this raises practical issues in relation to analysing such data. The log files contain huge amounts of data, creating a serious obstacle for researchers collecting learning evidence from learning environments [92]. For example, in relation to a basic puzzle game between a limited number of users, nearly 400,000 rows of log data were generated [93]. This kind of situation makes it very hard to capture an individual student’s learning, knowledge and skills, and also makes it difficult to identify the actions and performances that represent learning. Therefore, analysing various learners’ behaviour/data, identifying the meaningful actions and combining those actions into learning evidence to show learners' skills are very complex processes in such environments.

Another issue with identifying students’ skills is that simple technologies cannot capture all of the learners’ skills. Several skills could be gained from collaborative activities, but it is a complex matter to automatically detect evidence of all of them [90]. For example, the relational skills of teamwork, collaboration, negotiation and communication are hard to measure using conventional kinds of assessments. Identifying a mechanism that can record all of the pertinent knowledge and skills obtained by the learner in order to evaluate learning outcomes from immersive learning activities is essential.

2.6.Assessment in Learning Environments

Assessment should measure a number of different aspects of students’ learning such as learner performance, success, knowledge and skills. On the other hand, all these different aspects require different methods to be applied. For instance, knowledge can properly be assessed by traditional exams, but evaluating skills requires more complex activities and the use of techniques that enable the student to demonstrate them. MUVEs can facilitate the latter by creating complex situations and allowing for the exploration of phenomena that do not exist in the physical world [3]; this supports the enhancement of skills and competencies. Providing assessments and feedback as part of the learning process can enhance student learning and improve performance.

Several approaches have been used for the assessment of student learning within learning environments. A first approach is the traditional school test approach, which involves giving paper tests to students or generating automated questions with multiple choice answers – either during the learning activity or after. For instance, in Second Life, most educators use an extension of classroom summative tests to provide a final assessment [94]. Another example is the quizHUD project [95] in SLOODLE [96] that uses a multiple choice interface to assess students' knowledge. This test-based approach can be useful when an environment is used to host lectures, imitating the physical classroom setting, and where the evaluation objective is to assess the student's knowledge. However, traditional school tests should not be applied to measure learning outcomes when the learning environment offers hands-on, experiential or experimental activities in order to teach students. In these settings, summative tests do not provide a full perspective of the students' learning and cannot adapt to the needs of learners, nor provide them with immediate feedback while they are working. Immersive environments provide significant learning opportunities for distance learning through distributed systems offering collaborative and cooperative activities; these require new assessment methods to meet the complex learning requirements which are supported. Thompson and Markauskaite [97] stated that 'educators need to move beyond traditional forms of assessment and search for evidence of learning in the learner interactions with each other and the virtual environment, and artefacts created'.

A second approach that has been applied to assess learning is analysing the actions of students. Such techniques are usually based on the cognitive task analysis method which consists of creating logic rules to track students' behaviours and to distinguish between the specific levels of skills of learners [98]. An additional technique which can be used to analyse the actions of students is to extract the performances of students from generated log files by applying machine learning or data mining methods. For instance, Kerr and Chung [89] analysed the log data of a user by applying cluster analysis algorithms in order to define the key features of the performance of a student in educational game environments. In addition, Bernardini and Conati [99] applied cluster methods and class rules to the log data of users to find out the different models used by learners exhibiting successful as opposed to unsuccessful strategies within the learning environment. Even though these studies were investigating the behaviour of users, they are limited to studying the relationships between data and identifying the quality of learning outcomes only from the log files as they are constituted. Moreover, it is more challenging to identify learning evidence in relation to collaborative learning activities where there are many contributing users. The log files store all the actions students make in response to problems, and this generates a large amount of data; the quantity of data represents a serious obstacle for researchers when collecting learning evidence in relation to learning outcomes [92]. Capturing the data of users without identifying how they are to be scored is not an effective way of creating information for assessments. It is preferable to develop, from inception, facilities in the learning environment which gather learning evidence and assess students learning [100].

According to Gobert [101], educators can encounter serious issues when attempting to assess learning within immersive environments. The first issue is that there is an absence of theoretical guidance regarding how to analyse streams of data generated from the performances of learners. The second issue is the lack of theoretical foundations which can be discovered in the literature regarding learning assessment and assessment approaches. On the other hand, some studies have applied the Evidence-Centred Assessment Design (ECD) framework introduced by Mislevy and Riconscente [92, 102]. ECD is a general framework originally created to assist with the assessment of student learning, using computer-based tests. This framework consists of a number of different models: the ‘student model’ (which determines what is best to measure, skills, knowledge, and/or abilities), the ‘evidence model’ (which determines how to measure these things) and the ‘task model’ (what situations can be used for measurement). ECD has been used to assess students using simulation environments [103, 104]. For instance, Shute [105] developed the Stealth assessment, based on ECD, which applies Bayesian networks to model a student’s actions in a learning game and then infer their level of problem-solving skill. Shute found that the inferred learning events closely matched with the students’ actual learning. However, most of these studies apply assessments which are specific to the game context involved and which assess a specific competence or skill, based on the behaviour of the player. In contrast, there are no standardised assessment models or guidance on observing learning activities, generally, which cover all aspects of learning, including student interactions, success, knowledge and skills. Furthermore, most of these latter studies focus on assessing individuals, but one of the most important features of MUVEs is that it allows for collaborative activities and the sharing of knowledge. To measure collaborative learning activity, we need to evaluate the learning of the group as well as the learning of individuals. Nevertheless, there are few studies that provide theoretical guidance on assessing such things.

2.7. Agent and Multi-Agent Approaches

Nowadays, agent-based techniques have become an important and very useful technology employed in applications generally because it provides the ability to control distributed systems. In particular, computer agents can make decisions and solve problems autonomously — which benefits large systems. Thus, this section discusses the background of the techniques involved with agent and multi-agent systems and highlights the importance of these approaches and their wide application.

2.7.1. Agents Background

In 1988, Minsky proposed a significant theory in his book *The Society of Mind* [106]. After studying human and artificial intelligence for some years, he hypothesised that a mind, per-se, is made up of components which are assembled to create a 'society of mind'. He considered the simplest component of a mind to be an agent. When agents are connected to perform a particular task, they are called an agency. Agencies are assembled to create a society of minds that can think and work. Minsky's theory has been used in the construction of several software engineering and artificial intelligence (AI) systems. It has been widely applied to attempts to create AI minds and imitate human thinking, communication, negotiation and action.

Many computer applications have been created based on this theory of agents and societies of mind. An agent is a computer-based system, located in an environment, which has the ability to act in this environment in order to achieve the aim that it was designed for [107, 108], see Figure *2-7*. Some say agents should have the capacity to learn while others argue that learning is not important and is in fact an undesirable ability for individual agents to possess [107]. The agent definition used for this research does not specify that an agent must be intelligent and does not define the environment that the agent acts upon. Agents can be any computer application that takes input and performs in an environment to produce an output.

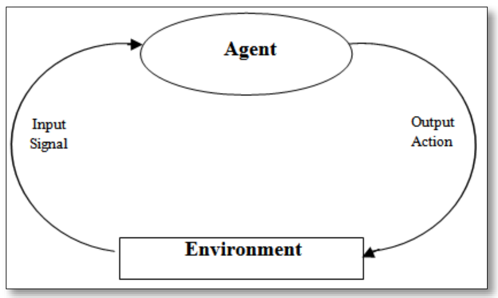

Figure 2-7: An Agent in its Environment *[107]*

One use of Minsky's theory is in the definition of multi-agent systems (MASs); these consist of a large number of agents that have differing roles and objectives, connected to each other in order to process input and produce output. The structure of a multi-agent system includes protocols for communication and interaction between agents. Communication protocols permit agents to understand and exchange messages, and interaction protocols facilitate agents in having conversations [107]. A MAS can also be defined as a group of problem-solvers that act collaboratively to solve problems which are beyond any single one of the agents' abilities or knowledge [109]. The significant advantages of such systems have led to the growth of research into their capabilities and the construction of applications using them.

2.7.2. Agents Classification

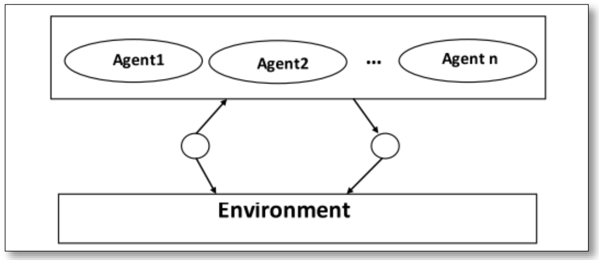

Figure 2-8: Multi-Agents System *[110]*

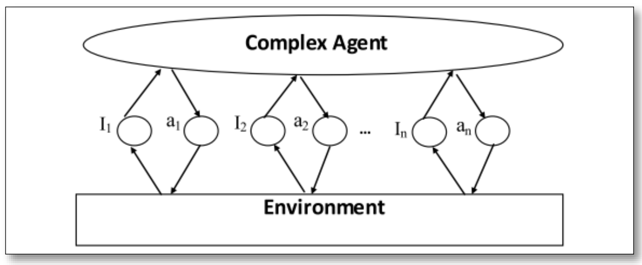

Figure 2-9: Complex Agent [110]

Figure *2-8* represents the structure of a multi-agent system; an agent can perform one or many tasks, and the data held by the agent, obtained from other agents, means that the agent also serves an information function [110]. An important characteristic of the agents is autonomy which means that agents are at least partially independent, autonomous or self-aware. In some systems, the making of individual, autonomous decisions by agents is avoided because of general concern about the input from environments so that decisions reflect a comprehensive view of the input; the agents collaboratively decide on the team's behaviour. They choose optimal actions according to rules of behaviour in an environment.

Another type of agent implemented within multi-agent systems is the complex agent (Figure *2-9*) [110], defined as an intelligent agent able to do numerous tasks concurrently. Such an agent receives input from sensors in the environment, stores this input, identifies the tasks which need to be executed at specific times, and then produces output actions via actuators.

Moreover, Sánchez [111] has proposed a different taxonomy of agents as shown in Figure *2-10*. He assumed that all agents are software agents and can be classified into different types according to the specific tasks they can perform and their capabilities. Programmer agents deal with complex software and hardware entities, network agents act across networks and distributed systems, and user agents are explicitly for utilization by end-users. User agents either collect users' data in order to produce personalised recommendations (information agents), provide visualised characters and objects for user interfaces (synthetic agents) or run synchronously with user applications to watch users' actions and automate specific actions (task agents). There are two classes of task agent: group agents which support computer collaborative tasks and personal agents which support tasks performed for each user individually.

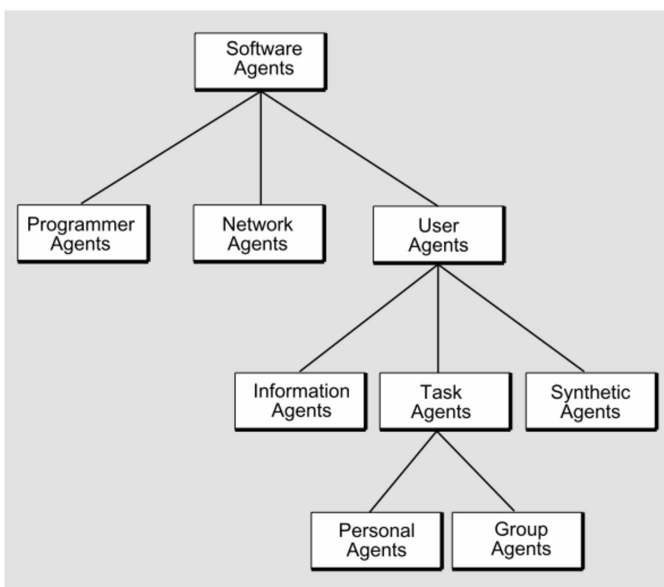

Figure 2-10: Agents Taxonomy [111]

2.7.3. The Application of Agents

One application which focuses on the use of user agents is the intelligent tutor system in COACH [112], which gives personalised recommendations to learners who use a computer to learn about specialist topic areas such as programming languages or operating systems. The note-taking system developed by [113] is another example of the use of a personal agent. The note-taking apprentice continuously attempts to predict the possible endings of notes taken via a touch-screen computer. Users can choose to use the predictions or reject them - so making their note-taking faster and more accurate.

On the other hand, Kautz et al. [114] constructed an environment that utilised a group task agent. The agent helped groups of users work together collaboratively. These researchers designed an environment in which user agents can communicate in order to seek out users who are experts on specific subjects – in order to solve shared issues within collaborative work situations.

An example of an environment that combines both personal agents and group agents is given in the study by Ikeda, Go and Mizoguchi [115]. They introduced a model called ‘Opportunistic Group Formation’ (OGF), based on artificial intelligence and multi-agent systems. Each learner is represented by an agent, an intelligent tutoring system, and this agent negotiates an effective learning group. The agent helps determine the situation of a learner and the extent of their need to join a group; thus, when the student faces difficulties in completing a task, the agent shifts from individual learning (personal agent) to collaborative learning mode (group agent). The study [116] developed a learning goal ontology that plays an important role in negotiating the group formation in OGF. The researchers used learning theories to classify learning goals and cause the formation of groups which were of maximum benefit to learners.

The study [117] created a methodology called the Adaptive Course Sequencing Approach (ACSA) which is a model for online tutoring systems with adjustable autonomy. The implementation of ACSA utilised a number of differing agents and a fuzzy logic analyser. The model includes human-agents, context agents and iTutor agents. The human-agents are the teachers/students - who can adjust and control the criteria/policies of the fuzzy rules and the autonomy level. An iTutor agent is a pedagogical agent that manages learning activities, leads learners according to their knowledge, identifies the autonomy level and enacts the active guidance rules. The context agents tracked learners' behaviours, sent information to student profiles and inferred the appropriate iTutor agent to notify of changes if there were any changes in a student's learning scenario or knowledge level. The system allowed the tutor agents and the human agents to collaborate in improving sequence rules of the tutoring system. Finally, a fuzzy logic analyser was utilised in order to map the input data sets, representing learner activities, to an appropriate class or lesson for the students. All these agents collaborated together in order to suggest appropriate classes for the various learners and courses that they needed to study – according to their knowledge level.

The cases outlined above have explored the use of agents in the MAS approach; this approach is effective in providing services for users. Agents solve problems, make decisions and communicate with each other. In addition, enabling the use of user data allows agents to manage or reason based on data/information stored in repositories.

2.8.Fuzzy Logic Approach

Fuzzy logic (FL) refers to some kinds of expert reasoning translated into a form that is understandable by computers. It is considered one of the techniques for artificial intelligence (AI), where the intelligent behaviour is achieved by creating fuzzy classes of parameters. Fuzzy logic has been applied in expert systems and is the basis of fuzzy expert systems [118]. It can work with logic representations containing linguistic variables and various values such as ‘poor’, ‘average’ and ‘good’, unlike classical logic which deals only with true or false values. The main drawback of classical logic is the limitation that it is constrained to deal only with these two values; these do not readily represent the complexity of the real, non-binary world. FL can be considered a multiple value logic, but with a reasoning logic purpose. Fuzzy reasoning represents the inference of an imprecise but possible deduction out of an initial set of fuzzy conditions.

Why Fuzzy Logic?

Fuzzy logic is able to model some common human reasoning mechanisms; these can be difficult to emulate using conventional approaches [118, 119]. This form of logic can be seen as closer to the way human brains work than strict Boolean logic. In this research, FL was applied because such logic is in some important ways similar to that used in human decision making and reasoning, and such a process was needed in order to simulate the way teachers make their assessments. The data we needed for this simulation was gathered from a large domain of experts (teachers) who provided information about their experiences of assessing students.

FL was the AI method of choice applied in this research as opposed to such methods as machine learning and deep learning. These latter approaches require extensive training sets so that they can build their classification models; however such training sets were unavailable to us in our situation because there were no data for the learning environment in question since it was created for this present research. Moreover, fuzzy logic is better for performing the kind of human reasoning our research requires [118, 119] while machine learning and deep learning perform better in prediction and classification tasks [120-122].

Statistical reasoning methods such as the Bayesian approaches were also considered to be less relevant to this present research than fuzzy logic because FL works better with data which presents the kind of uncertainty that our data does. Berenji [123] described uncertainty as "the lack of complete information." He also stated, "uncertainty may also reflect incompleteness, imprecision, missing information, or randomness in data and a process." [123]. FL can handle uncertainty, ambiguity and vagueness in data, and the presence of these in various kinds of real-life problems is a predictable situation [124, 125]. The study by Baraldi et al. [126], compared between Bayesian and FL approaches and found that FL can represent the uncertainty in input and output data better than Bayesian methods. Because of this strength, it was felt that FL was exactly what was needed for our framework — to represent the uncertainty in students' input data and the uncertainty of assessment (output data).

In addition, FL supports translating expert knowledge into a computer processable form (FL rules) which is nevertheless understandable also by humans. Hence, teachers can readily adjust and add to these rules whenever needed. All the above reasons led us to choose fuzzy logic, over all other methods, as the inference technique to be used within our framework.

Fuzzy Logic System

Mendel [127] presented an architecture for general-purpose fuzzy logic systems; it includes the following process (Figure 2-11):

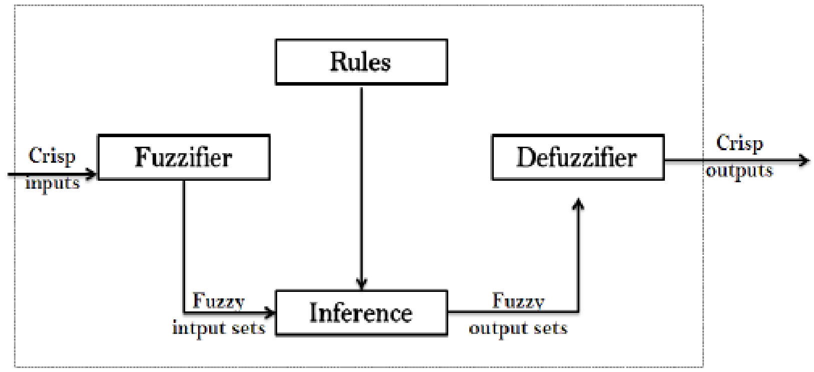

Figure 2-11: Fuzzy Logic System [127]

a) Crisp input: it is the non-fuzzy data which generates the fuzzy inputs.

b) Fuzzifier: it uses membership functions to convert the data (crisp values) into fuzzy values [127, 128]. A membership function (MF) is a curve by which a fuzzy set feature can be identified, giving each component a consistent membership degree or value. It links each element of the input to a membership value within a closed interval unit between 0 and 1 (Figure 2-12) [128].

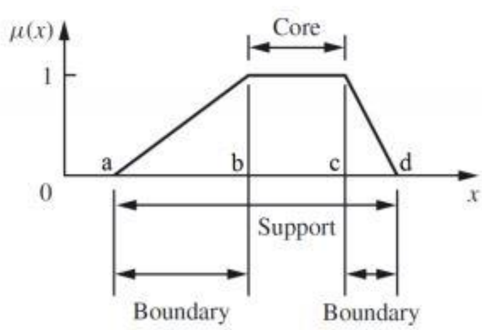

Figure 2-12: General Membership Function [128]

There are five MF shapes in general: triangular, Gaussian, trapezoidal, sigmoidal and generalized bell membership functions [128]. Usually a single membership function will define one single fuzzy set, and many MFs are defined in order to process an input variable comprehensively. For instance, to represent temperature data input, the fuzzy input sets could be: too-cold, cold, warm, hot, and too-hot (Figure 2-13).

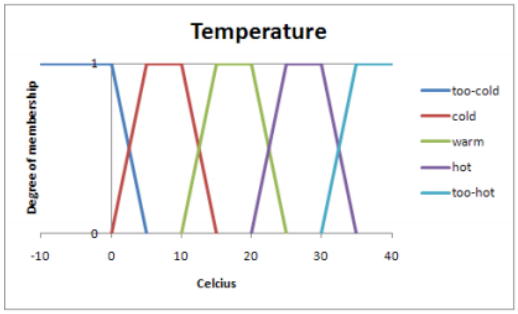

Figure 2-13: Membership Functions for T (temperature) = too-cold, cold, warm, hot, too-hot [129]

c) Rules: fuzzy rules are established using various linguistic labels. These rules take the form of IF-THEN statements which define the relationships between the inputs and outputs that are to be used in the inference process *[127]*.

d) Inference: the inference mechanism calculates the required firing strength of each rule for each circumstance, so that it can be decided whether a rule is to be considered “fired” in response to a specific input, generating fuzzy output sets.

e) Defuzzifier: this process converts the output of the inference into a numeric value. This is the process whereby the final output is determined, using a defuzzification technique. After completing the inference decision, the fuzzy number must be transformed into a “crisp” value; this process is called defuzzification.

f) Crisp output: it produces the final output data.

The Application of Fuzzy Logic to Student Assessment

Fuzzy logic has been applied in a number of learning environments in order to facilitate student learning and assessment. In [130], Subbotin and Voskoglou developed a new triangular fuzzy model (TFM) to assess the critical thinking skills of K12 students. Their model was based on the centre of gravity (COG) defuzzification technique. This fuzzy model replaced the rectangular membership functions of their previous COG model. The authors succeeded in solving cases where students' scores were ambiguous in relation to the two grades (A and B) available. Their experiments were undertaken in Los Angeles schools and they were focused on connecting students' CT skills with their language competencies.

Another example of using FL for the assessment of student performance is provided by the Yadav and Singh Fuzzy Expert System [118] which applied fuzzy inference mechanisms and rules to assess student academic performance. The authors proposed several approaches towards the creation of a practical technique for evaluating student academic performance using fuzzy logic methods. They compared their results with the classical statistical approach which is currently used in education. The statistical approach is based on mathematical rules for calculating learners' scores and grades while the fuzzy logic approach provides great flexibility and reliability. They had two fuzzy scenarios: the fuzzy-1 scenario whereby the membership functions were the same for all the semester exam results looked at. In contrast, in the fuzzy-2 scenario, the quality of the membership functions was improved to work better for the second semester's exam results. The study found that fuzzy logic is suitable for evaluating student performance in theoretical lessons and in lessons which used e-learning systems and not only for laboratory learning. However, the two scenarios looked at were limited since they dealt only with students’ scores in semester exams and they did not assess any other aspects of learning such as learning skills or competencies.

FL has also been applied in online learning environments to create adaptive systems that change the content of interfaces, based on students’ skills and/or knowledge. In the research [131, 132], the authors created an adaptive virtual learning environment. They applied quantitative measures, via an FL approach, to the assessment of students’ learning of skills and then changed the content of the 3D environment based on the assessment results, giving the students a sense of one to one tutoring. The system which was developed provided customised materials for each type of student - to improve learning. A function was created which calculated the level of skill that a student had attained based on the number of errors, the time taken on tasks and the students’ test scores, and then they provided the student’s grade in terms of good, average or weak. In order to determine a student’s level, the system gave the students a test at the end of each learning unit. Then the result of this was input to the FL model so that it could make decisions about the content provided to each student. The following Figure 2-14 describe their system architecture. While the authors showed the proposed framework was effective, there were some limitations to their research. The function model that they developed accepts only three quantitative variables - number of errors, time and test scores - to calculate the level of skills acquired in the learning environment; however, other variables are required if one wants to obtain greater insight into the quantity and quality of skills that students learn.

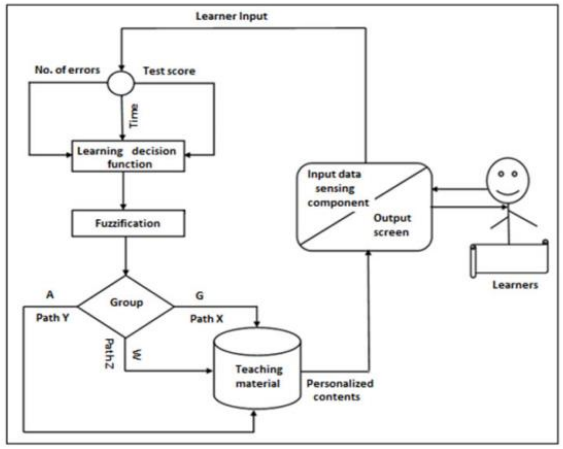

Figure 2-14: System Architecture [131, 132]

In addition, the research [133] built personalised 3D-MUVEs based on its model of learners. The application which was developed as an integration of the three systems: *OpenSim*, *Sloodle* and *Moodle*. To evaluate the type (model) of the students, a fuzzy logic approach was applied. It was found that there were 16 types of learners based on their learning models and each type can be divided into three sub-types or levels: high, medium and low. Also, there were learning materials created specifically for each of these levels. The fuzzy value (student type) was calculated from the answers to a questionnaire; from these answers, the specific student type of the respondent and their degree of membership (high, medium or low) was determined. However, the generation of the learning materials for this study was done manually, and this process should be automated if this kind of approach is to result in practical improvements to the learning environment. In addition, the evaluation of a student's type was based on questions and answers, while students' performance in the activities should also be evaluated when deciding on the learning material to be presented.

The study by Luis et al. [134] exhibits another example of the use of FL in learning. It suggests a model using expert systems including machine learning and fuzzy logic mechanisms for evaluating students' non-verbal avatar activities within a collaborative environment. The nonverbal interactions were looked at in order to understand the collaborative attributes which were being demonstrated: students' individual characteristics such as influence or involvement, and the group's characteristics such as cohesion - when completing a task. The research model was to analyse row data in log files using a multilayer process - data filtering, fuzzy classification, and then rule-based inference - in order to produce an assessment for each group. The study presented methods for managing, translating, analysing and inferring from data, using filtering, fuzzification and two semantic inference stages. The model was also able to apply customised rules, based on various different experts' opinions and knowledge, to increase the efficacy of the rule-based stages as required and to add semantic layers to the outcomes. However, the study presented this work in its early stages and the analysis was performed offline after collecting log files from the user accounts — whilst evaluating students in real-time is important in order to provide students with instant feedback about their work. Also, as has been said, capturing users' data without first identifying how this to be scored is not an effective process in terms of creating accurate assessments. According to Tesfazgi [100], it is better to build the learning environment from the start around the requirement to collect learning evidence and assess students' learning in real-time.

Chrysafiadi and Virvou [135] created a model that examines students' knowledge, applying a fuzzy logic approach in order to identify the knowledge level of each student (the topic being programming concepts) as users interacted with the environment. Their system was used by postgraduate students at the University of Piraeus to learn how to program in the C programming language. The authors evaluated the accuracy and effectiveness of this modelling of learners taking an online programming course. The authors argued that there are no clear guidelines in the literature in terms of evaluating student models in relation to Intelligent Tutoring Systems (ITSs), so they chose to apply well-known general evaluation methods to evaluate the fuzzy student model and to develop an accurate evaluation methodology. Thus the study chose to use Kirkpatrick's model and the layered evaluation method [135] to validate the effectiveness of the system and demonstrate the efficacy of the system's decision-making. The experiments involved separating the students into two groups, one group using the system without fuzzy logic and the other group using the system developed utilising FL. Subsequently, they gave the students questionnaires which asked about the effectiveness and adaptivity of the system. After performing their experiments and comparing the results, the researchers found that applying fuzzy logic in relation to a learners' model can lead to the enhancement of the students' performance and the system's adaptivity. It also increases the validity of the system's decisions.

From all the previous work that has taken place, we can conclude that fuzzy logic is a significant and effective approach to evaluating and assessing students across a number of different learning environments. It is also well-suited to the purpose of our current research which seeks to use the idea of agents to produce a system that reasons in something like the same manner that a human being does.

2.9. Summary of the Literature Review

Numerous technologies which have emerged, such as immersive environments, have been designed to support and enhance learning experiences. 3D VWs provide examples of immersive learning environments that have been widely applied to connect scholars and educators, to provide many different kinds of activities, and to enable participants to access learning sessions remotely. These kinds of educational spaces are increasing in popularity due to the many features that distinguish them from other online systems. They connect students in real-time and so enhance interactivity, exploration and engagement. Moreover, they facilitate the investigation of ideas, situations and places that cannot be accessed physically; they deliver learning processes; provide realism of interaction; enable discussions and activities around even the most complex of topics in a straightforward way with less cost. In addition, they facilitate collaborative learning which can help students to work with their peers, resulting in the improvement of learners' knowledge and experience, generally.

The assessment of students is a critical aspect to consider when organising learning activities in 3D VW spaces. Learners acquire new information or skills while practising individual or collaborative activities. There is extensive coverage in the empirical literature of the merits of appraising students in real world classrooms; however, there is a lack of research concerning observing and assessing students in virtual worlds. Thus, this thesis aims to overcome this limitation, exploit the affordances of 3D virtual worlds and investigate the ways in which students can be assessed, in VWs, in order to enhance their learning effectiveness.

An advantage of virtual worlds is that they are able to capture details concerning students' actions automatically, in a way that would not be possible in the real world. However, straightforward technologies used to collect learning evidence are not capable of tracing and capturing the entirety of a learner's knowledge and skills, so more sophisticated evidence collection methods are needed to record the real-time learning outcomes which are exhibited when learners behave collaboratively and to capture learners' performance from the activities they engage in within 3D-VWs. Consequently, this study looks to MAS and FL methods to enhance learning assessment in such environments.

MAS techniques provide greater services for users because agents, together, can solve problems, make decisions and communicate with each other. In addition, enabling the use of users' data allows agents to manage or reason based on this information. Also, fuzzy logic is an important approach to the evaluating and assessing of students in different learning environments. Determining and assessing students' learning and skills is not a straightforward task. Many factors that cannot be measured or observed directly must be taken into account. Students' learning and skills are not fixed variables, evaluating them mean that it is necessary to deal with human subjectivity and uncertainty. Therefore, FL is well-suited to the purpose of our current research which seeks to use agents to produce a system that deals with uncertain values and reasoning in the same manner that a human would. This thesis describes a novel approach to assessing learning taking place in collaborative groups within 3D-VWs. It introduces a computational mechanism that extends the MAS method by combining software agents and user evaluation with FL technology to support the identification of learning evidence from collaborative activities. Such a computational framework can help to solve the issues that may occur when collecting learning evidence and assessing learning outcomes.

Therefore, the next chapter (Chapter 3) presents this computational framework - which can dynamically recognise users' behaviour, collect learning evidence data and analyse events to measure the learning outcomes exhibited. This newly created framework is called Mixed Intelligent Virtual Observation (*MIVO*) which mixes various learning and computational models for observing and evaluating collaborative learning in immersive environments.

Chapter 3

3. Research Framework

"You can observe a lot just by watching."

—The philosopher Yogi Berra

The literature review of Chapter 2 (section 2.6) concluded that work on observing and measuring online collaborative learning outcomes dynamically and on the fly within 3D virtual worlds is scarce. Also, it concluded that there is a need to find event detection methods which can dynamically collect learning evidence and analyse events in order to measure the quality and quantity of learning taking place. Such a mechanism could help to identify proof of learning in collaborative activities within immersive environments and correlate the evidence with the specific learning objectives in place in order to assess the overall outcomes of the learning processes. Gardner and Elliott [13] stated that "learning within technology creates a pedagogical shift that requires teachers to think about measuring outcomes in non-traditional ways". As a result - by looking at the learning theories; the assessment and observation frameworks; and the multi-agent and inference approaches available (see Chapter 2) - we have been able to propose a Mixed Intelligent Virtual Observation conceptual framework (*MIVO*), as mentioned in our papers [136] [137]. This framework combines both learning and computational elements for the purpose of observing and evaluating collaborative learning in VWs. It consists of many components: user interface, pedagogical unit, virtual observation, inferencing, a data unit and assessment presentation (*Figure 3-1*). These components are discussed individually in the following section.

3.1. Conceptual Framework

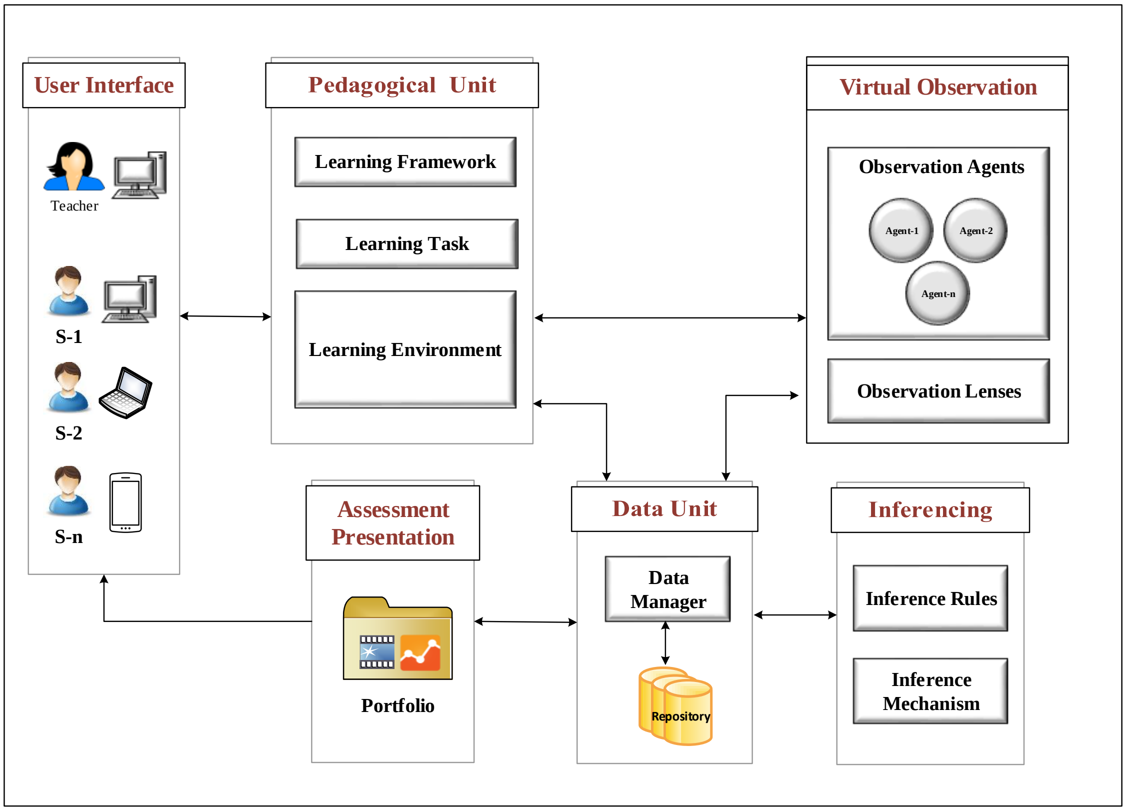

Figure 3-1: Mixed Intelligent Virtual Observation Conceptual Framework for Collaborative Learning Environments (*MIVO*)

3.1.1. User Interface

This element concerns who the users are and their roles in the learning environment. Users can be either learners or instructors, and the form of user interface displayed to them is based on their identities and roles. Instructors could have a customisable interface that allows them to design learning activities. Moreover, they can view learners' assessments and portfolios in order to evaluate their (the learners') performances and review their work.

From the learners' viewpoint, the user interface enables them to interact with the environment and with other students' avatars. All learner participants may work together on the simulation learning activities available in the 3D environment. They can participate in the activities, evaluate other learners, obtain learning feedback from the system and view their own portfolios.

3.1.2. Pedagogical Unit

The pedagogical unit used here consists of three components: the learning framework, the learning task and the environment within which the collaborative learning practices will take place.

- **Learning Framework:** When designing learning activities for virtual worlds it is important to base them on a well-known learning framework. In the pedagogical unit, teachers can specify the learning framework that should be followed in order to accomplish the learning objectives. Moreover, teachers can specify the observation criteria whereby student learning can be evaluated. A number of different learning frameworks have been developed, such as Bloom's taxonomy which describes the levels of learning objectives as a hierarchy [138]. Bloom's taxonomy is as follows: knowledge, comprehension, application, analysis, synthesis and evaluation. Another example of learning frameworks is the Mayes and Fowler framework [139] which defines the learning stages of courseware: *conceptualization, construction and dialogue*. Section 3.3 details the learning framework that has been applied in this research.
- **Learning task:** The learning task defines the learning scenarios and activities that can be performed by students in the learning environment. This task can be planned in advance and then adjusted by teachers. Some online learning systems have used the Instructional Management System (IMS) learning design [140] to support the production, design and sharing of online learning activities - that learners can work on within e-

learning sessions. IMS also allows teachers to describe the structure of the learning objects, including the learning objectives, activities, tasks, and learning outcomes.

- **The learning environment:** This is the environment that facilitates students to collaborate together to undertake the learning activities designed by the instructors. A learning environment may be immersive such as those represented by virtual worlds generally are, support collaborative learning and enable individual students to interact with their peers in ways which will allow them to develop new skills and share knowledge. Virtual worlds encourage students to engage with each other in ways that improve teamwork and decision-making by bringing them together in real-time [85]. Sharing ideas and knowledge between group members can help individuals to gain a better understanding of complex phenomena and the relationships between learning objects [3].

3.1.3. Virtual Observation

The observation unit determines the methods which are to be used to observe and evaluate students taking part in the virtual world. It encompasses two subsidiary models: the observation agents and the observation lenses models. In the following sub-sections, these models are briefly introduced - then 3.2 explains the usage of these models in relation to this research.

- **Observation Agents:** This defines the agents required to monitor students and observe their progress in the learning environment. The model describes the mechanism which will be used for collecting data in the VWs - to better understand the learning outcomes of groups and individuals. Multi-agent systems consist of several agents who act independently by using inputs from the environment to produce actions as outputs. The nature of these agents mostly dictates that they have specific tasks and capabilities. All the agents here have a shared goal, which is to provide learning data, but the differences between them reside in the fact that each agent sees and records the events which emerge from the collaborative activities based on their particular abilities. At the same time, all the agents collaborate together to achieve one goal and that is the capturing of learning evidence that supports the evaluation of learner performance. More details about the agents' model are given in section 3.2.1.

- **Observation Lenses:** The lenses determine how the data captured by agents can be analysed. It was seen that the literature has generally concluded that observing students in classrooms requires the consideration of numerous criteria, aspects and frames in order to acquire sufficient insights into the students' learning and so be able to improve the quality of their education. According to the observation framework [44], teachers can adopt particular 'lenses' in order to evaluate and observe students. So, to do this in VWs, the observation framework is designed and utilised to evaluate that which can be monitored in these environments. For example, the learning interactions lens, which is a "*frame*" or "*lens*" of the observation framework [44], looks at what social interactions take place between students in the classroom and evaluates the environmental interactions which occur between the students and their surroundings. Another lens that may be utilised is the student success lens; this assesses students' success by counting the number of correct answers, the number of these which have been reinforced or acknowledged, and the number of delayed corrections. Based on the assessment requirements and objectives of the learning activities, educators can adopt any of the observation lenses available and employ them in the classroom. This research has extended the use of these lenses by employing them in non-physical, 3D virtual environments in order to observe students' learning and assess this. More details concerning the immersive environment lenses can be found in section 3.2.2.

3.1.4. Inferencing

This item consists of two components: the inference mechanism and inference rules.

- **Inference mechanism:** It is necessary to adopt an inference method in order to make sense of the data collected by the observation agents. Such can be used to discover relationships between the data and its meaning in order to infer further evidence of learning. It will also facilitate the making of decisions regarding the assessment of students and their performances.
- **Inference rules:** The inference rules analyse and return conclusions about their data. They can translate the conditions returned by the observation lenses and also transfer the human experts' communications concerning their student evaluations – both into a logical form - so that the system can perform in the same way as a human expert.

Applying an inference mechanism and rules is essential to our observation framework. Chapter 5 details the inference method used in this research.

3.1.5. Data Unit

The data unit consists of a data manager and its repositories. The data manager controls the moving of data to/from the learning environment and other units. The immersive environment frequently generates data yielded from users and agents, and these data are subsequently sent to the data unit to be saved in a repository so that they can be processed to generate assessment scores. Also, the data is transferred to/from the inference unit when necessary. After the data is processed, it is sent to the assessment presentation unit and hence presented to users so that they can view student performance.

3.1.6. Assessment Presentation

The last unit contained in the *MIVO* framework is the assessment presentation. This describes how the learning outcomes are presented, for the purpose of demonstrating the assessments of individuals and groups, to both instructors and learners. Via the evidence gathered by agents and the inference methods applied, the observation model facilitates the analysis of the learning outcomes from the activities and from the correlation of these with the learners' feedback. This unit is also responsible for the representation of the assessment feedback via dashboards – which show when performance is high and when it is low. These dashboards can be accessible to the users (both instructors and students) so that group and the individual evaluations can be made. Furthermore, this unit underlies the presentation of the evidence of learning through video snaps that map video recordings to the timestamps of items of evidence. All these evidence can then be correlated to the learners' portfolios and gathered for further educational purposes. One way of assessing the students' learning is via their learning portfolios. A student's portfolio is an assemblage of work and linked materials that represent accomplishments and activities. The portfolios can contain learning evidence, self-evaluations, the strategies used (by the student) to choose content and indicate standards by which to evaluate the quality of the work [141]. The goal is to gather evidence that demonstrates the learners' skills, talents, capabilities and achievements. Thus, through the assessment presentation unit, students can build their online portfolios, and these will enhance the learning affordances of the immersive environment and facilitate the visualisation of learning evidence and outcomes.

3.2. The Virtual Observation Conceptual Models

The *MIVO* framework was proposed as a comprehensive system for observing students who are using online immersive learning. The most important component of the framework, and the contribution of this research, is the construction of the models which facilitate the observation and assessment of collaborative learners using online immersive environments. Therefore, the focus of this research is on this, the virtual observation unit, which in turn consists of the observation lenses and the observation agents. The following sections explain the methods used for each model and how this present research utilises them in the immersive environment.

3.2.1. The Agents Model - Mixed Agents Model (*MixAgent*)

The virtual observation model maps between the observation of learners (to assess how they perform) in conventional classroom settings and such observations made in 3D-VWs. It dictates that there be a way of combining computational agents in order to replicate how an instructor would monitor her or his students and observe progress in a classroom setting. This research demonstrates a mechanism whereby data relevant to this purpose can be collected in VWs - to better understand the learning outcomes of groups and individuals. In order to record learning events, and overcome some of the limitations discussed in the literature review, the MAS method is extended by adding ‘natural’ agents to the software agents. These natural agents consist of the learners taking part in the tasks. In the *MixAgent* model (Figure *3-2*), the data from the agents are sent to the inferencing system to identify the learning evidence they contain and so to assess the learning being achieved by each student. The *MixAgent* model was first introduced in our study [142] and subsequently it was discussed in detail in [143] [137]. The following section describes the capabilities of the agents, including their specific assessment roles:

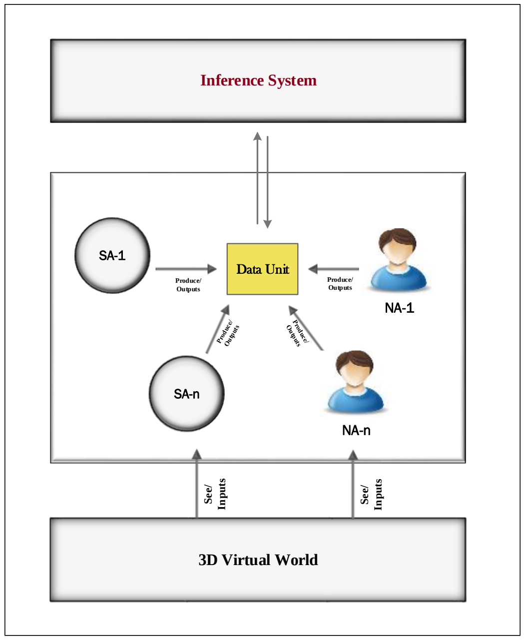

Figure 3-2: Mixed Agents Model (*MixAgent*).
Abbreviations: SA = software agent; NA = natural agent.

a. Software Agents (SA): When a user has received authorisation for the virtual world, a software agent (SA) is assigned to them. SAs are able to monitor the activities of individuals in real-time, collect information about learning events and send data to be saved in the repositories – so that it can be transferred to the inference system. Such an agent may also be called a user agent as mentioned in [111, 144], however, we chose to call it "software agent" to differentiate it from the following kind of agent.

b. Natural Agents (NA): In this study, we expand the idea of multi-agents from its original conception – of being only concerned with software agents - to include human agents. The learners' own intelligence should not be ignored as a means of producing better learning evidence for evaluation. For this reason, learners are regarded as natural agents who, among other things, can monitor the behaviour of other learner avatars and produce action output which represents evidence about learners' performance. Thus, they can behave in a similar way to software agents which observes users' actions and can provide information about those users' learning outcomes as output. Natural agents can assist in measuring the quality of learning outcomes; such measurements or evaluations can be difficult to achieve when relying solely on automated approaches.

Such agents are employed via a process of peer evaluation; this is particularly suitable in relation to group exercises [34] and can provide insights that conventional technology would struggle to pick up on [90]. The learners can rate the skills and qualities of the performance of other learners within the group setting. When learners are working together on assigned tasks, they are enabled to rate each other's performance via a rating tool. These quantitative scores are compiled and are then transferred to the system for reasoning. The data from the NA is recorded in the repositories and then the system uses them in the inference system.

Agents are given common objectives and they are then asked to collaborate in real-time so that data can be amassed and then be transferred for inferencing. To make sense of the data collected by the agents, one basis on which to build the inference system was fuzzy logic, and this was the basis used here. Once the data is collected from both types of agents and placed in the repositories, it can be analysed using inference rules in order to shed light on the performance of individual learners. Fuzzy logic methods can handle multiple values and perform human-like reasoning, going someway to providing a unified vision of agency within the agents' model. Also, using such method can act as a bridge between agents operating in the learning environment, so the data collected from agents can be analysed by employing fuzzy logical rules that are of use in retrieving learning evidence. Applying inferencing facilitates reasoning from the data which has been accumulated and enables the acquisition of more meaning from these data, gathered by each agent; the inferences made will help later when making decisions regarding the assessment of students and the improvements (or lack thereof) in their profiles. Chapter 5 provides more details on the fuzzy logic reasoning employed by the system in relation to practice in the research prototype.

3.2.2. Observation Lenses Model (The *OLens* Model)

The *OLens* model determines how the data captured by the agents can be analysed. The literature reviewed generally concluded that to observe students in classrooms, educators should consider numerous criteria, aspects and frames in order to gain more insight into the students' learning and so be in a position to improve their education. However, not all the learning outcomes and skills mentioned in Chapter 2 can be easily observed and identified in immersive environments. For example, creative thinking, body language and emotions are significant aspects that evidence the quality of learning taking place, but it would be very problematic to record and evaluate them in relation to virtual spaces.

Depending on the observation framework [44] used, we employed differing ‘lenses’ to the research model and extended them for use in 3D-VWs in order to evaluate what could be monitored in these environments. The *OLens* model defines the granularity levels employed when observing students and recording evidence of collaborative learning: from high-level to low-level observation (*Figure 3-3*). This *Olens* model was introduced and discussed in our papers [136] [137] [145].

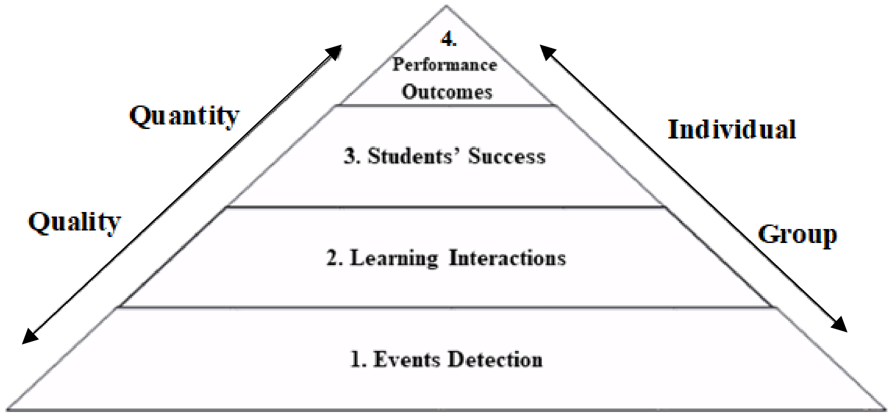

*Figure 3-3: Observation Lenses Model (*OLens *Model)*

1. The Events Detection Lens

This refers to 'arms-length' observations, such as when a teacher observes collaboration between students without paying great attention to what is occurring in relation to a specific task. In the virtual world, this equates to recording a sequence of events without making any attempt to interpret these processes. This lens is providing the interface to the other *OLens* layers and it is structuring the events data so that this can then be analysed by the other layers. The aim is to acquire the details of all implicit and explicit learning events relating to the students. This implies the recording of actions and the storage of user logs for use with other lenses where greater attention is paid to analysing and interpreting learner actions.

In consequence, the *MixAgent* model illustrated previously is employed by the Event Detection Lens because it facilitates the detection of the learning events in real-time and the accumulation of evidence of learning in order to grade how students perform in terms of quality and quantity when participating in virtual worlds. Software agents automatically observe how students behave, capturing their behaviour events in system logs and converting evidence of actions into data that can be stored in an underlying repository. Students perform the role of natural agents; each student is able to observe the quality of their peers' performance and provide their opinion on each other's performance and skills. These data are also stored in the repository, and the data gathered by both the natural and the software agents are subsequently interpreted within the *OLens* model. The inference system interprets the data, inferring how each individual user is performing. Crucial learning evidence can be identified based on these inferences, and this can lead to a clarification of the relationship between the data and its underlying meaning. (In this context this is the associated learning performance of each user.)

2. Learning Interactions Lens

More thorough observations are made by this lens. Its purview includes observations of social interactions between students and the environmental interactions between the students and the virtual world. Not only do these observations relate to the quantity of interactions but also to their qualities; thus making it possible to infer which students are making the most valuable contribution to the group. The number of interactions is recorded, in terms of the number of interactions by each individual and also the number of interactions by the group as a whole.

Since this lens focuses on more thorough observations, it needs to query the data amassed, in order to evaluate each participants' interaction when engaging in the immersive environment. Table *3-1* below offers examples of the quantity and quality indicators that can be used to assess the contributions of participants and their interactions with other students in the virtual world.

	Quantity Indicators	Quality Indicator
Individual	-The volume of communications and actions of a student.	-The average rating scores for a student, as received from other members, re the quality of interactions.
Group	-The sum of all the communications and actions undertaken by all the members of a group.	-The average rating score relating to all members of one group, concerning the quality of the interactions.

Table 3-1: Interactions Indicators

3. Students' Success Lens

Teachers are able to observe and assess individual learners within real world classrooms; this lens extends this behaviour to the virtual world. Success can be interpreted as the ratio of correct to incorrect responses that learners provide to a series of exercises, questions or assignments [44]. Therefore, the success lens aims to mimic the ability of teachers in the real world to evaluate the learners' success via their (the learners') responses. Table 3-2 below provides examples of the indicators that can be used to determine the success of a group or an individual when attempting a task.

	Quantity Indicator	Quality Indicator
Individual	The number of completed/ incomplete tasks relating to an individual.	The average rating scores concerning the quality of a student's work when completing a task.
Group	The number of completed/ incomplete tasks attempted by the group.	The average of the quality rating scores relating to the members of the group.

Table 3-2: Task Success Indicators

4. Performance Outcomes Lens

Using this lens, learners are observed in greater detail in order to identify the results of them taking part in learning activities. There are various types of performance outcomes that can be assessed. That is to say, performance outcomes can be interpreted as what a student knows, understands, or can do to perform a learning task [146]. As such, these outcomes relate to the student's overall knowledge, skill and competence levels. It is possible to interpret student data in order to evaluate certain skills and competencies. This lens goes beyond merely counting the number of correct answers and instead provides a summative evaluation of the quality and quantity of performance outcomes. Multi-user virtual environments (MUVE) are usually used for collaborative learning exercises and, therefore, due consideration should be given to the participants' collaborative skills. There are a number of distinct collaborative skills which can be examined, such as leadership, communication, creative conflict and the maintenance of trust [147]. Table 3-3 offers instances of learning outcome indicators that can be used for evaluating the skills of individuals by this lens.

	Quantity Indicator	**Quality Indicator**
Individual	The quantitative element of a student's collaborative skill - such as participation and/or communication.	The student rating scores of the collaborative skills related to qualities such as negotiation and reasonability.
Group	The sum and/or average of the above quantitative element, relating to all the members of the group.	The sum and/or average of such scores, relating to all the members of the group.

Table 3-3: Performance Outcomes Indicators

Each level on the *OLens* is explored to determine how physical observation can be simulated in immersive environments. These frames should measure the individual's and the group's performance and the quality and quantity of each learning outcome. In effect, using the *OLens* facilitates the measurement of student performance across a variety of learning outcomes. The next chapter (Chapter 4) provides more details about the research prototype that was developed to implement the *OLens* and *MixAgent* models.

3.3. Learning Framework

When dealing with learning activities in immersive environments, it is very important to have a learning model that can assist in designing and organising learning activities within such spaces. It is essential to follow a learning framework in order to build student activities. One of the well-known learning frameworks is the Mayes and Fowler framework [139]. Mayes and Fowler introduced an important courseware framework for identifying the cycle of learning phases (Figure 3-4) [139]. The identification of these phases helps us to understand the steps that a learning process goes through and enables us to comprehend how to structure learning activities which are based on the use of technology. The learning cycle processes are as follows: conceptualisation, construction and dialogue [139]. The acknowledgement of these stages can assist in organising learning activities which are to take place in a virtual environment as follows:

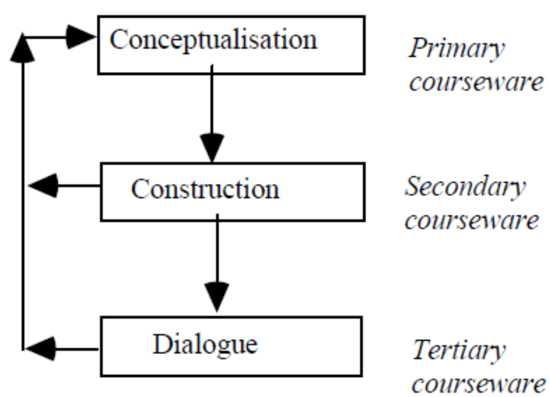

Figure 3-4**:** A Framework for Understanding Courseware [139]

- **Conceptualisation:** This relates to the learner’s first communication with other users’ concepts. It includes the first contact between the student’s own framework of understanding and a new explanation from others. In terms of the learning activities, instructors first briefly explain to the students the collaborative environment and the tasks that they need to complete.

- **Construction:** This denotes the procedure of combining and creating concepts for use when undertaking meaningful practical tasks. Usually, these are tasks like writing essays, laboratory activities, programming projects, etc. The results of this process are products such as notes, essays, programs or laboratory reports. In the learning activities they have chosen to involve themselves with, the learners will start to work together to solve the tasks.
- **Dialogue:** At this stage, conceptualisations are verified by testing and are developed through discussions between educators and learners and between learners and their peers.

We designed the learning activities in the immersive environment which were based on the Mayes and Fowler framework; more details about the learning tasks will be given later in Chapter 6, section (6.4.2).

3.4. The Elements of *MIVO*

This conceptual framework has been created in order to support the observation of collaborative groups in 3D virtual environments. In order to assess collaborative learning activities, the following elements should be taken into account when utilising the *MIVO* framework:

- Environment - this refers to the learning space that users perform within, this can also be designated as the 3D virtual simulation or virtual reality space.
- The learning design - this refers to the type of learning that the students are going to receive (formal or informal), and the learning tasks, objectives, and learning outcomes that the instructors/teachers must construct.

- Collaboration - this refers to the number of users who will participate in the learning task - two learners or more. It also refers to the collaboration procedure and the communication tool which can be employed for communication between learners.
- Agents - this refers to the types of agent that will be utilised in the environment in order to collect the evidence of learning.
- Context - this refers to the data and evidence which is observed by agents: actions, conversations, movements or answers.
- Inferencing – this refers to the inference method that is applied to make sense of the data collected by the observation agents and to enable decision making regarding student assessment.
- Assessment - this refers to the assessment indicators and lenses defined by the educators in order to evaluate individual and group performance.
- Assessment presentation – this denotes the presentation of the feedback derived from the system. This could be instant feedback, summative feedback at the end of the learning activity, or it could be in the form of learning portfolios wherein learning evidence has been collected.

3.5. Chapter Summary

Chapter 3 introduced the novel conceptual framework, the Mixed Intelligent Virtual Observation framework (*MIVO*), which combines learning and computational elements in order to support the observation of, and the evaluation of, collaborative learning in immersive environments. The framework comprises six components: the user interface, the pedagogical unit, the virtual observation, inferencing, the data unit and the assessment presentation. In particular, this chapter has focused on the virtual observation model which includes two significant components: the *MixAgent* and the *OLens* models. The *MixAgent* model defines all the types of agents which can be used to recognise events in real-time – to gather learning evidence and assess student performance in collaborative learning environments. It also utilises a fuzzy reasoning approach as a mechanism to combine the generated data from the agents and infer the learning outcomes which have been achieved as a result of the collaborative activities.

In addition, the *OLens* model comprises four lenses; these lenses appraise the learners' performance from a variety of perspectives. *OLens* provides the granularity levels and the details of what it is possible to observe and assess in VWs. It also specifies the learning evidence which can be derived from collaborative learning and identifies the indicators whereby the learners can be assessed (in relation to each lens). First, the definition of the event detection lens identifies the methods for gathering information about events by the automated and the natural agents - to mimic teacher observations which take place from a high level of generality. Secondly, the learning interactions lens concerns the environmental and social interactions between learners and the virtual world. Thirdly, the students' success lens focuses on the degree to which students have succeeded in doing collaborative tasks. Finally, the performance outcomes lens mimics the more detailed kind of teachers' observations which can examine more thoroughly the results of learning activities. Many types of learning outcomes can be assessed but this research evaluates, in particular, certain collaborative skills and competencies.

It is important to monitor the progress being made by individual students in order to confirm whether or not their learning objectives are being achieved. Moreover, this monitoring process may inform the instructor about any improvements that could be made in relation to the collaborative learning task itself. For these reasons, the *MIVO* components have been applied in a 3D virtual environment as a proof-of-concept prototype to provide superior insight into the application of the models.

Accordingly, the next chapter (Chapter 4) provides further details regarding how these lenses and agents can be applied in practice in virtual worlds. It also offers a number of examples to illustrate how the pedagogical lenses can be implemented in order to gauge the performance of students in virtual worlds.

Chapter 4

4. Mixed Intelligent Virtual Observation Prototype

"The computer is incredibly fast, accurate, and stupid. Man is incredibly slow, inaccurate, and brilliant. The marriage of the two is a force beyond calculation."

—Albert Einstein

The previous chapter introduced the conceptual framework to be used for this research and the research models that have been created for it, aimed at observing collaborative learning activities which take place in immersive environments. This chapter describes the experimental work relating to the creation of the proof-of-concept prototype. The chapter also discusses the practical work which has been carried out within the learning environment, incorporating the research models via three experimental phases. The first phase was to develop a 3D virtual world which incorporated evidence collection agents. The second phase was to implement the observation rules which were to be used; these are based on the *OLens* model. This was connected with the system agents so that the lenses could then be examined. The third phase was to use this constructed virtual observation system to observe collaborative learning activities and then to construct the assessment presentation unit, including the feedback dashboard interface.

4.1. Phase 1: The Implementation of the 3D Virtual Environment with Agents

The first phase comprised of creating the learning environment interfaces to enable users to practice collaborative activities. This phase also involved implementing the agents in the *MixAgent* model to collect learning evidence and assess students' performance.

4.1.1. The 3D Virtual Environment

To illustrate how the *OLens* model can be applied in practice, we have used the Interreality Portal [52]. This is a 3D VW created at the University of Essex as a PhD project. The Interreality Portal was originally developed in order to allow students to engage in collaborative learning activities; the subject area within this system operates the use of embedded systems in smart homes. This learning environment was created using *Unity*[9], a cross-platform game engine that can be used to build 2D and 3D VWs including multi-user games; the environment also supports JavaScript and C# routines. *Unity3D* was used because of its flexibility and its functionality in terms of creating a virtual world.

The Interreality Portal is an ideal platform for supporting collaborative learning endeavours; it requires participants to become involved in 'hands-on' learning activities. The original portal uses a 3D BuzzBox model (Figure 4-1) which has various sensors and actuators, together simulating a smart house. Typically, students are required to engage with their fellow participants in programming the actuators and sensors by establishing IF-THEN-ELSE rules in real-time. To do so, they are provided with access to a collaborative programming board and a number of icons; each of the latter represents either a part of the IF-THEN-ELSE rule or a sensor or actuator (Figure 4-1).

[9] https://unity3d.com/

Figure 4-1:Graphical User Interface (GUI) – InterReality Portal *[52]*

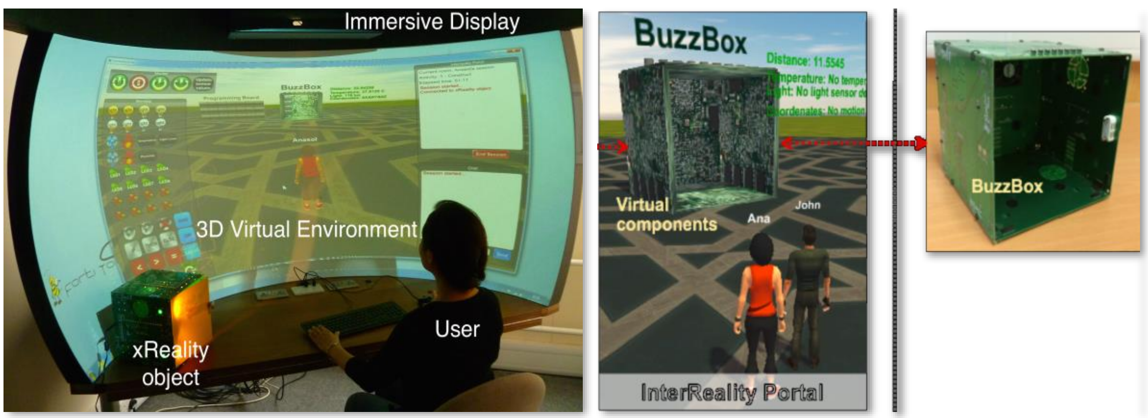

Figure 4-2: The BReal Lab *[52]*

The virtual BuzzBox was designed as a simulation of the real-world BuzzBox for the research purposes described in [52] (see Figure 4-2). However, in this present research, to maximise the affordances of the 3D virtual world and engage students more thoroughly in the learning activity, we exchanged the virtual BuzzBox model with a smart house model. Replacing the 3D BuzzBox model with this newly designed virtual 3D house helped users to immerse themselves more thoroughly in the virtual world. Users can view their avatars acting inside a virtual smart house and they can control the house's appliances. This graphical simulation also assists students to understand the important concept of smart house sensors and actuators, especially participants who were not from the computer science department.

The 3D model of the smart home was designed using 3D Sketchup Warehouse[10] and then imported to the learning environment (*the InterReality Portal*). The necessary alterations to the virtual world were made by changing the game scripts. Virtual sensor values such as home temperature, light level and avatar distance are displayed to participants so that they can use these in managing the home appliances. These latter are controlled by various virtual actuators, distributed around the environment in order to make the learning activities more realistic. The virtual actuators are to control LEDs, lights, heaters, fans and sound systems. The presence of these sensors and actuators enabled the students to create and run programs and then see the effects that these programs had on the virtual home's devices. Figure 4-3 illustrates the 3D modelling of the smart home.

Figure 4-3: Screenshots of the 3D Virtual Smart Home

[10] https://3dwarehouse.sketchup.com/?hl=en

The Virtual Observation Portal Interface

After modifying the virtual world in order to make the learning activity more enjoyable for the student participants and creating mechanisms to track these participants, we re-named the virtual world "*Observe Portal*" since, within it, the students are observed and assessed by the environment. When users initiate the system, a login window is then displayed so that they can enter their names and connect with the game server (Figure 4-4).

Figure 4-4: Login Window

The following window (Figure 4-5) is then activated; this enables the user to choose an avatar type (teacher or student) and create a new learning session or, alternatively, to join a session which has already been created.

Figure 4-5: Session Window

Joining an existing session assists student participants in starting to learn how to perform tasks within the VW. The GUI contains the same programming board and services window as does the Interreality Portal. This GUI also includes a modified chat window that students can use to communicate with each other while they are performing tasks (Figure 4-6). The chat window has new (i.e., added for the new portal) buttons that permit students to classify their sentences; this facility helps in evaluating their communication skills and is explained later in Chapter 5 (section 5.4). Moreover, another new feature has been added to the interface, called the rating tool; this tool enables each student to rate the performance of the other participants (those in the same session/group), see Figure 4-6.

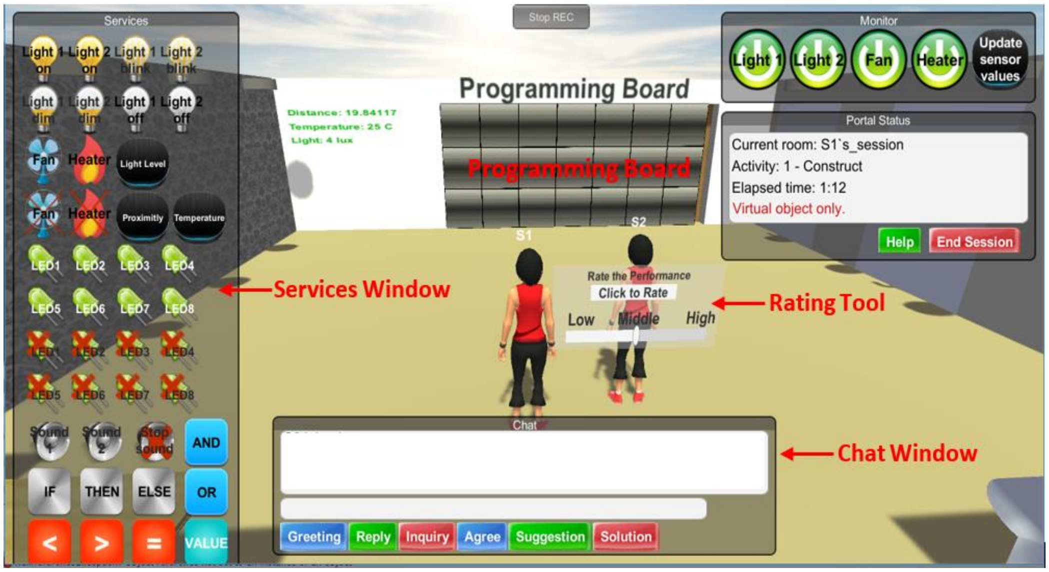

Figure 4-6: *Observe Portal* Interface

In addition, expert participants can join the learning session as teachers to observe and assess students while they are collaborating together (Figure 4-7).

Figure 4-7: The Teacher Observes the Students in *Observe Portal*

The technology also monitors the actions of each participant and accumulates details of events which have been triggered; information about such events is saved into the repository. As the students complete various tasks, the virtual platform automatically stores evidence of their learning and evaluates each individual's actions. As soon as a task is completed, the participants are presented with a dashboard giving details of how they and their group performed, summarising their performance and what has been learnt. Recorded videos of each student's work are also available so that their performance can be compared with the provided assessments. Student assessment and the associated dashboards are explained in detail in section (4.3).

4.1.2. Mixed Agents (*MixAgent*) Implementation

As mentioned in Chapter 3, a combination of software agents and natural agents (human users) are employed in the *MixAgent* model. Software agents automatically observe how students behave, capturing their behavioural events in system logs, and so converting actions into data that can be stored in an underlying repository. Students perform the role of natural agents. They appraise other students and accumulate implicit evidence that could not otherwise be easily determined via a purely technological approach [90]. Each student can observe the quality of their peers' performance and provide their opinion on each other's collaborative skills.

Natural Agents (NA)

In order to employ the natural agent (NA) concept in practice and allow students to provide their evaluation of other students in relation to the learning environment, a rating tool was developed (Figure 4-8). The rating tool is intended to assist students in rating the quality of their peers' performance in terms of their activities working on the learning tasks. The tool has a simple design and is based on a scale of three performance levels: low, middle and high. Low performance is coded in the system as a zero (the student does not receive any score), middle performance is coded as one (the student acquires just one point) and high performance equals two (the student obtains two points). Within the VW, the display related to this tool appears behind each student; however, users cannot see how others are rating them.

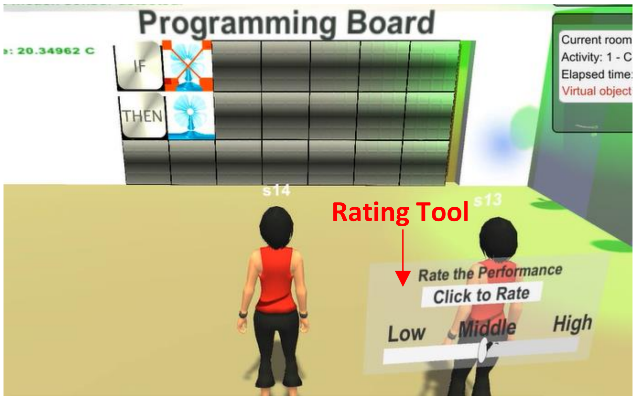

Figure 4-8: Natural Agent Rating Tool

All the ratings entered are captured from the instances of the rating tools in real-time and sent to the repository. A class script in C# has been created to manage the NA tool. The rating data are sent to the server so that they can be saved into the database which is controlled separately. Then, these data are used to perform calculations and analyses at runtime so that the quality of performance of each user can be determined. NA data which is stored in the repository is subsequently interpreted by the *OLens* model.

Software Agents (SA)

When a user signs into the virtual world, a software agent (SA) is assigned to them. The SA monitors the student's activities in real-time. It is basically a C# class that observes any new events triggered in the system by the user, then dynamically sends information relating to these events to be recorded in the repositories in the server.

All the agents work towards the same goals, interacting in real-time to accumulate data which can subsequently be used in an inference mechanism. We applied a fuzzy logic method to amalgamate the data produced by both the natural and the automated agents — in order to make decisions about students' performance and assess them. Chapter 5 presents more detail about the fuzzy logic systems created for assessing students.

4.1.3. System Architecture

The system architecture developed for the research prototype for the purposes of supporting the research model is illustrated in Figure 4-9 below:

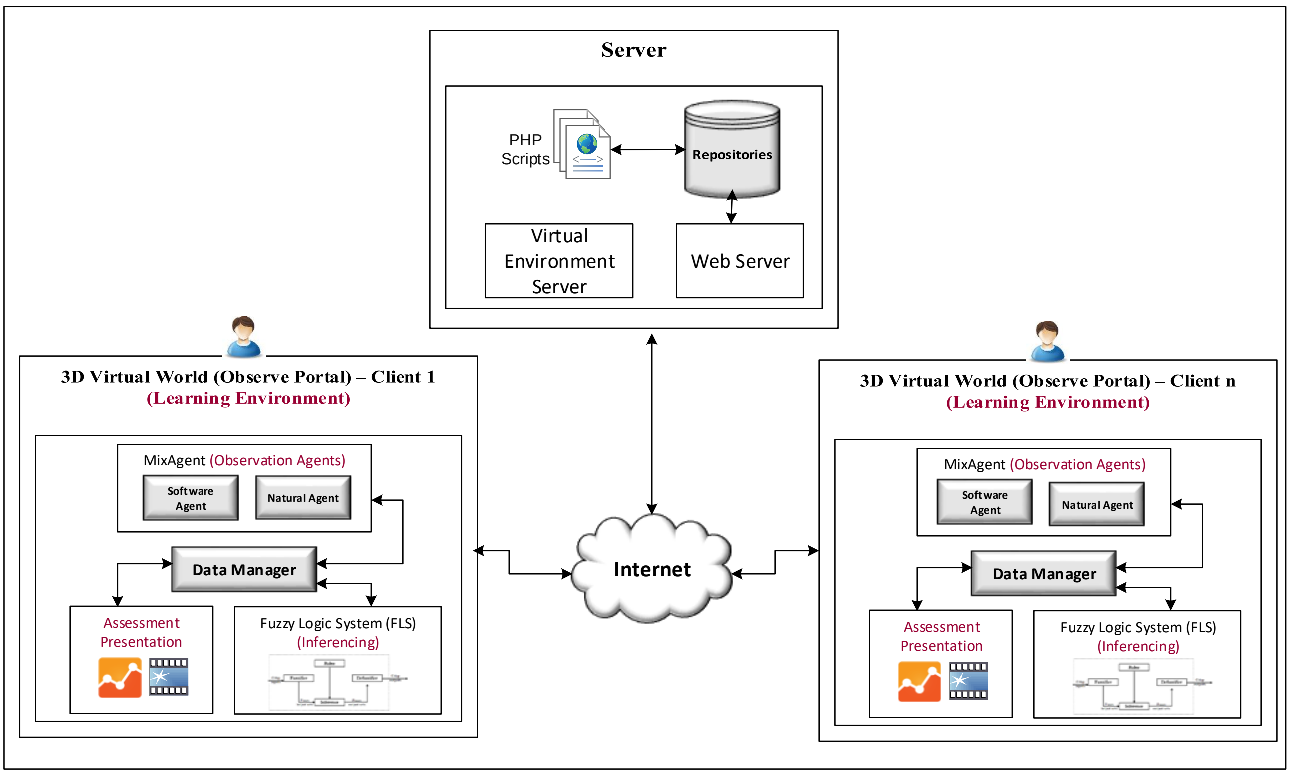

Figure 4-9: *Observe Portal* — System Architecture

These components of the system architecture implemented the elements of the *MIVO* conceptual framework (Chapter 3); the system architecture is divided into client and server elements as follows:

a) Client

- 3D Virtual World (Learning Environment): Includes, first, the login system interface which identifies the learners and teachers, and what roles the user can fulfil. These roles can then be played out in the virtual 3D environment. For instance, students are free to engage with their peers in order to achieve educational goals and also to appraise each other's progress. Teachers use the same interface to observe the educational activities being undertaken, and they can see the students' evaluations at the end of each session.

- MixAgent: This is the set of observation agents that was mentioned earlier in relation to the *MIVO* framework. The students' actions are automatically observed by software agents in real-time. In addition, the students perform the role of natural agents by rating how well their fellow students perform. This insight data will be used in the subsequent evaluation processes and so it is sent to the data manager to be stored.
- Data Manager: This is part of the data unit which is included in the *MIVO* conceptual framework. The data manager has access to all of the data produced, retrieving it from the repositories as required. Client agents submit data via the data manager, and these data are subsequently sent to the server and stored in the repositories. Similarly, the data manager is responsible for transferring data to and from the fuzzy logic system when necessary.
- Fuzzy Logic System: This is the inferencing mechanism that is employed in the research prototype and mentioned in *MIVO* framework. This model processes the raw data in order to derive outputs which are usable for evaluation purposes. This process involves fuzzification, inferencing and defuzzification activities. This fuzzy system also contains the rules that execute the conditions provided by the *OLens* frames within the *MIVO* framework.

- Assessment Presentation: This is the assessment interface which displays the students' performance evaluation to teachers and learners. It illustrates the assessment feedback via dashboards, showing when performance is high and when it is low. Furthermore, the interface can provide the students with video recordings of themselves — to themselves and others. Thus, the students' work can be reviewed.

b) Server

The server is a component of the data unit within the *MIVO* conceptual framework. It includes the data storage and the elements required to process these data. The implementation of the server was carried out using *SmartFoxServer X2 (SFS2X)*[11]. *SFS2X* is a platform optimized for online and multiplayer games. It offers an API for connecting several clients to a server. The server also includes repositories that store data in real-time, accumulating details about events and the actions of participants. The repository contains a *MySQL*[12] database which can be used to save users data. *MySQL* is a relational database management system (*RDBMS*) for online systems. To connect the client *Unity3D* with the database security, *PHP* scripts were developed and *MAMP PRO*[13] was used to make the connection on the server side (Figure 4-9). Furthermore, web server requests are triggered via WWW classes in Unity. *MAMP PRO* is a combination of web server software systems, including *PhpMyAdmin*, *Php*, *MySQL* and *Apache*. Every time a new session is created on the server, the group and individual data are stored in the database tables.

[11] http://www.smartfoxserver.com
[12] https://www.mysql.com/
[13] https://www.mamp.info/en/

Figure 4-10 represents the user interactions and the information flows which exist in the research prototype (*Observe Portal*).

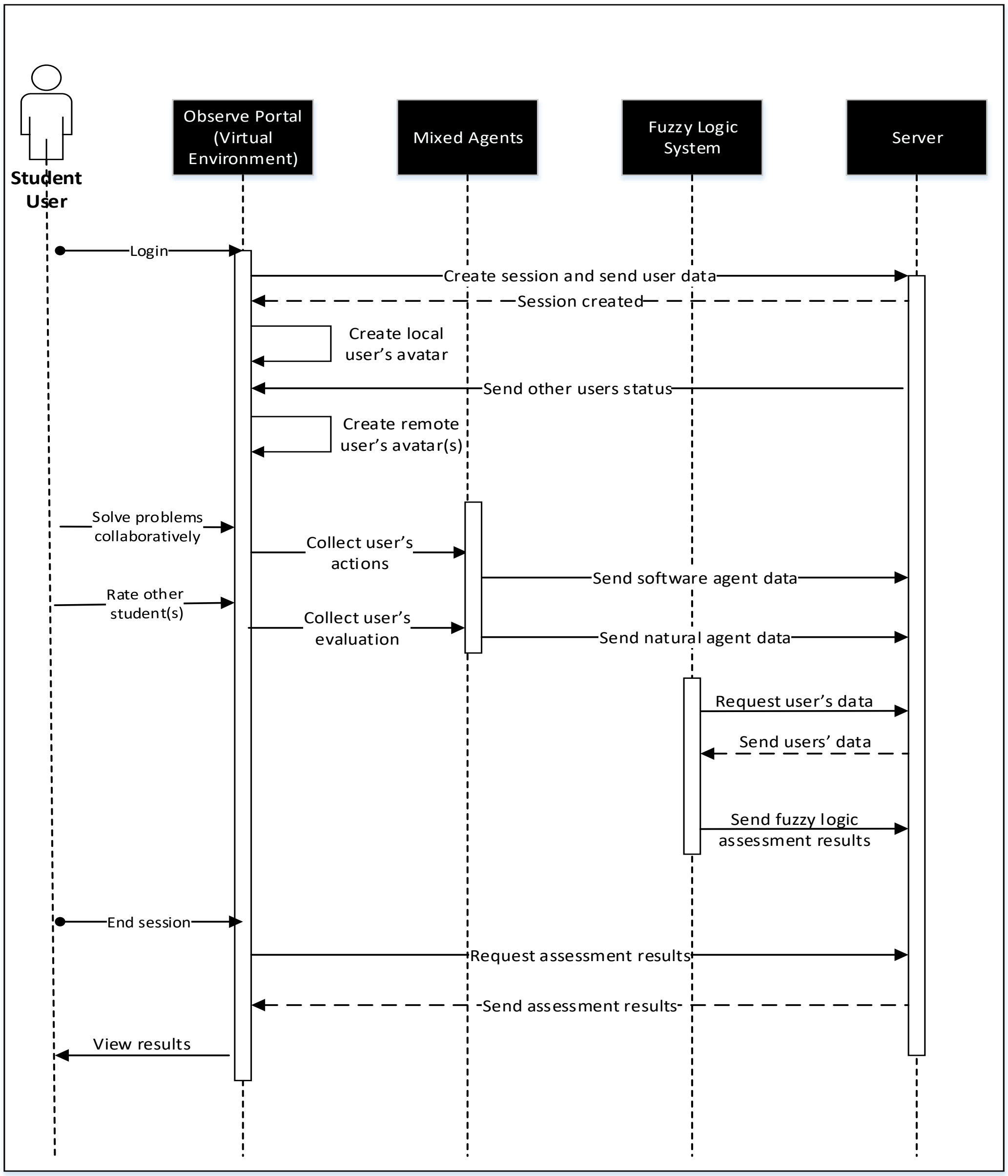

Figure 4-10: *Observe Portal* Interaction Diagram

The next section (4.1.4) defines the data and the events for which such data will be saved in the server.

4.1.4. Events and Data

In our prototype, the ability to capture events which occur in the virtual world is important because the assessment decisions about learners are based on the data concerning these learners and events. This section describes the events data the system captures and saves in the database. Also, it describes the assessment data that are produced by the system to be used in the assessment presentations. Figure 4-11 presents the *MySQL* database tables that hold all the data and entities. The events which can be captured from the virtual world are:

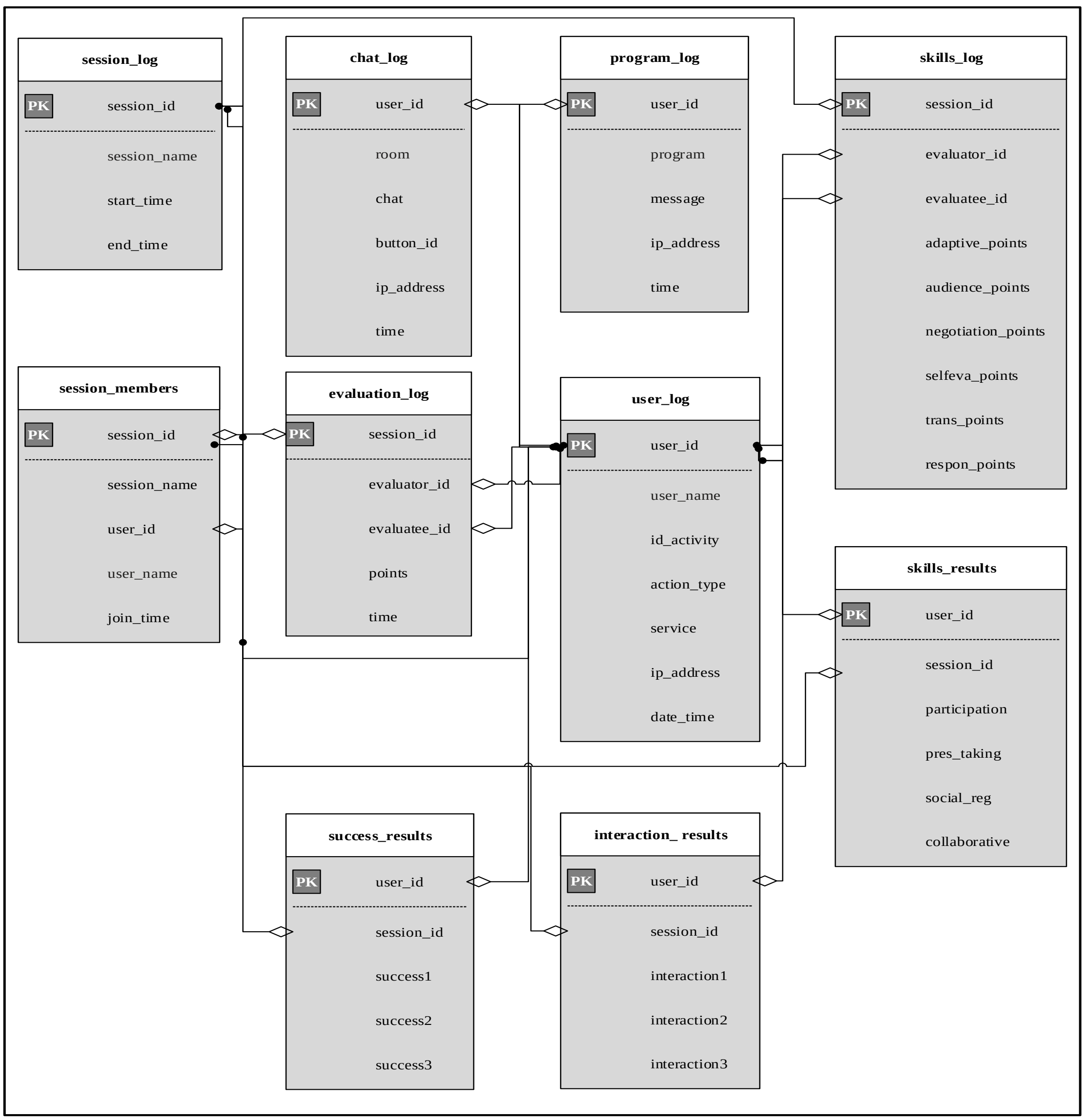

Figure 4-11: *Observe Portal* Relational Database Structure

- User events (User_Log Table): user events are captured by the system by saving information concerning all the user actions that occur in the learning environment; these events include: the virtual object used, the user activity, the service used and the time of each action. These data are saved onto the user_log table in the database (Figure 4-11); this has the following fields: user_id, user_name, id_activity, action_type, service, ip_address, date_time.

  The reason for capturing these events is to use them in assessing the level of student interaction with the learning environment; this relates to the *'Learning Interactions'* lens.

- Program events (Program_Log Table): these events are saved in the program_log table, shown in Figure 4-11. This table stores all the programs created by the students. The system saves the following data: user_id, program, message, ip_address, time.

  The recording of these events is essential in order to facilitate the evaluation of student task success - which is part of the third lens '*Task Success*'.

- Chat events (Chat_Logs Table): these constitute all the interactions that occur between students via the chat windows. These events are stored in the chat_logs table (Figure 4-11). This table stores the following data: user_id, room, chat, button_id, ip_address, time. The chat logs are saved in the repository to facilitate the evaluation of students interactions with their colleagues in the learning sessions; this is part of the second lens *'Learning Interactions*'. Furthermore, saving these events contributes to the assessing of learner collaborative skills.

- Rating events (Evaluation_Log Table): the rating data that a user produces when occupied in the learning tasks, aimed at rating the other student(s) who are working collaboratively with her/him, are stored in the evaluation_log table (Figure 4-11). This table stores the following data: session_id, evaluator_id, evaluatee_id, points, time. These data are used as the natural agent data and are used in the system to assess the student's level of success and the quality of their interaction.

- Skill rating (Skills_Log Table): this is the rating a user gives to other student(s) when she/he evaluates their collaboration skills. These data are saved in the skills_log table (Figure 4-11); the table fields are: session_id, evaluator_id, evaluatee_id, adaptive_points, audience_points, negotiation_points, selfeva_points, trans_points, respon_points.

  All these data are used to evaluate the level of collaborative skills demonstrated by the students - this concerns the quality of collaboration.

- Session events (Session_Log and Session_Members Tables): when a session is created, the session_log table records information about it (Figure 4-11) and saves: session_id, session_name, start_time, end_time. Moreover, the session members are saved in the session_members table, this includes: session_id, session_name, user_id, user_name, join_time.

  These events are recorded in order to identify the learning sessions and group members involved in each session.

When the captured events are processed by the fuzzy logic system, this produces the result data that is then presented to users via the assessment interface. These result data are saved in the following tables:

- Interaction_Results Table: it stores the students' interaction assessment results as yielded after system processing. It holds the following data: user_id, session_id, interaction1, interaction2, interaction3.
- Success_Results Table: it stores the assessments of the students' levels of success. The table holds the following: user_id, session_id, success1, success2, success3.
- Skills_Results Table: it saves the assessments of the students' skill levels. The skill_results table records the following data: user_id, session_id, participation, pres_taking, social_reg, collaborative_skills.

4.2.Phase 2: Application of the Observation Lenses (*OLens* Model)

This section illustrates how we applied the *OLens* model to the learning environment, in practical terms, and clarifies the various methods used for applying and pedagogically mapping the lenses — in relation to accumulating the data and establishing the rules for its use in the virtual worlds.

4.2.1. Event Detection Lens

As previously pointed out in Chapter 3, this layer is more about providing the data interface from the wider education/computing environment. Once the participants are active the system starts collecting and saving data related to their actions and so mimicking the way in which a teacher might observe students from a high-level perspective without providing any detailed assessment of how they (the students) are performing. All the events which are captured from the virtual world have been defined earlier, in section (4.1.4). The *MixAgent* model described in section (4.1.2) is the model used to monitor the learning activities in real-time; this accumulates evidence of learning in order to grade how the students perform, in terms of quality and quantity of actions, while participating in their virtual worlds. All the agents work towards the same goals, interacting in real-time to accumulate data which can subsequently be used in the fuzzy inference mechanism.

4.2.2. Learning Interaction Lens

This performs observations of the social interactions between the students and the environmental interactions between the students and the virtual world. This lens extends the teachers' abilities to judge the activities of students to include the group interactions. This, additionally, allows for the inference of the quantity and quality of the learners' interactions by creating rules based on the teachers' viewpoints. Table 4-1, below, shows the quantity and quality indicators that have been used to assess both the contributions of participants and their interactions with other students in the virtual world.

	Quantity Indicators	**Quality Indicator**
Individual	-The number of actions recorded in the chat log in the session of the monitored period. -The number of actions taken to use the virtual objects. -The number of rules constructed in order to create programs for the smart home.	-The average rating given by students to others in the group in the course of the monitored period. Rating scores: Low= 0, Middle = 1, High= 2
Group	-The total number of all the actions of all the members in a group. Also the actions derived from the chat log, user log, and program log during a defined period.	-The average rating — calculated from all the ratings given to all the members of a group in a defined period. Rating scores: Low= 0, Middle = 1, High= 2

Table 4-1: Interaction Indicators

Since this lens focuses on thorough observations, it requires common interfaces in order to query the data amassed for the purpose of evaluating each participant’s interactions when engaging in *Observe Portal*. Table 4-1 shows that to evaluate the students’ interactions which took place in a learning activity, data were acquired from the software agent about the number of actions which occurred during the chat session and the number of rules which were created for the smart house. While the natural agents' data provides information about the quality of interactions — the average rating for a student as provided by other users of the environment is calculated in order to obtain an average indicating the quality of a student's performance, as related to a task. Examples of queries used to obtain SA data about the quantity of interactions from the repository are as follows:

- SELECT COUNT ('USER_ID') AS CHATS FROM `CHAT_LOG` WHERE `USER_ID` = '“USER _ID"' AND `TIME `>= '00:00:00' AND `TIME `< '05:00:00'

- SELECT COUNT ('USER_ID') AS PROGRAMS FROM `PROGRAM_LOG` WHERE `USER_ID` = '“USER_ID"' AND `TIME `>= '00:00:00' AND `TIME `< '05:00:00'

An example query which obtains the average rating (NA data) from the repository in order to understand the quality of interactions is as follows:

- SELECT AVERAGE 'POINTS' FROM `EVALUATION_LOG` WHERE `EVALUATEE_ID` = '“EVALUATEE_ID"' AND `TIME `>= '00:00:00' AND `TIME `< '05:00:00'

When obtaining these data from the SA and NA, the system sent them to the fuzzy logic system to make decisions and assess students' interactions. Chapter 5 provides more detail about the FLS created for this lens.

4.2.3. Student Success Lens

This lens aims to mimic the ability of teachers in real-world classrooms to evaluate the overall success of students. In this context, student success can be inferred from the fraction of correct responses made to the various questions and tasks [44]. Table *4-2* below provides the indicators used to determine the success of a group or an individual when attempting a task.

	Quantity Indicator	Quality Indicator
Individual	-The number of correct responses made by the student in a defined period. -The number of completed tasks in a set amount of time.	-The average ratings provided by other members (of the group) concerning the quality of a student's work when completing a task.
Group	- The number of correct responses made by the group in a defined period. -The number of completed tasks made by the group in a set amount of time.	-The average of all the ratings (of other members) provided by members about the quality of the group's work when completing a task.

Table 4-2: Task Success Indicators

As for the interaction lens, for the success lens, we created common interfaces which enabled us to query the data amassed (i.e., APIs). From Table 4-2, we can see that the quantitative data was obtained from the software agents by counting the number of correct answers and complete tasks which occurred in a defined period. The natural agents' data was used to determine the quality of success encountered; hence, the average rating for a student —given by other users — was calculated in order to determine the performance quality of a student in relation to a task. The following is an example of a query used to obtain, from the repository, SA (quantitative) data concerning successes:

- SELECT COUNT ('USER_ID') AS PROGRAMS FROM `PROGRAM_LOG` WHERE `USER_ID` = '“USER_ID"' AND `MESSAGE' ='CONDITION-TRUE' AND `TIME`>= '00:00:00' AND `TIME`< '05:00:00';"

The following is an example of a query used to obtain an average rating (NA data) from the repository, in order to evaluate the quality of successes:

- SELECT AVERAGE 'POINTS' FROM `EVALUATION_LOG` WHERE `EVALUATEE_ID` = '“EVALUATEE_ID"' AND `TIME`>= '00:00:00' AND `TIME`< '05:00:00';"

After obtaining these data from the SA and NA, the system sent them to the fuzzy logic system to make decisions and assess students' success. Chapter 5 provides more detail about the FLS created for the student success lens.

4.2.4. Performance Outcomes Lens

This lens goes beyond merely counting the number of correct answers and instead provides a summative evaluation of both the quality and the quantity of the performance outcomes achieved. Multi-users virtual environments are often used for collaborative learning exercises and, in relation to these, due consideration should be given to the participants' collaborative skills. Hesse [14] proposed a skills taxonomy which suggested a variety of skills that should be measured specifically when students are engaged in collaborative activities (Figure *4-12*). The framework proposed by Hesse is particularly well suited for assessing the cognitive and social skills of students. What is more, this framework is able to distinguish between various collaborative skill levels (high-middle-low). For this reason, Hesse's framework has been utilised, here, to assess the collaborative problem-solving skill levels exhibited via the version of the *OLens* model used here. Figure 4-13 presents an example of a set of social skill classifications based on Hesse taxonomy[14].

Element	Indicator	Low	Middle	High
Participation				
Action	Activity within environment	No or very little activity	Activity in familiar contexts	Activity in familiar and unfamiliar contexts
Interaction	Interacting with, prompting and responding to the contributions of others	Acknowledges communication directly or indirectly	Responds to cues in communication	Initiates and promotes interaction or activity
Task completion/ perseverance	Undertaking and completing a task or part of a task individually	Maintains presence only	Identifies and attempts the task	Perseveres in task as indicated by repeated attempts or multiple strategies
Perspective taking				
Adaptive responsiveness	Ignoring, accepting or adapting contributions of others	Contributions or prompts from others are taken into account	Contributions or prompts of others are adapted and incorporated	Contributions or prompts of others are used to suggest possible solution paths
Audience awareness (Mutual modelling)	Awareness of how to adapt behaviour to increase suitability for others	Contributions are not tailored to participants	Contributions are modified for recipient understanding in the light of deliberate feedback	Contributions are tailored to recipients based on interpretation of recipients' understanding
Social regulation				
Negotiation	Achieving a resolution or reaching compromise	Comments on differences	Attempts to reach a common understanding	Achieves resolution of differences
Self evaluation (Metamemory)	Recognising own strengths and weaknesses	Notes own performance	Comments on own performance in terms of appropriateness or adequacy	Infers a level of capability based on own performance
Transactive memory	Recognising strengths and weaknesses of others	Notes performance of others	Comments on performance of others in terms of appropriateness or adequacy	Comments on expertise available based on performance history
Responsibility initiative	Assuming responsibility for ensuring parts of task are completed by the group	Undertakes activities largely independently of others	Completes activities and reports to others	Assumes group responsibility as indicated by use of first person plural

Figure 4-12: Hesse's Social Collaborative Skills and their Levels [14]

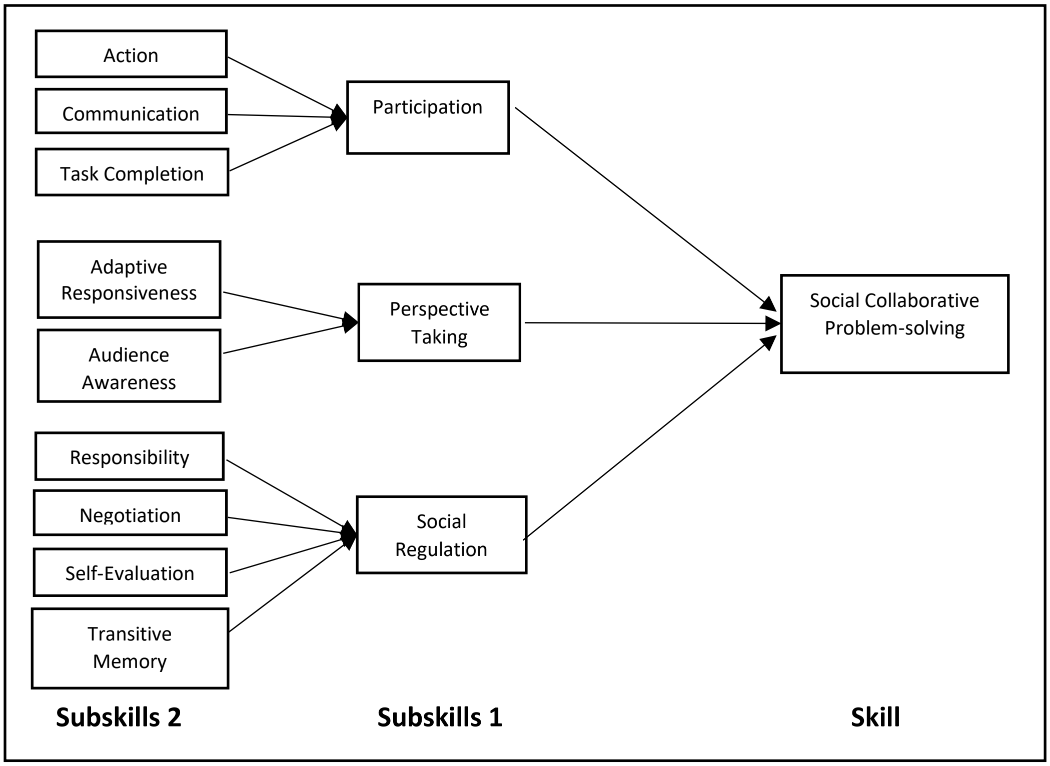

Figure 4-13: Collaborative Social Problem-Solving Skills

Table 4-3 offers instances of learning outcome indicators used for evaluating the skills of individuals via this lens. According to Hesse [14], the participation skills concern the quantity of collaboration: number of actions, number of communications and number of completed tasks, while the perspective taking and social regulation skills revolve more around the quality of collaboration and interaction.

The collaborative skills that concern quantity are evaluated by the system (software agents). However, the data that the system itself accumulates as a result of the students' actions does not naturally lend itself to being used for measuring the collaborative skills that relate to quality. This measurement can only be meaningfully achieved with the addition of feedback from the students (natural agents) themselves. It has been shown that peer evaluation techniques can be used to measure the skills of students [38]. Indeed, such an approach can provide very thorough insights into the quality of the collaborative skills demonstrated by students when engaged in collaborative tasks. Thus, all of the participating students are required to offer their opinions on their peers' skills. At the end of each session, a rating panel appears on the screen, allowing each student to rate the skills that related to the collaboration quality of the other participants such as the perspective taking subskills (adaptive responsiveness and audience awareness) and the social regulation subskills (negotiation, self-evaluation, responsibility initiative and transitive memory) (Figure 4-14).

	Quantity Indicator	**Quality Indicator**
Individual	Evaluate the number of individual participation skills and sub-skills: -actions -communications -task completions	The student rating given for the following collaborative skills: 1-Perspective Taking - Adaptive Responsiveness - Audience Awareness 2-Social Regulation -Negotiation -Self-evaluation -Responsibility initiative -Transitive Memory
Group	Evaluate the group participation skills and sub-skills: actions, communications and task completions for the group.	The group rating scores for the collaborative skills: perspective taking and social regulation.

Table 4-3: Learning Outcome Indicators

When obtaining these data from the learning environment, the system sent them to the fuzzy logic system (FLS) to make decisions and assess students' collaborative skills. Chapter 5 also provides more detail about the FLS created for this lens.

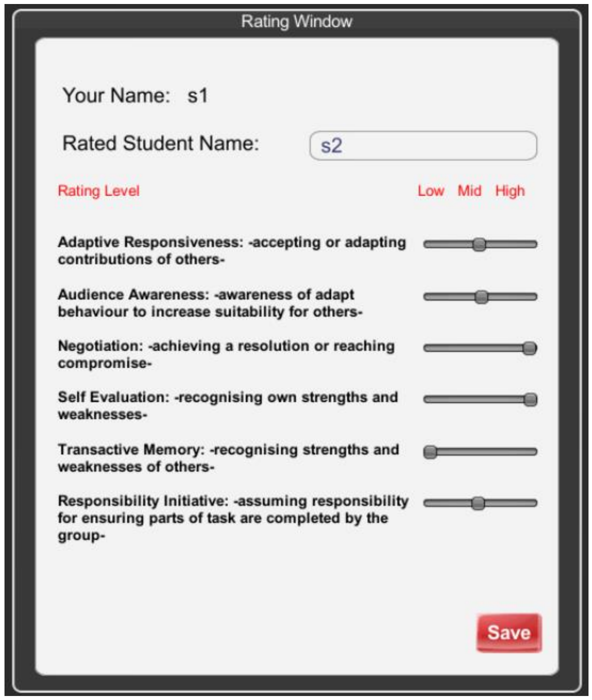

Figure 4-14: Skills Rating Window

4.3.Phase 3: Using the *Observe Portal* in Collaborative Activities and Constructing the Assessment Presentation Interface

This section briefly describes the collaborative learning scenario that the students performed within when they participated in the *Observe Portal*. Moreover, it explains the assessment interface and gives examples of the evaluation outputs that were presented to the students after they finished their learning sessions.

4.3.1. The Collaborative Learning Activities

Students were grouped together, each group consisting of two or more students. Then, the students were given several collaborative programming activities to undertake; they were required to program actuators and sensors in order to teach them (the students) the functionality of embedded systems. The programs that the students construct are then implemented in the virtual smart home — provided that the students generate syntactically correct rules (Figure 4-15). Also, the graphical user interface (GUI) provided to the students permits collaboration between users as well as communication through the messaging tool. Data concerning all of the students' actions and events are saved in the repository to be retrieved in real-time.

When the students perform activities and collaborate with their peers, we expect variations in their actions and performance. According to Bartle [148], there are a number of particular types of players, and these types can be delineated based on their preferred actions in multiplayer games, the player classifications are: a) achievers who they like to show off their skills and get points and so succeed in the game; b) explorers who like to explore new things and be immersed in the game; c) socialisers who enjoy the game mainly through interacting with others and making friends; d) killers who like to kill enemies and compete in this specific way with other players. The types of players who are most likely to engage in a useful way with our learning environment are: the achievers, the explorers, and the socialisers. The student tasks involved require the attainment of success within each learning session, the exploring of the virtual home and learning environment, and socialising in order to collaborate with other learners. These three types of players can fit within our learning environment and can collaborate, as needed, in order to complete the learning tasks.

However, some learners may be clearly keen to contribute, but others may be less inclined to do so. In order for this kind of situation to be recorded, the GUI offers a rating tool (shown in Figure 4-15) which allows collaborators to repeatedly score each other's quality of performance. Rating each other while working on the task activities may be a cause of interruption in terms of student learning. However, according to [149], an easy to use design in terms of the feedback tool can serve to avoid significant learning interruption. Accordingly, the purpose of the simple rating mechanism used in our system is to provide for synchronous feedback in the course of learning activity - without the need to interrupt the students' activities significantly. Additionally, in order to prevent possible significant interruption, before starting the learning sessions we asked the students only to rate others when they are free to do so: for example, after completing a program, after a conversation or after completing a learning task. Furthermore, the researcher observed the learning sessions at all times to make sure that the students were not disturbed. At the end of each task, if a student forgot to rate their classmates, the system displays a reminder that they should do so. Thus, a student is likely to rate others at least once each task.

Figure 4-15: Students' Collaboration in the *Observe Portal*

As mentioned earlier, the system gathers data from both user activity and student ratings and then deploys the fuzzy systems in each lens. After finishing the tasks, the teachers and students can choose to receive dashboard reports assessing each individual's contributions and giving details of the learning outcomes and skills that have been demonstrated.

4.3.2. The Assessment Presentation Interface

By applying the *OLens* model in this system, we aimed to better assess the students' overall performance in relation to their collaborative learning activities in the virtual world, providing users with more useful feedback. The process of collecting, analysing and reporting student data for the purposes of recognising when the learning occurs is called Learning Analytics (LA) [150]. LA focuses on gathering data from a variety of different sources in order to provide valuable information on learning [151]. Additionally, an important feature of LA is the visualisation of data. Users should be able to review the analysed results and relate them to the learning objectives involved, directly or indirectly [152]. LA can support educational organizations in improving their quality of learning [151]. As a result, many studies have applied LA and have incorporated graphical dashboards in order to report student assessment and provide visualisations of their performances with respect to learning activities [151, 153-155].

Our work follows this LA research stream by presenting analysed learner data using graphical feedback presentations to report on learning outcomes from collaborative activities in virtual worlds. Accordingly, when an individual learning session has been completed in the virtual world, the progress which has been made by each student can be viewed on dashboards and via video recordings. The dashboard is customised for each learner so that they can only see their own review and the accumulated results of their group as a whole. On the other hand, teachers can review all students' assessments from the system. Figure 4-16 is the window that appears to students after they finish a learning session; it enables the students to select which assessment they would like to view.

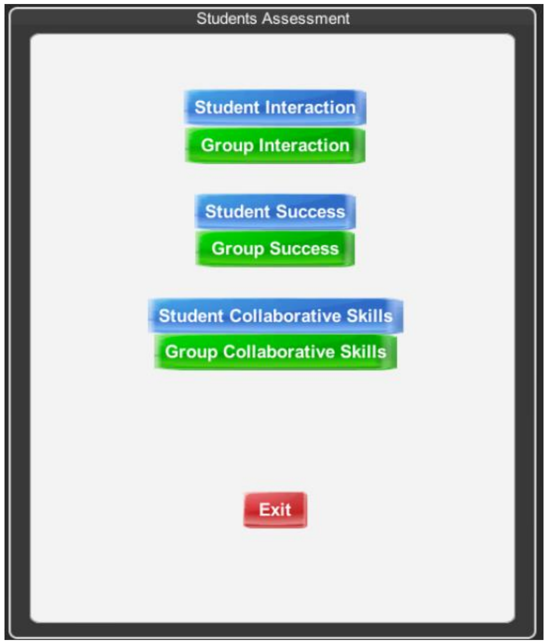

Figure 4-16: Screenshot of the Assessment Window

Figure 4-17 presents an image of the assessment screen that appears once a student has completed a session and has clicked on the "Student Interaction" button in the earlier window (Figure *4-16*). The screen pictured enables the users to view details of the student interactions that took place during the session; these details are derived from the underlying agents and inferred by the fuzzy logic system. Moreover, users are also able to review their performance by watching a recorded video (Figure 4-18). The videos are constructed by using *AVFX Movie Recorder*[14], a Unity asset that helps to capture audio and video in real-time from Unity scenes and applications.

Users are able to review the recorded video with reference also to the assessment dashboard. If the student has been criticised in some way at a certain stage in the proceedings, they can watch that particular section of the session and then, perhaps, better appreciate why they have received such feedback. For instance, Figure 4-17 shows that student X was

[14] https://assetstore.unity.com/packages/tools/input-management/avfx-movie-recorder-8010

assessed during Task 2 (5-10 min) with a high level of interactions, while during Task 3 (10-15 min) the student was assessed as exhibiting middle-level interaction. Thus, this student, or alternatively the teacher involved, can go back through the video and review the performance in order to understand how these marks came about (Figure 4-18).

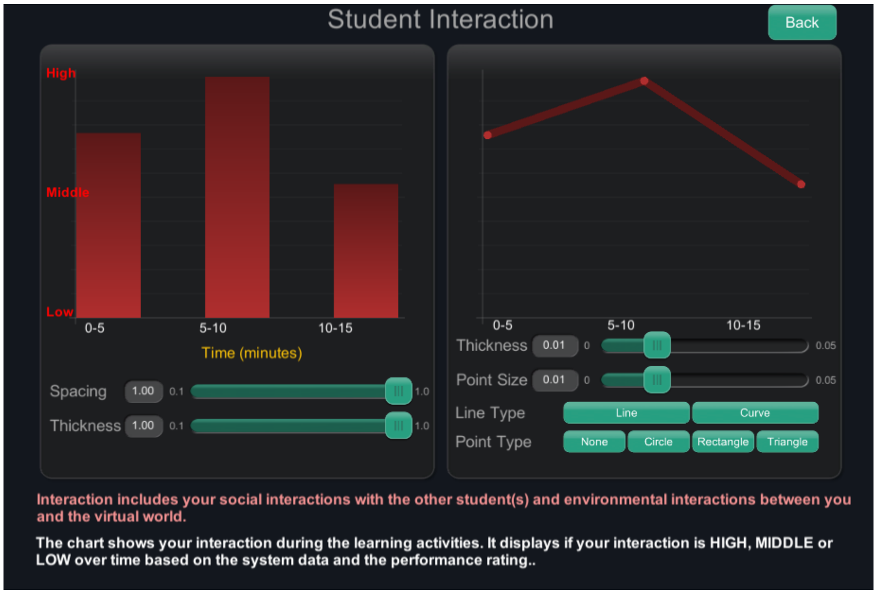

Figure 4-17: Student's Interactions by the Task

Figure 4-18: Video Recording to Review Student Performance

In addition, users are able to view the group's interactions chart (Figure 4-19).

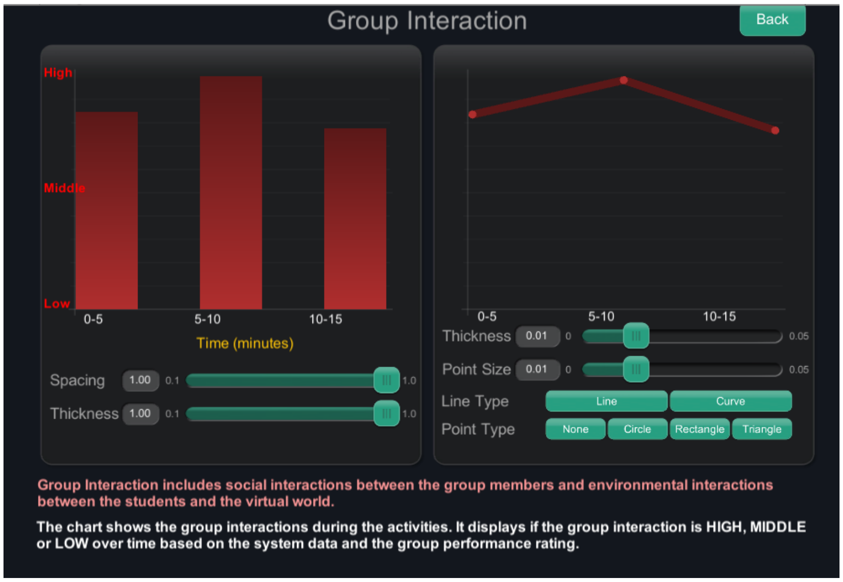

Figure 4-19: Group Interactions by the Task

In relation to reviewing task success, Figure 4-20 presents a dashboard that can be used to evaluate a particular student's success. This dashboard shows the progress that has been made in the learning tasks. As demonstrated in terms of the success lens, both the number of tasks completed over the duration of the learning activity and the ratings obtained from/given to the members concerning the quality of students' activities when completing tasks are used as yardsticks for quantifying progress. The quantitative and qualitative measures available concerning student performance have been used as crisp values by which the FLS infers the degree of task success. For example, the image below, Figure 4-20, shows that student Y completed Task 3 very successfully. Also, users are able to review their success levels in regard to each task by looking at the group dashboard shown in Figure 4-21.

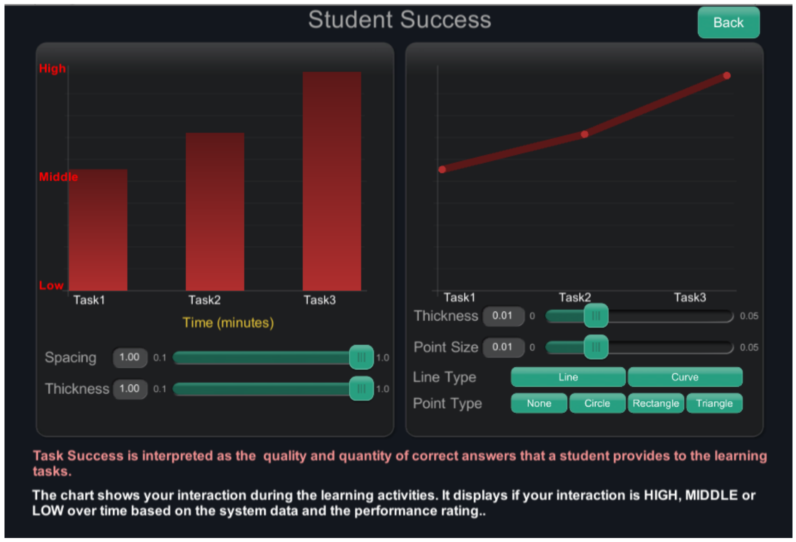

Figure 4-20: Student's Task Success Dashboard

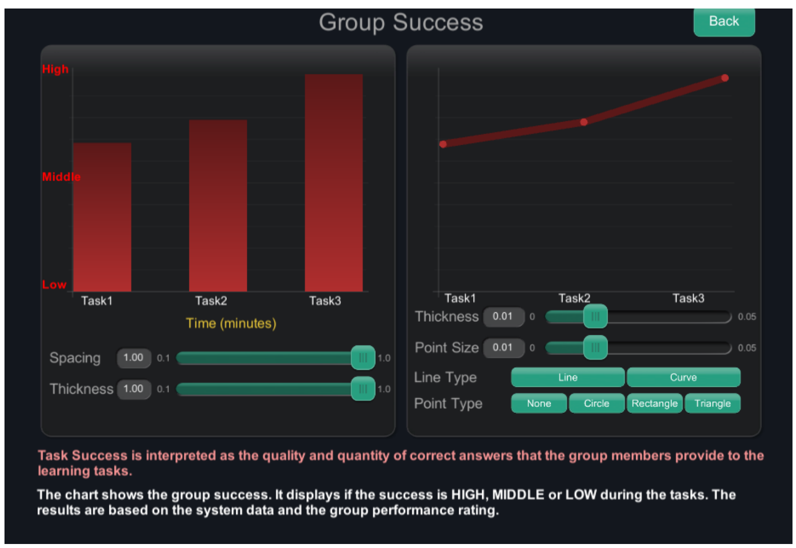

Figure 4-21: Group's Task Success Dashboard

Figure *4-22* shows the dashboard that illustrates how students' social collaborative skills have been rated in the learning activities. These results are based on observations gathered by the natural agents (the fellow students) and also by the automated agents. These data are also interpreted by the FLS in order to provide a thorough assessment of each student's skills — which are then each marked as high, middle or low. The dashboard shows participation, perspective taking and social regulation skill levels; in addition, the last column of the dashboard shown in Figure *4-22* represents the level of social collaborative skill overall. Furthermore, the lower half of the display explains to users how each collaborative skill is calculated. Finally, Figure 4-23 displays the level of the group's collaborative skills as a whole.

Figure 4-22: Student's collaborative skill level dashboard

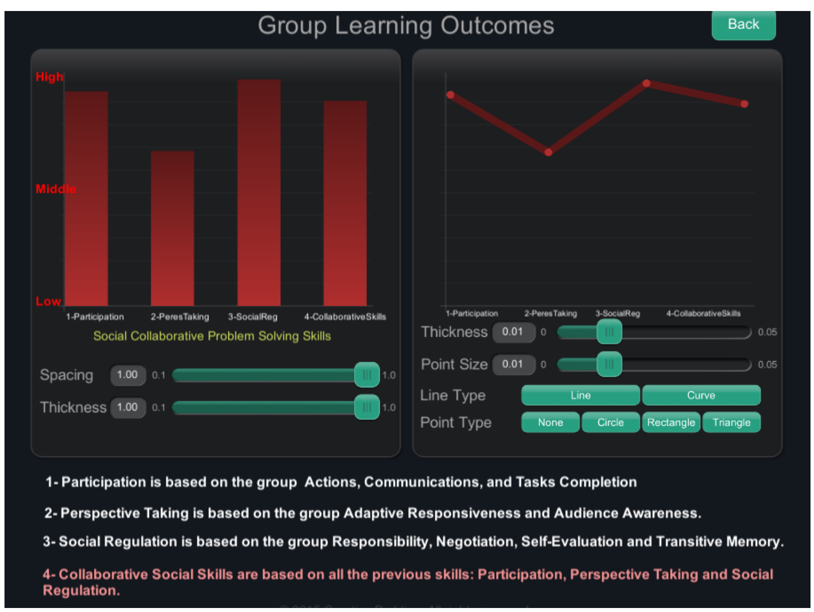

Figure 4-23: Group Collaborative Skill Level Dashboard

It is necessary to continually monitor each student's progress because otherwise, it would not be possible to establish whether or not they are achieving what should be expected of them. This monitoring feature has proven to be valuable to teachers because it enables them to evaluate the students' results and amend the learning activities in the virtual world accordingly. The process is substantially enhanced by the feedback received from peers about how the other students are performing. We anticipate that all of the progress reports put together would help individuals and groups to identify their areas of weakness and what should be focused upon in order to facilitate further improvements in performance.

4.4.Chapter Summary

This chapter described the application of *MIVO* framework to an immersive environment. It explained the experimental and practical work which was required in order to construct *Observe Portal* (proof-of-concepts prototype). The experimental work consisted of three phases. The first phase was to develop the 3D virtual world by implementing the evidence collection agents — the software and the natural agents (*MixAgent* model). Both system and peer monitoring play central roles in improving awareness of how students are performing. Thus we presented in this chapter how we applied the *MixAgent* model in order to be able to recognise events in real-time and gather learning evidence so that student performance in the collaborative learning environment could be assessed.

The second experimental phase was to implement the *OLens* model in the learning environment. The *OLens* model comprises four lenses: events detection, learning interactions, the success of students and performance outcomes. These lenses have different indicators and they appraise student performance from a variety of perspectives.

Lastly, the third phase was to apply the observation methods to the collaborative learning activities and to construct the assessment presentation unit and the assessment feedback displays etc. In order to do this, the observation models were applied in real-time during collaborative learning activities taking place in *Observe Portal* and the results from the observation lenses were presented by creating dashboard interfaces and video recording/playback facilities. The assessment charts were customised for each learner so that a student could see only their own evaluation but could also review the accumulated results of their group as a whole, however, teachers/experts can review assessments of all the participants.

Measuring the performance of individual students can determine whether a student has achieved the desired learning objectives. Such an approach has also proved to be highly valuable for teachers when reviewing learners' work, enabling them to further enhance the learning activities offered in the VWs. In addition, the feedback generated by the system highlighted to the individual students their weakest areas, so that they could work on these and improve their overall performance.

Chapter 5 explains the application of the fuzzy logic approach to the assessing of students' performance in *Observe Portal*. The chapter illustrates the fuzzy systems created for each *OLens* lens - to simplify the analysis and interpretation of the agents' data.

Chapter 5

5. Fuzzy Logic System

"A logical picture of facts is a thought."

— Ludwig Wittgenstein

In this present study, we have applied a fuzzy logic (FL) method to the amalgamation of all the data produced by the software and the natural agents, as described in Chapter 4, in order to make decisions about students' performance and to assess their interactions, success and skill levels. The value of fuzzy logic and the reasons for using it lie in its ability to accommodate many types of values and a wide definition of what it is to reason about this kind of data, much in the same way as a human would. Moreover, fuzzy logic can deal with uncertainty in data and is able to model the common human reasoning mechanisms that are often difficult to emulate using conventional computing approaches. The major drawback of classical logic is its limitation that it is constrained to dealing with just two values; such cannot comprehensively represent the complexity of the real, non-binary world. FL can be considered as being a multi-value logic with a reasoning logic purpose.

Fuzzy reasoning means the inference of an imprecise and merely possible deduction from an initial set of conditions. FL is especially compatible with our model since human beings (and experts) generally do reason in a fuzzy manner, and, here, we are trying to combine natural and software agents within a single system. The use of FL for building our artificial agents means that there is more consistency across the differing agents. The fuzzy model created in this research for evaluating students' performance in relation to each lens was based on Mendel's [127] FL system (Figure 5-1):

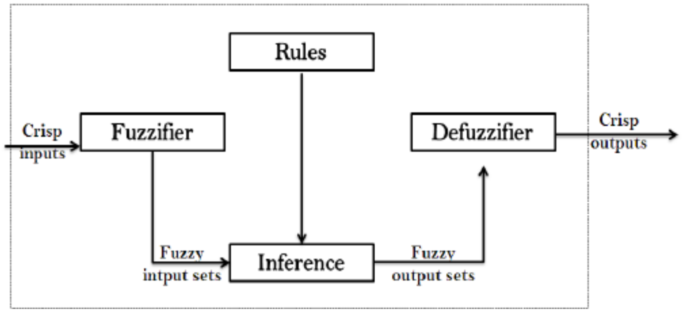

Figure 5-1: The Fuzzy Model Used for Students' Evaluation [127]

a) <u>Crisp Values:</u> Crisp values represent the raw data obtained by the *Observe Portal* system from both the software and the natural agents.

b) <u>Fuzzification:</u> This is the process of changing the crisp values (the students' actions and ratings) to fuzzy input values, using an appropriate membership function (MF). In the current study, a trapezoidal membership function is utilised because such MF shape has been widely used in fuzzy systems [128]. A trapezoidal function is defined by a lower limit a, an upper limit d, a lower support limit b, and an upper support limit c, where a < b < c < d (Figure 5-2). Thus, four parameters are used to specify each membership function (a, b, c, d) and each membership function is used to represent one item of student data. More specifically, the system gathers each user's clicks, communications and completed tasks via the software agent and gathers the ratings from the natural agents, and translates each of these kinds of data into fuzzy sets. Then the fuzzy sets are sent to the inference engine as inputs.

$$\mu_A(x) = \begin{cases} 0, & (x < a) \text{ or } (x > d) \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \dfrac{d-x}{d-c}, & c \le x \le d \end{cases}$$

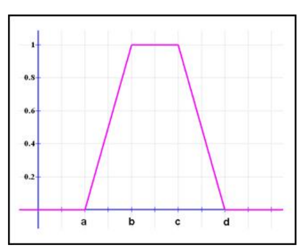

Figure 5-2: Trapezoidal membership function [128]

c) Rules: Fuzzy rules are established, using various linguistic values, to facilitate student evaluation. Fuzzy rules interpret the data stored in the repository, inferring from these data how each individual user is performing. Crucial learning evidence can be identified based on these inferences, and this can lead to a clarification of the relationship between the data and its underlying meaning (in this context this underlying element is the learning performance of each user).

d) Inference: The inference mechanism calculates the firing strengths of each rule in relation to each item of data in order to decide whether a rule is to be considered fired, or not, in response to a specific input; this process generates fuzzy output sets.

e) Defuzzification (of the assessment of a student): This is the process of determining the final output (the assessments of the students) by using a defuzzification technique. After completing the inference decision, the resultant fuzzy number must be transformed into a crisp value; this process is called defuzzification. Many defuzzification methods have been developed for calculating the final output of a fuzzy logic process. One category of these is the max criterion class of methods; these work variously using three measures: the largest of maxima (LOM), the mean of maxima (MOM), and the smallest of maxima (SOM) [156]. This type of defuzzification employs one of these measures to select that the value of membership reaches the maximum. Other defuzzification techniques may be based on the centre of area (COA), the centre of gravity (COG), or centroid. A centroid-based method is employed in this research because techniques of this type are the most commonly applied in expert systems to support the imitation of expert inferencing [128, 157, 158]. Such systems are similar to our application.

The centroid method determines “the output value by calculating the centre of gravity of the possibility distribution of the outputs” [156]. The output value produced by the centroid method is calculated by using the following formula:

$$Z = \frac{\int_{\mu} \mu(x) x dx}{\int_{\mu} \mu(x) dx}$$

In *Observe Portal*, different fuzzy logic subsystems (FLSs) have been created for each level of the *OLens* model in order to facilitate the analysis and interpretation of agents’ data. The FLSs take data from the repository and provide an evaluation of the students in terms of the levels of their performance. The FLSs were created using an open-source C# library for fuzzy logic (DotFuzzy) [159]. Each FLS also contains a number of logical rules for making inferences. The following sections describe how the FLSs in the *OLens* lenses were created.

5.1.Initial Learning Activities – The Physical Classroom Observation

These learning sessions were run to use them as the expert bases for student assessment and they helped to develop the fuzzy expert system that been applied in *Observe Portal* to assess students learning. This experiment included 15 learning sessions, each session has two students and one expert participants. The total number of experts were 8 experts, the experts participated many times in the sessions. Students were divided into groups of two students to collaborate and work in the learning tasks. They were seated and provided with a PC and a physical BuzzBox as a simulation of a physical smart room. The BuzzBox contained sensors and actuators that users can program to control. Learners were asked to collaborate to configure the BuzzBox through a programming board. Also, a video camera was set up in the room to record the collaboration and discussions between students. Chapter 6 includes more details about the physical classroom experiments and the learning activities that students performed.

Additionally, experts observed the students while they (the students) were collaborating; the experts were provided with manual sheets to evaluate the students by observing. The assessments from the experts besides the video recording from the learning sessions were used to count students actions in order to build the membership functions in each *OLens* lens.

The learning task carried out in the physical BuzzBox is the same as the learning task used with the virtual smart home later on in the evaluation experiments; the difference between them being just the learning environment. The objective here is to observe how human experts evaluate student learning both in the physical and in the learning environment. Therefore, the fuzzy logic system created by using the expert assessment carried out with respect to the physical environment was transferred for use in *Observe Portal*.

5.2.FL for Students' Interaction Lens

The fuzzy logic technique was applied to combine the data produced by the agents so that the system make decisions about students' levels of interaction. This section illustrates the creation of the FLS used by the Students' Interaction lens. Figure 5-3 shows the diagram of the interaction FLS, it comprises the following elements:

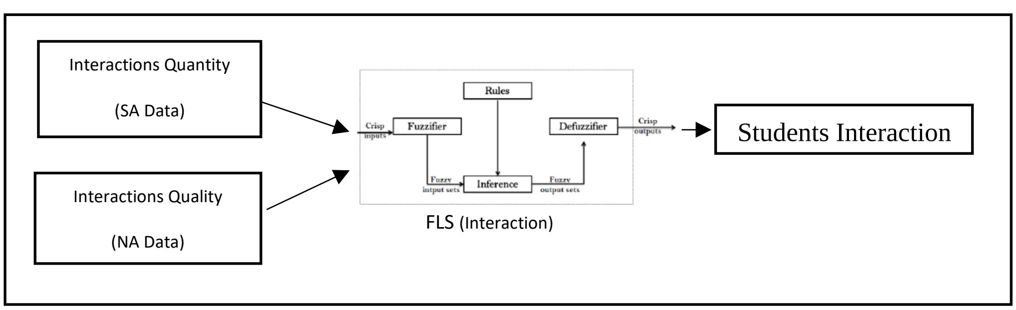

Figure 5-3: FLS for the Learning Interaction Lens

a) Crisp inputs: These represent the data, relating to the students, which must be evaluated; these data are acquired via both the natural (NA) and the software agents (SA) in the learning activities. The crisp inputs here consist of interaction quantity (SA data) and interaction quality (NA data). As explained in Chapter 4, the quantity of interactions comes from students' activities (created rules and communications) saved in the repositories, while the quality of interaction received from other students rating.

b) Fuzzifier: In the implemented system, in order to evaluate the level of a student's interaction, the student's data are evaluated against particular fuzzy sets using a trapezoidal membership function (MF). In order to develop the MF of the input data, as mentioned before, we used the initial experiments in section (5.1). The students were observed while collaborating and were provided with input from experts about whether their level of interaction can be considered high, middle or low; this expert information is then transferred to the ranges in the MFs. For example, experts rated students' as having high interaction levels when those students interacted with other students and/or with the environment more than eight times for one task. They considered the interaction level to be in the middle range when between four and ten such interactions took place, and to be low when fewer than six interactions were made per task. Based on these data, the MF of the interaction inputs was developed to be used as an input for the FLS shown in Table 5-1; Figure 5-4 shows the shape of the MF. Each input has its own MF which is based on teachers' evaluations.

Linguistic Variable	Interval
Low	[0 0 4 6]
Middle	[4 6 8 10]
High	[8 10 30 30]

Table 5-1: Fuzzy Input Set for Interaction Quantity

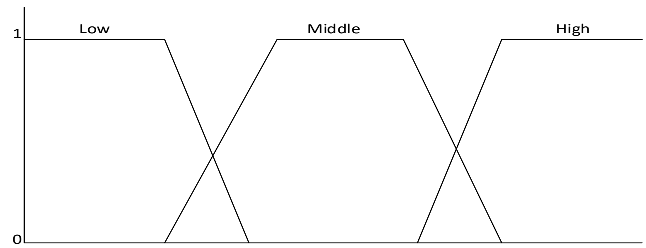

Figure 5-4: Trapezoidal MF (Membership Function) for Interaction Quantity and Interaction Quality

The following table (Table *5-2*) illustrates the fuzzy set related to the rating of students by each other; this is considered to be the interaction quality input. Each student rating takes one of three values, 0 for low, 1 for middle and 2 for high; these values have been transferred to a fuzzy set to be used as a fuzzy input (see Table 5-2).

Linguistic Variable	Interval
Low	[0 0 0.5 0.7]
Medium	[0.5 1 1.5 1.7]
High	[1.5 1.7 2 2]

Table 5-2: Fuzzy Input Set Representing Interaction Quality

The methods of evaluating the individuals and the groups

The fuzzy input sets so far described are used to assess the interaction levels of individual students, not those of groups. In order to evaluate the interaction level of a group as a whole, we need to count the number of interactions made by all the members in the group; thus the system sums the number of interactions made by a group, and these data are used as a crisp input. In addition, each interval present in the fuzzy input set is amplified by multiplying its frequency by the number of students in the group (as a whole) — as shown in Table 5-3; this is in order to create the membership functions on a normalised basis. For instance, in Table 5-3, each value is multiplied by n, where n is the number of group members. If the system is evaluating an individual student, n=1 and the values stay the same. On the other hand, if the system is evaluating the performance of a group of 2 or 4 students, n=2 or n=4 — each interval is multiplied by the appropriate n value. We use this method for all the lenses so that we can assess the quantity of performance related to individual students or groups as a whole, using the same mechanisms.

Linguistic Variable	**Interval**
Low	*[0 0 ٤n ٦n]*
Middle	*[4n 6n 8n 10n]*
High	*[8n 10n 30n 30n]*

Table 5-3: Fuzzy Input Set Representing Interaction Quantity — For Individuals and Groups

In order to assess the interaction quality (rather than the quantity, as above) of a group, the mean average rating of the members of the group is calculated and used as a crisp value, indicating the group's interaction quality. This average rating will lie between 0 and 2, and it is transferred to the fuzzy input set as in Table 5-2.

c) Fuzzy Rules: The linguistic rules for the interaction FLS were also excluded from experts evaluations. These rules are:

1. IF (Quantity_Interactions IS Low) AND (Quality_Interaction IS Low) THEN Interaction IS Low
2. IF (Quantity_Interactions IS Low) AND (Quality_Interaction IS Middle) THEN Interaction IS Middle
3. IF (Quantity_Interactions IS Low) AND (Quality_Interaction IS High) THEN Interaction IS Middle
4. IF (Quantity_Interactions IS Middle) AND (Quality_Interaction IS Low) THEN Interaction IS Middle
5. IF (Quantity_Interactions IS Middle) AND (Quality_Interaction IS Middle) THEN Interaction IS Middle
6. IF (Quantity_Interactions IS Middle) AND (Quality_Interaction IS High) THEN Interaction IS High
7. IF (Quantity_Interactions IS High) AND (Quality_Interaction IS Low) THEN Interaction IS Middle
8. IF (Quantity_Interactions IS High) AND (Quality_Interaction IS Middle) THEN Interaction IS High
9. IF (Quantity_Interactions IS High) AND (Quality_Interaction IS High) THEN Interaction IS High

d) Defuzzification: Table 5-4 represents the fuzzy output set of the interaction lens. Our implementation uses the centroid defuzzification method as this has been widely applied in the literature [157].

Linguistic Variable	Interval
Low-Interaction	[0 0 30 40]
Middle-Interaction	[30 40 60 70]
High-Interaction	[60 70 100 100]

Table 5-4: Fuzzy Output Set for Interaction

5.3.FL for Students’ Success Lens

In the students’ success lens, w**e** also applied the fuzzy logic technique in order to combine the data produced by the agents so that we could make decisions about students’ levels of success. This section explains the FLS used by this lens.

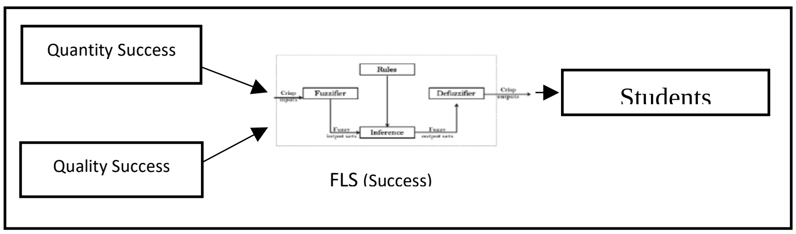

Figure 5-5: FLS for Students’ Success Lens

The FLS illustrated by this diagram (Figure 5-5) uses the following elements:

a) <u>Crisp inputs</u>: The crisp inputs here are the SA and NA data. The quantitative data were obtained from SA by counting the number of correct and completed tasks which occurred in a defined period. The natural agents’ data was used to determine the quality of success encountered; hence, the average rating for a student —given by other users — was calculated to determine the performance quality of a student in relation to a task.

b) <u>Fuzzifier</u>: As mentioned earlier, experts observed students while they were collaborating; thus, we obtained input from them (the experts) about student success. The experts were asked to rate the students’ levels of success: high, middle or low. This was in order to enable us to develop the MF. The experts marked students as having a high level of success when they (the students) developed four or more correct rules for the virtual smart home, a middle level of success level when they created between 2 to 5 correct rules, and a low level of success if they were only able to develop three correct rules or less. Based on these data, the MF for the success inputs was developed (Figure 5-6) and each input received its own MF which was based on observing the experts' (teachers) evaluations. Additionally, as with the previous FLS, in order to assess the success of both individual students and groups using the same fuzzy input sets, the system sums the numbers representing an individual or a group success to use these as crisp inputs. Further, as before, each interval value in the fuzzy input set is multiplied with n=1 if the assessment is for an individual or n= the number of students if the assessment is for a group (Table 5-5).

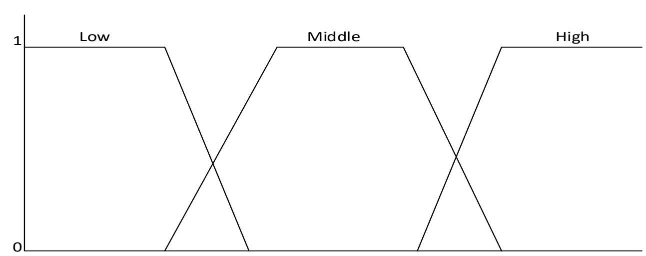

Figure 5-6: Trapezoidal Membership Function for Success Input and Output

Linguistic Variable	**Interval**
Low	[0 0 n 2n]
Medium	[n 2n 3n 4n]
High	[4n 5n 6n 6n]

Table 5-5: The Fuzzy Input Set Representing Success Quantity

As with the previous lens, to assess the qualitative level of success of an individual student or a group, the average rating, calculated for the individual or across all the members of the group for each task and then used as crisp values representing the individual or the group's qualitative success level. Then the student generated ratings (the qualitative success level input) is transferred to a fuzzy set to be used as a fuzzy input (Table 5-6).

Linguistic Variable	Interval
Low	[0 0 0.5 0.7]
Middle	[0.5 1 1.5 1.7]
High	[1.5 1.7 2 2]

Table 5-6: The Fuzzy Input Set Representing Qualitative Success

c) Fuzzy Rules: The linguistic rules which have been developed for the success FLS are:

1. If (SuccessQuantity is Low) and (SuccessQuality is Low) then (SuccessLevel is Low)
2. If (SuccessQuantity is Low) and (SuccessQuality is Middle) then (SuccessLevel is Middle)
3. If (SuccessQuantity is Low) and (SuccessQuality is High) then (SuccessLevel is Middle)
4. If (SuccessQuantity is Middle) and (SuccessQuality is Low) then (SuccessLevel is Middle)
5. If (SuccessQuantity is Middle) and (SuccessQuality is Middle) then (SuccessLevel is Middle)
6. If (SuccessQuantity is Middle) and (SuccessQuality is High) then (SuccessLevel is High)
7. If (SuccessQuantity is High) and (SuccessQuality is Low) then (SuccessLevel is Middle)
8. If (SuccessQuantity is High) and (SuccessQuality is Middle) then (SuccessLevel is High)
9. If (SuccessQuantity is High) and (SuccessQuality is High) then (SuccessLevel is High)

d) Defuzzification: Table 5-7 represents the fuzzy output set used to represent the students' success levels.

Linguistic Variable	Interval
Low-Success	[0 0 30 40]
Middle –Success	[30 40 60 70]
High-Success	[60 70 100 100]

Table 5-7: The Fuzzy Output Set of Success Level

5.4. FL for Collaborative Skills (Performance Outcomes Lens)

Although evaluating student skill levels can greatly enhance a learning process and help in the understanding of the students' weaknesses and strengths, assessing such skills as exhibited in an immersive environment, based on the students' actions, is complex. Thus, assessing the performance outcomes and the level of skill displayed by each student has been undertaken by using fuzzy logic (FL) approach to inferring the student's skills. The fuzzy logic method was used to combine all the data produced by both the natural agents and the automated agents so that determinations of the students' collaborative skill levels could be made. According to Hesse [14], students social collaborative problem-solving skills can be evaluated by determining three skills: participation, perspective taking and social regulation skills. In order to evaluate the students' participation skill in relation to the learning activity, we had to evaluate three sub-skills: actions, communications and task completion. In addition, to evaluate the perspective taking skill, the adaptive responsiveness and audience awareness sub-skills had to be evaluated. Furthermore, in order to evaluate the social regulation skill, four sub-skills had to be looked at: responsibility, negotiation, self-evaluation and transitive memory. The associated FLSs were developed for the purpose of feeding data into the FLS responsible for evaluating the exhibited social collaborative problem-solving skills overall (see Figure 5-7).

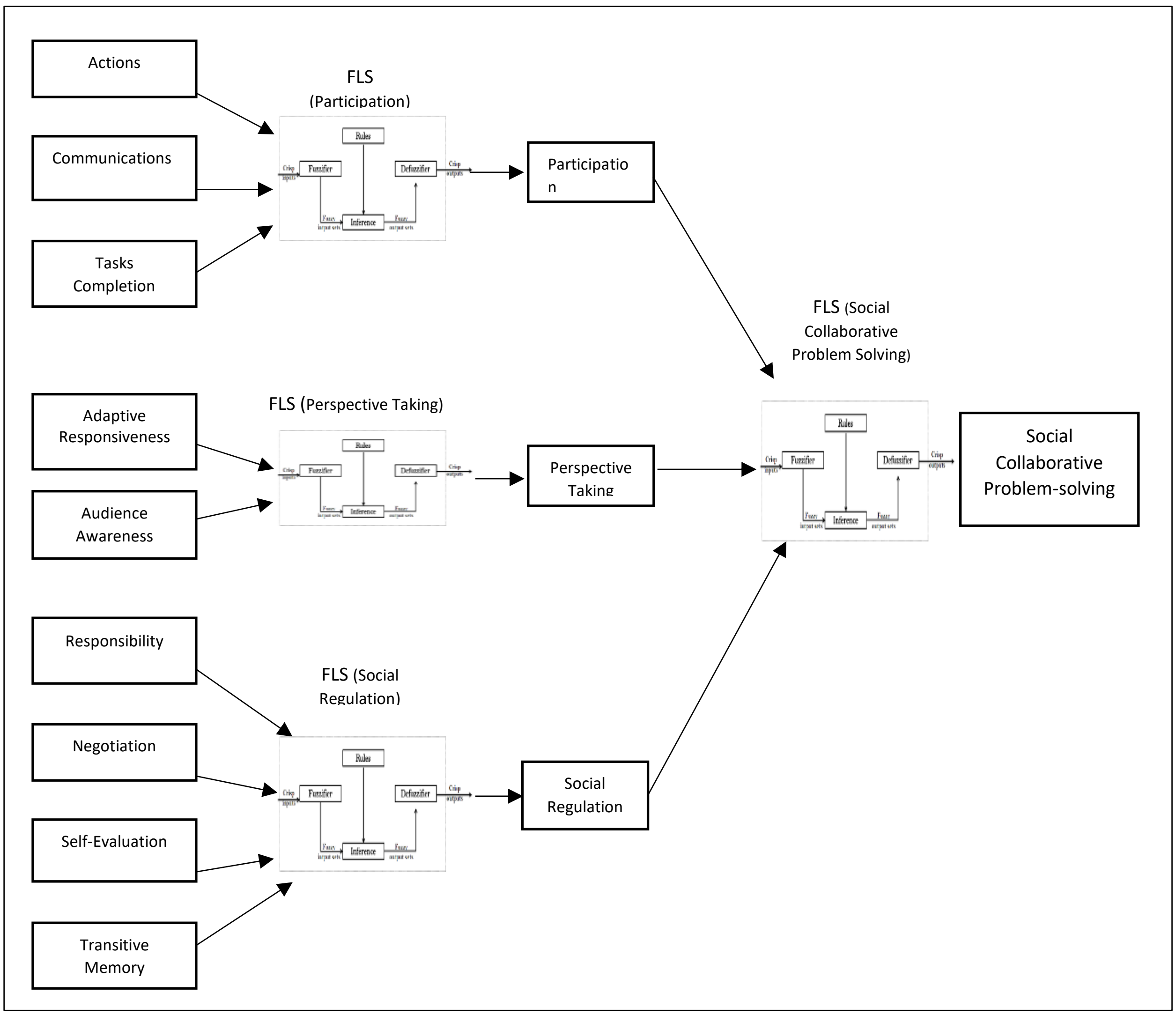

Figure 5-7: FLSs for Social Collaborative Problem-Solving Skills and Sub-Skills

In this diagram (Figure 5-7), the system is shown as a hierarchal FLS which encompasses multiple FL sub-systems: 1) the participation FLS, 2) the perspective taking FLS, 3) the social regulation FLS, and 4) the social and collaborative problem-solving FLS. Descriptions of each of these FLSs follow:

1) Participation FLS

Participation defines the minimum requirements for interaction. It refers to the willingness of students to share ideas and information, and to involve in the problem-solving steps [14]. Participation distinguishes between three quantity subskills: action, communication and task completion. These subskills are used as crisp inputs in the participation FLS (Figure 5-8).

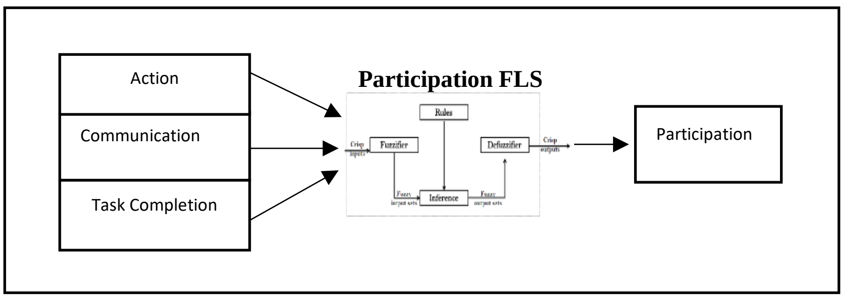

Figure 5-8: Participation FLS

The participation FLS crisp inputs are explained as follows:

- **Actions**

In order to obtain data about the students' actions from the system, we looked at the skills framework and tried to encode the different levels of action related to the system. As shown in Figure 5-9, Hesse [14] defined the action indicator as representing activity within the environment. He also classified action subskill as occurring at three levels: low, middle and high. Student action is low when the student exhibits little or no activity, middle when action is performed only in familiar contexts and high when actions are performed in both familiar and unfamiliar contexts.

Element	Indicator	Low	Middle	High
Participation				
Action	Activity within environment	No or very little activity	Activity in familiar contexts	Activity in familiar and unfamiliar contexts

Figure 5-9: Action Indicator Definition [14]

Therefore, the agent data concerned with student activity are calculated in order to understand the quantity of student activities related to both familiar and unfamiliar contexts. In addition, the action FLS (*Figure 5-10*) was created to evaluate the significance of each action sub-skill, using a number of different fuzzy rules — in order to simulate expert observation.

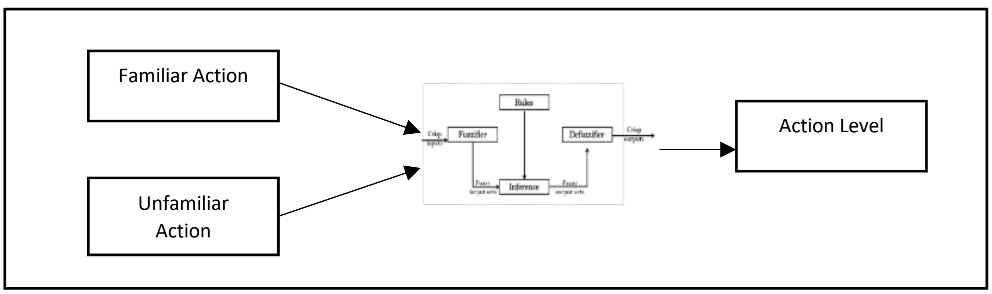

Figure 5-10: Action FLS

To evaluate whether students are performing unfamiliar actions or are only performing familiar actions before they start working on the learning task, the students receive a brief description of the learning environment. The researcher first explained to the students the concepts of sensors and smart homes and how they (the students) can use the facilities available in *Observe Portal* to program the smart home presented to them. This explanation includes only direct descriptions of the light sensors; this is so that the students can be assessed later with regard to whether or not they program any other type of sensor and/or explore the other kinds of virtual home appliance — so that their actions can be evaluated.

For instance, if a student uses the light sensors in creating rules along with the same IF rule that has actually been shown to them (before they entered the environment), the system considers that they (the student) is performing a familiar action. However, if a learner uses a different syntax for an IF rules and/or tries to control one of the other home appliances or sensors, the system counts this as an unfamiliar action. All these information have been explained to the experts before evaluating the students, so they understand the differences between familiar and unfamiliar action.

a) Crisp Input

Based on the previous action indicator definition, it can be seen that the crisp inputs for the action FLS are "familiar action," and "unfamiliar action," as shown in Figure 5-10.

b) Fuzzification:

The students were observed by the experts and the expert information about assessing the action subskill is then transferred to the ranges in the MFs. Accordingly, we divided actions to familiar MF and unfamiliar MF, and each MF has two linguistic variables "small amount" and "large amount".

The fuzzy set input related to familiar actions and its two linguistic variables is shown in Table 5-8. From experts' observation, the "small amount" value pertains if the user creates only a few "IF" rules in a familiar context (between 0 and 4) and the "large amount" value pertains if the user creates several familiar rules (more than 3). Further, Table *5-9* represents the fuzzy set input related to unfamiliar actions; like the fuzzy set related to familiar actions, it contains two linguistic variables "small amount" and "large amount" — for students who create a small or a large number of unfamiliar IF rules, respectively. In the same way that the previous FLSs were set up so that they were able to assess both groups and individuals using the same fuzzy input sets, here, the system sums the number of familiar/unfamiliar actions of the individual or the group and use these as crisp inputs. In addition, each interval value in the fuzzy input set is multiplied with n=1 if the assessment is for an individual student or n= the number of students if the assessment is for a group.

Linguistic Variable	Interval
Small Amount	[0n 0n 2n 4n]
Large Amount	[3n 10n 15n 15n]

Table 5-8: Familiar Action Fuzzy Set Input

Linguistic Variable	Interval
Small Amount	[0n 0n 2n 4n]
Large Amount	[3n 10n 15n 15n]

Table 5-9: Unfamiliar Action Fuzzy Set Input

c) Fuzzy Rules:

1. If (FamActions is SmallAmount) and (UnFamActions is SmallAmount) then (ActionsLevel is Low)
2. If (FamActions is LargeAmount) and (UnFamActions is SmallAmount) then (ActionsLevel is Middle)
3. If (FamActions is LargeAmount) and (UnFamActions is LargeAmount) then (ActionsLevel is High)
4. If (FamActions is SmallAmount) and (UnFamActions is LargeAmount) then (ActionsLevel is High)

d) Defuzzification:

The fuzzy output set related to the action levels given in Table 5-10.

Linguistic Variable	Interval
Low-Action	[0 0 30 40]
Medium-Action	[30 40 60 70]
High-Action	[60 70 100 100]

Table 5-10: Fuzzy Output Set for Action FLS

- **Communication**

The second input used by the participation FLS is communication. The skills framework constructed by Hesse [14] defined interaction, or communication, as the process of "interacting with, prompting and responding to the contributions of others". We have called the interaction indicator "communication" because here it is mostly concerned with communications between collaborative group members and also so that this indicator name is not confused with the interaction lens.

Student communication levels are classified as either low, middle or high. Student communication is low when it consists only of acknowledging communication directly or indirectly, such communication is classed as middle when it includes responding to cues and high when it includes the promoting and initiating of activity or interaction (Figure 5-11) *[14]*.

Element	Indicator	Low	Middle	High
Participation				
Interaction (Communication)	Interacting with, prompting and responding to the contributions of others	Acknowledges communication directly or indirectly	Responds to cues in communication	Initiates and promotes interaction or activity

Figure 5-11: Definitions for Communication (Interaction) Indicator [14]

Consequently, the agent data relating to communication is collected in order to understand the quantitative aspects of the students' communications. In addition, the communication FLS (Figure 5-12) was created to evaluate the sub-skill represented by the level of action exhibited by the students; the FLS operates using a number of different fuzzy rules — in order to simulate expert observation.

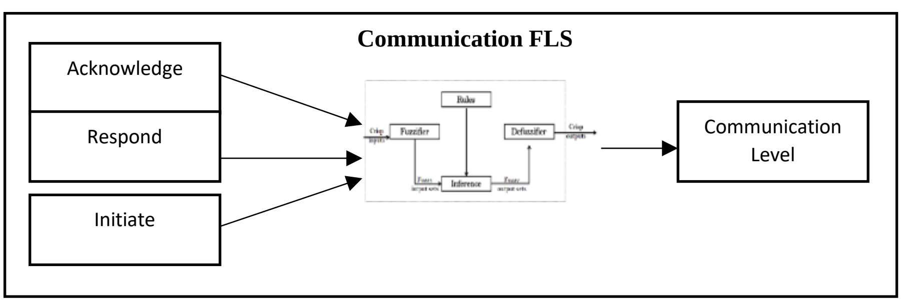

Figure 5-12: Communication FLS

However, it is complicated to develop a chat tool that can analyse students' communication automatically in the chatbox and classify conversation. Consequently, to obtain data from the system and evaluate whether students, via the chat facility, are simply acknowledging a communication, or are responding to cues, or are initiating interaction or activity, the chatbox itself presents a number of buttons each of which represents a pre-classified form of communication. Thus once a student has constructed a message, they classify it in order to send it. Figure 5-13 is a screenshot of the classified chat window; this has six buttons: Greeting, Reply, Inquiry, Agree, Suggestion and Solution.

Figure 5-13: Screenshot of the Classified Chat Box

Button	Button Code	Classification
Greeting	1	Acknowledgement
Reply	2	Responding
Agree	3	Responding
Inquiry	4	Initiating
Suggestion	5	Initiating
Solution	6	Initiating

Table 5-11: Classification of the Chat Buttons

Table 5-11 shows the levels associated with the chat buttons used to categorised users' messages: the system codes each button and saves, in the repository, which button was used. Then these data are used as inputs to the communication FLS. For instance, if a user writes "hi" in the chat and then clicks the "Greeting" button, the system simply saves the greeting button's code (1) in the database. Then, the system counts the number of times this button is pressed in a session and uses this statistic as a crisp value input to the FLS — in order to evaluate student communication.

a) <u>Crisp Inputs</u>

The values used for the communication FLS crisp inputs are: acknowledge, response and initiate.

b) <u>Fuzzification</u>

The expert information about assessing the communication subskill is transferred to the ranges in the MFs. The "acknowledge" fuzzy input set is divided into two linguistic variables "small amount" and "large amount" (

Table *5-12*). The "small amount" value represents the situation where the user clicks on the acknowledgement button between 0 and 4 times, and the "large amount" value represents the situation where the user clicks on it more than 3 times. Likewise,

Table *5-13* represents the "respond" fuzzy input set which also contains two linguistic variables, "small amount" and "large amount", for students who click the respond button a small or a large number of times respectively. Finally, the last table (Table 5-14) illustrates the initiate fuzzy input set, also containing two linguistic variables: small and large amount. Moreover, to be able to assess an individual student and also a group using the same fuzzy input sets, the system sums the number of communications made by each individual or group and uses these summed values as crisp inputs. In addition, each interval value in the fuzzy input set is multiplied with n=1 if the assessment is for a student or n= the number of students if the assessment is for a whole group.

Linguistic Variable	Interval
Small Amount	[0 0 2*n* 4*n*]
Large Amount	[3*n* 10*n* 50*n* 50*n*]

Table 5-12: Acknowledge Fuzzy Input Set

Linguistic Variable	Interval
Small Amount	[0 0 2*n* 4*n*]
Large Amount	[3*n* 10*n* 50*n* 50*n*]

Table 5-13: Response Fuzzy Input Set

Linguistic Variable	Interval
Small Amount	[0 0 2*n* 4*n*]
Large Amount	[3*n* 10*n* 50*n* 50*n*]

Table 5-14: Initiate Fuzzy Input Set

c) Fuzzy Rules:

The fuzzy rules for the communication FLS are:

1. If (AchknowldgeB1 is SmallAmount) and (ResponseB2 is SmallAmount) and (InitiateB3 is SmallAmount) then (CommunicationLevel is Low)
2. If (AchknowldgeB1 is SmallAmount) and (ResponseB2 is LargeAmount) and (InitiateB3 is SmallAmount) then (CommunicationLevel is Middle)
3. If (AchknowldgeB1 is LargeAmount) and (ResponseB2 is LargeAmount) and (InitiateB3 is SmallAmount) then (CommunicationLevel is Middle)
4. If (AchknowldgeB1 is LargeAmount) and (ResponseB2 is SmallAmount) and (InitiateB3 is SmallAmount) then (CommunicationLevel is Low)
5. If (AchknowldgeB1 is SmallAmount) and (ResponseB2 is SmallAmount) and (InitiateB3 is LargeAmount) then (CommunicationLevel is High)
6. If (AchknowldgeB1 is LargeAmount) and (ResponseB2 is SmallAmount) and (InitiateB3 is LargeAmount) then (CommunicationLevel is High)
7. If (AchknowldgeB1 is SmallAmount) and (ResponseB2 is LargeAmount) and (InitiateB3 is LargeAmount) then (CommunicationLevel is High)
8. If (AchknowldgeB1 is LargeAmount) and (ResponseB2 is LargeAmount) and (InitiateB3 is LargeAmount) then (CommunicationLevel is High)

d) Defuzzification:

The fuzzy output set of the communication level is in Table *5-15*.

Linguistic Variable	Interval
Low-communication	[0 0 30 40]
Middle- communication	[30 40 60 70]
High- communication	[60 70 100 100]

Table 5-15: Fuzzy Output Set for Communication

- **Tasks Completion**

Task completion is defined as "undertaking and completing a task or part of a task individually" [14]. The task completion evaluation is considered low when a student maintains a presence but does not do anything, medium when the student identifies and attempts a task and high when the student completes a task and/or uses multiple different strategies (Figure 5-14).

Element	Indicator	Low	Middle	High
Participation				
Task completion/ perseverance	Undertaking and completing a task or part of a task individually	Maintains presence only	Identifies and attempts the task	Perseveres in task as indicated by repeated attempts or multiple strategies

Figure 5-14: Task Completion Indicator Definition [14]

Based on the definition given above and based on the experts' observation, the task completion fuzzy set was developed so that a value of "low" is yielded when a student completes between 0 and 2 rules, middle when the student attempts between 1 and 5 rules, and high when the student attempts more than 4 rules. Also, the fuzzy input set, as before, is constructed taking into account whether it is for an individual (n=1) or for a group (n=the number of members of the group), see Table 5-16.

Linguistic Variable	**Interval**
Low	$[0 \;\; 0 \;\; 1n \;\; 2n]$
Middle	$[1n \;\; 2n \;\; 4n \;\; 5n]$
High	$[4n \;\; 5n \;\; 30n \;\; 30n]$

Table 5-16: Task Completion Fuzzy Input Set

Participation FLS:

Finally, the previously described outputs, from the action FLS and the communication FLS, along with the task completion fuzzy output set are all used as inputs to the participation FLS (Figure 5-15).

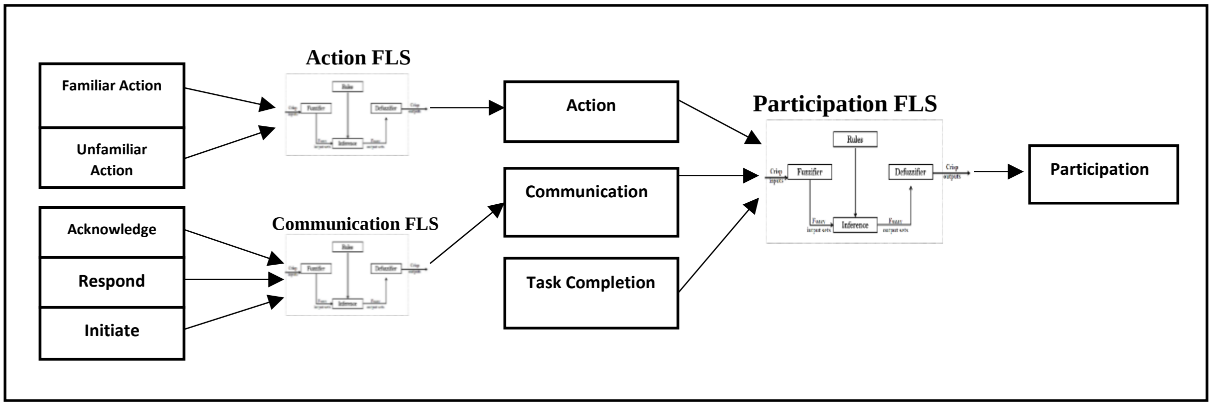

Figure 5-15: Inputs for Participation FLS

The linguistic rules which have been developed for the Participation FLS system are:

1. If (Actions is High) and (Communication is High) and (TaskComp is High) then (Participation is High)
2. If (Actions is High) and (Communication is High) and (TaskComp is Middle) then (Participation is High)
3. If (Actions is High) and (Communication is High) and (TaskComp is Low) then (Participation is High)
4. If (Actions is High) and (Communication is Middle) and (TaskComp is High) then (Participation is High)
5. If (Actions is High) and (Communication is Middle) and (TaskComp is Middle) then (Participation is Middle)
6. If (Actions is High) and (Communication is Middle) and (TaskComp is Low) then (Participation is Middle)
7. If (Actions is High) and (Communication is Low) and (TaskComp is High) then (Participation is Middle)
8. If (Actions is High) and (Communication is Low) and (TaskComp is Middle) then (Participation is Middle)
9. If (Actions is High) and (Communication is Low) and (TaskComp is Low) then (Participation is Low)
10. If (Actions is Middle) and (Communication is High) and (TaskComp is High) then (Participation is High)
11. If (Actions is Middle) and (Communication is High) and (TaskComp is Middle) then (Participation is Middle)
12. If (Actions is Middle) and (Communication is High) and (TaskComp is Low) then (Participation is Middle)
13. If (Actions is Middle) and (Communication is Middle) and (TaskComp is High) then (Participation is Middle)
14. If (Actions is Middle) and (Communication is Middle) and (TaskComp is Middle) then (Participation is Middle)
15. If (Actions is Middle) and (Communication is Middle) and (TaskComp is Low) then (Participation is Middle)
16. If (Actions is Middle) and (Communication is Low) and (TaskComp is High) then (Participation is Middle)
17. If (Actions is Middle) and (Communication is Low) and (TaskComp is Middle) then (Participation is Middle)
18. If (Actions is Middle) and (Communication is Low) and (TaskComp is Low) then (Participation is Low)
19. If (Actions is Low) and (Communication is High) and (TaskComp is High) then (Participation is Middle)
20. If (Actions is Low) and (Communication is High) and (TaskComp is Middle) then (Participation is Middle)
21. If (Actions is Low) and (Communication is High) and (TaskComp is Low) then (Participation is Middle)
22. If (Actions is Low) and (Communication is Middle) and (TaskComp is High) then (Participation is Middle)
23. If (Actions is Low) and (Communication is Middle) and (TaskComp is Middle) then (Participation is Middle)
24. If (Actions is Low) and (Communication is Middle) and (TaskComp is Low) then (Participation is Low)
25. If (Actions is Low) and (Communication is Low) and (TaskComp is High) then (Participation is Middle)
26. If (Actions is Low) and (Communication is Low) and (TaskComp is Middle) then (Participation is Low)
27. If (Actions is Low) and (Communication is Low) and (TaskComp is Low) then (Participation is Low)

2) Perspective Taking FLS

Perspective taking skill refers to the ability to view a problem via the other group member eyes; sometimes a group cannot find a solution without understanding the actual situation of the collaborators [160]. Perspective taking sub-skills are: adaptive responsiveness and audience awareness. The perspective taking FLS in Figure 5-16 includes the following:

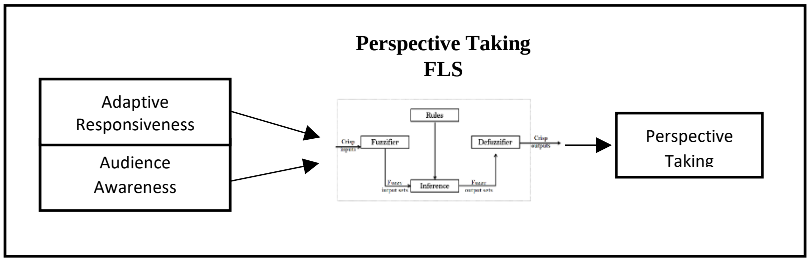

Figure 5-16: Perspective Taking FLS

a) Crisp inputs: The crisp inputs for the perspective taking FLS are: data relating to the adaptive responsiveness and audience awareness sub-skills. The crisp inputs data for this FLS are taken from the ratings that the students have acquired by the end of the learning session (each group member rates the others) because the perspective taking sub-skills refer more about the quality of interaction and the natural agents in this research are utilised for the quality of collaboration. The rating from the other students is always from 0 to 2 (Low (0), Middle (1), High (2)).

b) Fuzzifier: the fuzzy input sets representing the adaptive responsiveness and audience awareness sub-skills are as follows — shown in Table 5-17 and Table 5-18:

Linguistic Variable	**Interval**
Low	[0 0 0.5 0.7]
Middle	[0.5 1 1.5 1.7]
High	[1.5 1.7 2 2]

Table 5-17: Fuzzy Input Set for Adaptive Responsiveness

Linguistic Variable	**Interval**
Low	[0 0 0.5 0.7]
Middle	[0.5 1 1.5 1.7]
High	[1.5 1.7 2 2]

Table 5-18: Fuzzy Input Set for Audience Awareness

c) Fuzzy Rules:

The linguistic rules for the perspective taking FLS are:

1. If (AdaptiveRespons is High) and (Awareness is High) then (PerspectiveTalking is High)
2. If (AdaptiveRespons is High) and (Awareness is Middle) then (PerspectiveTalking is High)
3. If (AdaptiveRespons is High) and (Awareness is Low) then (PerspectiveTalking is Middle)
4. If (AdaptiveRespons is Middle) and (Awareness is High) then (PerspectiveTalking is High)
5. If (AdaptiveRespons is Middle) and (Awareness is Middle) then (PerspectiveTalking is Middle)

6. If (AdaptiveRespons is Middle) and (Awareness is Low) then (PerspectiveTalking is Middle)
7. If (AdaptiveRespons is Low) and (Awareness is High) then (PerspectiveTalking is Middle)
8. If (AdaptiveRespons is Low) and (Awareness is Middle) then (PerspectiveTalking is Middle)
9. If (AdaptiveRespons is Low) and (Awareness is Low) then (PerspectiveTalking is Low)

d) Defuzzification: This converts the fuzzy output sets to crisp output values (*Table 5-19*), using the centroid defuzzification method.

Linguistic Variable	**Interval**
Low- Perspective Taking	[0 0 30 40]
Middle- Perspective Taking	[30 40 60 70]
High- Perspective Taking	[60 70 100 100]

Table 5-19: Perspective Taking Fuzzy Output Set

3) **Social Regulation FLS**

The social regulation skills refer to the strategic part of collaborative problem-solving [161]. Generally, students use the awareness of the other group members strengths and weaknesses, to resolve differences in strategies or viewpoints. The social regulations sub-skills are: responsibility, negotiation, transactive memory and self-evaluation. Figure *5-17* illustrates the fuzzy logic system for social regulation.

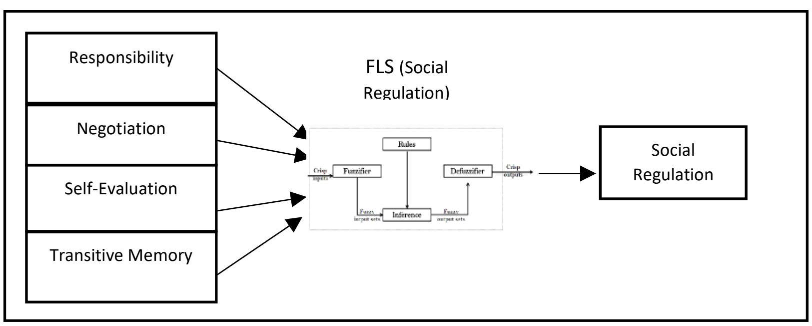

Figure 5-17: Social Regulation FLS

a) <u>Crisp inputs</u>: The crisp inputs for the Social regulation FLS are data representing students': responsibility, negotiation, self-evaluation and transitive memory. Data which evaluates the strengths of these sub-skills are taken from the ratings which have been acquired by students by the end of the learning session because these subskills also refer to the quality of collaboration. The rating ranges from 0 to 2 (Low (0), Middle (1), High (2)).

b) <u>Fuzzifier</u>: The fuzzy input sets representing responsibility, negotiation, self-evaluation and transitive memory all have the same linguistic values because all of them are based on rating data from the students (Table 5-20).

Linguistic Variable	Interval
Low	[0 0 0.5 0.7]
Medium	[0.5 1 1.5 1.7]
High	[1.5 1.7 2 2]

Table 5-20: The Form of The Fuzzy Input Sets for Responsibility, Negotiation, Self-Evaluation, and Transitive Memory

c) <u>Fuzzy Rules</u>: The linguistic rules used by the Social Regulation FLS are:

1. If (Response is Low) and (Negotiation is Low) and (SelfEvalu is Low) and (Transitive is Low) then (SocialReg is Low)
2. If (Response is Low) and (Negotiation is Low) and (SelfEvalu is Low) and (Transitive is Middle) then (SocialReg is Low)
3. If (Response is Low) and (Negotiation is Low) and (SelfEvalu is Low) and (Transitive is High) then (SocialReg is Middle)
4. If (Response is Low) and (Negotiation is Low) and (SelfEvalu is Middle) and (Transitive is Low) then (SocialReg is Low)
5. If (Response is Low) and (Negotiation is Low) and (SelfEvalu is Middle) and (Transitive is Middle) then (SocialReg is Middle)
6. If (Response is Low) and (Negotiation is Low) and (SelfEvalu is Middle) and (Transitive is High) then (SocialReg is Middle)
7. If (Response is Low) and (Negotiation is Low) and (SelfEvalu is High) and (Transitive is Low) then (SocialReg is Middle)
8. If (Response is Low) and (Negotiation is Low) and (SelfEvalu is High) and (Transitive is Middle) then (SocialReg is Middle)
9. If (Response is Low) and (Negotiation is Low) and (SelfEvalu is High) and (Transitive is High) then (SocialReg is Middle)
10. If (Response is Low) and (Negotiation is Middle) and (SelfEvalu is Low) and (Transitive is Low) then (SocialReg is Low)
11. If (Response is Low) and (Negotiation is Middle) and (SelfEvalu is Low) and (Transitive is Middle) then (SocialReg is Middle)
12. If (Response is Low) and (Negotiation is Middle) and (SelfEvalu is Low) and (Transitive is High) then (SocialReg is Middle)
13. If (Response is Low) and (Negotiation is Middle) and (SelfEvalu is Middle) and (Transitive is Low) then (SocialReg is Middle)
14. If (Response is Low) and (Negotiation is Middle) and (SelfEvalu is Middle) and (Transitive is Middle) then (SocialReg is Middle)
15. If (Response is Low) and (Negotiation is Middle) and (SelfEvalu is Middle) and (Transitive is High) then (SocialReg is Middle)
16. If (Response is Low) and (Negotiation is Middle) and (SelfEvalu is High) and (Transitive is Low) then (SocialReg is Middle)
17. If (Response is Low) and (Negotiation is Middle) and (SelfEvalu is High) and (Transitive is Middle) then (SocialReg is Middle)
18. If (Response is Low) and (Negotiation is Middle) and (SelfEvalu is High) and (Transitive is High) then (SocialReg is Middle)
19. If (Response is Low) and (Negotiation is High) and (SelfEvalu is Low) and (Transitive is Low) then (SocialReg is Middle)
20. If (Response is Low) and (Negotiation is High) and (SelfEvalu is Low) and (Transitive is Middle) then (SocialReg is Middle)
21. If (Response is Low) and (Negotiation is High) and (SelfEvalu is Low) and (Transitive is High) then (SocialReg is Middle)
22. If (Response is Low) and (Negotiation is High) and (SelfEvalu is Middle) and (Transitive is Low) then (SocialReg is Middle)
23. If (Response is Low) and (Negotiation is High) and (SelfEvalu is Middle) and (Transitive is Middle) then (SocialReg is Middle)
24. If (Response is Low) and (Negotiation is High) and (SelfEvalu is Middle) and (Transitive is High) then (SocialReg is Middle)
25. If (Response is Low) and (Negotiation is High) and (SelfEvalu is High) and (Transitive is Low) then (SocialReg is Middle)
26. If (Response is Low) and (Negotiation is High) and (SelfEvalu is High) and (Transitive is Middle) then (SocialReg is Middle)
27. If (Response is Low) and (Negotiation is High) and (SelfEvalu is High) and (Transitive is High) then (SocialReg is High)
28. If (Response is Middle) and (Negotiation is Low) and (SelfEvalu is Low) and (Transitive is Low) then (SocialReg is Low)
29. If (Response is Middle) and (Negotiation is Low) and (SelfEvalu is Low) and (Transitive is Middle) then (SocialReg is Middle)
30. If (Response is Middle) and (Negotiation is Low) and (SelfEvalu is Low) and (Transitive is High) then (SocialReg is Middle)
31. If (Response is Middle) and (Negotiation is Low) and (SelfEvalu is Middle) and (Transitive is Low) then (SocialReg is Middle)
32. If (Response is Middle) and (Negotiation is Low) and (SelfEvalu is Middle) and (Transitive is Middle) then (SocialReg is Middle)

33. If (Response is Middle) and (Negotiation is Low) and (SelfEvalu is Middle) and (Transitive is High) then (SocialReg is Middle)
34. If (Response is Middle) and (Negotiation is Low) and (SelfEvalu is High) and (Transitive is Low) then (SocialReg is Middle)
35. If (Response is Middle) and (Negotiation is Low) and (SelfEvalu is High) and (Transitive is Middle) then (SocialReg is Middle)
36. If (Response is Middle) and (Negotiation is Low) and (SelfEvalu is High) and (Transitive is High) then (SocialReg is Middle)
37. If (Response is Middle) and (Negotiation is Middle) and (SelfEvalu is Low) and (Transitive is Low) then (SocialReg is Middle)
38. If (Response is Middle) and (Negotiation is Middle) and (SelfEvalu is Low) and (Transitive is Middle) then (SocialReg is Middle)
39. If (Response is Middle) and (Negotiation is Middle) and (SelfEvalu is Low) and (Transitive is High) then (SocialReg is Middle)
40. If (Response is Middle) and (Negotiation is Middle) and (SelfEvalu is Middle) and (Transitive is Low) then (SocialReg is Middle)
41. If (Response is Middle) and (Negotiation is Middle) and (SelfEvalu is Middle) and (Transitive is Middle) then (SocialReg is Middle)
42. If (Response is Middle) and (Negotiation is Middle) and (SelfEvalu is Middle) and (Transitive is High) then (SocialReg is Middle)
43. If (Response is Middle) and (Negotiation is Middle) and (SelfEvalu is High) and (Transitive is Low) then (SocialReg is Middle)
44. If (Response is Middle) and (Negotiation is Middle) and (SelfEvalu is High) and (Transitive is Middle) then (SocialReg is Middle)
45. If (Response is Middle) and (Negotiation is Middle) and (SelfEvalu is High) and (Transitive is High) then (SocialReg is Middle)
46. If (Response is Middle) and (Negotiation is High) and (SelfEvalu is Low) and (Transitive is Low) then (SocialReg is Low)
47. If (Response is Middle) and (Negotiation is High) and (SelfEvalu is Low) and (Transitive is Middle) then (SocialReg is Middle)
48. If (Response is Middle) and (Negotiation is High) and (SelfEvalu is Low) and (Transitive is High) then (SocialReg is High)
49. If (Response is Middle) and (Negotiation is High) and (SelfEvalu is Middle) and (Transitive is Low) then (SocialReg is Middle)
50. If (Response is Middle) and (Negotiation is High) and (SelfEvalu is Middle) and (Transitive is Middle) then (SocialReg is Middle)
51. If (Response is Middle) and (Negotiation is High) and (SelfEvalu is Middle) and (Transitive is High) then (SocialReg is High)
52. If (Response is Middle) and (Negotiation is High) and (SelfEvalu is High) and (Transitive is Low) then (SocialReg is Middle)
53. If (Response is Middle) and (Negotiation is High) and (SelfEvalu is High) and (Transitive is Middle) then (SocialReg is High)
54. If (Response is Middle) and (Negotiation is High) and (SelfEvalu is High) and (Transitive is High) then (SocialReg is High)
55. If (Response is High) and (Negotiation is Low) and (SelfEvalu is Low) and (Transitive is Low) then (SocialReg is Middle)
56. If (Response is High) and (Negotiation is Low) and (SelfEvalu is Low) and (Transitive is Middle) then (SocialReg is Middle)
57. If (Response is High) and (Negotiation is Low) and (SelfEvalu is Low) and (Transitive is High) then (SocialReg is Middle)
58. If (Response is High) and (Negotiation is Low) and (SelfEvalu is Middle) and (Transitive is Low) then (SocialReg is Middle)
59. If (Response is High) and (Negotiation is Low) and (SelfEvalu is Middle) and (Transitive is Middle) then (SocialReg is Middle)
60. If (Response is High) and (Negotiation is Low) and (SelfEvalu is Middle) and (Transitive is High) then (SocialReg is Middle)
61. If (Response is High) and (Negotiation is Low) and (SelfEvalu is High) and (Transitive is Low) then (SocialReg is Middle)
62. If (Response is High) and (Negotiation is Low) and (SelfEvalu is High) and (Transitive is Middle) then (SocialReg is Middle)
63. If (Response is High) and (Negotiation is Low) and (SelfEvalu is High) and (Transitive is High) then (SocialReg is High)
64. If (Response is High) and (Negotiation is Middle) and (SelfEvalu is Low) and (Transitive is Low) then (SocialReg is Middle)
65. If (Response is High) and (Negotiation is Middle) and (SelfEvalu is Low) and (Transitive is Middle) then (SocialReg is Middle)
66. If (Response is High) and (Negotiation is Middle) and (SelfEvalu is Low) and (Transitive is High) then (SocialReg is Middle)
67. If (Response is High) and (Negotiation is Middle) and (SelfEvalu is Middle) and (Transitive is Low) then (SocialReg is Middle)
68. If (Response is High) and (Negotiation is Middle) and (SelfEvalu is Middle) and (Transitive is Middle) then (SocialReg is Middle)
69. If (Response is High) and (Negotiation is Middle) and (SelfEvalu is Middle) and (Transitive is High) then (SocialReg is High)
70. If (Response is High) and (Negotiation is Middle) and (SelfEvalu is High) and (Transitive is Low) then (SocialReg is Middle)
71. If (Response is High) and (Negotiation is Middle) and (SelfEvalu is High) and (Transitive is Middle) then (SocialReg is High)
72. If (Response is High) and (Negotiation is Middle) and (SelfEvalu is High) and (Transitive is High) then (SocialReg is High)
73. If (Response is High) and (Negotiation is High) and (SelfEvalu is Low) and (Transitive is Low) then (SocialReg is Middle)
74. If (Response is High) and (Negotiation is High) and (SelfEvalu is Low) and (Transitive is Middle) then (SocialReg is Middle)
75. If (Response is High) and (Negotiation is High) and (SelfEvalu is Low) and (Transitive is High) then (SocialReg is High)
76. If (Response is High) and (Negotiation is High) and (SelfEvalu is Middle) and (Transitive is Low) then (SocialReg is Middle)
77. If (Response is High) and (Negotiation is High) and (SelfEvalu is Middle) and (Transitive is Middle) then (SocialReg is High)
78. If (Response is High) and (Negotiation is High) and (SelfEvalu is Middle) and (Transitive is High) then (SocialReg is Middle)
79. If (Response is High) and (Negotiation is High) and (SelfEvalu is High) and (Transitive is Low) then (SocialReg is High)
80. If (Response is High) and (Negotiation is High) and (SelfEvalu is High) and (Transitive is Middle) then (SocialReg is High)
81. If (Response is High) and (Negotiation is High) and (SelfEvalu is High) and (Transitive is High) then (SocialReg is High)

d) Defuzzification: This converts fuzzy output sets into crisp output values (Table 5-21), using the centroid defuzzification method.

Linguistic Variable	Interval
Low	[0 0 30 40]
Middle	[30 40 60 70]
High	[60 70 100 100]

Table 5-21: Social Regulation Fuzzy Output Set

4) Social collaborative problem-solving FLS

1. Crisp inputs: The crisp inputs for the social collaborative problem-solving FLS consist of outputs from previously described FLSs: the participation FLS, the perspective taking FLS, and the social regulation FLS. These inputs are combined and processed in order to make the required decisions concerning student social collaborative skills levels (Figure 5-18).

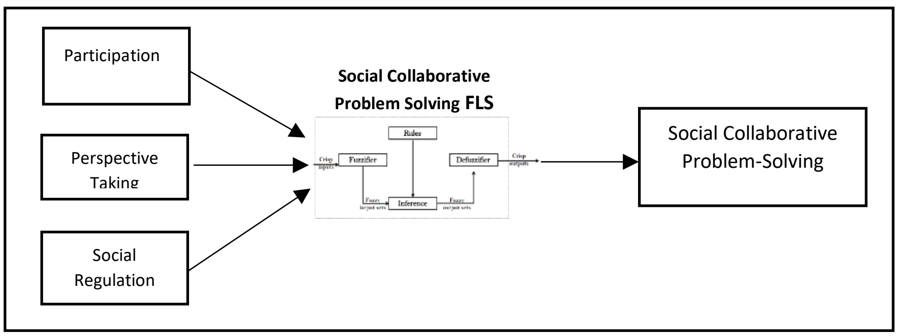

Figure 5-18: Social Collaborative Problem-Solving FLS

2. Fuzzifier: The fuzzy inputs sets, participation, perspective taking, and social regulation, all use the same linguistic variables (low, medium, high) and the same numerical values, which range from 0 to 100 (Table 5-22).

Linguistic Variable	Interval
Low	[0 0 30 40]
Medium	[30 40 60 70]
High	[60 70 100 100]

Table 5-22: The Form of the Fuzzy Inputs Set Representing Participation, Perspective Taking and Social Regulation

3. Fuzzy Rules: The fuzzy rules used by the social collaborative FLS are:

1. If (Participation is High) and (PeresTaking is High) and (SocialReg is High) then (Collaborative_Skill_Level is High)
2. If (Participation is High) and (PeresTaking is High) and (SocialReg is Middle) then (Collaborative_Skill_Level is High)
3. If (Participation is High) and (PeresTaking is High) and (SocialReg is Low) then (Collaborative_Skill_Level is Middle)
4. If (Participation is High) and (PeresTaking is Middle) and (SocialReg is High) then (Collaborative_Skill_Level is High)
5. If (Participation is High) and (PeresTaking is Middle) and (SocialReg is Middle) then (Collaborative_Skill_Level is Middle)
6. If (Participation is High) and (PeresTaking is Middle) and (SocialReg is Low) then (Collaborative_Skill_Level is Middle)
7. If (Participation is High) and (PeresTaking is Low) and (SocialReg is High) then (Collaborative_Skill_Level is Middle)
8. If (Participation is High) and (PeresTaking is Low) and (SocialReg is Middle) then (Collaborative_Skill_Level is Middle)
9. If (Participation is High) and (PeresTaking is Low) and (SocialReg is Low) then (Collaborative_Skill_Level is Middle)
10. If (Participation is Middle) and (PeresTaking is High) and (SocialReg is High) then (Collaborative_Skill_Level is High)
11. If (Participation is Middle) and (PeresTaking is High) and (SocialReg is Middle) then (Collaborative_Skill_Level is Middle)
12. If (Participation is Middle) and (PeresTaking is High) and (SocialReg is Low) then (Collaborative_Skill_Level is Middle)
13. If (Participation is Middle) and (PeresTaking is Middle) and (SocialReg is High) then (Collaborative_Skill_Level is Middle)
14. If (Participation is Middle) and (PeresTaking is Middle) and (SocialReg is Middle) then (Collaborative_Skill_Level is Middle)
15. If (Participation is Middle) and (PeresTaking is Middle) and (SocialReg is Low) then (Collaborative_Skill_Level is Middle)
16. If (Participation is Middle) and (PeresTaking is Low) and (SocialReg is High) then (Collaborative_Skill_Level is Middle)
17. If (Participation is Middle) and (PeresTaking is Low) and (SocialReg is Middle) then (Collaborative_Skill_Level is Middle)
18. If (Participation is Middle) and (PeresTaking is Low) and (SocialReg is Low) then (Collaborative_Skill_Level is Low)
19. If (Participation is Low) and (PeresTaking is High) and (SocialReg is High) then (Collaborative_Skill_Level is Middle)
20. If (Participation is Low) and (PeresTaking is High) and (SocialReg is Middle) then (Collaborative_Skill_Level is Middle)
21. If (Participation is Low) and (PeresTaking is High) and (SocialReg is Low) then (Collaborative_Skill_Level is Middle)
22. If (Participation is Low) and (PeresTaking is Middle) and (SocialReg is High) then (Collaborative_Skill_Level is Middle)
23. If (Participation is Low) and (PeresTaking is Middle) and (SocialReg is Middle) then (Collaborative_Skill_Level is Middle)
24. If (Participation is Low) and (PeresTaking is Middle) and (SocialReg is Low) then (Collaborative_Skill_Level is Low)
25. If (Participation is Low) and (PeresTaking is Low) and (SocialReg is High) then (Collaborative_Skill_Level is Middle)
26. If (Participation is Low) and (PeresTaking is Low) and (SocialReg is Middle) then (Collaborative_Skill_Level is Low)
27. If (Participation is Low) and (PeresTaking is Low) and (SocialReg is Low) then (Collaborative_Skill_Level is Low)

4. Defuzzification: This converts the output fuzzy set into crisp output values, using the centroid defuzzification method (Table *5-23*).

Linguistic Variable	Interval
Low	[0 0 30 40]
Medium	[30 40 60 70]
High	[60 70 100 100]

Table 5-23: Social Collaborative Problem-Solving Output Fuzzy Set

All these FLSs were applied in real-time in *Observe Portal* to evaluate the skill and sub-skills levels being demonstrated by the students.

5.5. Chapter Summary

This chapter has presented details of the fuzzy logic systems created to evaluate students in relation to each lens. Determining and assessing the quality and quantity of students' interactions, success and skills is not a straightforward task. Such assessments depend on many elements that cannot be directly measured or observed. The factors involve variables that are not fixed and evaluating them requires the acknowledgement of human subjectivity and uncertainty. Therefore, FL is well-suited to the purpose of our current prototype; the FL inference system is used to deal with uncertainty values and reason with data in the same manner that a human being would. Data from the *MixAgent* model was processed using the fuzzy reasoning approach, combining the data variously generated by the different agents. In addition, a hierarchical FLS has been developed to assess the learners in terms of each lens; this assessment includes looking at students' interactions, success and collaborative skills. The FLS also facilitated inferring about and assessing the learning outcomes obtained from taking part in the collaborative activities - by individual students and by groups. The results from the FLSs implemented were shown to students and teachers graphically via flow charts, as described in Chapter 4. These charts show the level of students' interactions, successes and collaborative skills (low, middle, or high) to help in understanding the various learners' outcomes.

Evaluating the research prototype is of great importance because only with relevant feedback can the research progress. Thus, the next chapter presents the evaluations of the research. These consist of both user (students) and expert (teachers) evaluations of the validity of the models. The evaluations were constructed as the result of a large number of educational sessions carried out in the virtual world. Each session used a group of students to imitate collaboration between learners and one computer science expert to imitate a teacher's classroom observations. Completing the evaluation validates the framework used in this research. The research evaluation experiments are demonstrated in detail in Chapter 6.

Chapter Six

6. Experimental Design and Evaluation

"Experience is the teacher of all things."

— Julius Caesar

Earlier chapters have described the research framework and models, the system architecture and the implementation of the proof-of-concept prototype (*Observe Portal*). The current chapter explains our evaluation of the *Observe Portal* assessment system through an experiment employing hands-on evaluation activities with real users. It begins with an introduction to the evaluation methods and the user studies which have been employed for 3D virtual environments, then explains the evaluation methods and measures adopted in the present study, including the experimental design developed to assemble evidence of the success of the concepts and models implemented in this research.

6.1.Evaluation Methods

We must first consider how it is appropriate to evaluate any software with educational applications, such as ours, especially in the early stages of its development. One way that has often been used is been by measuring, specifically, the user acceptance of such new applications, or measuring variables which are claimed to predict future acceptance. Thus, user acceptance is briefly discussed in this section.

In addition, most e-learning systems are aiming at improving the ease with which learning can be achieved by students - by measuring learning effectiveness to evaluate the success of the system. However, our research is less concerned with learning effectiveness measures, and more concerned with how we can enable more effective feedback and assessment; this represents a different measure of success. Effective feedback is a factor that strongly influences learning and impacts student performance [162-164]. And so, because our prototype mostly focuses on student assessment, we need to discuss the methods used to measure student experience with the assessment feedback. Therefore, user acceptance and student experiences are briefly reviewed in the following sub-sections. Both of these factors must be considered in order to answer our research hypothesis.

6.2.Research Hypotheses

Here is a list of the hypotheses presented in Chapter 1. The hypotheses are relevant to evaluating the research framework and prototype under investigation in the present study. If confirmed, they would indicate a highly positive evaluation of that framework and prototype. The hypotheses are:

It is possible to create a computational observation framework which can support the gathering of the learning evidence from collaborative distance students using immersive environments – and which is capable of being used as the basis for student assessment.

1. *Such an observation framework will be able to integrate between software and human agents - such that users will have positive attitudes towards their roles as human agents when performing distance-learning tasks in the virtual world (**H1**).*
2. *Systems based on this observation framework will provide collaborative distance learners with assessment feedback, and users will report positive experiences of such assessments (**H2**).*

3. *Systems based on this observation framework will provide assessments that are very similar to human-expert assessments; these system assessments will be produced using less effort overall (H3).*
4. *The assessment results and feedback from such an observation system will be preferred by users over and above that yielded from human experts (**H4**).*
5. *Users of the observation system will express their acceptance of it (**H5**).*

6.2.1. Measuring Users' Acceptance

Numerous studies have stated the importance of understanding and measuring technology acceptance because the success of any new technology is largely based on its adoption by users [165, 166]. To this we would add that successful use is also important, although it is not measured by typical acceptance models. On the other hand, it is regarded as critical for organizations to identify the acceptance level of innovations before spending time and money on their implementation. Consequently, many models have been developed in the field of technology to help understand the acceptance of new technologies, including the following: the Theory of Planned Behaviour [167], the Theory of Reasoned Action (TRA) [168], Innovation Diffusion Theory (IDT) [169], the Unified Theory of Acceptance and Use of Technology (UTAUT)[170], the System Usability Scale [171] and the Technology Acceptance Model (TAM) [165].

TAM is the approach which has been most widely used to verify the acceptance of new technologies. Some of the technologies to which TAM has been applied in order to evaluate the users' acceptance are: multimedia learning systems [172], PC or microcomputers [173, 174], database programs [175], blackboards [176], expert support systems [177], 3D virtual environments [178-180] and mixed-reality environments [52, 181]. TAM had been also employed to measure teachers' acceptance of 3D-VW assessment systems such as *WorldOfQuestions* [182].

TAM identifies two specific, essential variables which must be assessed: the perceived usefulness (PU) of a new technology, and its perceived ease of use (PEOU). Davis [165] explained that high levels of both perceived usefulness and perceived ease of use will positively affect the utilisation of a new technology. His original model also draws attention to other variables and relationships which need to be measured in order to gain the fullest understanding of this issue, as shown in Figure 6-1. The model claims that actual system use will be highly predictable, in the way mapped out in Figure 6-1, provided PU, PEOU, certain external variables and also behavioural intention to use are all measured. Thus, the model can help us to determine the future acceptance or rejection of a technology. Additionally, it can assist in predicting users' behavioural intentions to use a technology [183, 184].

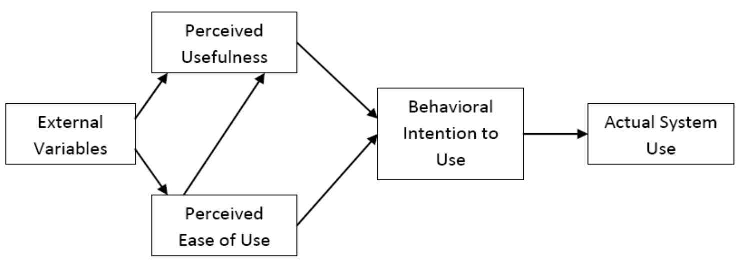

Figure 6-1: The Technology Acceptance Model [165]

TAM Variables

Perceived ease of use (PEOU) is a major element in the acceptance of a technology. Davis [165] defined PEOU as "the degree to which a person believes that using a particular system would be free of effort". Perceived ease of use is therefore particularly important in evaluating a system. Usually, people may reject a technology, or might apply it ineffectively, if a system is useful but hard to use. Systems which are easier to use are expected to be more acceptable to users [185].

In addition, perceived usefulness (PU) refers to “the degree to which a person believes that using a particular system would enhance his or her job performance” [165]. People tend to apply a system if they think it will assist them in accomplishing their work efficiently. They also express more interest in enhancing their performance if the usefulness of a technology intended to help them do this is apparent. Perceived usefulness is expected to be affected by perceived ease of use, but not the reverse (Figure 6-1): the easier a system is to use, the more useful it can be perceived to be [183, 186].

Finally, the intention to use (IU) is a significant variable in evaluating technology adoption and potential usage. Many studies have reported that the use of a system is determined by the intention to use it [187-189]. Therefore, the intention to use a system should be considered as an important indicator of the value of the implementation and actual use of a technology.

Following the TAM principles, both experts and students in this study were asked several questions related to perceived usefulness, perceived ease of use and the intention to use after some experience actually using the research assessment prototype.

6.2.2. Measuring Student Experiences with the Assessment Feedback

Effective feedback is essential for learning because it helps students to improve their performance. The quality of feedback is the factor that strongly distinguishes between good and bad courses [162, 163]. Black and Wiliam [164] also stated that the nature and quality of feedback can have either a positive or negative effect on learning and on students’ performance. Accordingly, many studies have been carried out to measure the quality of feedback. For instance, Hounsell, McCune, Hounsell and Litjens [190] studied learners' experiences of feedback in relation to six bioscience course units. As a result, they proposed a guide for teachers which outlined six steps and a feedback loop. The study data was gathered via interviews and surveys with students who were in either first or final year courses. However, the results of that study are mostly applicable to the measuring of students' experience in relation to courses and semesters, and not in relation to just one learning session. Our present study, of course, is focused on just one learning session.

Nicol and Macfarlane-Dick [191] provided a model which encompassed seven principles of good feedback practice. It helped teachers to examine their assessment practices in relation to a self-regulation model and to these seven principles. Furthermore, Gibb and Simpson [192, 193] created the Assessment Experience Questionnaire (AEQ). The AEQ has been employed to evaluate assessment in science classes by measuring participants opinions of six suggested aspects of assessment: time demands and student effort, assignments and learning, quantity and timing of feedback, quality of feedback, use of feedback and the examination and the effects of this on learning.

In this research, to measure students' experiences with, and perceptions of, both the human (expert) assessment and the system assessment feedback provided in the study, the Assessment Experience Questionnaire (AEQ) [192, 193] was employed because it has been developed precisely to measure this sort of student experience of assessment and it has clear instruments for measuring such student experiences. Three items (time demand and student effort, assignments and learning, and the examination and learning) are concerned, most specifically, with student learning, assignments and examinations over a semester of study.

Thus, these three items are not applicable in the current case because this research assesses students based on their performance on only one learning occasion. Therefore, for the current research, following Gibbs [194], just three measures (opinions regarding the quantity and timing of feedback, the quality of feedback and the use of feedback) are employed to measure students' experiences with both system and the human (expert) feedback. Hence, we included only the AEQ questions which were relevant to the measuring of these items.

Aside from the TAM and AEQ based measures, other research instruments have been employed to evaluate broader aspects of our systems. We next present a full list of such that we have used.

6.3.The Research Instruments

In this research, a combination of qualitative and quantitative measures was used to assess aspects other than just acceptance in the TAM sense, and to evaluate the hypotheses (section 6.2).

6.3.1. User Questionnaire Measures

First, all student and expert participants provided background information about themselves through questionnaires. This covered basic demographic information as well as their experience of virtual worlds, educational software, etc. Other questionnaires differed between those for students and those for experts and in relation to the differing experimental conditions which will be described more fully in section 6.4 (phase 1 versus phase 2 conditions (2-1 and 2-2)).

Many of the subsets of the questionnaire items described below included one or more open response questions. These then led to the questionnaires yielding some qualitative data alongside the quantitative data from the rating scale response items. The questions were checked by an expert in designing questionnaires, Dr Desmond Thomas, who works for the University of Essex and provides courses in designing questionnaires for doctoral research. In addition, the questionnaires were double-checked by Dr Joy Helvert at the University of Essex.

A. Students

After performing all the required tasks in just one of the experimental scenarios (see section 6.4), each student participant then responded to a post-questionnaire, which included both open response and closed items, covering some or all of the following areas (dependent on experimental conditions) relevant to the testing of the hypotheses:

- **Students' acceptance of the system (SA):** To measure the users' acceptance of the *Observe Portal* assessment system, the Technology Acceptance Model (TAM) [184], as described in section 6.2, was employed. This model measures the acceptance of technology based on relationships between perceived usefulness (PU), perceived ease of use (PEOU) and the intention to use the system (IU). Questions relating to each of these models are included in the student post-questionnaire shown in Appendix A (A.5.2).
- **Students' experiences of the assessment feedback (AEQ):** To measure students' experiences of (or rather, attitudes to) the assessment results, the Assessment Experience Questionnaire (AEQ) [193] was used, as described in 6.2.2. We included the AEQ questions which measure the student's perceptions of the quantity and timing of the feedback (QTF), the quality of the feedback (QF) and the use of the feedback (UF), see Appendix A (A.5).

- **Students' preferred approach (PA):** These questions elicit the student's preferred assessment approach; these questions were asked in order to compare the students' responses to the assessment feedback provided by a traditional expert (teacher) with that from the system. These questions are shown as part of the student post-questionnaire given in Appendix A (A.5.2).
- **Perception of chat communication facility (COMM):** The students' perceptions of the chat communication facility were also measured; this was to evaluate the students' attitudes to the use of the classified chatting tool within the immersive environment. The questions related to this are included in the student post-questionnaire shown in Appendix A (A.5.2).
- **Perception of natural agent rating procedure (NA):** Students were asked about their attitude to the experience of working as natural agents required to rate their partner students. These questions are also included in the student post-questionnaire shown in Appendix A (A.5.2).

B. Experts

The experts, who played the role of teacher-like assessors in the study, also responded to a post-questionnaire with both open and closed response items covering the following areas (relevant to the hypotheses in section 6.2):

- **Experts' acceptance of the system (EA):** The Technology Acceptance Model (TAM) [184] was, again, chosen as a basis for measuring expert acceptance of the system. The questions employed are included in the expert post-questionnaire shown in Appendix A (A.6).

- **Experts' preferred approach (E-PA):** These questions ask the experts about their preferred approach; this was in order to compare their attitudes to traditional expert assessment with those they had concerning the system assessment, using the same virtual world observation and rating data. The questions relating to this are included in the expert post-questionnaire in Appendix A (A.6).
- **Experts' assessment experience (EXP):** These questions ask the experts about the experiences they encountered in the course of both the physical and the virtual assessment processes and their opinion on each of these, in order to compare experts' experiences in relation to both the traditional observation process and virtual observation. The questions employed for this are included in the expert post-questionnaire shown in Appendix A (A.6).
- **Experts' reflection on the assessment processes (REF):** This variable records the experts' reflections on the effectiveness of the two assessment methods. The questions employed for this are included in the expert post-questionnaire shown in Appendix A (A.6).

6.3.2. Human Expert and *Observe Portal* Assessment

Here, the instrument consists of the assessment results which were completed in the course of the various different experimental phases/conditions - either by the system, as described in chapter 5, or by the experts/teachers using their assessment sheets, or by both. When carrying out these assessments, both the experts and the system had access to the same information that they could use as the basis of their observations: i.e., the on-screen actions, the chat and the ratings provided by the students undertaking the tasks. The students received scores for their performance with regard to the house programming tasks from both teachers and system. These results provide some information regarding the assessment systems indirectly - when we compare the scores given by the human experts with those given by the *Observe Portal* system, for the same students undertaking the same tasks in the VW.

6.3.3. Data from Student Performance

To measure students' performance in the course of the learning activities, we also analysed the data logs that were saved in the server repository of the immersive environment experiment and the video recordings made in the physical environment experiment, as is explained in detail in section 7.3. The quantitative measures derived from this, again, could be used to evaluate the system assessment.

- **Data logs.** In the virtual world scenarios, all the students' data logs regarding their actions, dialogues and tasks completed in the learning environment were saved with timestamps in a database. These data were used to count the number of 'actions', i.e. the amount of dialogue and the number of tasks completed in a given time, so as to compare the results obtained from the two scenarios, e.g. as between working in the physical, and the totally virtual, environments. A large number of logged events meant a high level of collaboration and a low number indicated limited collaboration.
- **Chat logs.** The chat logs were quantitatively analysed to measure the frequencies of occurrence of various different categories of student chats, as distinguished by the classified chatting tool in *Observe Portal*. This measure, along with the questionnaire items about the perceptions regarding student communication via the chatting tool (COMM), facilitated a better understanding of the students' experiences and their acceptance of the use of this chat tool.

- **Rating logs.** All rating data (of other students) elicited from student users were saved in the database. These data were also analysed in order to measure the extent to which the rating tool was used in *Observe Portal*. This measure, in addition to the questionnaire items on student perceptions of working as natural agents (NA), assisted the researcher in ascertaining the degree of acceptance by the students of working as natural agents when collaborating in the learning activity. This, in turn, helped with the evaluation of the Mixed Agents approach adopted in the software design of this present system.
- **Video recording.** A video camera was set up to record the student actions and the conversations which took place in the physical classroom scenario. These videos were analysed to measure the level of the students' interactions, quantities of dialogue engaged in and the number of completed tasks so as to be able to compare the physical form of collaboration with its virtual equivalent (wherein all the relevant information was automatically logged as described above). Again, the number of actions and completed tasks were used as the basis of the settings of the input ranges accepted by the fuzzy logic system.

6.4. Experimental Design

Software evaluation, in support of proof-of-concept, may involve student-based, teacher/expert-based and system-based evaluations, and here we are reporting on the first two. In order to enable us to evaluate different aspects of the assessment system, it was necessary to arrange for students and experts to have some experience of using the portal, in its current prototype form. To achieve this, groups of students and experts were invited to participate in the research experiment. Students participated in group activities (mostly undertaken in pairs) either in the real or in the virtual world, and the experts, in the role of teachers/assessors, observed the learning session either physically or virtually – in order to evaluate the students; in the virtual world scenarios, the students were also evaluated by the system. Using the instruments described in section 6.3, quantitative and qualitative data was gathered to throw light on the stated hypotheses (section 6.2). This section presents the experimental approach used by this research, the learning activity which was designed in relation to this and the experimental phases required in order to prove the research hypothesis.

Chapter 3 presented the *MIVO* framework and the Virtual Observation models that embed learning components with computational objects. This was followed by a description of the implementation of the key elements that contributed to building a prototype that included *MIVO* and the observation models in a VW, as discussed in Chapter 4 and Chapter 5. The developed system (*Observe Portal*) was used in the experimental evaluation of the collaborative activities to prove the hypotheses stated in Chapter 1 and earlier in section 6.2.

The experimental evaluation made use of the proposed learning activity scenario, discussed in Chapter 4, particularly in relation to students' collaborating in learning tasks in *Observe Portal*. Since our research focuses on assessing collaborative students, the learning scenario is designed to allow students to work together on learning tasks. It permits collaboration between users as well as communication through a chat system. In addition, the learning activities in the virtual world offer a rating tool which allows collaborators to score each other's quality of performance on a continuing basis. The learning tasks required students to program actuators and sensors; this task was set in order to teach the functionality of embedded systems. The programs that the students constructed were then implemented in a simulation of a smart home. Two different types of observers (human experts and the system) watched the students in order to assess them. Once the session has finished, the learning activity participants received assessment reports from one of the observers (either the system or an expert) about that participant‘s contributions in terms of learning outcomes.

The experimental work for this research was conducted in two main phases, with the second phase being undertaken under two conditions, separately. The phases were as follows:

Phase 1- This consisted of experiments conducted in a physical classroom situation; the students performed the learning tasks involved with this research in a real-world classroom.

Phase 2- In contrast, consisted of experiments conducted within an immersive virtual environment where students collaborated together in a virtual world (*Observe Portal*). This second phase presented two different situations: one in which the system was the sole observer of the students, and one where both the system and experts (teachers) observed.

The use of all these various different phases/conditions was necessary in order to prove the research hypotheses. Section 6.4.5 details the research phases and the relationships between them and the research hypotheses.

6.4.1. Ethical Approval

Before starting the evaluation trials, the researcher obtained approval from the University of Essex Ethical Committee to conduct the experiments involving human participants. A copy of the research proposal; the participant consent form (Appendix A (A.3)), the research questionnaires (Appendix A); a description of the recruiting process; and the data storage protocol had been submitted for approval.

In terms of data access and security, the participants’ data were stored separately on a secure local server and were protected with passwords such that they could only be accessed by the researcher. These data were not sharable with others and were deleted from the server as soon as there was no longer any need to keep them.

After obtaining ethical approval from the university, the researcher began conducting the experiments, including the hands-on trials with users.

6.4.2. The Overall Experimental Approach

Each of the research experimental phases involved entirely different participants (so demonstrating a 'between group' design). The general experimental approach followed in the phases was:

A. <u>Before each experimental scenario was presented:</u>

Before starting the hands-on activities, we needed to collect participants background information before and then analyse these in relation to the research hypotheses. At this stage, the researcher undertook the following:

- The participants were separated into groups. Each group (mostly of two, selected for collaborative working) included one student who had some computer programming knowledge and another student with little experience in programming - to facilitate effective knowledge sharing between them. Most of the student participants were undergraduates or masters' students; they had a variety of backgrounds and expertise. The experts were PhD researchers who had experience of teaching and also of observing students taking computer science and programming courses.
- The students were provided with a link to an online survey so that they could provide relevant background information. The student pre-questionnaire asked about personal demographics, and also the students' backgrounds in computing, 3D virtual worlds,

gaming, and intelligent environments. A copy of the questionnaire is included in Appendix A (A.1).

- The experts were also provided with a pre-questionnaire which gathered similar background information from the experts, asking about their teaching expertise, and any teaching experience they had in 3D virtual worlds, intelligent environments and computing. The expert pre-session survey is shown in Appendix A (A.2).

The students' and the experts' background information are described and analysed in detail in section (6.5).

B. In the course of the experiment (the learning activity):

Participation in the learning activity was an essential element of our experiments; here, students worked together and experts or the system or both observed them in order to produce assessment results. Then, student experiences relating to the tasks and the assessment results were analysed with respect to the research hypotheses.

- **Student participants:** the experimental steps experienced by the students were as follows:
  1- Student participants first signed a consent form (see Appendix A (A.3)).
  2- Next, the researcher gave a 10 min. presentation concerning the learning environment and the learning tasks that the student pairs had to collaborate to accomplish (the tasks are described in section 6.4.3). In addition, in relation to the virtual world sessions (phase 2), the students were informed that the task was not just to program the house but also to evaluate each other by using the rating tool - and to communicate effectively, using the chat window.

3- After these preparations, the students began the process of collaborating to complete the assigned tasks.

- **Expert participants:** The experiment steps experienced by the experts were as follows:

1- Before the sessions began, the expert participants were provided with assessment sheets of the form shown in Appendix A (A.4).

2- Next, the researcher explained to the experts the procedures whereby they could observe the students and use these sheets to record their assessments - since they had not previously been made familiar with these aspects.

3- On the day of the experiment, the experts first signed a consent form, see A.3 in Appendix A.

4- In the course of the learning sessions themselves, the experts observed the students, either physically (phase 1) or through their access to the online virtual world (phase 2). In either case, they then used the manual assessment sheets, A.4, as shown in Appendix A, to record their assessments of the students regarding their performance in the course of the programming tasks. They (the experts) were also asked to use the collaborative skills sheet, see A.4, to evaluate the students' collaborative skills.

All the participants (both the experts and the students) were also observed by the researcher while performing their tasks and notes were taken to document any data of interest in relation to the research. The data of this kind which that were documented by the researcher were: the time taken to evaluate the students by teachers, the time taken to finish each task by the students, data regarding the flaws in the collaboration between students which emerged, and data regarding the mistakes made in the course of each task.

C. <u>After the experiment</u>:

Finally, it was necessary to collect the participants' (student and expert) experiences and their opinion about the assessment. In particular, it was necessary to obtain participants' responses based on the evaluation measures, given in section 6.3.

Thus, after they finished their learning activities, the assessment from the system and/or from the experts was provided to the student participants in order to enable them to review their evaluation and examine their performance in relation to the collaborative activity. Then both experts and students filled out their post-session surveys (A.5 and A.6). All the questionnaires are included in Appendix A.

6.4.3. The Learning Activity

The learning tasks were designed to facilitate communication and collaboration between students in order to evaluate them based on the *OLens* levels. The learning activity as a whole was inspired by an earlier pioneering work [52] and it has also been evaluated by Dr Rasha Alruwilli of the psychology department of the University of Essex. She validated that the activity could be used to measure social collaborative skills, and of course, it was part of the aim of the present research to evaluate the assessment with regard to these skills. Figure 6-2 describes the student learning activity; this was essentially the same across all the conditions/phases.

Observed Collaborative Learning Activity

Learning Objectives

1- Understand the concept of intelligent environments.
2- Create rules that change the functionality of sensors of the smart home.
3- Collaborate and work together as a team.

Conditions

1- Participants are in the same physical location or in different locations using VW.
2- Access to PCs.
3- Access to Internet.
4- The client software has been installed.
5- Use physical BuzzBox or virtual smart home.

1 : 1

Activities

1- Familiarise with the learning environment.
2- Understand the learning task.
3- Work with other members and agree about the created rules.
4- Do the tasks and program the rules using the programing board.
5- Execute the rules.

Figure 6-2: Learning Activity Used in the Study

Mayes and Fowler proposed a framework for courseware which followed what they saw as the essential cycle of three learning phases [139]. Keeping these stages in mind helps one to understand the steps that the learning process generally goes through and indicates how to structure learning activities using technology appropriately. The learning cycle processes which they proposed are: conceptualisation, construction and dialogue [139]. Thus, based on the description given above of the objectives and nature of the learning activity which we selected, it may be seen that we implemented the learning activity using Mayes and Fowler's framework steps as follows:

- **Presentation (conceptualisation stage):** The researcher first explained to the students the concepts involved with sensors, installed either in physical or virtual smart homes, presenting the ways in which they can use the facilities provided to program them. The explanation gave them examples of the use of rules which can be employed to program lights. Then, once the session had been completed, learners were assessed based on whether or not they had successfully been able to program various types of sensors and create rules other than the examples that were given to them in the presentation stage. The assessment also measured student's collaborative skills.
- **Practical (construction stage):** After completing the presentation, the researcher gave the students three subtasks based on the unit of learning illustrated in *Figure 6-2*, and asked them to collaborate and solve the problems involved. The learning tasks were:

  "*Imagine this is your smart house, collaborate and do the following*:

  1. *Formulate rules for the house appliances*
  2. *Create rules for hot weather situations*
  3. *Create rules for fire situations*

- **Discussion (dialogue):** Students could discuss their work face-to-face in the physical activity scenario (phase 1 – section 6.4.5.1) or via the chat facility in the virtual activity scenario (phase 2 – section 0) to reflect on their learning and their solutions.

6.4.4. Recruiting Participants

Participants of all these experimental phases were recruited from the University of Essex through an opportunistic sampling approach, and the students were paid £5 per an hour to participate in the experiments. Students were undergraduate or master's students representing a number of different areas of expertise. For these experiments, the aim was to recruit typical tertiary level students across a range of specialisms who were not already fully familiar with the programming concepts that were required to complete the tasks. It was, however, necessary to obtain a mixture of students, some of whom knew something about programming while others did not. This was because, as described above, the study design required pairs (groups of two) to be set up in which one member had more relevant knowledge than the other.

In terms of the expert participants, the aim was to recruit experts who fully understood the topic being learned via the tasks, and had experience of teaching, and especially of assessing, students. Thus, the experts were PhD students in their final years who had experience of teaching and observing students taking part in computer science and programming courses and labs.

6.4.5. Research Phases/Conditions

The above sections have described what was common to all the experiment phases that we conducted. All the research phases utilised the same procedures in terms of the experimental approach (section 6.4.2), the learning activity (section 6.4.3) and the method whereby the participants were recruited (section 6.4.4). Table 6-1 provides a summary description of the differences between the experimental phases/conditions, which are then described in detail. The table also shows the measurements and other data that were collected in each phase and for each condition.

Phase	Phase 1	Phase 2	
Condition		**Cond 2-1**	**Cond 2-2**
Model of house programming activity	Programming functions in a physical model of a house	Programming functions in a virtual house	
Location of students	Together in the same room	Online in the virtual environment	
Medium of student communication	Face to face oral	Via an online typed chat facility only	
Role of student	Collaboration, completing the learning tasks	Collaboration, completing the learning tasks, rating each other	
Researcher’s sources for objective measures: numbers of completed programs, time spent chatting, etc.	Video recording and computer logs	Screen recording and computer logs	
Teachers’/experts’ data sources for observation and assessment of students	Direct observations of the students working	None	By observation of student actions and chat in the virtual world
System’s data source for assessment of students	None	Observations of student actions, chats and ratings within the VW.	
Presentation of assessment to students	Expert’s assessment sheets	Computer graphical charts and also video recordings	Expert’s assessment sheets, computer graphical charts and also video recordings
Observation by researcher	Direct observation of the students and teachers/experts	Observation of students via the computer	Observation of students and teachers/experts via the computer
Measures and data collected from participants	**Student:** -Assessment experience questionnaire (AEQ) **Expert:** -Experts’ assessment experience (EXP) -Experts’ reflection (REF) **Other measures:** -Expert’s assessment	**Student**: -Assessment experience questionnaire (AEQ) -Technology acceptance (TAM) -Perception of chat communication (COMM) -Perception of natural agent procedure (NA) **Other measures:** -Chat logs -Rating logs -System’s assessment	**Student:** -Assessment experience questionnaire (AEQ) -Technology acceptance (TAM) -Perception of chat communication (COMM) -Perception of natural agent procedure (NA) -Students’ preferred approach (PA) **Expert:** -Technology acceptance (TAM) -Experts’ preferred approach (E-PA) -Experts’ assessment experience (EXP) -Experts’ reflection (REF) **Other measures:** -Chat logs -Rating logs -System’s assessment -Expert’s assessment

Table 6-1: The experimental phases/conditions

6.4.5.1.Phase 1: Physical Classroom Observation

The physical classroom phase was undertaken in order to use its results as a reference in relation to later student assessment in the virtual world; also, this activity helped in the development of the expert assessment model and the fuzzy system that was then applied in the virtual observation phase (phase 2) to assess student performance. Developing the fuzzy logic system based on expert evaluation was explained in Chapter 5. Furthermore, this (physical classroom) phase enabled a comparison between users' reports concerning their real-world experiences and their assessments with those obtained later in the virtual world, to reveal any significant differences in terms of these assessments and user responses between the two types of environment - as required in order to prove the research hypotheses.

This phase (phase 1) consisted of 15 learning sessions. Each session had two students and one expert/teacher participants. We recruited the participants by employing the method discussed in section 6.4.4. The total number of student participants in phase 1 was 30 and the total number of experts was 8 - experts participated several times in the sessions. The sessions employed pairs of students to enable collaboration between students and one expert to imitate classroom observation with one teacher assessor.

- **During the experiments (the learning activity):**

The experimental approach shown in section 6.4.2 and the learning task described in section 6.4.3 were employed in this phase. The physical classroom experiments took place in room (5A.544) of the computer science building, at the University of Essex, see Figure 6-5. The experiments were as follows:

1- Student participants were divided into two-student groups to collaborate and work on the learning tasks. Each pair was seated at a table and provided with a PC and a BuzzBox (Figure 6-3) to be used as a simulation of a physical smart room. The BuzzBox contained sensors and actuators that users may program to control the functions (Figure 6-4).

2- The learner participants were asked to work collaboratively to configure the physical BuzzBox through a programming board (Figure 6-5). A video camera was set up in the room to record this collaboration and discussions between students.

3- The expert participants' role was to observe the students while they were collaborating (Figure *6-6*) and assess them using the manual sheets and the collaborative skills sheet, shown in Appendix A (A.4). The experts did not give any hints to, or talk to, students during the sessions.

4- After students had finished the learning activities, the experts gave to each student their assessment sheets so that they could review their results.

Figure 6-3: Physical BuzzBox [52]

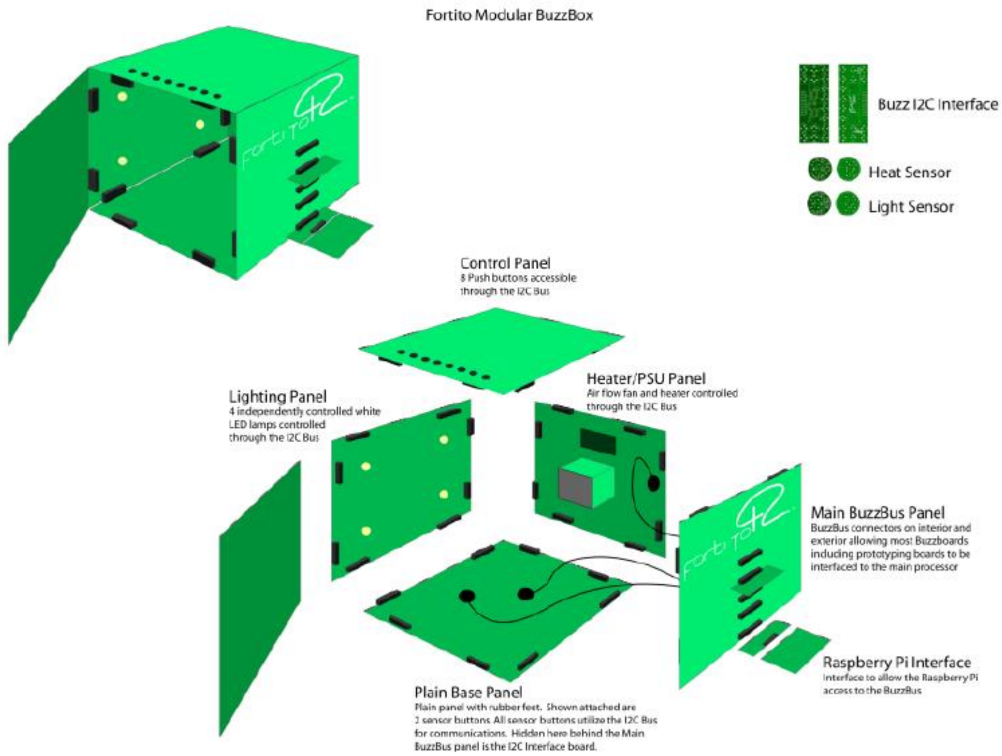

Figure 6-4: Fortito's BuzzBox diagram [52]

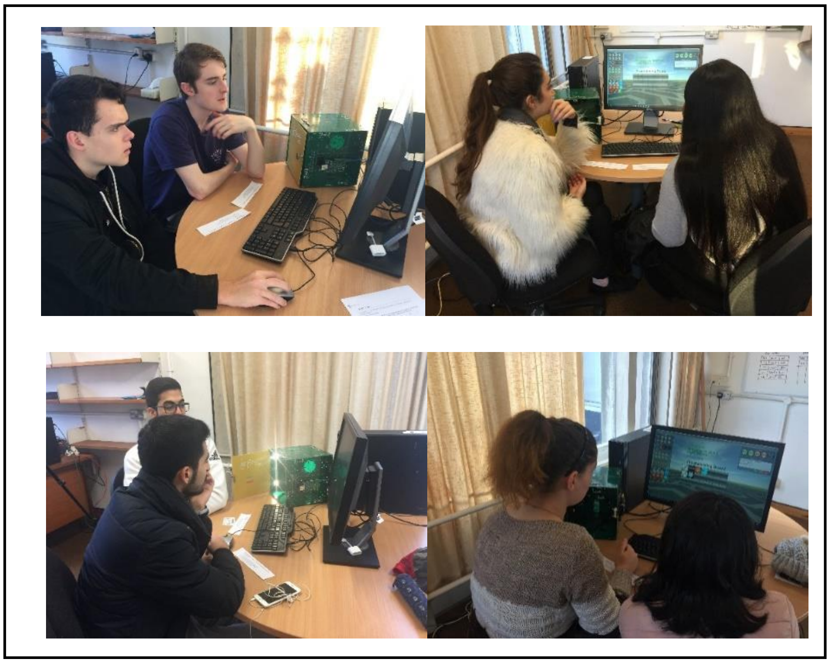

Figure 6-5: Students collaborating in the Phase 1 experiments

Figure 6-6: Experts observing the students undertaking the Phase 1 experiments

6.4.4.2. Phase 2: Virtual World Observation

This phase includes two conditions: the use of virtual but no human expert observation (condition 2-1) and the use of virtual observation along with human expert observation (condition 2-2). This phase was split into these two conditions in order to compare learner experiences in the learning environment with and without expert observation - to reveal any significant differences in student responses in relation to these two conditions.

In addition, these situations were necessary to undertake in order to prove the research hypothesis. Both phase-conditions (condition 2-1 & condition 2-2) support the following: First, the virtual world activities undertaken under the various conditions applied can be used to validate the effectiveness of the *MixAgent* model and its tools by measuring user attitudes to their embodying of the human agent roles when performing tasks in the virtual world (H1).

Second, *Observe Portal* is able to provide users with assessment feedback, and this means that student responses to this assessment can be measured (H2). Third, the virtual world activities facilitate, in general, the measurement of the acceptance of the system (H5).

Furthermore, conducting virtual observations in conjunction with the human expert observation condition (2-2) permits a comparison between the results of the expert assessment those of the system assessment (H3); this also supports the discovery of which of the assessment approaches is the one preferred by participants (H4). The following sections detail each virtual world condition.

Condition 2-1: Virtual observation without experts

With regard to this condition, 17 sessions were held, 15 with pairs of students and 2 sessions with groups of four students.

- **During the experiments (the learning activity):**

  The experimental approach described in section 6.4.2 and the learning task described in section 6.4.3 were the ones utilised in this phase. The virtual world observation sessions took place in rooms 3A.524, 3A.526 and Lab 3 of the computer science building, the University of Essex. The learning sessions conducted were as follows:

1- Each session involved a group of two or, alternatively, four students sitting in different locations so that they could only communicate with each other through the chat facility provided on screen. Each student had access to a personal computer displaying the learning interface (Figure 6-7).

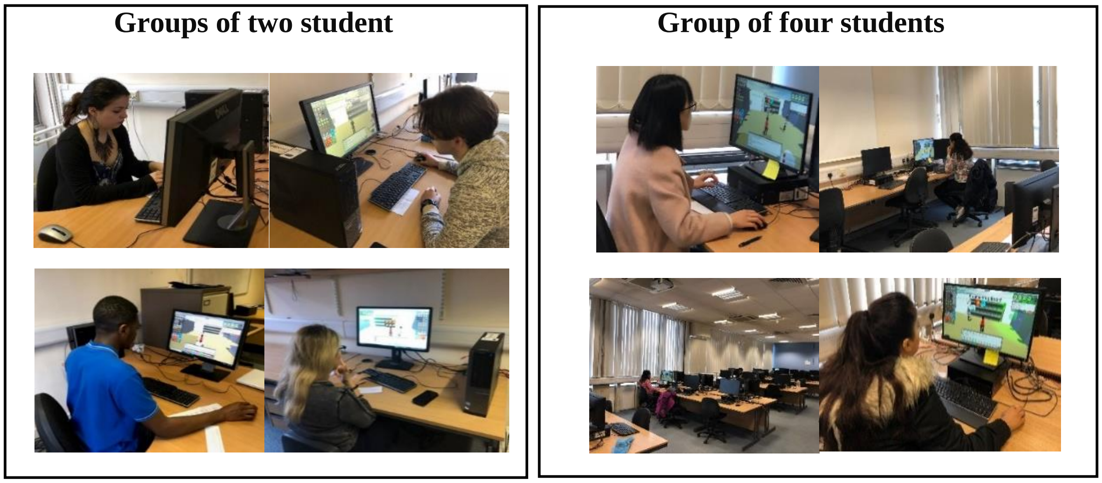

Figure 6-7: Phase 2 experiments - Virtual observation without experts

2- The activity required the group (of either 2 or 4) to work together in the virtual world to find solutions to the learning tasks and create IF-ELSE rules for the virtual house sensors and actuators; this was all just as in phase 1, but in a VW, on screen. Additionally, learners were asked to work as natural agents and rate each other during the learning activity, through the on screen rating tool (*Figure 6-8*).

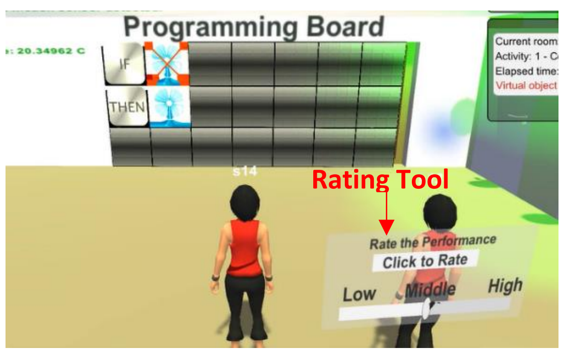

Figure 6-8: Students rating tool

3- At the end of each session, the system assessed the students virtually and, to provide feedback, showed the student their dashboards and the video recordings; these represented the system's evaluation of their performance and allowed them to view both the individual and group assessments (the student dashboards were described earlier in Chapter 4).

Condition 2-2: Virtual observation with expert observation

In relation to this condition/situation, 15 sessions were held, each session using a group of two students, to enable collaboration, and one expert - to imitate classroom observation with one teacher.

- **During the experiments (the learning activity)**

The experimental approach described in section 6.4.2 and the learning task described in section 6.4.3 were utilised for this phase. The learning activity employed in cond. 2-2 was similar to that employed in cond. 2-1 but with the key difference that the latter included an expert participant to observe students through the interface to the VW, allowing access to the same information that the system used to make its assessment. The experiments conducted were as follows:

1- Each session used a group of two students plus one expert, and each participant was sitting in a different location (Figure 6-9). The students had access to personal computers and so were able to communicate with each other only through the virtual world.

2- The students were asked to collaborate in order to solve the learning tasks and to rate each other's performance in the virtual world, as in cond. 2-1.

3- The experts were asked to log into the learning environment so that they appeared as teacher avatars in the virtual world. Their role was to observe and assess the students manually by noting the learning evidence and using the assessment sheets (see Appendix A (A.4)).

This phase/condition, therefore, supported two means by which the students were evaluated the students: the system assessed the students virtually with respect to their activities in the VW, while the experts assessed the students manually based on the same information – which was made available to them.

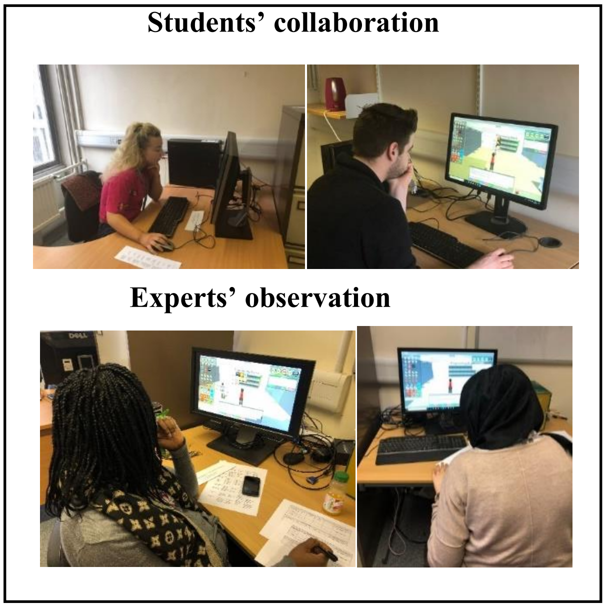

Figure 6-9: Phase 2 experiments – Virtual observation alongside expert observation

4- At the end of each session, the system presented the students with their assessment dashboards and video recordings to illustrate the students' individual performance and in addition the group's performance; the experts were also given this information. Furthermore, the human expert evaluation sheets were given to the students to allow them to compare the system's assessment with the expert's assessment.

6.5. Participants Background Information

Both the students and the experts were recruited from the University of Essex through an opportunistic sampling. We next describe the samples in more detail.

6.5.1. Students

Across all the experimental phases, 98 students participated in the learning activities, as described in *Table 6-2*.

Phase	Number of Participants	Participants per Group
Phase 1	30	2
Phase 2 – Condition 2-1	30	2
	8	4
Phase 2 – Condition 2-2	30	2
Total	98	

Table 6-2: Number of students participating in each phase/condition

The student demographic information indicated that 53% of the participating students were males and 47% were female, of varied nationalities (British, Emirati, German, Indian, Bahraini, Pakistani, Italian, Turkish, Bulgarian, Romanian, Maltese, Latvian, Greek and Cypriot, Saudi). 81% of the students were in the age range of 18-24 years, 16% were between 25-30 and 3% were aged between 31-35 years old. The reported levels of English proficiency were 57% native or bilingual proficiency, 32% full professional proficiency, 4% working professional, 6% limited working proficiency and 1% elementary proficiency. At the time that the experiments were conducted, most of the participants were undergraduate or master's students and their course subjects ranged from computer science, data science and computer engineering to criminology, economics, business, government, speech and language therapy, accounting, actuarial science, law and philosophy and finance. A summary of the students' preliminary, general, information is shown in the following Table 6-3:

	Total	
Categories	**Number of students**	**Percentage**
Total Number of students	**98**	**100%**
General Information (GI)		
Age		
18-24	79	81%
25-30	16	16%
31-35	3	3%
Gender		
Male	52	53%
Female	46	47%
Nationalities		
Emirate, British, Indian, German, Bahraini, Pakistani, Italian, Turkish, Bulgarian, Romanian, Maltese, Latvian, Greek, Cypriot, Saudi		
Self-reported Levels of English Language		
Elementary proficiency	1	1%
Limited working proficiency	6	6%
Professional working	4	4%
Full professional proficiency	31	32%
Native or bilingual proficiency	56	57%
Level of studies		
First-year undergraduate	40	41%
Second-year undergraduate	26	26%
Third-year undergraduate	14	14%
Postgraduate (Master)	18	18%
Subject of Study		
Computer science, Computer engineering, Criminology, Economics, Business, Government, Speech and language therapy, Accounting, Actuarial Science, Law and Philosophy, Finance, Data Science		

Table 6-3: General student information

Table 6-4 presents the students' reported expertise in computer use (CU) and programming (PE). It shows that almost all the participants owned personal computers and, in terms of computing experience, 26% rated themselves as beginners, 51% intermediate, and 23% claimed to be experts. The table also shows that 50% stated that they were beginner programmers, 39% were intermediate and 11% were experts in programming.

Computer Use (CU)		
Own a personal computer?		
Yes	97	99%
No	1	1%
Computing expertise?		
Beginner	26	26%
Intermediate	49	51%
Expert	23	23%
Programming Experience (PE)		
Programming expertise		
Beginner	49	50%
Intermediate	38	39%
Expert	11	11%

Table 6-4: Student reported computing and programming experience

Additionally, about half of the participants said that they had experience in using virtual worlds and computer games (Table 6-5). 51% of the students were familiar with virtual worlds and 49% were not, but a few (17%) of the latter had used virtual worlds before. Moreover, 49% of the learners played video games several times a week, while 51% did not play such games at all.

Virtual World Experience (VWE)		
Are you familiar with virtual worlds?		
Yes	50	51%
No	48	49%

How often do you use virtual worlds?		
Not at all	81	83%
once or twice per week	17	17%
4-5 times per week	0	0%
Every day	0	0%
Please select the virtual worlds that you have used or heard of		
Second Life, RealXtend, Meshmoon, Open Wonderland, IMVU, Club Penguin, Habbo, The Sims, School of Dragon		
Video Games Experience (VGE)		
Playing video games		
Not at all	50	51%
Once or twice per week	27	28%
4-5 times per week	10	10%
Every day	12	12%
If you play video games please name the ones you use		
Fortnite, RuneScape, Puzzle games on mobile, RPG's on PC, Life is Strange, League of Legends, Yu-gi-oh! Duel Links, The Elder Scrolls V Skyrim, Dragon Ball Xenoverse, GTA, Fifa, Assassin's Creed, WoW, Hearthstone, GTA, Grand Theft Auto 5, Mass Effect, Football Manager, WWE 2k among, Nba, dota 2, League of Legends, Humans of Might and Magic, Barbara Ainola, Need for speedCounter strike, generally fps		

Table 6-5: Student Virtual World and Computer Games Experience

With respect to knowledge of intelligent environments (IE), 55% of the participants were familiar with smart houses and intelligent spaces and 45% were not. Notably, 54% had previously undertaken practical activities in a computer engineering lab (Table 6-6).

Intelligent Environments Knowledge (IE)		
Are you familiar with smart houses/intelligent spaces?		
Yes	54	55%
No	44	45%
Are involved in practical activities/assignments in a computer engineering lab?		
Yes	53	54%
No	39	40%

Have you used or heard of technology to make your house "smart"?		
Yes	78	80%
No	20	20%

Table 6-6: Student Knowledge of Intelligent Environments (IE)

The student group working (GW) information in Table 6-7 shows that 77% of the students liked to work in groups while 23% did not like group activities. Also, most of the students rated themselves as having high or middle collaborative skills.

Group Working (GW)		
Do you like to work in groups?		
Yes	75	77%
No	23	23%
Rate your collaborative skill level		
I have High collaborative skills	47	48%
I have Middle collaborative skills	49	50%
I have Low collaborative skills	2	2%
I prefer to work alone on projects		
Strongly agree	14	14%
Agree	35	36%
Disagree	36	37%
Strongly disagree	2	2%

Table 6-7: Students' Self-Reported Group Working (GW)

It can be seen, from Tables 6.5 and 6.6, that the students who participated in our experiments were certainly not unusual in terms of their age, gender, and nationality mix, nor in their range of subject specialisms. It is also perhaps to be expected of educated young people today that they almost all owned computers and claimed at least moderate familiarity with them, that around half played video games, though only a minority played games involving virtual worlds, around half claimed to know what a smart house was, and around three quarters claimed to like working in groups. With respect to variables more closely connected with the activities in the study, it is clear that a slight majority could claim some knowledge of programming. However, overall it could be said that about half the students had some relevant programming knowledge while about half did not. This served the purposes of the study since what we needed was mixed ability pairs/groups of students who could be put together in order to perform the tasks collaboratively — hence, meaningful communication with respect to the tasks could then occur between these students.

6.5.2. Experts

Overall, eighteen experts participated in the experiments, some of these participated twice. Table 6-8 shows that 72% of the expert participants were females and 28% were males, and that these experts came from a variety of nationalities (British, Iraqi, Nigerian, Libyan and Saudi). The experts' age demographics at the time of the experiment were: 6% between 46-50 years old, 33% between 41-45 years old, 22% between 36-40 years, 22% between 31-35 years old and 17% of the participants' ages were from 25-30. The claimed levels of English proficiency were: 17% native or bilingual proficiency, 61% full professional and 22% working professional proficiency. At the time that the experiment was conducted, most of the participants were PhD students in their third or fourth years, and 11% of them already had their PhD degree. Their majors were 72% computer sciences and 28% computer engineering.

	Total	
Categories	**Number of experts**	**Percentage**
Total	18	100%
General Information (GI)		
Age		
20-25	0	0%

26-30	3	17%
31-35	4	22%
36-40	4	22%
41-45	6	33%
46-50	1	6%
51+	0	0%
Gender		
Male	5	28%
Female	13	72%
Nationality		
Saudi	7	39%
Iraqi	3	17%
Nigerian	4	22%
British	2	11%
Libyan	2	11%
Level of English		
Elementary proficiency	0	0%
Limited working proficiency	0	0%
Professional working	4	22%
Full professional proficiency	11	61%
Native or bilingual proficiency	3	17%
Level of education		
Master's Degree	0	0%
First year PhD	0	0%
Second year PhD	0	**0%**
Third year PhD	8	44%
Fourth year PhD	8	44%
PhD Degree	2	11%
Major		
Computer sciences	13	72%
Computer engineering	5	28%

Table 6-8: Experts' General Information

All the expert participants had experience of teaching computing subjects such as programming languages, networks or databases, or of teaching electronic engineering topics such as analogue circuit design. 39% of the expert participants had 1-5 years experience in teaching, 39% had been teaching for 6-10 years, 17% had been teaching for 11-15 years and 5% had 16-20 years teaching experience (Figure 6-10). All the participants (100%) had taught and observed students in computer labs, and all of them were experts in teaching programming courses.

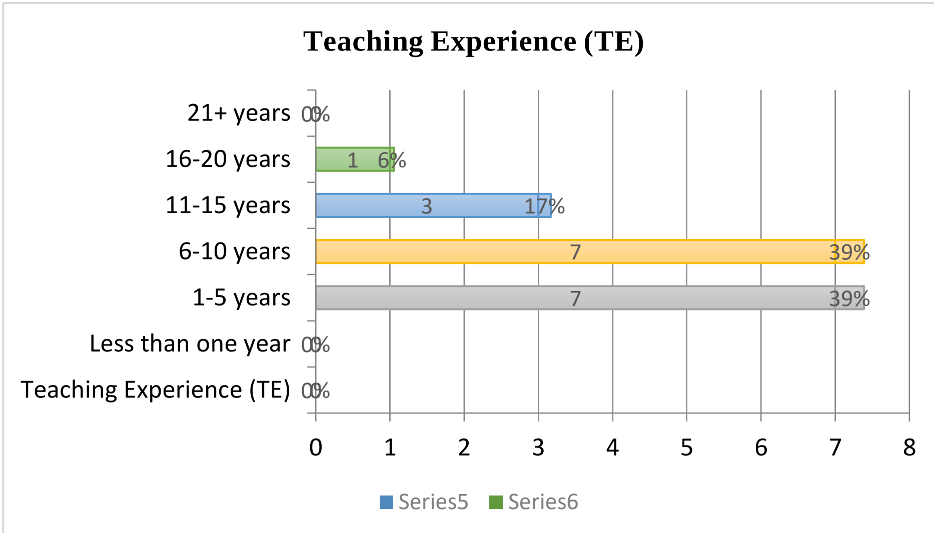

Figure 6-10: Experts Teaching Experience (TE)

In addition, the participants all had experience of assessing student learning by different methods such as tests, assignments, group projects, individual projects and observation. Most of the experts (94%) had assigned group projects to their own students, and they had also assessed these group projects by different methods: self-evaluation, peer-evaluation, final product evaluation or teacher observation evaluation.

Several expert participants (67%) were already using educational software which supports lecturers in teaching and assessing students, while 33% were not. Some of the teaching-related systems in use were *Blackboard*, *Moodle*, *Chatbots*, *QA systems*, and *MATLAB*. In addition, most of the participants liked using new technology in teaching and they considered that understanding new technologies was relatively easy for them.

Most of the participants (83%) were also familiar with virtual worlds, while 17% were not. 17% of them used virtual worlds on a daily basis, 22% used VWs 4-5 times per week, 11% used VWs twice per week, and 33% did not use them regularly. Some of the virtual worlds that they reported that they had used were: Second Life, RealXtend, or their own (constructed) virtual world. However, most of them did not use VWs for teaching and learning. Finally, most of the experts, i.e., 89% of them, were familiar with smart houses/intelligent spaces, while only 11% were not.

Overall then, the experts in our sample all possessed the qualities that we needed for participation in the experiment. The one element that they all lacked was expertise in using the assessment rating scales that the researcher chose. Hence the researcher explained to them all the use of these rating scales - before they (the experts) started observing the students.

6.6.Chapter Summary

This chapter has described in detail the numerous instruments which have been employed in this research to gather data relevant to the evaluation of the assessment system, the hypotheses which have been posited, the experimental design employed, and the participants involved. In addition, this chapter has looked at the experimental phases/conditions in which the experiments were performed (for the purpose of proving the research hypotheses). Finally, the chapter delineated the research samples in more detail by analyzing the background data gathered from the participants. The next chapter (Chapter 7) will present the analysis of the data and the determining of the experimental results.

Chapter 7

7. Results and Analyses

The previous chapter (Chapter 6) demonstrated the experimental design and the evaluation methods and measures adopted in this study. In addition, it described the experimental phases/conditions that have been applied to evaluate the *Observe Portal* system. The current chapter interprets the user evaluation results. It begins by detailing the mapping between the research hypotheses and the research instruments. Then it describes the data analysis procedures. Lastly, it demonstrates the experiment results as yielded by the participants' questionnaire, the user logs, the system scores and the assessment sheets.

7.1.Mapping Hypotheses and Instruments

The hypotheses (H1-H5) mentioned in Chapter 1 have been reworded and divided into sub-hypotheses as shown in Table 7-1; this was in order to derive some more testable hypotheses. Table 7-1 also indicates what instruments must be used to test each hypothesis.

Code	Hypotheses	Phase/ Condition	Associated Instruments	Other Instruments
H1-main	Users express positive attitudes towards their roles as human agents when performing distance-learning tasks in the virtual world.		1- NA 2- COMM	Data logs
H1.1	The students express positive attitudes to the activity of rating each other when performing collaborative learning tasks in the virtual world.	Phase 2 (Cond 2-1 & 2-2)	Student NA	Rating logs
				Student background variables
H1.2	The students express positive attitudes to the online classified chat facility which they can use when performing collaborative learning tasks in the virtual world.	Phase 2 (Cond 2-1 & 2-2)	Student COMM	Chat logs
				Student background variables
		Phase1 and Phase2(Cond 2-1)	Comparison of COMM	

H2 - main	The *Observe Portal* system provides collaborative distance learners with assessment feedback, and users report positive experiences of such assessment feedback.		1-AEQ: (QTF- QF- UF) 2- REF	
H2.1	The students express positive attitudes to the speed and amount of feedback they can obtain from the *Observe Portal*.	Phase 2 (Cond 2-1 & 2-2)	Student QTF	
		Phase1 and Phase2(Cond 2-1)	Comparison of QTF	
H2.2	The students believe that the *Observe Portal* system provides very useful information about interaction, success, and collaborative skills	Phase 2 (Cond 2-1 & 2-2)	Student QF	
			Comparison of QF	
H2.3	The students report that they have a good understanding of the *Observe Portal* assessment.	Phase 2 (Cond 2-1 & 2-2)	Student QF	
		Phase1 and Phase2(Cond 2-1)	Comparison of QF	
H2.4	The students claim to make extensive use of the feedback from the *Observe Portal*.	Phase 2 (Cond 2-1 & 2-2)	Student UF	
		Phase1 and Phase2(Cond 2-1)	Comparison of UF	
H2.5	The experts report that they have positive attitudes to the value of the *Observe Portal* assessment.	Phase 2 (Cond 2-2)	Expert REF	
		Phase1 and Phase2(Cond 2-1)	Comparison of REF	
H3- main	The *Observe Portal* provides assessments that are very similar to human-expert assessments; these Observe Portal assessments are produced using less effort overall.		EXP	System & expert assessments
H3.1	The human experts find that making their own assessments of the students' activities in the virtual world is a difficult task.	Phase 2 (Cond 2-2)	Expert EXP	Successful assessment completion by experts
		Phase1 and Phase2(Cond 2-1)	Comparison of EXP	
H3.2	The system matches the human experts' performance closely in relation to the assessment of all measures of student task performance in the VW.	Phase 2 (Cond 2-2)		Student scores provided by the system and human expert

H4-main	Students and experts prefer the *Observe Portal*'s assessment feedback over and above that yielded from the human experts.	Phase 2 (Cond 2-2)	1-Student PA	
			2-Expert PA	
			Comparison of student and expert PA	
H5-main	Students and experts express their acceptance of using the *Observe Portal* assessment system.		TAM: (PU-PEOU-IU)	
H5.1	Students and experts alike find that the *Observe Portal* system is useful.	Phase 2 (Cond 2-1 & 2-2)	Student PU	
		Phase 2 (Cond 2-2)	Expert E.PU	
H5.2	Students and experts alike find that the *Observe Portal* system is easy to use.	Phase 2 (Cond 2-1 & 2-2)	Student PEOU	
		Phase 2 (Cond 2-2)	Expert E.PEOU	
H5.3	Students and experts express an intention to use the *Observe Portal* system in the future.	Phase 2 (Cond 2-1 & 2-2)	Student IU	
		Phase 2 (Cond 2-2)	Expert E.IU	

Table 7-1: Operationalization of the Hypotheses Mapped onto Instruments

7.2. Data Analysis Procedures

This section describes three important aspects relating to how the data was analysed, prior to the determining of the results.

7.2.1. Reliability of Questionnaire Responses

Many of the key student and expert variables, as listed in Chapter 6 (section 6.3.1), were measured via subsets of questionnaire items - rather than having a single item for each variable. Leaving aside any open-response items (see 7.2.3), the items were typically responded to on a 4-point Likert scale which had different labels for measuring different constructs (e.g. 1= Strongly Disagree, 2=Disagree, 3=Agree, 4= Strongly Agree / 1=Very Poor, 2=Poor, 3=Good, 4=Very Good). This enabled us to check for the internal reliability of each subset of such items, using Cronbach's alpha coefficient.

The first stage in the analysis was as follows: since the questionnaire included both positively and negatively worded items, the negative item ratings were transformed into positive ones. This is required before calculating Cronbach's alpha, and also makes interpreting the results simpler (section 7.3).

Table 7-2 shows that the reliability coefficients for all the student variables surpassed the value of 0.70, which is regarded as entirely satisfactory [195, 196]. Thus, each subset of items is confirmed as measuring one construct, as expected.

Variable	Cronbach's Alpha	No. of Items
Students perception of communication (COMM)	.703	3
Student perception of natural agent role (NA)	.882	6
Students assessment experience (AEQ)	.795	18
Student acceptance of the system (SA)	.748	10
Student preferred approach (PA)	.826	8

Table 7-2: Student Post-Survey Construct Reliability

Table 7-3 presents Cronbach's alpha in the same way, but for the expert questionnaire items. The results for the constructs REF, EA and EXP indicated that the expert questionnaire responses had high reliability [195, 196]. However, the expert PA construct was measured mostly through open-ended questions, and had no multiple items with the same rating response scale. Therefore, alpha was inapplicable.

Variable	Cronbach's Alpha	No. of Items
Expert reflection about system assessment (REF)	.792	7
Experts' Acceptance of the System (EA)	.728	13
Expert observation experiences (EXP)	.820	12
Expert preferred approach (E.PA)	N/A	N/A

Table 7-3: Experts post-survey construct reliability

7.2.2. Normal Distribution Check

It is also important to test if the research data is normally distributed before starting the statistical analysis. In order to test the rating scale response data for normal distribution, the one-sample Kolmogorov-Smirnov test [197] was used as follows:

- H0: The data is normally distributed (Null hypothesis)
- Ha: The data is not normally distributed (Alternative hypothesis)
- P= 0.05 (Significance level)
- D=the upper bound of absolute differences between the observed data and the normal distribution
- Critical value of D= 0.04301

The results from the Kolmogorov-Smirnov test are included in Appendix B (B.1 and B.2). They show that D is greater than the critical value (0.04301) for all items (Appendix B.1 and B.2), thus the data is not normally distributed. Therefore, the null hypothesis (H0) is rejected. For this reason, nonparametric techniques are used for inferential quantitative data analysis in this research. In addition, the non-parametric techniques are employed because of the use of Likert scales in the questionnaires. Usually, data on Likert-type scales are ordinal data which does not measure the distance between points, just that each point is higher than the last; the distance between points cannot be assumed to be equal [198]. In such cases, the median should be used instead of the mean to summarise the data. In addition, Kruskal-Wallis, Friedman, Wilcoxon, Mann-Whitney and Binomial tests were applied to the research data because they do not require normal distributions [199].

7.2.3. Handling of Open Response Questionnaire Items

The open-response items yielded qualitative data, i.e., words, rather than quantitative data. Hence, they were submitted to qualitative analysis. Specifically, the responses were sorted into themes, so as to ascertain the various kinds of points that the participants made - once differences purely of wording were discounted.

7.3. Experiment Results

This section presents and explains the quantitative and qualitative findings in turn for each hypothesis as listed in Table 7-1.

7.3.1. H1: Users express positive attitudes towards their roles as human agents when performing distance-learning tasks in the virtual world

As described in Table 7-1, this is to be judged by the student NA questionnaire response data and the rating log data (H1.1). In addition, it was addressed by examining the student COMM questionnaire responses and the chat logs in phase 2 (H1.2).

Perception of natural agent rating (NA)

The rating of the perception of the natural agent role (NA) measures students' positive attitudes to their required additional task of rating other student(s) when performing the programming tasks in *Observe Portal*, in all conditions of phase 2. Table 7-4 lists the questions used in the questionnaires.

NA1	Using the rating tools to evaluate the other student(s) while working or at the end were EASY
NA2	Using the rating tools to evaluate the other student(s) while working or at the end were FUN
NA3	Using the rating tools to evaluate the other student(s) while working or at the end were USEFUL
NA4	Using the rating tools to evaluate the other student(s) while working or at the end were INTERESTING
NA5 (r)	Using the rating tools to evaluate the other student(s) while working or at the end were DIFFICULT
NA6 (r)	Using the rating tools to evaluate the other student(s) while working or at the end were ANNOYING
NA7 (r)	Using the rating tools to evaluate the other student(s) while working or at the end were BORING
NA8	Explain the reasons why it was comfortable (or not) to rate the other student(s) through the rating tools
NA9	Do you have any additional comments of the overall experience when rating others in the virtual world?

Table 7-4: Perception of natural agent rating (NA)

The pairs of items (NA1 – NA5), (NA2 – NA6) and (NA4 – NA7) were designed as matching positively and negatively worded questions to reduce the probability of participants biasing a response by responding to the scale values automatically regardless of the content of the item [200]. Appendix B (B.3) shows the tables of all the NA items that have been transformed so that higher numbers indicate more positive attitudes to the rating of other students.

In fact, 89% of participants, in phase 2, found working as natural agents and rating the other students through the rating tool was easy (NA1/NA5 composite, median = 3, “Agree"). In addition, 78% of the students found it fun (NA2/NA6 composite, median = 4, “Strongly Agree"), and 86% found it interesting (NA4/NA7 composite, median = 3, “Agree"). Also, 82% of the participants found it useful (NA3, median = 3, “Agree").

In addition, a one-sample binomial test was run to determine if there was a statistically significant positive response. This is achieved by testing if the responses differ significantly from the midpoint response (2.5 on the 1-4 response scale). The quantitative data showed that distance-learning students working in the virtual world (phase 2) reported significantly positive attitudes ($p < .001$) to the rating tool which they used to rate others, way above the neutral midpoint of the 1-4 response scale (Table *7-5*).

		N	Min	Max	Median	Mean	SD	Binomial test p
NA1	Using the rating tools to evaluate the other student(s) while working or at the end were EASY	68	1	4	3	3.22	.666	<.001
NA2	Using the rating tools to evaluate the other student(s) while working or at the end were FUN	68	1	4	3	3.12	.890	<.001
NA3	Using the rating tools to evaluate the other student(s) while working or at the end were USEFUL	68	1	4	3	3.07	.779	<.001
NA4	Using the rating tools to evaluate the other student(s) while working or at the end were INTERESTING	68	1	4	3	3.19	.758	<.001
NA5 (r)	Using the rating tools to evaluate the other student(s) while working or at the end were DIFFICULT	68	1	4	3	3.35	.768	<.001
NA6 (r)	Using the rating tools to evaluate the other student(s) while working or at the end were ANNOYING	68	1	4	3	3.12	.890	<.001
NA7 (r)	Using the rating tools to evaluate the other student(s) while working or at the end were BORING	68	1	4	3	3.15	.718	<.001
NA items	Overall mean of NA items (phase 2)	68	1	4	3	3.174	.616	<.001

Table 7-5: Student NA result with one-sample binomial test

In general, these results demonstrate positive attitudes and acceptance by users of working as natural agents and using the rating tool to rate other students in the virtual world, so the results support H1.1.

This was further shown by the qualitative data from the open-response item (NA8) which asked 'why it was comfortable (or not) to rate the other student(s) through the rating tools'. The main points made were as follows:

Students reported that rating each other was comfortable because the rating was anonymous and not face-to-face, so they could be honest when they rated other students (who could not find what scores were given to them by whom) e.g.: *“The rating tool was comfortable to use because it was anonymous to the other student I was with”* s18.

Numerous participants also described the rating tool as very easy to use for a variety of reasons:

1- The rating tool was directly related to the learning topic: "*directly related to the skills we were supposed to show while completing the task”* s31.

2- The tool was easy to use because of its clarity/transparency of design, for example one student said: "*easy to use and clear instructions together with reminders were given”* s46, *“There were just three 'options'* (high – middle – low), *so it was easier”* s49, *"You only got like one button to press, so it makes it pretty easy”* s56.

3- It was noted that sufficient tie had been allowed for participants to register their ratings, e.g.: "*there was enough time between each task, I didn't have to rate the other student frequently”* s5 -- although in fact they were encouraged by reminders to rate each other in each task.

4- Students also found that rating someone else had benefits to themselves, e.g.: "*because you can give constructive feedback and receive it,”* s10, and as a consequence “*because in this way we can evaluate each other and work on the weaknesses*,” s7.

A few participants did complain that rating was not comfortable, and one objected to having to "*rate the other student using the 3 different ratings. There is no room for* *feedback"* s24. This student clearly wanted some space for giving fuller, qualitative written feedback, as a teacher might do.

The final optional student question in this set (NA9) elicited additional comments. These were generally favourable about the overall experience: "*It is an interesting way of giving feedback"* s10; "*It's a process useful for people to interact and be true about each other's skills*" s34.

One suggested an improvement which, like the comment, s24, above, suggested that the format did not allow for enough richness of feedback to be given: *"The rating scale should be increased. Maybe to a 5-star system"* s5.

From this, we can see that although some participants found that the three-point rating scale (low-middle-high) made the evaluation much easier for them and more comfortable, other students suggested making the response options more elaborate. Increasing the scale could allow for more accurate ratings especially for people who are criticised, however, it would inevitably make the rating task more onerous and could overload the students, given that they are doing collaborative tasks, classifying their chat sentences, and rating others all at the same time. In our future work, we could, however, include an open feedback option at the end of each session to enable more extensive student expressions of opinion about their peers' performance.

In addition to the above results, the Kruskal-Wallis test (Table 7-6) was employed to determine if some key features of the students' backgrounds influenced their positive rating of using the rating tool. The results demonstrated that the users' computing expertise level (CU2) and knowledge of virtual worlds (VW1) did not have a significant influence on the overall positive attitude of students towards using the rating tool (NA1). These results can perhaps be explained by the straightforward design of the natural agent tool, with a scroll bar and a rating button, which meant that no prior knowledge was really needed.

Dependent variable	Independent variable	Chi-square	DF	P
Student NA1	Computing expertise (CU2)	2.517	2	0.284
Student NA1	Virtual worlds experience (VW1)	0.591	1	0.442

Table 7-6: Kruskal-Wallis results: positive attitude to the rating tool

Rating Logs

In order to gain a different perspective on students' attitudes to rating during the learning activities, the logs saved in the database by the system were analysed. These comprise a record of all the rating scores that students had given to each other while working in the sessions. The frequency of students rating other students was determined by counting the number of times a student rated their group member(s) in each task separately: recall that each session involved three tasks to be completed. Results from the rating analysis are seen in Figure 7-1.

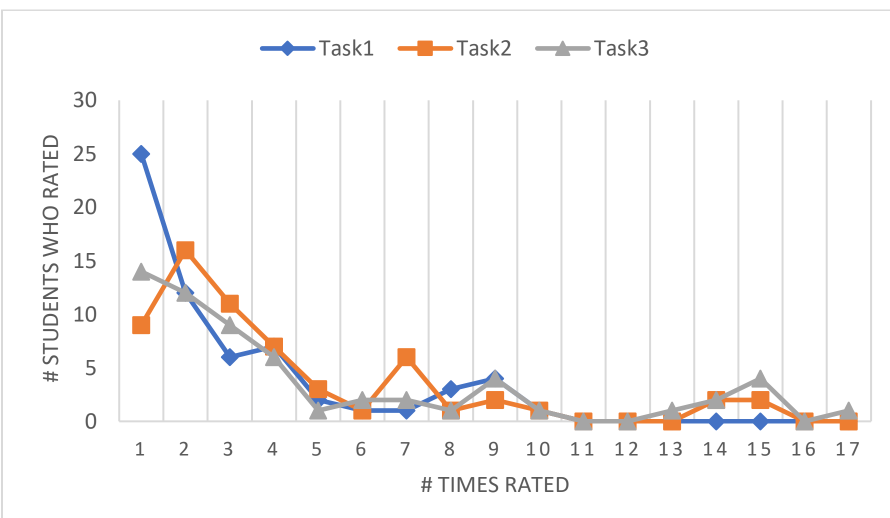

Figure 7-1: Frequency of rating by students

During Task 1, 37% of the students rated other students once, 18% rated twice, 9% rated three times, 10% gave four ratings and the others rated a greater number of times. The maximum frequency with which a student rated another student during task 1 was 10 times.

During Task 2, 13% of the participants rated others once, 24% rated others twice, 16% rated others three times, 10% rated others four times and the others rated a greater number of times. The maximum number of ratings was 15 times.

During Task 3, 21% of the students rated the other student one time, 18% rated twice, 13% rated three times, 9% four times and the others rated a greater number of times. The maximum number of ratings was 17 times.

Generally, even though the rating frequencies varied between the groups/pairs, students gave a rating to the other group members at least once during each subtask of a learning task. Furthermore, it is notable that the frequencies of rating in fact increased across the three subtasks, suggesting that students liked doing it more as they got used to it rather than quickly getting bored with it and doing it less: for instance, the percent of students rating more than 4 times rose across tasks from 26% to 37% to 39%, and the maximum number of ratings by one person rose from 10 to 15 to 17.

This shows that students found rating easy and/or valuable and they did the rating mostly multiple times during each task while collaborating with the other students. In addition, the findings from the analysis of the participants' attitudes to the NA rating tool and the student comments have shown a positive attitude of the students to working as natural agents and evaluating the other student(s) through the rating tool.

All these findings support the hypothesis *"The students express positive attitudes to the activity of rating each other when performing collaborative learning tasks in the virtual world"* (H1.1). In turn, this constitutes one positive evaluation point for our *Observe Portal* system.

Student perception of chat communication (COMM)

The COMM questionnaire items measured students' attitudes when using the classified chat window in the virtual environment (phase 2). Table 7-7 lists the questions used in the questionnaires:

COMM1**(r)**	I found difficulties when communicating with the other student(s) via the multi-buttons chat window
COMM2	It was comfortable to communicate with the other student(s) through the virtual interface (i.e. using the chat window with classifying buttons)
COMM3	Explain the reasons why it was comfortable (or not) to communicate with the other student(s) through the chat window
COMM4	How would you rate your experience of collaborating with students in other location(s) using the classified chat
COMM5	Please provide any extra comment you have on your experience working with the other student(s) in the virtual world.

Table 7-7: Perception of chat communication (COMM)

Items COMM1 and COMM2 were designed as positively and negatively wordings of the same question - to reduce the probability of participants biasing a response by responding to the scale values automatically regardless of the content of the item [200]. In order to calculate the aggregate results from the questions, the responses to negatively worded items were transposed into their corresponding positive equivalents before the composite values were calculated.

86% of participants reported that communication through the classified chat window was easy (COMM1/COMM2 composite, median = 3, “Agree"). In answer to COMM4, which elicited a rating of the experience of collaborating via chat, 94% of the students reported that their experience was good or very good (COMM4, median=4, “Very Good”). All the result tables relating to the COMM items are included in Appendix B4.

In addition, the one-sample binomial test was run to determine if the positive ratings were significantly above the midpoint rating on the scale (2.5) (*Table 7-8*). The quantitative data showed that distance-learning students working in the virtual world (phase 2) reported significantly positive attitudes ($p < .001$) to the chat process which they engaged in (way above the neutral midpoint of the 1-4 response scale).

		N	Min	Max	Median	Mean	SD	Binomial test p
COMM1 (r)	I found difficulties when communicating with the other student(s) via the chat window	68	2	4	3	3.31	.675	<.001
COMM2	Communicating with the other student(s) through the virtual interface (i.e., using the chat window with classifying buttons) was comfortable.	68	1	4	3	3.18	.809	<.001
COMM4	How would you rate your experience of collaborating with students in other location(s) using the classified chat?	68	1	4	4	3.44	.655	<.001
COMM	Mean of all COMM items (phase 2)	68	2	4	3.33	3.31	.479	<.001

Table 7-8: Student COMM descriptive statistics and one-sample binomial test (phase 2)

Moreover, when we turn to the qualitative responses of the students in phase 2 to the open-response item (COMM3), which asked participants to 'explain the reasons why it was comfortable (or not) to communicate with the other student(s)', we again found mainly positive views along with some negative points about some aspects of the chat.

Some student comments showed that using the chat was comfortable due to it resembling real-life conversation in real-time or at least social media communication: "*It was done with precision that it seemed like it was almost real. .... I was interacting with the other student live. Almost similar to the chat system in social media, except it was more professional and clearly defined*” s2. Others drew attention to the well-designed nature of the chat facility: “*The chat window is very well structured*” s32. “*It was comfortable to communicate with the other student due to the well programmed chat*” s54, although one was slightly less positive about this aspect: “*A bit complicated but overall easy*” s62.

As with open responses to NA, a number pointed out how the chat helped the process of doing the programming task: “*It was very comfortable because it gave me the confidence to interact with the virtual environment*” s2. Another said explicitly that “*the chat room made it easy to talk to the other student therefore I felt that more suggestions were contributed towards the task*” s19. Indeed, s20 implied further that this might not have occurred in face to face communication *“She was calm and waited for my responses”*.

However, there were two areas of negative comment. Instead of chat, some participants simply preferred to have voice communication *"sitting next to the person, or talking on the phone*” s43. The reasons given were "*I'm slow at typing*" s18, or just that it was "*less convenient*" s10.

The other negative comments referred to the requirement to select a category describing the nature of each chat communication. “*Considering the way communications happen in this day and age everyone is used to chat boxes. It was a little bothersome to always have to press the correct button for the correct intention*” s9. In addition, s52 had not understood the purpose of the classification requirement, which was in fact to assist the system in exploiting the chat data to help make its assessment of students, hence s52 quite reasonably did not see any need for it: “*from my perspective the use of categories for the conversation is unneeded as the person I am communicating with most likely understands the statement and doesn't need clarification on what category it comes*”.

Next, in order to ascertain whether liking the chat was sensitive to individual characteristics of participants, we additionally used some of the background variables which we had measured to make comparisons on relevant variables. Clearly, the system is better if attitudes to it are not seriously affected by prior computer-related experience of the student, so we hoped in this case for nonsignificant results. Thus, we ran the Kruskal-Wallis test (Table 7-9). The results showed that users’ computing expertise level (CU2) and knowledge of virtual worlds (VW1) did not have a significant influence on the reported comfortableness of using the classified communication chat (COMM2). These results could be related to the simple design of the classified chat which makes it easy to understand and use.

Dependent variable	Independent variable	Chi-square	DF	P
COMM2	Computing expertise (CU2)	4.842	2	0.089
COMM2	Virtual worlds experience (VW1)	2.242	1	0.134

Table 7-9: Kruskal-Wallis results: Ease of use of chat

Communication in the physical and virtual world

The COMM questionnaire items had not only been used in phase 2, where students worked entirely in the VW, as reported above, but also in phase 1, the physical condition where all inter-student communication was live face to face, rather than via a chat facility. Hence, we could assess whether students actually preferred either mode of communication to the other.

Descriptively, face to face communication was preferred, slightly, to chat communication on two of the three relevant questionnaire items (*Table 7-10*). However, the Mann-Whitney test showed that the difference was not significant (p>.05).

		Condition	N	Mean	M-W Z	P
COMM1 (r)	I found difficulties when communicating with the other student(s)…	1	30	3.50	-.509	.611
		2-1	30	3.60		
COMM2	It was comfortable to communicate with the other student(s) ….	1	30	3.53	-1.143	.253
		2-1	30	3.37		
COMM4	How would you rate your experience of collaborating with students in other location(s) ….	1	30	3.57	-.581	.561
		2-1	30	3.40		

Table 7-10: Mann-Whitney test of COMM difference between phase 1 and 2-1

Thus, these results further support H1.2 confirming that communication and collaboration between students in the virtual world by using the chat box with multiple classified buttons did not significantly reduce their reported comfortableness from that experienced in normal face to face conversations. This supports that collaborating via the chat window was easy to use.

Chat logs

Additionally, to gain a different, more objective, perspective on students' communication and collaboration through the classified chat box in the virtual world, the chat logs recorded in both phase 2 conditions were analysed. Due to time and space limitations, we restricted the analysis to following up on the issues raised in the open responses concerning the classification of the communications which participants had to undertake, and so we've reported only on the analysis of the classification of the communications. If their accuracy is high, we can argue that this requirement probably did not impose a great burden on the students nor create a source of difficulty which might have distracted them from the main goal of programming the house. Hence this would support H1.2.

As described in Chapter 5 Section 5.4, the chat messages could be sent into 6 type classifications, each with a button to be selected each time a student used the chat (Table *7-11*). However, for its internal purposes in making the assessments, the system reduced the classification to three communication skill categories (as classified in the collaboration skills) and we use those for the analysis here.

Student Response Category	Button Code	System Button Type	System Classification
Greeting	1	B1	Acknowledgement
Reply	2	B2	Responding
Agree	3	B2	Responding
Inquiry	4	B3	Initiating
Suggestion	5	B3	Initiating
Solution	6	B3	Initiating

Table 7-11: Classification of the chat buttons

The frequencies of using the three types of chat buttons by users are shown in *Table 7-12*.

No. of Users	Total use B1	B1 per user	Total use B2	B2 per user	Total use B3	B3 per user	Total messages
68.00	78.00	1.15	1259.0	18.51	507.0	7.46	1844.00

Table 7-12: Frequencies of use of the chat button types by students

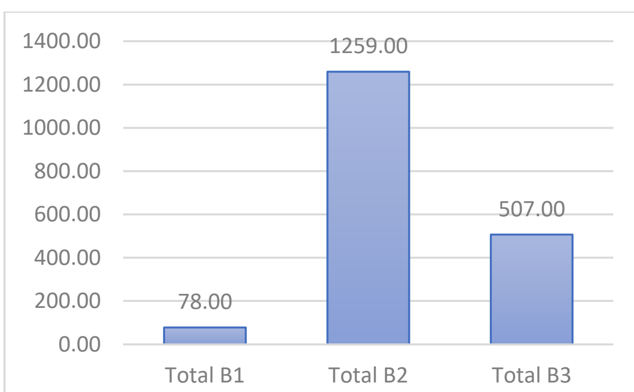

Figure 7-2: Buttons used by participants

Of the 1844 messages that were classified, the vast majority involved the two 'responding' buttons 2 and 3 (68.3%), followed by the three 'initiating' buttons 4, 5, 6 (27.5%), with fewer instances of button 1 'acknowledgement' (4.2%) (Figure 7-2). It would be expected that greetings would only occur once or twice per session, and the predominance of responses over initiations shows that each problem that was identified typically generated multiple messages to resolve it.

To determine whether the students chose the buttons suitably when they were communicating, the chat logs also were categorised by the researcher into correct use or wrong use of the buttons, in terms of the six-way classification. The overall rate of the correct choice of the buttons was 89.8%. The mistakes were largely due to the REPLY button being used where AGREE was more appropriate, which in terms of the three-way classification would be counted correctly, and could have been made due to the limited time allowed for the tasks, and because students perceived the goal as being to complete the house programming tasks well rather than communicate well via the chat and use the correct buttons. The fact that the majority of the communications were classified correctly, however, could be taken as evidence that they did not find this task too difficult, even though some did not like having to do it, and, strictly, it was irrelevant to the target activity of programming functions in a house. Hence, we may speculate that it did not interfere much with students' performance of the main tasks, the assessment of which was the object of the research to evaluate.

In summary, the student results related to the chat in phase 2 conditions combine to show that participants were predominantly very positive about the chat and found it not too hard to use, despite the requirement to classify every communication, which they mostly did correctly. Furthermore, users' positive attitudes were not dependent on the background or experience of individual students. Thus, H1.2 was supported: "*The students express positive attitudes to the online classified chat facility which they can use when performing collaborative learning tasks in the virtual world"* and again this adds to the positive evaluation of our *Observe Portal* system and *MIVO* framework.

7.3.2. H2: The *Observe Portal* system provides collaborative distance learners with assessment feedback, and users report positive experiences of such assessment feedback

We next move to H2 which concerns student attitudes to the feedback provided by the *Observe Portal* system, which is the core of its function. This was evaluated primarily using the three student-focused assessment experience (AEQ) constructs presented in the student questionnaire, in phase 2, and the REF construct measuring the expert attitude to it, which we report on in subsequent sections. We begin with AEQ which covers quantity and timing of feedback (QTF**)**, quality of feedback (QF) and use of feedback (UF) [194].

Student judgment of quantity and timing of feedback (QTF)

Table 7-13 lists the items that measure the quantity and timing of the system feedback and summarises the student response.

QTF1	On this activity I get plenty of feedback from the system on how I did.
QTF2	The system assessment comes back very quickly.
QTF3	There is hardly any feedback on my performance when I finish.
QTF4	I would learn more if I received more feedback from the system.
QTF5	Whatever assessment I get comes too late to be useful.

Table 7-13: Quantity and timing of feedback (QTF)

Some items supporting the QTF variable were reverse worded, so as with previous constructs, the ratings were changed so that high values always indicate a high level of satisfaction. In addition, we calculated separate composite scores for feedback quantity and feedback timing (Appendix B.3). The results showed that 85% of the students gave positive responses about the quantity of the system feedback on their performance (QTF1/QTF3/QTF4 composite value, median=3, “Agree”). Also, 89% of the participants found that the timing of receiving the assessment back from the system was good (QTF2/QTF5 composite value, median=3, “Agree”).

In addition, one-sample binomial tests were run (Table *7-14*) to determine if there were statistically significant differences between responses on QTF items and the midpoint scale value (2.5). Most of the quantitative data showed that distance-learning students working in the virtual world (phase 2) reported significantly positive attitudes to the speed and quantity of feedback - above the neutral midpoint of the 1-4 response scale. These results support the hypothesis “*The students express positive attitudes to the speed and amount of feedback they can obtain from the Observe Portal*” (H2.1).

The one exception was the individual item QTF4(r) asking whether more feedback would help to learn more. The students denied this only at a level not significantly above the 2.5 midpoint of the scale (mean 2.68). We do not know of course whether some thought it was more quantity of feedback that would be better, e.g. before the end of all the tasks, or whether they were thinking of more different kinds of feedback than those that the system provided. Still, it implies a possible area for improvement of our assessment system.

		N	Min	Max	Median	Mean	SD	Binomial test p
QTF1	On this activity I get plenty of feedback from the system on how I did	68	1	4	3	3.13	.710	<.001
QTF2	The system assessment comes back very quickly.	68	1	4	3	3.34	.637	<.001
QTF3 (r)	There is hardly any feedback on my performance when I finish.	68	2	4	3	3.40	.577	<.001
QTF4 (r)	I would learn more if I received more feedback from the system.	68	1	4	3	2.68	.837	.396
QTF5 (r)	Whatever assessment I get comes too late to be useful.	68	1	4	3	3.18	.732	<.001
Feedback Quantity	Mean of QTF1/ QTF3/ QTF4 (phase 2)	68	2	4	3	3.07	.557	<.001
Feedback Timing	Mean of QTF2/ QTF5 (phase 2)	68	2	4	3	3.26	.550	<.001

Table 7-14: Student QTF descriptive statistics and one-sample binomial test

Student judgment of the quality of feedback (QF)

The following list includes questions to measure the quality of the system feedback (*Table 7-15*).

QF1	The system assessment mostly tells me how well I am doing in relation to others
QF2	The assessment from the system is useful to understand the individual and the group INTERACTION LEVEL in the virtual world
QF3	The assessment from the system is useful to understand the individual and the group SUCCESS LEVEL in the virtual world
QF4	The assessment from the system is useful to understand the individual and the group COLLABORATIVE SKILLS LEVEL
QF5	The assessment from the system helps me to understand things better
QF6	The assessment from the system shows me how to do better next time
QF7	Once I have read the assessment from the system, I understand what I did
QF8 (r)	I don’t understand some of the system assessment .
QF9	I understand what the assessment is saying.

Table 7-15: Quality of feedback

Six of the QF items (QF1, QF5, QF6, QF7, QF8 and QF9) asked about the quality of the assessment in general and if the student understood the assessment from the system, in phase 2, and consequently the composite values were calculated for these items. Since QF8 was a negatively worded item, it was transformed as with the previous constructs. 77.53% of the students claimed that they understood the system assessment and that it helped them to recognise their performance level as regards the learning activity (*QF1/* QF5/ *QF6/* QF7/ QF8/ QF9 composite median=3; “Agree”) (see Appendix B5).

Furthermore, the one-sample binomial test in Table 7-16 illustrates that all these QF items were agreed with at a level significantly higher than the midpoint of the response scale (2.5). These results from the quantitative data, therefore, support the hypothesis “*The students report that they have a good understanding of the Observe Portal system assessment”* (H2.3).

		N	Min	Max	Median	Mean	SD	Binomial test p
QF1	The system assessment mostly tells me how well I am doing in relation to others	68	2	4	3	2.99	.586	<.001
QF5	The assessment from the system helps me to understand things better	68	1	4	3	3.21	.682	<.001
QF6	The assessment from the system shows me how to do better next time	68	1	4	3	3.19	.778	<.001
QF7	Once I have read the assessment from the system, I understand what I did	68	1	4	3	3.13	.790	<.001
QF9	I understand what the assessment is saying	68	1	4	3	3.18	.752	<.001
QF8 (r)	I don’t understand some of the system assessment	68	1	4	3	3.19	.674	<.001
QF	Mean of the above QF items (phase 2)	68	1.83	3.83	3	3.15	.388	<.001

Table 7-16: Student QF (general understanding) descriptive statistics and one-sample binomial test

Additionally, more specifically, most of the participants (93%) agreed that the assessment from the system was useful to understand their individual and the group interaction level

(QF2), success level (QF3) and collaborative skills level (QF4). Again these are all significantly positive (Table 7-17), so these results support the hypothesis that "*The students believe that the Observe Portal system provides very useful information about interaction, success and collaborative skills*" (H2.2).

		N	Min	Max	Median	Mean	SD	Binomial test p
QF2	The assessment from the system is useful to understand the individual and the group INTERACTION LEVEL in the virtual world	68	1	4	3	3.34	.704	<.001
QF3	The assessment from the system is useful to understand the individual and the group SUCCESS LEVEL in the virtual world	68	1	4	3	3.34	.660	<.001
QF4	The assessment from the system is useful to understand the individual and the group COLLABORATIVE SKILLS LEVEL	68	1	4	3	3.35	.707	<.001

Table 7-17: Student QF (three specific measures) descriptive statistics and one-sample binomial test

In addition, in order to compare students' reported usefulness of system assessment of the three kinds of information - interaction level, success level and collaborative skills level - and determine if there are significant differences between them, we employed the Friedman test. The Friedman test showed there is no statistically significant difference in perceived usefulness of the system feedback depending on the type of assessment, χ^2 (2) = 0.197, p = 0.909.

Student report of the utilization of the feedback (UF)

The following list Table 7-18 contains questions to measure the utilization of the feedback (UF).

UF1	I use the assessment feedback to go back over what I have done in the learning activity
UF2	I think the system assessment will help me with any subsequent activity.
UF3	The system assessment prompts me to go back over video material recorded earlier to understand the scores
UF4 (R)	I tend to only read the marks from the system without watching the recorded video.

Table 7-18: Utilization of the feedback (UF)

79.41% of the participants claimed that they used the assessment feedback to go back over what they had done in the learning activity (UF1 median=3; “Agree”), 85.29% of the students think the system assessment will help them with any subsequent activity (UF2 median=3; “Agree”) and 69.12% agreed that the system prompted them to use the recorded video material to understand their assessment scores (UF3 median=3; “Agree”), however, 48.53% of the participants stated that they just read the marks from the system without watching the videos (UF4 median=3; “Agree”).

		N	Min	Max	Median	Mean	SD	Binomial Test p
UF1	I use the assessment feedback to go back over what I have done in the learning activity	68	1	4	3	2.94	.689	<.001
UF2	I think the system assessment will help me with any subsequent activity	68	1	4	3	3.15	.697	<.001
UF3	The system assessment prompts me to go back over video material recorded earlier to understand the scores	68	1	4	3	2.93	.886	<.005
UF4 (r)	I tend to only read the marks from the system without watching the recorded video	68	1	4	3	2.57	.698	.904
UF	Mean of all UF items (phase 2)	68	1.75	4	3	2.90	.428	<.001

Table 7-19: Student UF descriptive statistics and one-sample binomial test

The one-sample binomial test (Table 7-19) showed that utilisation of feedback was claimed at a level significantly above the midpoint of the scale on all items except UF4(r), where students denied watching the recorded video only at a middling level (around 2.5 on the scale). This result could be because of the limited time students had to finish the learning activity or because that the experiment was not part of a real course, so students had no expectation that the expertise they had gained would be needed again in future.

Overall, however, the UF results were significantly positive. Hence they moderately support the hypothesis "*The students claim to make extensive use of the feedback from the Observe Portal*" (H2.4).

Student attitudes to assessments in the physical and virtual world

We compared the findings for student QTF, QF and UF with respect to the system feedback in phase 2 (cond. 2-1) with those for the same questionnaire items in relation to the human expert feedback in phase 1, to see if there was a significant preference for one or the other. Clearly, it would add to the positive evaluation of the system feedback if it was in fact preferred to human feedback, across a range of feedback constructs.

We employed the Mann-Whitney test to evaluate the significance of the differences between responses to the groups of QTF, QF and UF items in phase 1 (the physical classroom with expert assessment based on direct observation) and those in phase 2 (cond. 2-1, the virtual classroom with only system assessment) (Table 7-20). Overall, QTF and QF showed no significant differences between the system assessment and expert assessment. On the other hand, the UF mean demonstrated a highly significant difference between the groups ($p<.001$). The students in the virtual sessions reported greater utilization of feedback than those in the physical classroom.

	Condition	N	Mean	M-W Z	P
QTF	1	30	2.49	-.734	.463
	2-1	30	2.43		
QF	1	30	3.06	-.015	.988

	2-1	30	3.07		
UF	1	30	2.34	-4.499	<.001
	2-1	30	2.86		

Table 7-20: Mann-Whitney test of AEQ differences between phase 1 and 2-1

Expert reflection about system assessment (REF)

Hypothesis 2.5 concerns the experts' views of the system assessment. The expert reflection items (REF) were reported in Table 7-21.

REF1	Learners understand the purpose of the system assessment.
REF2	Learners receive useful feedback from the system after the activity.
REF3	The marking of the learners' performance from the system is helpful.
REF4	The provided system assessment can help learners to improve their performance
REF5	I think that the ObservePortal assessment is reliable
REF6	Why do you think the assessment is reliable or not?
REF7	I think that the ObservePortal assessment is valid
REF8	Why do you think the assessment is valid or not?
REF9	Peer evaluation is a good approach to assess the quality of student performance in VWs.
REF10	Using the rating tool to assess the other group member(s) is an appropriate method in the collaborative activity.
REF11	Based on the previous question, explain the reason for your choice?

Table 7-21: Experts' reflection about the assessment

The experts' REF responses to the first four items were concerned with how far they thought the system assessment was helpful to students. It showed that 80% of the participants agreed that learners understood the purpose of the system assessment (REF1, median=3, "Agree"). Additionally, 100% of the experts argued that the system provided learners with useful post-activity feedback which could be used to improve their performance (REF2/REF3/REF4 composite value, median=4, "Strongly Agree"). It should be noted, however, that, on the binomial test, the item, REF1, (Table 7-22) failed to reach significance. Thus, it could be said that the experts were not all convinced that the students understood the assessments.

		N	Min	Max	Median	Mean	SD	Binomial test p
REF1	Learners understand the purpose of the system assessment	10	2	4	3	3.20	.789	.109
REF2	Learners receive useful feedback from the system after the activity	10	3	4	4	3.60	.516	.002
REF3	The marking of the learners' performance from the system is helpful.	10	3	4	4	3.60	.516	.002
REF4	The provided system assessment can help learners to improve their performance	10	3	4	4	3.60	.516	.002
REF5	I think that the *Observe Portal* assessment is reliable	10	3	4	4	3.70	.483	.002
REF7	I think that the *Observe Portal* assessment is valid	10	3	4	4	3.70	.483	.002
REF9	Peer evaluation is a good approach to assess the quality of student performance in VWs	10	2	4	4	3.50	.707	.021
REF10	Using the rating tool to assess the other group member(s) is an appropriate method in the collaborative activity	10	2	4	4	3.40	.843	.109
REF	Mean of all REF items	10	3	4	4	3.53	.286	.002

Table 7-22: Expert REF descriptive statistics and binomial test

Next, the results for REF5 and REF7 suggested that 100% of the participants thought that the *Observe Portal* assessment was reliable (REF5) and valid (REF7), (REF5, REF7, median=4, "Strongly Agree"). The ratings provided by experts for these two items were almost identical. The open response comments to items REF6 and REF8, eliciting further explanation of responses to REF5 and REF7 respectively, support this, in that they are few, and several of them are identical. For instance, *"Because it is instant and gives 2 Assessments individual and group"* was offered for both, and the results being instant concerns convenience rather than either reliability or validity. In fact, the only items clearly supporting reliability was *"The system can collect more learning evidence than people"*, since it is known that measurement or assessment based on more data points is typically more reliable than that based on fewer. Perhaps the meaning was intended to be that the system can process more data, with equal attention to everything, unlike a human.

Other comments were, in reality, more about validity. Sometimes the criterion used was unspecified: *"It gives a good picture of the users' activities on the portal"* does not say what exactly the expert regarded as *good*. Possibly this expert meant what others said explicitly: "*covers most of the assessment requirements*" and *"gives details about users' response"*. In other words, the criterion for being good was that a rich range of information about student performance was provided to the researcher and/or the student being assessed. This then perhaps refers to the fact that the system provided a video record, as well as scores on the three sets of scales that the human experts also scored students on.

It is very noticeable that in their open responses the experts mostly supported their ideas of validity through a comparison of the system with human assessors. Several approved of the system's assessment because they felt it was very like the human expert's: *"Because it simulates human evaluation"*; *"Comparing to my assessment, I conclude that the system is valid and useful for these activities"*. Another made clearer that the precise basis for comparison was that the results were close. For him/her the system was valid because the scores it came up with were close to those that a human expert would make: *"The results of the assessment by the system is almost matching my physical assessment".* Indeed, probably all experts who made human comparisons were referring mainly to the output feedback since they had not had explained to them by the researcher the details of the algorithms used by the system to arrive at these scores, so they could not judge how human-like the decision making process was. Interestingly only one suggested that the system might be valid by being different from the human, indeed better *"I believe the system might be more effective than me"*. However, the expert did not go on to say in what exact respect he/she thought the system was better.

Finally, three REF items focused on the involvement of student rating in the assessment. 80% of the participants found that peer evaluation with the rating tool was a good approach to assess the quality of student performance in VWs (REF9/REF10 composite value, median=4, "Strongly Agree"). However, we must note that on the binomial test of REF10 item failed to reach significance. Hence it seems that while the experts definitely approved of peer feedback in general, they were not so clearly in favour of the rating tool used in the study, when their ratings were not significantly above the scale midpoint (2.5) (Table 7-22).

They further offered some open response qualitative data in explanation of the latter finding (in REF11). There was one response unambiguously in favour of the rating scale as part of the assessment: *"The rating tool was efficient to assess the other group member in collaborative activity."* Another indicated that such rating was appropriate "*Because it simulates human evaluation",* although in fact the peer evaluation was human evaluation: it did not simulate it. Perhaps he/she meant to say that it simulated evaluation made by peers of each other in natural conversation. Another approved of it as contributing to the learning process, which of course does not require it to be also part of the assessment: *"Being able to reflect on the individual performance as well as that of the others is a key in enhancing learning experience."* Peer-assessment enables students to improve their judgement skills and it promotes lifelong learning by not just relying on teacher evaluation [201].

However, a number expressed mixed feelings, which account for the binomial test result (Table *7-22*) for REF11: e.g. *“sometimes it's good and sometimes not”.* Only one offered an explanation, however: *“I actually partially agree on this. As student sometimes could judge based on emotions regardless of true performance.”* This expert felt that student ratings might not be sufficiently objective, but rather based on feelings. For example, if the students in a group are friends, they might give high scores to their peers even if these are not contributing to the learning activity. Overall, most of the expert REF results support the hypothesis that *“The experts report that they have positive attitudes to the value of the Observe Portal assessment”* (H2.5). Hence the finding is consistent with the student results for H2 in providing a positive attitude to the system assessment.

In conclusion on this issue, similar to what we did for the student AEQ measures above, we compared the 10 human experts' summary ratings over all the REF issues for the system assessment above (phase 2 (cond. 2-2)) with 8 experts' overall ratings of parallel aspects of their assessment made through physical observation (in phase 1). The difference was highly significant (Mann-Whitney Z = -3.003, p=.003). This, therefore, demonstrates that, collectively on the aspects measured by the REF items, the experts reported considerably more positive experiences of, and attitudes to, the *Observe Portal* system assessment feedback than to feedback from teachers physically present as students worked. This, therefore, further reinforces the positive evaluation reported earlier in this section, based on an analysis of the REF responses in cond. 2-2 alone.

7.3.3. H3: The *Observe Portal* provides assessments that are very similar to human-expert assessments; these *Observe Portal* assessments are produced using less effort overall

In order to support this hypothesis, we need to ascertain, in terms of condition 2-2, whether in fact the experts did find that a lot of effort/difficulty was involved in assessing pairs/groups of students working collaboratively on house programming tasks, and whether indeed they and the system arrived at much the same scores. For the first, we present primarily the findings from the EXP subset of items on the teacher questionnaire; for the latter, we compare the scores given by the human experts and the system.

Expert observation experience (EXP)

This set of items (Table *7-23*) measures a set of variables concerning experts' experience of, and reflection about, the assessment process they followed in phase 2 (cond. 2-2), in order to find out if doing a manual (hardcopy) assessment based on observation of the virtual world was difficult or easy for them.

EXP1 (r)	It is EASY to observe students’ performance in the virtual environment using the manual sheets
EXP2	It is DIFFICULT to observe students’ performance in the virtual environment using the manual sheets
EXP3	It is INCONVENIENT to observe students’ performance in the virtual environment using the manual sheets
EXP4	It is TIME-CONSUMING to observe students’ performance in the virtual environment using the manual sheets
EXP5	Explain the reasons why it was difficult (or not) observing and assessing the students' performance using the manual sheets
EXP6 (r)	If there were more than two students, I'll observe and assess them easily.
EXP7	Based on the previous question (EXP6**)**, explain the reason for your choice.
EXP8	At some point, I got lost while I'm observing the students.
EXP9	Collecting learning evidence from collaborative activities in VW was difficult

EXP10 (r)	Assessing the group and the individual's INTERACTIONS was easy
EXP11	Assessing the group and individuals TASK SUCCESS was difficult
EXP12 (r)	Assessing the COLLABORATIVE SOCIAL SKILLS of the group and individuals was easy
EXP13	Do you have any additional comments about the overall experience using the manual sheets for the assessment?

Table 7-23: Expert observation experience (EXP)

As with previous constructs, EXP items were worded the opposite way to others, so we have reversed the ratings so that high ratings consistently indicate greater difficulty in observing and assessing students in condition 2-2. Results from the analysed data showed that 53% of participants agreed that it was difficult to observe students' performance in *Observe Portal* using the manual sheets (EXP1-r/EXP2/EXP6-r/EXP8/EXP9/EXP10-r/EXP11/EXP12-r composite value; median = 3, “Agree") (Appendix B.6). In addition, 60% of the experts found observing students' performance in the learning environment by using the manual sheets was inconvenient (EXP3, median = 3, “Agree ") and 60% found it time-consuming (EXP4, median = 3, “Agree").

Additionally, the one-sample binomial test yielded only one statistically significant difference from the midpoint among all the EXP items Table 7-24. In general, the quantitative data showed a number of responses near or even just below the midpoint of the scale, however, EXP6 is significantly high. In other words, there is some evidence that the experts found the manual assessment process was not particularly easy, but fell near to the neutral midpoint of the 1-4 response scale in terms of ease; this result is only moderately consistent with H3.1.

		N	Min	Max	Median	Mean	SD	Binomial Test p
EXP1 (r)	It is EASY to observe students' performance in the virtual environment using the manual sheets	10	1	3	1	٢	.966	.344
EXP2	It is DIFFICULT to observe students' performance in the virtual environment using the manual sheets	10	1	4	2	2.5	1.080	.344
EXP3	It is INCONVENIENT to observe students' performance in the virtual environment using the manual sheets	10	2	4	2	2.8	1.033	.754
EXP4	It is TIME-CONSUMING to observe students' performance in the virtual environment using the manual sheets	10	2	4	3	2.9	.994	1.00
EXP6 (r)	If there were more than two students, I'll observe and assess them easily.	10	1	4	3	٣	.816	.021
EXP8	At some point, I got lost while I'm observing the students.	10	1	4	3	2.9	1.197	1.00
EXP9	Collecting learning evidence from collaborative activities in VW was difficult	10	2	4	3	3	1.054	1.00
EXP10 (r)	Assessing the group and the individuals INTERACTIONS was easy	10	1	3	2	2.3	.994	.754
EXP11	Assessing the group and individuals TASK SUCCESS was difficult	10	1	4	2	2.5	1.080	.344
EXP12 (r)	Assessing the COLLABORATIVE SOCIAL SKILLS of the group and individuals was easy	10	1	3	3	2.4	.949	.754
EXP	Mean of all EXP items (condition 2-2)	10	1.83	3.25	2.20	2.50	.٤٩٢	.344

Table 7-24: Expert EXP descriptive statistics and binomial test

In order to understand the reasons given for finding the assessment difficult or easy, we looked first at the experts' comments on EXP5 – "'Explain the reasons why it was difficult (or not) observing and assessing the students' performance using the manual sheets". A variety of considerations were mentioned, some focusing more on the ease or not of the fact that observation was via the VW on the computer rather than of students in a physical class, others on the matter of the recording of the assessment being on paper (manual sheets) rather than on an electronic form, both of which were referred to implicitly in various ways in the other EXP items.

One expert pointed out the benefits of observation in the VW rather than in a physical classroom: *"It was useful and fair I believe. Because I could see the performance of individual student unlike in the class room".* However, as another pointed out, the ease of that depended on the expert having some skills: *"It is easy if the person can observe and deal with the system."* Another expert added that assessing students via the portal into the VW in which the tasks were performed would in fact be more difficult if the expert had to record his/her assessment ratings electronically, rather than on paper - as was actually the method used in condition 2-2. Such an 'automated sheet' would mean navigating away from the display of the students working to enter scores in an assessment sheet window, so something might be missed: *"better to have a manual sheet than having an automated sheet as navigating away from the virtual environment could be disturbing."* Finally, one expert summed up her view quite clearly by saying: "*It is not difficult but this auto method is better and faster".* We take this to mean that she thinks that the *Observe Portal* assessment was superior even though the process was not overly difficult for human experts/teachers.

The above points were supplemented by open responses to EXP7 which referred to EXP6, the closed item where the binomial test showed opinions to be significantly negative (above 2.5) about ease: 90% of the experts agreed and saw that observing and assessing more than two students will be a difficult task (EXP6 (r), median=3, "Agree"). EXP7 asked the experts to explain the reason for their choice.

In fact the experts' comments mostly just responded that higher numbers were a problem simply due to the need to look at several students at once: "*I have to observe each individual* *this will not be easy in manual process*"; "*It might be harder to track more than two students and cross match those skills manually*." There were however some suggestions about how to make the observation of larger numbers of students easier for the human experts. One suggested "*more students, need more people to observe*", i.e. several experts would watch, each focusing on different people in a group or class. That is however usually impractical in the real world. Another suggested: "*In the current situation I think it is hard to observe more than 2 students. There should be a way of letting the teacher know who, what and when to assess*". This suggested that the observer should not watch everybody all the time and focus on several different measures at once, instead observing many students should have a specific system to follow when watching. Finally, one possibly referred to the idea that accessing a recording afterwards could help: *"It could be easy if the system documents the students' performances."* However, this expert might have been simply referring again to the system assessment being better than that of a human expert.

In conclusion, once again several commented unambiguously that the system would clearly be superior with larger numbers of students to observe and assess: *"This would be difficult if I have more than two students. Therefore the system assessment would be more accurate."*

The last piece of evidence that we present from the EXP questionnaire data is a comparison between experts' responses to the EXP items in phase1 (physical classroom) and parallel items from phase 2 (cond.2-2). The Mann-Whitney test showed that there was no significant difference between the overall ratings based on EXP items in phase 1 (mean=2.4) and phase 2 (mean=2.5): Z=-.179, p=.858. The quantitative results, therefore, indicated that experts' ratings in phase 1 and phase 2 were close. Neither of these measures indicated that observing the students was easy. This therefore again provides only moderate support for H3.1.

Analysis of human experts' manual assessment sheets

In order to gain more understanding about the experts' experience of ease or difficulty, the expert manual assessment sheets from phase 2 (cond. 2-2) were also analysed in depth.

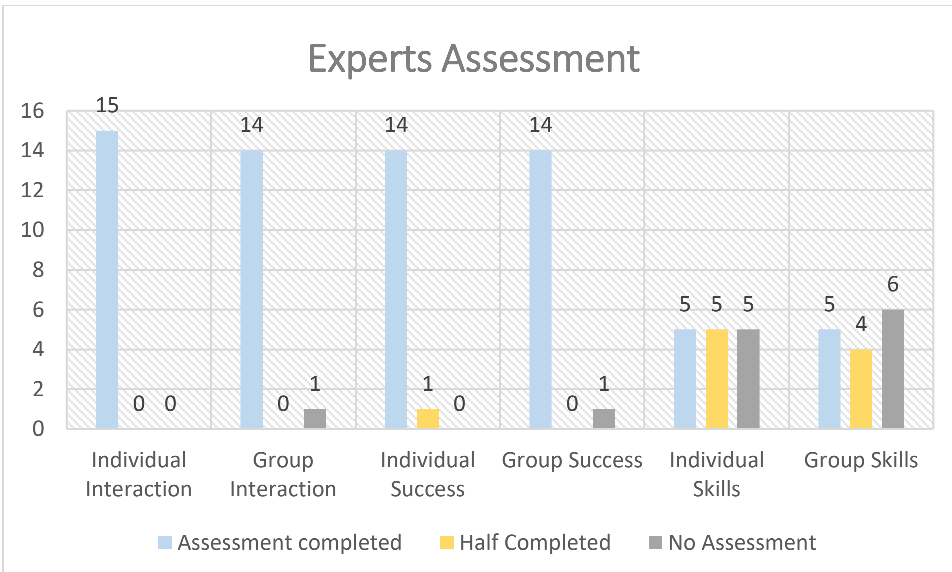

Figure 7-3: Experts' manual assessment sheets

Table 7-4 displays, for each main category of assessment, the frequencies of the experts completing their manual assessment, not completing it, or providing no assessment. The human experts mostly managed to complete the interaction and success assessments of the groups and individuals, whereas the collaborative problem-solving skills assessments were only 33% completed for the individuals and groups and so 66% were half or not completed. Accordingly, it might seem that the experts spent more time and effort in evaluating students' interactions and success, but most of them did not assess the social collaborative skills. However, a simple explanation for this is that the conscious work involved in assessing collaborative problem-solving skills was far greater. For both interaction and success, there was just one three-point scale used for both individual and group assessment, and no definitions were provided for any of the points on these scale points. This made the assessment quick, but intuitive rather than explicit in its basis. By contrast, social collaborative problem-solving skills involved 3 separate three point subscales, each with further three point scales within it, and all the scales had verbal definitions of each point which made it very explicit, but also a lot more time consuming to use, as the expert might well need to reread the definitions when applying all these scales. Hence it is perhaps not surprising that often the experts rarely managed to complete scoring on the social collaborative problem-solving skills scale.

In addition, based on the researcher observation, most of the experts had taken 7 to 15 mins to return the assessment sheets after the end of the learning activities in *Observe Portal*. On the other hand, the system assessments were instantly displayed to users when they completed the leaning tasks. Furthermore, 100% of the interaction, success and social collaborative skill assessment results for groups and individuals were presented to students by the system.

In summary, although some experts reported assessing students by using the manual sheets was an easy task, their ease ratings were often not significantly above the midpoint of the EXP rating scale on the binomial test, and many of the experts could not actually complete the assessment sheets, especially assessing the collaborative skill level. The findings from the expert open responses when observing the VW and using the manual sheets show in the end that most of the experts found observing and assessing students in *Observe Portal* was easy only because they were observing just two students: they judged that observing more than two students would be a very difficult task, tracking many users and assessing them at the same time. All these results, therefore, give some support to the hypothesis that "*The human experts find that making their own assessments of the students' activities in the virtual world is a difficult task"* (H3.1).

System assessment scores compared with human expert assessment scores

In order to test H3 fully, we also compared the scores awarded by the system with those awarded by the human experts. Since we built the *Observe Portal* system and created its fuzzy system based on human experts' evaluations, we expected that the system should match the human expert evaluation. Thus, student assessment scores in phase2 (cond. 2-2) were used to make the comparison because that is the condition where students were assessed by both the system and human experts on the same performance. The scores included were scores on the three score scales: task success, learning interaction and social collaborative problem solving. The mean scores may be seen in along with the results of the Wilcoxon test to compare the pairs of scores for each student participant.

Measure		Assessor	Mean	Wilcoxon Z	P
Task Success	Success 1 (Task1)	Human	2.40	-2.556	.011
		System	2.00		
	Success 2 (Task2)	Human	2.07	-1.342	.180
		System	1.97		
	Success 3 (Task3)	Human	2.38	-.471	.637
		System	2.30		
Learning Interaction	Interaction 1 (Task1)	Human	2.73	-1.291	.197
		System	2.51		
	Interaction 2 (Task2)	Human	2.47	-.440	.660
		System	2.53		
	Interaction 3 (Task3)	Human	2.80	-1.147	.251
		System	2.63		
Social Collaborative Problem Solving	Participation	Human	2.57	-.٢٦٥	.7٩١
		System	2.60		

	Perspective taking	Human	2.73	-١.300	.١٩٤
		System	2.60		
	Social regulation	Human	2.60	.000	1.000
		System	2.50		
	Skills	Human	2.85	-1.633	.102
		System	2.43		

Table 7-25: Wilcoxon test of score differences between human experts and *Observe Portal* system

The quantitative findings show no significant differences between system assessments and expert assessments on any of the interaction and collaborative skills measures, which means that system and expert closely matched on these measures for individuals. However, the data did show significant differences in the “Success 1” assessment (relating to task 1). When we examined the task 1 data, we found that the system rated students on average around 4 points lower than did the experts, which, on a three-point scale, is quite a sizeable amount. For this task, the human experts awarded 3 much more often than the system did, and more often than they awarded 2, while the system was more inclined to award the mid score of 2. For example, in success1 assessment, the system gave 11 students score 2 (middle) while the experts gave all these score 3 (high). This pattern did not continue in success 2 and success 3, however, where the two types of assessor scored students identically. This indicates that the system was more critical than the experts in evaluating the students' task success, at least on the first assessment occasion and so H3.2 is partially supported.

7.3.4. H4: Students and experts prefer the *Observe Portal*’s assessment feedback over and above that yielded from human experts

This hypothesis was tested by considering the PA questionnaire responses of both students and experts, in relation to condition 2-2 where both human experts and the system provided the students with assessments.

Student preferred approach (PA)

In order to compare human expert assessment with system assessment, the PA items asked the students about which method they preferred to assess their performance. Since the participants in phase2 (cond2-2) were the only students who experienced both assessment approaches, they were the only ones asked the PA questions in *Table 7-26*.

PA1	I think the use of the ObservePortal system has a significant advantage over traditional methods (teacher observation) for assessing students' performance in the virtual world
PA2	Which method provides more information about the group and the students' performance?
PA3	Which assessment method provides USEFUL information about the group and the student's performance?
PA4	Which method assesses better the individual and the group INTERACTION level?
PA5	Which method assesses better the individual and the group SUCCESS level?
PA6	Which method assesses better the individual and the group COLLABORATIVE SKILLS level?
PA7	Which approach would you prefer to use for assessing the group and individual performance?
PA8	Based on the previous question (PA7), explain the reason for your choice?
PA9	I believe that the assessment provided by the .... is more accurate.
PA10	Based on the previous question (PA9), explain the reason for your choice?

Table 7-26: Students' preferred approach

From the analysis of the PA items, we concluded first that 87% of the students thought that the *Observe Portal* system had an advantage over traditional methods (teacher observation) for assessing students' performance in the virtual world (PA1 median=3, "Agree"). The remainder of the PA items asked participants either simply to choose between the two assessment methods or give open response answers. 60% found the *Observe Portal* provided more information than the expert about the group and individual student performance (PA2). 67% expressed the opinion that the system assessment provided more USEFUL assessment information (PA3). Additionally, for assessing the interaction level, 57% found the system better than teacher assessment (PA4), 73% found the system was better in assessing the success level (PA5) and 80% of the participants found the system was much better in assessing the collaborative skills level (PA6). Furthermore, 53% of the students preferred the system to assess the group and individual performance (PA7), and 67% believed that the assessment from the system was more accurate than teacher assessment (PA9), see *Figure 7-4*.

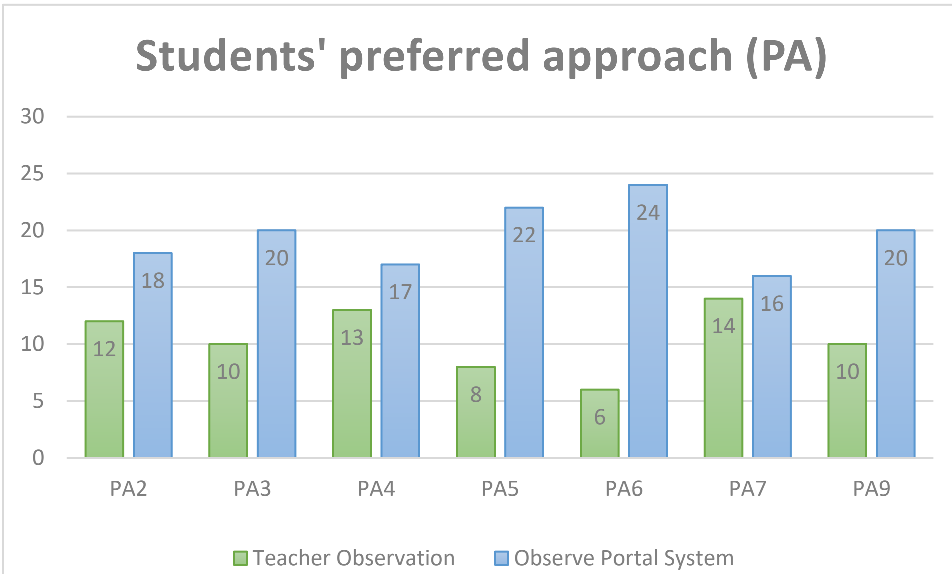

Figure 7-4: Evaluation of students preferred approach (PA)

The findings of the one-sample binomial test (Table 7-27) showed that the students believed that the *Observe Portal* system had very significant advantages over traditional methods (teacher observation) for assessing their performance in virtual worlds (PA1, $p<.001$), way above the neutral midpoint of the 1-4 response scale. On the two-choice PA items, however, the results were less simple. Students found that the *Observe Portal* was significantly better in assessing only the success level (PA5, $p= 0.016$) and the collaborative skills level (PA6, $p= 0.001$) compared with the traditional method (teacher observation). On the other hand, although the system assessment was descriptively always preferred, none of the other comparisons were significant. The assessment information (PA2, *p*= 0.362), the usefulness of the assessment information (PA3, *p*= 0.099), assessment of the interaction level (PA4, *p*= 0.585), the chosen approach to assess student performance (PA7, *p*= 1.00) and the accuracy of the assessment (PA9, *p=0.099*) did not yield significant differences between preference for the assessment from the human expert and preference for that from the *Observe Portal*.

		N	Min	Max	Median	Mean	SD	Binomial Test p
PA1	I think the use of the *Observe Portal* system has a significant advantage over traditional methods (teacher observation) for assessing students' performance in the virtual world	30	2	4	3	3.07	.583	<.001
		N	Min	Max	% Choosing System			Binomial Test p
PA2	Which method provides more information about the group and the students' performance?	30	1	2	%60			.362
PA3	Which assessment method provides USEFUL information about the group and the student's performance?	30	1	2	%67			.099
PA4	Which method assesses better the individual and group INTERACTION level?	30	1	2	%57			.585
PA5	Which method assesses better the individual and group SUCCESS level?	30	1	2	%73			.016
PA6	Which method assesses better the individual and group COLLABORATIVE SKILLS level?	30	1	2	%80			.001
PA7	Which approach would you prefer to use for assessing group and individual performance?	30	1	2	%52			1.000
PA9	I believe that the assessment provided by the............. is more accurate.	30	1	2	%67			.099

Table 7-27: Student PA descriptive statistics and binomial test

The reasons for this variation in response are not immediately apparent. Possibly they arise from differing interpretations by students of the detailed wording of the items. Or perhaps while the first item was given less thought, and so seemed to them to have a clear answer, the experience of answering the following items (PA2 onwards) raised students' awareness that the issue was not so clear cut as they initially thought.

We, therefore, examined the open questions associated with PA8 and PA10 which asked for the reasons behind the responses to the closed question. These comments provided some evidence of attitudes which existed that fell short of a total preference for the system assessment but did not entirely explain why the detailed responses to the closed items differed across items in the way that they did.

Many presented reasons for favouring the system assessments. Some were rather vague, such as: "*it's more advanced"* (s55). The most popular specific one was that it was claimed to be objective and unbiased and so more accurate. One stated this in an interesting way, s56: " *There's less bias using the Portal, it measures things without actually seeing what you're doing."* This seemed to reflect that idea that although the same data was available both to the system and the human assessor, in fact the former did not truly see the student actions in the VW in any human sense of the word. Five others expressed a similar general view of the system's merit of objectivity, but this was stated in most detail by s53: *"Since it is a computer program, there is no way it could be biased or judgmental when assessing any type of group or individual performance. It also brings up criticisms and points of information that are strictly related to the task at hand, which makes it much more objectively efficient."*

Another claim for the system was that it was more complete: *"because everything you do u can see the results"; "More data is being processed allowing better overall results to be produced"*. Possibly the fact that the system, but not the human, presented participants with a video of their session added to the idea that the human was not seeing/using all the data. In fact, however, the same data was available to both the system and the human. The effect of completeness was judged to be that this contributed to accuracy: *"The system is more accurate in the sense that it prevents ones' personal biases and also help to analyse students' performances throughout the whole processes rather than traditional method that only allows one to access on a particular time and place."* Once again, however, in the present study, the human expert was observing the students continuously all through their performance of the tasks. Finally, understanding was also seen as helped by the system: "*the system.... give more opportunities to understand what you do"* s60; *"Because the data is easier to understand and harder to misjudge than human observation" s45.*

By contrast, there were also some views that human assessment was better: *"I believe that even if technology develops and reaches extremely high standards, the experience and the knowledge of humans cannot be easily replaced. That's why I trust the traditional methods more."* s35. Another said: *"Teachers maybe see some weak points that the system would not"* s55. Here the comments seem to suggest that there are valuable human abilities that cannot be mimicked by a system. In this sense, then, contrary to the claims described above which were in favour of the system, some thought that human assessment would be more complete: *"I think the teacher observation would be more complete than the system" s52.* There has long been a debate on that and there are people who think human minds are always better than computers or the opposite [202, 203], hence some people, even if the system gives them good results, still do not believe in machines.

Again, on understanding, some seemed to believe that the human expert assessments had the advantage: *"I think that feedback from teacher is better because she/he can explain it more."* s42. That is an interesting recognition of the idea that although the system seemed to give more feedback, because of the video and assessment charts, possibly that extra material was not actually considered very helpful compared with a human giving an additional explanation of student performance. Notably, however, nobody suggested that the human would be less biased.

A number of comments attempted to recognise a balance where the system and the human each had strengths: *"most of the things are better with the Observe Portal, but for success of the task it is best if there is an external observer*" s37. That comment did however run counter to the result for PA5, where the responses were significantly in favour of the system in assessing task success. Another appeared to generalise his/her balanced response to other kinds of task than those in the experiment: *"I think traditional methods are important in some situations in the sense that more feedback can be given regarding the presentations, while the Observe Portal system can help to assess interaction and collaborations that may be difficult for to assess via single presentation or through reading the group project/ essay."* s46. This seems to claim that when student work can be assessed via productions, such as presentations or essays which represent the culmination of some aspect of the students' work (but no such production was involved in the researcher's experiment), human assessment might be superior as richer feedback can be given. Where, however, the focus is, as in the present case, on the *process* of doing some piece of work, especially involving interaction and collaboration, this commentator saw the system as having the advantage. According to Wells [8], educators should evaluate the whole learning process when leading learning activities rather than just look at the final artefact as evidence of learning. Overall, we can say that the comments suggest a view that was summed up neatly by one participant: *"I believe there are more things that the Observe Portal has gotten right" s56*.

Generally, the previous results on balance support the first part of hypothesis 4: "*Students and experts, both, prefer the Observe Portal system's assessment feedback over the traditional teacher/expert's assessment feedback*."

Expert Preferred Approach (E-PA)

In order to deal with the part of H4 which referred to the experts, we gathered data with a set of questionnaire items especially for the experts (E-PA) (Table 7-28).

E-PA1	I think the use of the *Observe Portal* system has a significant advantage over traditional methods (teacher observation) for assessing students' performance
E-PA2	Which approach would you prefer to use for assessing collaborative students in virtual world?
E-PA3	Based on your previous Q, explain the reason for your choice?
E-PA4	Could you give us your view of the *Observe Portal* system?
E-PA5	What aspects of the *Observe Portal* do you think helps most in assessing students' learning in the virtual world?

Table 7-28: Expert preferred approach

The analysis of the experts' responses showed that all participants (100%) agreed or strongly agreed that using the *Observe Portal* system had a significant advantage over the traditional method (teacher observation) for assessing students' performance (E-PA1, median=4, "Strongly Agree"). Likewise, all the experts said they preferred to use the *Observe Portal* system for assessing collaborative students in a virtual world over the traditional method, i.e. doing it themselves in the way they had in condition 2-2 (E-PA2, median=4, "Strongly Agree") (Figure 7-5).

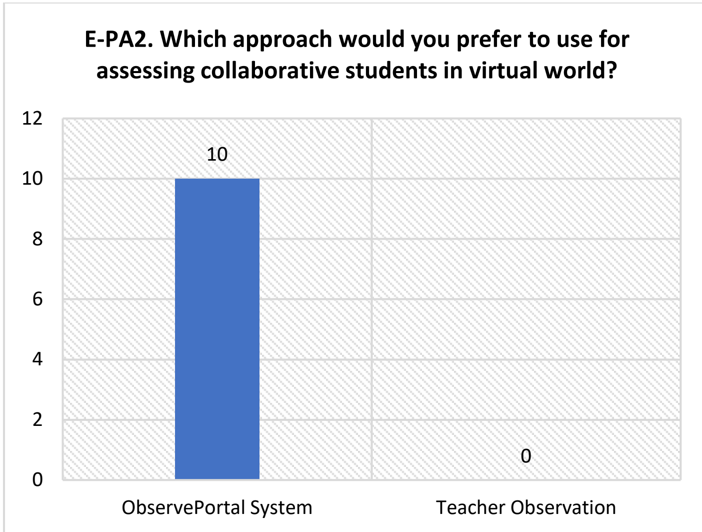

Figure 7-5: Evaluation of E-PA2

Furthermore, the one-sample binomial test showed that on these two closed items the experts responded statistically significantly way above the neutral midpoint of the 1-4 response scale, Table 7-29.

		N	Min	Max	Median	Mean	SD	Binomial Test p
PA1	I think the use of the *Observe Portal* system has a significant advantage over traditional methods (teacher observation) for assessing students' performance	10	3	4	4	3.60	.516	.002
		N	Min	Max	% Choosing System			Binomial Test p
PA2	Which approach would you prefer to use for assessing collaborative students in the virtual world?	10	1	1	100			.002

Table 7-29: Expert PA descriptive statistics and binomial test

To further understand the positive results with regard to the previous items, we analysed the experts' comments in the open-response items E-PA3, E-PA4, and E-PA5, which asked for further comments and reasons. Note that, at the time of making these comments, the teachers had the opportunity to compare the system feedback with their own – with respect to the students that they had assessed.

While some experts just praised the system as "*a good innovation that will be useful with the education sector*", a "*great system to use in learning*" or "*very effective*", some also provided several reasons for their preference for the system feedback. A few said it was "*more accurate*" than their own expert feedback, presumably having compared the two for some example students. A specific advantage claimed was that accuracy was maintained even when there were several students working in a group: *"It is good to be able to observe the performance of all students individually and track them which is hard to achieve in reality."*

However, the most popular reason given was that it was "*easier*" and would "*save time and effort*". Nevertheless, one admitted that they would also themselves assess if feasible: "*It saves time and effort. I might use both where possible but it is handy to have an automated system*." Some again further pointed out that the ease of use became a more important factor when more students had to be assessed: "*for more than 2 students it will be very efficient as the environment will not be too crowded compared to the real world and the activities of each student can be easily monitored*"; "*Especially for large number of students working in VW*."

Experts' wider comments also drew attention to two important provisos. First, one said that the system was "*very efficient for these activities*". This implies that it might not be so efficient for other activities, as indeed was suggested by one of the students as noted above, who in effect referred to it being good for the assessment of process rather than of product. We cannot tell if the expert here had the same distinction in mind, however. Another expert further commented that it was "*a good idea to do such system especially that Immersive learning need more improvements to meet learning requirements*".

Finally, the above refers to the system needing to "*meet learning requirements*". Similarly, another says: "*It can give a good assessment if it matches the assessment's requirements*". This then draws attention to the fact that, in order to be useful, what the system does in making its assessments must agree with what those in charge of the learning require students to learn, and expect to assess.

The last expert PA item, E-PA5, asked an open question about which aspect of the *Observe Portal* the experts thought helped them most in assessing students' learning in the virtual world. In their responses, speed and accuracy were widely mentioned and these were also mentioned in other comments made by the experts. The aspects covered by the experts' statements included almost every feature of the *Observe Portal* system. These aspects were as follows:

a) All the system features
b) Speed and accuracy
c) Assessing collaborative skills
d) Observing the number of attempts, as well as the rating tool
e) The time taken to perform an action by each student
f) The record of the actions performed by students
g) The chat box and the design of the environment

All the above results for the E-PA items, therefore, support the expert related element of the hypothesis: "*Students and experts prefer the Observe Portal's assessment feedback over and above that yielded from human experts"* (H 4).

Comparison between Preferred Approach of Experts and Students

Finally, the Mann-Whitney test was used to compare experts and students in phase 2-2 (the virtual classroom with expert) on the two parallel PA quantitative items which both groups responded to (PA1, PA2).

		Group	N	Mean	M-W Z	P
PA1	I think the use of the *Observe Portal* system has a significant advantage over traditional methods (teacher observation) for assessing students' performance.	Students in 2-2	30	3.10		.015
		Experts in 2-2	10	3.60		
		Group	**N**	**% Choice of System**	**Chi sq.**	**P**
PA2	Which approach would you prefer to use for assessing collaborative students in virtual world?	Students in 2-2	30	%60	5.714	.019
		Experts in 2-2	10	%100		

Table 7-30: Significance tests of PA difference in phase 2-2 between students and experts

As Table 7-30 shows, there were significant differences between experts and students ($p < 0.05$) on both PA items. In both cases, the experts approved the *Observe Portal* assessment significantly more strongly than the students did.

7.3.5. H5: Students and experts express their acceptance of using the *Observe Portal* assessment system.

Finally, we turn to the measurement of acceptance as defined by the three core variables of the TAM evaluation model (see 6.2.1) [20]. In this way, we hope to confirm H5. Below, we report the results in relation to the core TAM variables in turn - both for the students via the three SA variables and for the experts via the corresponding three EA variables (6.2).

Student perceived usefulness (PU)

This primarily measures how far a student user considers that using the *Observe Portal* is useful for assessing performance. *Table 7-31* lists the questions used to measure the usefulness of the system.

PU1	The use of *Observe Portal* is useful for assessment in 3D-VW collaborative activities
PU2 (r)	The use of the *Observe Portal* system is not suitable for assessing students' performance
PU3	The use of the *Observe Portal* system allows me to get a deeper understanding of the individual and group performance
PU4	I find that *Observe Portal* is useful in assessing students' interactions, success, and collaborative skills
PU5 (r)	I don't see that *Observe Portal* makes any difference in assessing students' interactions, success, or collaborative skills

Table 7-31: Student perceived usefulness (PU)

The three items PU1, PU2, and PU3 asked if the system was suitable for assessing students operating within 3D environments or not (PU2 negative worded). As for previous constructs, the composite values were calculated for these items together. PU4 and PU5 (PU5 is negatively worded) asked separately if the *Observe Portal* was useful for assessing students' interactions, success levels, and collaborative skills. 87% of the participants found that the *Observe Portal* system was suitable for assessing distance students in a virtual world (PU1/PU2/PU3 median=3, "Agree"). Similarly, 86% indicated that the system was useful for assessing students' interactions, success levels and collaborative skills (PU4/PU5 median=3, "Agree") (Appendix B.7).

		N	Min	Max	Median	Mean	SD	Binomial Test p
PU1	The use of *Observe Portal* is useful for assessment in 3D-VW collaborative activities	68	1	4	3	3.15	.605	<.001

PU2 (r)	The use of the *Observe Portal* system is not suitable for assessing students' performance	60	2	4	3	3.05	.699	<.001
PU3	The use of the *Observe Portal* system allows me to get a deeper understanding of the individual and the group performance	68	1	4	3	3.16	.589	<.001
PU4	I find that *Observe Portal* is useful in assessing students' interactions, success, and collaborative skills	68	1	4	3	3.10	.736	<.001
PU5 (r)	I don't see that *Observe Portal* makes any difference in assessing students' interactions, success, and collaborative skills	60	1	4	3	3.15	.709	<.001
PU	Mean of all PU items	68	1.80	4.00	3	3.13	.523	<.001

Table 7-32: Student PU descriptive statistics and one-sample binomial test

In addition, the one-sample binomial test yielded ($p<.001$) for all items (Table *7-32*). The quantitative data confirm that the students found the *Observe Portal* system significantly useful for assessing distance students operating within a virtual world, way above the neutral midpoint of the 1-4 response scale. These results, therefore, support the first part of the hypothesis (H5.1): "*Students and experts alike find the Observe Portal system is useful.*"

Student perceived ease of use (PEOU)

Perceived ease of use is "the degree to which a person believes that using a particular system would be free of effort". Table *7-33* lists the questions used to measure perceived ease of use of the *Observe Portal* system.

PEOU1	I find the Observe Portal is easy to use in virtual world collaborative activities
PEOU2	The feedback obtained from the system is clear and understandable
PEOU3	It is difficult to use the Observe Portal assessment system.
PEOU4	Students assessment through Observe Portal is easy

Table 7-33: Perceived ease of use (PEOU)

As with the previous constructs, items were designed both in positive and negative worded forms, so were reversed as needed: in the present case PEOU3. Composite values were also calculated for the mean of all four items (Appendix B.7). 92% of the participants found that the *Observe Portal* is easy to use in virtual world collaborative activities (median = 3, “Agree").

Furthermore, the one-sample binomial test showed that, on all items as well as the means for the whole set, the students found the *Observe Portal* system to be significantly useful for assessing distance students in the virtual world, considerably above the neutral midpoint of the 1-4 response scale (p<.001) (Table *7-34*).

		N	Min	Max	Median	Mean	SD	Binomial Test p
PEOU1	I find the *Observe Portal* is easy to use in virtual world collaborative activities	68	1	4	3	3.24	.672	<.001
PEOU2	The feedback obtained from the system is clear and understandable	68	1	4	3	3.25	.720	<.001
PEOU3 (r)	It is difficult to use the *Observe Portal* assessment system.	68	1	4	3	3.23	.673	<.001
PEOU4	Student assessment through *Observe Portal* is easy	68	1	4	3	3.29	.600	<.001
PEOU	Mean of all PEOU items	68	1.75	4.00	3	3.26	.452	<.001

Table 7-34: Student PEOU descriptive statistics and one-sample binomial test

In addition, Kruskal-Wallis results (*Table 7-35*) revealed that users’ computing expertise level (CU2), knowledge of virtual worlds (VW1), intelligent environment knowledge (IE1) and programming experience (PE) did not have a significant influence on perceived ease of use of the *Observe Portal* system (PEOU1). These results could be related to the simple techniques adopted to design the system and the assessment flowcharts. These results, therefore, support the hypothesis “*Students and experts find the Observe Portal system is easy to use*” (H5.2).

Dependent variable	Independent variable	Chi-square	DF	P
PEOU1	Computing expertise (CU2)	0.111	2	0.111
PEOU1	Virtual worlds experience (VW1)	1.643	1	٠.١51
PEOU1	Intelligent environment knowledge (IE1)	0.365	1	0.546
PEOU1	Programming Experience (PE)	4.029	2	0.133

Table 7-35: Kruskal-Wallis results for ease of use of the *Observe Portal* system

Student intention to use (IU)

Intention to use is the variable which, in the TAM model, has the most immediate impact on acceptance and mediates the effects of PU and PEOU. In our study, it was represented by three positive worded items, whose composite we also calculated to represent the whole scale. *Table 7-36* lists the questions used in the questionnaires:

IU1	I would use *Observe Portal* to assess my performance
IU2	I would use *Observe Port*al to understand my performance
IU3	Assume that I had access to the *Observe Portal*. I intend to use it to understand the individual and the group performance

Table 7-36: Intention to use

Overall, 86% of the students reported intending to use the *Observe Portal* to assess their individual and group performance (median = 3, “Agree") (Appendix B.7). Furthermore, the one-sample Binomial test (Table *7-37*) showed that the students in future significantly intended to use the *Observe Portal* system for assessing their performance in a virtual world, assuming it was available, way above the neutral midpoint of the 1-4 response scale. These results support the hypothesis “*Students and experts express an intention to use the Observe Portal system in the future.”*

		N	Min	Max	Median	Mean	SD	Binomial Test p
IU1	I would use *Observe Portal* to assess my performance	68	1	4	3	3.15	.797	<.001
IU2	I would use *Observe Portal* to understand my performance	68	1	4	3	3.24	.715	<.001
IU3	Assuming that I had access to the *Observe Portal*. I intend to use it to understand individual and group performance	68	1	4	3	3.18	.732	<.001
IU	Mean of all IU items	68	1.00	4.00	3	3.19	.68	<.001

Table 7-37: Student IU descriptive statistics and one-sample binomial test

Expert perceived usefulness (E.PU)

We next consider the same three TAM3 components of acceptance again, but this time in relation to the experts [184]. Again, here TAM3 included the three variables: perceived usefulness (E.PU), perceived ease of use (E.PEOU) and intention to use (E.IU). E.PU is designed to measure whether the experts believed that the *Observe Portal* system was useful for student assessment in virtual worlds (Table 7-38). Like other sets of items, E.PU3 and E.PU5 were negatively worded so had their scores reversed.

E.PU1	The *Observe Portal* is useful for assessment in 3D-VW collaborative activities
E.PU2	The use of the *Observe Portal* system allows me to get a better understanding of the individuals' and groups' performance
E.PU3 (r)	I think the *Observe Portal* system is not suitable for assessing students' performance
E.PU4	I find that *Observe Portal* is useful for assessing students' interactions, success, and collaborative skills.
E.PU5 (r)	I don't see that *Observe Portal* makes any difference in assessing students' interactions, success, and collaborative skills

Table 7-38: Expert perceived usefulness (E.PU)

E.PU1, E.PU2, and E.PU3 asked generally about whether the system was useful for making assessments in 3D environments or not while E.PU4 and E.PU5 asked about whether the *Observe Portal* was useful specifically for assessing students' interactions, success levels and collaborative skills. The results table in Appendix B.7 shows that all the experts found the *Observe Portal* system was useful for assessment in the collaborative virtual world, (E.PU1/E.PU2/E.PU3(r) median=4, “Strongly Agree”). Also, ٩5% of the experts expressed the opinion that the system was useful for assessing students’ interactions, success levels and collaborative skills (E.PU4/E.PU5(r) median=3, “Agree”).

The one-sample binomial test was applied to determine whether there was a statistically significant difference from the scale midpoint rating in responses to the E.PU items (Table 7-39). The quantitative data showed that the experts believed that the *Observe Portal* system is, in all the areas that were asked about, significantly useful for assessing distance students in a virtual world, considerably above the neutral midpoint of the 1-4 response scale. This was particularly prominent for E.PU1, 3, 4 where no rating fell below 3.

		N	Min	Max	Median	Mean	SD	Binomial Test p
E.PU1	The *Observe Portal* is useful for assessment in 3D-VW collaborative activities	10	3	4	4	3.70	.483	.002
E.PU2	The use of the *Observe Portal* system allows me to get a better understanding of the individuals’ and groups’ performance	10	2	4	4	3.30	.675	.021
E.PU3 (r)	I think the *Observe Portal* system is not suitable for assessing students’ performance	10	3	4	4	3.30	.483	.002
E.PU4	I find that *Observe Portal* is useful for assessing students' interactions, success, and collaborative skills.	10	3	4	3.5	3.50	.527	.002
E.PU5 (r)	I don’t see that *Observe Portal* makes any difference in assessing students’ interactions, success, or collaborative skills	10	2	4	3	3.30	.675	.021
E.PU	Mean of all E.PU items (condition 2-2)	10	2.5	4	4	2.95	.261	.021

Table 7-39: Expert PU descriptive statistics and one-sample binomial test

These quantitative results clearly support the hypothesis: *“Students and experts alike find the Observe Portal system is useful”* (H5.1).

Expert perceived ease of use (E.PEOU)

The perceived ease of use is defined as “the degree to which a person believes that using a particular system would be free of effort”. Table *7-40* lists the questions used to measure the perceived ease of use of the *Observe Portal* system.

E.PEOU1	I find the ObservePortal is easy to use in virtual worlds
E.PEOU2	Using the ObservePortal system is clear and understandable
E.PEOU3	It is difficult to use the ObservePortal assessment system.
E.PEOU4	Assessment through the ObservePortal system is easy

Table 7-40: Expert perceived ease of use (E.PEOU)

As with the previous constructs, the E.PEOU items were phrased both positively and negatively, so reversed values were calculated and used where necessary. The results showed that 93% of the experts significantly agreed that the *Observe Portal* was easy to use in virtual worlds (median = 4, “Strongly Agree") (Appendix B.7).

Next, the one-sample binomial test showed that the experts found the *Observe Portal* system significantly easy to use for assessing distance students in virtual worlds, way above the neutral midpoint of the 1-4 response scale (Table 7-41).

		N	Min	Max	Median	Mean	SD	Binomial Test p
E.PEOU1	I find the *Observe Portal* is easy to use in virtual worlds.	10	1	4	4	3.60	.966	.021
E.PEOU2	Using the *Observe Portal* system is clear and understandable.	10	3	4	4	3.80	.422	.002
E.PEOU3 (r)	It is difficult to use the *Observe Portal* assessment system.	10	1	4	3.5	3.10	1.197	.109
E.PEOU4	Assessment through the *Observe Portal* system is easy.	10	3	4	3	3.40	.516	.002
E.PEOU	Mean of all E.PEOU items.	10	2	4	3.75	3.50	.553	.021

Table 7-41: Expert PEOU descriptive statistics and binomial test

These quantitative results again support the hypothesis: *“Students and experts find the Observe Portal system is easy to use”* (H5.2).

Expert intention to use (E.IU)

Intention to use is the third core TAM variable, intended here to measure the intention to use the *Observe Portal* system. *Table 7-42* lists the questions used in the questionnaires.

E.IU1	I would use ObservePortal to assess students’ performance in collaborative activities
E.IU2	I would use the system for assessing students’ interactions, success, and collaborative skills.
E.IU3	Assume that I had access to the ObservePortal. I intend to use it to assess student performance.

Table 7-42: Intension to use

Items E.IU1, E.IU2, and E.IU3 asked about the experts' intention to use *Observe Portal* to assess student performance. All items were positively worded and the result shows that, overall, 90% of the experts would intend to use the *Observe Portal* to assess individual and group performance in virtual worlds, where it was available (median = 4, “Strongly Agree") (Appendix B.7).

Next, the one-sample binomial test (Table *7-43*) showed that on all items the experts expressed a significant intention to use the *Observe Portal* assessment system in the future (if available) - way above the neutral midpoint of the 1-4 response scale. This result, therefore, supports the hypothesis *“Students and experts express an intention to use the Observe Portal system”* (H5.3).

		N	Min	Max	Median	Mean	SD	Binomial Test p
E.IU1	I would use *Observe Portal* to assess students' performance in collaborative activities	10	2	4	4	3.60	.699	.021

E.IU2	I would use the system for assessing students' interactions, success, and collaborative skills	10	2	4	4	3.60	.699	.021
E.IU3	Assuming that I had access to the *Observe Portal*, I intend to use it to assess student performance	10	2	4	3	3.30	.675	.021
E.IU	Mean of all E.IU items	10	2	4	4	3.50	.652	.021

Table 7-43: Expert IU descriptive statistics and binomial test

Overall, the evaluation based on student and expert responses to the TAM variables, perceived usefulness (PU), the perceived ease of use (PEOU) and the intention to use (IU), yielded positive results. Most of the participants agreed that the system was easy to use, useful, and they would want to use it, where available. Accordingly, following the TAM model, we conclude that they accept the use of the *Observe Portal* as a tool for assessing individual student and group performance. These results then support hypothesis 5, *"Students and experts express their acceptance of using the Observe Portal assessment system."* This, in turn, constitutes another positive evaluation point with regard to the system.

7.3.6. Final comments

Finally, we may note that the students showed to be cooperative participants, producing a number of responses to the open questionnaire items, and generally they seemed to enjoy taking part in the experiment. The students' optional additional comments on their overall experience were as follows:

1. *"it was a nice experience which allowed me to use my past knowledge of C programing."*
2. *"was very interesting and exciting, would certainly do it again"*
3. *"Thank you it was very interesting!"*
4. *"it was perfect"*

7.4.Chapter Conclusion

This chapter has presented the results arising from the results yielded by the surveys aimed both at the experts and the students, and analyses of the human and system scores, the experts' assessment sheets and the student chat and ratings. Overall, the data strongly supports the effectiveness of the research framework, and the prototype superiority over human expert assessment, for the kind of assessment in question (real-time assessment relating to multiple scales of students working on learning tasks). This finding will be further considered and discussed in the next chapter (Chapter 8).

Chapter 8

8. Discussion

Chapter 7 described the empirical evaluation of the prototype that we created based on *MIVO* framework to facilitate the observing, recording, assessing and giving feedback on collaborative distance learning activities performed by participants in a 3D virtual world environment. The evaluation of these activities was achieved by setting up three experimental situations in which students used the system; this enabled a comparison between the virtual observation/assessment system and observation/assessment by a human expert, performing in a role similar to that of a teacher. This chapter (Chapter 8) discusses further the findings from these experimental set-ups and their broader significance for this research area. The chapter opens with a summary of how far the research hypotheses were supported. Then it discusses the evaluation results concerning the effectiveness of the *MIVO* framework and the *Observe Portal* system.

8.1.Research Aim Revisited

The overall research aim, as stated in Chapter 1, was to expand the ability of 3D virtual environments to support learning by improving the method of assessing learning in such spaces. The research focused on the creation of a novel computational architecture framework that improved upon previous frameworks for assessing and giving feedback on collaborative distance learning taking place in a VW (*Observe Portal*). Apropos this, the conceptual *MIVO* framework was proposed in Chapter 3, drawing upon parallels between the observing and assessing of students in physical settings and the observing and assessing of them in VWs. To do this (observing etc., in VWs) successfully, a *MixAgent* mechanism was provided which combines natural agents and software agents within the same environment – to support the recording learning evidence from virtual activities. Also, the *OLens* model was constructed to simulate teachers' observation and assessment of collaborative students in immersive environments, as explained in Chapter 3. In addition, the research made a significant contribution by demonstrating the implementation of the *Observe Portal* as a proof-of-concept prototype applying the fuzzy logic approach described in Chapter 4 and Chapter 5. Such a computational-architectural framework could be used to fill the gap between the collecting of learning evidence from technological sources and the collecting of such from human teachers etc., and may well be able to help solve the issues that arise when evaluating such aspects of learning as interaction, task success and the social collaborative skills which come into play when students work collaboratively in a distance-learning environment. Another contribution of the present research is the experimental results yielded by the learning activities created specifically to enable the researcher to empirically evaluate the *Observe Portal*. Research hypotheses were stated which could be tested through data gathered from experts and students using the system (Chapter 6 and Chapter 7).

8.2. Discussion of the Research Results

In Chapter 6, the research hypotheses are re-expressed in more practical terms that allow for their empirical testing, and, indeed, these hypotheses have all been evaluated through the empirical work described in Chapter 6. Chapter 7 restates the hypotheses with an indication of how well supported each was (Table 7-1). This section (8.2) discusses further the findings from the evaluation experiments and their broader significance.

Hypothesis 1

An essential aspect of this research was to propose a computational framework to enhance the existing methods of observing and assessing distance-learning students working collaboratively in VWs. Our framework was aimed at improving the collection of learning evidence from distance collaborative students by the means of integrating software and natural agents. In effect, the students were required not only to communicate with each other while they performed the learning tasks collaboratively, as would happen also in face to face pair work in a physical classroom, they additionally were required, for the purposes of assisting the system to make better assessments, to categorise each message they sent (to other students) into one of six types, and to engage in repeatedly rating each other's quality of performance. Neither of these activities usually takes place in the course of collaborative work where a human teacher is engaged in assessing the students - whether in a physical or virtual learning space - since they would not usually be regarded as activities that were necessary for learning to occur. Hence it was essential to ascertain if, in fact, students acquired positive attitudes towards these natural agent tools when using them while performing distance-learning tasks in the virtual world (*H 1)*.

To evaluate this hypothesis, firstly, we measured whether the students reported positive attitudes towards rating each other when performing distance-learning tasks in the virtual world, and whether their use of the rating function as they worked suggested that they did not find it onerous (*H 1.1)*. Secondly, we measured whether the students had positive attitudes to the online classified chat facility which they have to use when performing tasks in the virtual world, and whether their accuracy of classification of their messages signalled any difficulty with the classification requirement (*H 1.2)*. To our knowledge, since natural agents such as the ones proposed in this present study are not widely used in assessment systems, these issues have not been researched before, so my results here constitute new findings.

The essential instruments needed to measure students’ attitudes to their roles as natural agents, rating each other while performing tasks in VWs, were the natural agent rating perception questionnaire items (NA). The students' responses to every closed-response NA item were significantly positive overall. 89% of the participants found that working as natural agents and rating other students through the rating tool was easy to do. In addition, 78% of the users reported that they found it fun, 86%, interesting, and 82%, useful. These results demonstrated the acceptance of the student users of working as natural agents and using the rating tool to rate other students in the course of, and at the end of, the learning activity. On the other hand, we must remember that the participants had, in all likelihood, not been called upon ever before to provide such ratings of their peers while undertaking learning tasks, or at any rate not with such frequency. Hence the whole process would have had a certain 'novelty effect'. We cannot be certain that if they were called upon to do this over an extended period of weeks, each and every time they worked collaboratively on learning tasks online, they would still maintain the same level of enthusiasm throughout.

In the open response data, some students reported that they felt that assessing each other was comfortable because there were no face-to-face evaluations and the ratings were anonymous, so they could be honest when they rated each other. This supports the findings of Lu and Bol [204] who studied the effect of anonymous electronic peer review. Those authors found that it was more effective for students to give anonymous scores and feedback to their peers than it was to provide students with reviews from identifiable other students. This was because an anonymous assessment system makes participants more amenable to giving an honest rating and providing their classmates with valuable remarks which were intended to improve their (the classmates) performance.

Other participants described the rating tool as very easy to use because it had a simple design, was easy to navigate and had a straightforward method for rating - through a scroll bar with three options (high – middle – low) and a button press. This result is in agreement with that of Darejeh and Singh [205], who stated that simple designs help users to understand system structure and improve their interest in using a particular software system. Also, the rating timing system and the reminders this produced helped participants to remember to evaluate other students on a continuing basis; this meant that the rating process became more likely to be used. In addition, according to [149], a simple and user-friendly student-originated feedback tool allows the creation of such feedback in a way which avoids the untoward interruption of a learning session. Therefore, another purpose of the simple rating mechanism in our system was to provide synchronous feedback during the learning activity without imposing significant interruption on the students. Furthermore, asking students before starting the learning session to rate others only when they are free to do so – i.e., after completing a program, after a conversation or after completing a learning task – also helped to prevent the feedback requirement negatively affecting the flow of student learning.

However, a few students suggested widening the rating scale to one of five points instead of just three. This draws attention to an important aspect of peer review, which is to provide learners with sufficient options by which to express their feedback and allow them to be more informative with their review [204]. However, although widening the scale would be more useful for people who are more cognisant of the educational process, and so more able and willing to give accurate ratings, this could overload the students' thinking processes, since they would be having to undertake collaborative tasks, classify their chat sentences and rate other students all at the same time.

A further, more significant, point, however, is that most of these students in their open responses, whether positive or not, seemed to be interpreting the rating activity as having a pedagogical purpose - to aid the learning process. Clearly requests for a wider scale, and more room for feedback, with the aim of working on each other's weaknesses, all relate to a learning purpose. However, the researcher's actual main purpose in providing the student rating system was, in fact, to provide the necessary information so that the system could work better (in terms of providing assessments), which is quite different. For that purpose, a wider scale and an opportunity for the students' to provide qualitative feedback would not be more helpful. Hence what is brought to light here is a conflict between the needs of the learning task and the needs of the system. Whether or not the students had been told what the purpose of the rating system was, they clearly focused on the possible pedagogical purpose, as more meaningful to them, but this conflicted with the researcher's main purpose.

Furthermore, the analysis of the frequencies of rating saved in the database by the system showed wide variation between students but a median frequency of twice in subtask 1 and three times each in subtasks 2 and 3. This, therefore, supports the conclusion that the simple form of rating implemented in this study did not represent too onerous a task and became routine over time. Kruskal-Wallis test results showed that the users' level of computing expertise level or knowledge of virtual worlds did not have a significant influence on the ease with which they believed that the rating could be performed. These results could be related to the simple design of the method offered by the natural agent tool - a scroll bar and rating button interaction. This, as has been said, was all that was needed to enable the natural agent input to the system, although the facilities provided were not really adequate to properly fulfil a pedagogical function in the task completion process. Overall, then, all the data related to H1.1 predominantly (though not without some provisos) support the hypothesis that *the students express positive attitudes to the activity of rating each other when performing collaborative learning tasks in the virtual world* (*H 1.1).*

Secondly, we needed to measure students' attitudes to the online classified chat system which could be used when performing tasks collaboratively in the virtual world. The COMM subset of items in the questionnaire which concerned this all returned significantly positive ratings. 86% of participants found that communicating via the classified chat window was relatively easy to use. Also, the majority of students found that their experiences of collaborating with the other student(s) in other location(s) using the classified chat were generally good.

Participants' qualitative comments about the chat system mostly concerned points which could be made about online written chat in general, regardless of any requirement to classify each message into one of six categories. For instance, students commented that using the chat was easy because it was similar to their daily interactions through chat boxes in social media. Others found that the chat helped them to collaborate better and contribute more. Others suggested reasons why the implementation made them feel personally good about the chat because it was very comfortable, easy to use and well structured. On the other hand, one found having to type troublesome as compared to face to face talking. This is in line with the results obtained by Pena-Rios [52] whose participants considered voice communication to be much easier than using chat. The use of text-based chat in this study was however required because of the facility to save chat content in user logs and analyse it later for research purposes. Furthermore, it would have been difficult if not impossible to elicit a classification of each message if the messages had all been spoken.

More of interest to us, however, was that some students did report that they found using the multiple buttons to classify their chat a little bit bothersome and indeed, for their purposes, unnecessary. Once again they were thinking of what the pedagogical function might be of the chat classification in terms of what they were required to do in relation to the tasks they were set (programming a house), and not surprisingly failed to see any such function in performing this classification. They had forgotten, that the function was actually to assist the system in making its assessment of them (the students). So, rather as with the rating system discussed above, we again see a slight conflict here between what serves a direct learning-related purpose and what is needed to make the expert system work as required.

On the other hand, the analysis undertaken of the classified chat, stored in the chat logs, revealed that 90% of the chat messages had been categorised properly and therefore that the users had pressed the correct buttons, classifying the discourse function of each of their written utterances into one of the six categories provided. The students' used the response categories most, but where a distinction was required between REPLY and AGREE, REPLY was over-used - in instances where the purpose of the utterance was really for AGREE.

This general pattern of error, where a more general category (here, REPLY) was used in place of a more specific one (here, AGREE, which is really just one kind of REPLY) is not unusual. It may be a sign that some students felt pressured by the limited time allowed for each subtask and the fact that they were being observed (at least in condition 2-2, where the expert observer was present on screen as an avatar). In such a situation students will naturally allocate cognitive resources primarily to what they see as the main task, programming the house, and this might in some instances leave insufficient time or cognitive resources for the identification of a message to be something more specific than a REPLY. However, it is perhaps remarkable that there were only around 10% erroneous categorisations - which shows therefore that this issue did not constitute a major problem.

Furthermore, students' positive attitudes were again found not to be dependent on their particular backgrounds or experiences. Thus, the hypothesis was predominantly supported: *"The students express positive attitudes to the online classified chat facility which they can use when performing collaborative learning tasks in the virtual world"* (H 1.2).

Putting together all the above findings from the student's NA and COMM questionnaire responses and their rating and chat logs we feel that the evidence overwhelmingly shows that there was an acceptance by the students of their roles as natural agents. All these results together support hypothesis 1, "*Users express positive attitudes towards their roles as human agents when performing distance-learning tasks in the virtual world"* (H1), and also support an overall positive evaluation of the *Observe Portal* system. These results also point to the effectiveness of our computational framework (*MIVO*) which was involved, right from the start, in the proposal of the *MixAgent* mechanism - to integrate software and natural agents in order to collect learners' data, identify the relevant learning evidence and so assess the learning achieved by groups and individuals. This does not however mean that we think no improvements can usefully be made in response to the relatively small number of negative comments yielded by the open questions. These we will address in the conclusion chapter.

Hypothesis 2

Another important evaluation-related objective of this research was to assess whether the *Observe Portal* system provided distance collaborative learners with valuable feedback, as judged by whether the users reported positive experiences with, and positive attitudes towards, the quality, quantity etc., of the assessment feedback provided (H2). This construct was evaluated in relation to the students by the use of a version of the AEQ questionnaire instruments (QTF, QF, UF) and in relation to the teachers, through the REF set of questionnaire items, as summarised in Table 7-1. Positiveness of response was measured in two ways: first, whether the users recorded positive ratings relative to the response scale itself (significantly above the midpoint), and second, whether more positive ratings were reported in the situation where the system provided feedback derived from students working totally online in a VW in condition 2 than were reported where the feedback came from a human expert observing students working face to face in a physical classroom (so taking on a traditional teacher's role) in condition 1.

As regards the student measures, the QTF, quantity and timing of feedback, results revealed that 91% of the students had positive experiences of this and reported that they had received plentiful feedback from the system in relation to their (the students') performance. Correspondingly, 89% of the participants reported that the assessments were received back from the system in a timely fashion. The students were significantly in agreement with all the QTF items except one: only 44.1% responded that they would have learned more (about their performance) if they had received more feedback from the system, so implying that they mostly considered that what they had in fact received was adequate. One way of interpreting this result is that, although the system did not provide any qualitative feedback, the assessment charts that it did provide were judged to be sufficient by the students for their own purposes. These results, therefore, support the hypothesis "*The students express positive attitudes to the speed and amount of feedback they can obtain from the Observe Portal*" (H 2.1).

Another aspect of the students' assessment experience which was measured was the quality of feedback (QF). The results demonstrated again that most of the student participants (93%) agreed that the feedback the system provided was useful in helping them to understand the reasons behind the assessments that were made of their individual and their group's interaction levels, of their task success level and of their social collaborative skills level. Indeed, no significant difference was found between their reported satisfaction with the system feedback on each of the three measures. In addition, 77.53% of the students reported that the feedback helped them to recognise their performance level in relation to the learning activity. All these items were responded to positively – as indicated by a binomial test. These results, therefore, support the hypotheses "*The students believe that the Observe Portal system provides very useful information about interaction, success and collaborative skills*" (H 2.2) and "*The students report that they have a good understanding of the Observe Portal assessment*" (H 2.3).

Finally, among the results reported via the AEQ, students widely claimed in response to the UF items presented to them that they did utilise the feedback provided by the system. The AEQs yielded that 79.41% of the participants said they used the assessment feedback in order to help them go back over what they had done in the learning activity. 85.29% of the students thought the availability of such a system assessment would be of use to them in relation to any such activity in the future. 69.12% of the students agreed that the system prompted them to use the recorded video material in order to understand the reasons for their assessment scores. Responses to all the items were significantly positive except in relation to one particular item to which learners (48.53%) responded that they had merely read the marks from the system, and had not in fact, as a result, watched the video of the whole session (which the system also provided), despite being prompted to do so.

This latter result might have been because the students wanted to finish the experiment early and did not wish to spend more time in watching the video of what they had just done, since of course the experimental sessions were not part of any ‘real’ course that they were taking. Hence the tasks they had undertaken and the skills/knowledge they had acquired would not need to be remembered or be revised in the way that tasks set in a course pursuant of qualification would be. This illustrates a largely unavoidable limitation of using experimental trials of new educational software which are undertaken outside of any real learning context. However pedagogically realistic the tasks are, the students will not be participating with the same motivations and sense of purpose that they would have when performing tasks that are part of an assessed (e.g., degree) course [206].

Still, overall the student participants mentioned did claim to exploit the system feedback after working in the VW significantly more than they exploited the human (teachers’) feedback after working in the physical classroom. Hence, overall, we regard these results from the quantitative data as supporting the hypothesis “*The students claim to make extensive use of the feedback from the Observe Portal*”(H 2.4). It must be noted, however, that the hypothesis is itself necessarily limited to be concerned with what they claim, rather than with what, in fact, happened since we did not gather data about any use they actually made of the feedback.

Moving now to the expert participants, some similar themes to those which emerged from the AEQ given to students were also demonstrated by expert's responses to the REF list of items, provided to them in order to evaluate the assessment feedback from the human expert point of view. The evaluation results showed significantly positive responses to all but two items. One of these two was concerned with whether they (the experts) believed that the learners understood the purpose of the system assessment. The students had claimed, to a significant extent, that they had, but the experts seemed to doubt that they (the students) actually had. However, such differences of opinion are to be expected when self-report instruments such as questionnaires are used rather than objective measures of understanding. All the experts (100%), however, found that the system provided learners with useful feedback subsequent to the activity, as well as that the scoring of the learners' performance by the system was helpful. Likewise, the instructors reported that the assessment provided by the system could help learners improve their performances.

In addition, 100% of the participants strongly agreed that the *Observe Portal* assessments were reliable and valid. They did not show much evidence of understanding the difference between those terms, which perhaps were too technical for such a questionnaire. However, they did evidence that one of the strongest criteria many of them used for endorsing the quality of the system feedback was that they felt it resembled their own in some way. They differed however in whether they thought the system actually did better than they did, e.g. in being able to process more information faster, or less well, in that some insights might be missed by it.

In response to other REF items, the experts strongly agreed (80%) that peer evaluation was a good approach to assess the quality of student performance in VWs but failed to significantly approve of using the rating tool to assess fellow group member(s) during the collaborative activities. In open responses, one voiced the view that students' ratings of each other could be too influenced by emotion while another regarded the authenticity of human peer ratings as adding strength to the assessment. Furthermore, it could add also to the depth of reflection and hence the learning process. According to Kyllonen [38], students collaborative skills cannot be effectively accomplished by simple assessment approaches. These skills require more advanced means of measurement such as peer-rating [38]. Peer-rating assessments are more predictive and accurate than other assessment methods when evaluating student skills [39]. Additionally, it enhances student decision skills and it supports lifelong learning by not just relying on instructors assessment [201].

Finally, as did the students, the experts judged the value of the system feedback via the VW in phase 2 (cond 2-2) as more valuable than which humans gave in phase 1, the directly observed physical classroom. Hence overall, but not without reservations, the expert views on the assessment were positive and so favourably evaluated the expert assessment system with respect to the hypothesis: *"The experts report positive attitudes to the value of the system assessment"* H 2.5.

In total, all the above results then for the most part support the main hypothesis *"The Observe Portal system provides collaborative distance learners with assessment feedback, and users report positive experiences of such assessment feedback"* (H2). Generally, this is a positive evaluation of the research prototype. The results also demonstrated the effectiveness of our *MIVO* framework for observing collaborative distance-learning students and its ability to be used as a basis of student assessment. In particular, the results advocate for the *OLens* model within the *MIVO* framework. *OLens* is the observation model for identifying and structuring evidence of student learning which has taken place in a 3D virtual environment as part of an e-learning assessment process. *OLens* was implemented by creating a multi-level assessment interface within the *Observe Portal* prototype. Each assessment window (interaction, success and collaborative skills) in the prototype corresponds to a level of the *OLens* model. The positive results yielded in relation to the previous hypotheses support the effectiveness of the model and its value in assessing learners' performance from various different perspectives.

Hypothesis 3

An important aspect of the expert assessment system to evaluate was whether it actually presented an improvement in assessing collaborative students over traditional teachers' observation (H 3). This aspect was tested in three different ways: by obtaining experts' reports on how difficult they found it to assess students as they worked at a distance online in phase 2 (cond 2-2) (EXP), by analysing the extent to which they actually managed to complete the assessment sheets in phase 2 (cond 2-2) and by comparing the scores awarded by the system with those provided by human experts.

The expert observation experience (EXP) questionnaire items measured the experts' experiences and their thinking about their manual observation process when assessing the students in the virtual world - with access to the same data that was used by the system to arrive at its assessment. We had expected that the experts would report high levels of difficulty in relation to assessing the students in the virtual world themselves - which would by implication enhance the evaluation of the system assessment. In the event, one aspect of expert assessment (in condition 2-2) was rated significantly difficult: this aspect concerned the difficulty involved when there were more than two students to observe and assess.

Otherwise, the results reflected a view that assessment by the experts themselves, of the activities in the VW, was only moderately easy. 53% of participants argued that it was difficult to observe students' performance in *Observe Portal* using the manual sheets. An examination of the experts' comments revealed why some, but not others, of the experts found observing the students relatively easy. Those who did find this observation task quite straightforward were those who were proficient at tracking student actions and chat history via the facilities offered by the environment. Conversely, others reported that they found collecting learning evidence difficult because this had to be done in real-time and they often lost track of who did what. On the issue of assessing more than two students, 90% of the experts judged that this, in particular, was not an easy task because there were even greater problems associated with tracking and assessing large numbers of students simultaneously in real-time and recording such assessments on the manual sheets. This finding echoes others' views to the effect that observing and taking notes is a challenging task, especially where teachers are required to observe many students at the same time with respect to differing activities within an educational setting [207].

Many of the experts agreed that assessing individual student's interactions and task successes was easier than assessing their collaborative social skills. However, most of the points made applied to the difficulty of trying to assess multiple students involved simultaneously with the same task (and regardless of whether a physical classroom or a VW setting was used) - e.g., losing track while observing the students, and finding assessment time-consuming and inconvenient. The same was true of a number of the other possible remedies (i.e., other than this assessment system) suggested in response to the issues encountered in teacher/expert assessment [41, 207].

In addition to the experts' own reported opinions concerning the ease and otherwise of undertaking assessments in condition 2-2, we also used evidence of their actual performance - in the form of their success rate in terms of completing the manual assessments. The results showed that, mostly, the assessments of the interactions and of the task success levels of groups and individuals were completed successfully by the experts; however, only 33% of the required assessments of the social collaborative skills displayed by the groups and students were completed successfully, leaving 66% of the assessment skills sheets left half or entirely uncompleted, while of course the *Observe Portal* system provided 100% of its required assessment feedback (and, in addition, without any time delay). An interesting finding was that the experts spent most of their time and effort evaluating students' interactions and success, and most could not easily assess the social collaborative skills dimension.

In summary, although some of the experts reported that assessing students via the use of the manual sheets was a relatively easy task, provided that there were no more than two students to assess simultaneously, many, nevertheless, found that they could not complete all the assessment sheets required - especially in terms of assessing the social collaborative skill aspects of the activities. In contrast, the system, of course, presented all the assessment results it was required to, to the users. This illustrates, among other things, the limitations of self-reported evidence in relation to evaluation studies, and indeed, research in general. Clearly, the teachers over-stated their abilities, in terms of their being able to make comprehensive assessments, and it is possible that they simply did not want to admit to the researcher that they had experienced the amount of difficulty that they had in fact experienced, even when assessing only two students – specifically as regards using the more complex social collaborative skills scales. Indeed, some experts may just have wanted to please the researcher, and for this set of items believed that reporting that the assessment was not too difficult for a human expert represented what the researcher wanted to hear (although actually in this case that was not the opinion that would most strongly support the value of the expert system). Even taking all this into consideration, however, we nevertheless feel that these outcomes do support H 3.1, *the human experts find that making their own assessments of the students' activities in the virtual world is a difficult task*.

The other essential evaluation point relevant to H3 was that of comparing the *Observe Portal* assessments to the expert assessments in terms of the actual scores awarded. The *Observe Portal* assessment had been built on the basis of human experts' evaluations, thus it was expected that "*The system matches the human experts' performance closely in relation to the assessment of all measures of student task performance in the VW* " H 3.2. The quantitative data and the tests used demonstrated that there were no significant differences between the system assessments and the expert assessments as regards all of the interaction and collaborative skills measures; this meant that the system and the expert assessments were closely matched on these measures. However, the data showed that there were significant differences in terms of the success1 in task1 assessment.

In order to understand these differences, we looked closely at the data and the rules programmed into our system. As explained in Chapter 5 - Section 5.3, we built the FL system on the basis of human experts' assessment beliefs employed in relation to learning activities. Both the system and the experts were required to score students as low, middle or high based on their (the students') performance. When comparing between the scores generated from the system and those given by the experts, we concluded that the system was more critical than the experts when evaluating the students' task success on the first assessment occasion. The following are some possible explanations for these differences between the system’s and the experts’ scoring in relation to success1/task1.

First, it was usually in the course of task 1 that the students started slowly to discover the learning environment and begun to understand how it worked. Once the students had progressed to task 2 or task 3, they understood the functionality of the system and therefore did better on these tasks than on the first. However, the *Observe Portal* did not consider this kind of factor when evaluating the students. It treated all three tasks in the same way when counting the number of correct programs and when considering whether each student’s success score should be low, middle or high. On the other hand, some of the human experts might have scored the students at a higher rating than they otherwise would have done, in relation to task 1, because they took the students’ unfamiliarity with the system, at that point, into account.

A second explanation for these differences in scoring is that the experts’ evaluations are based on their own opinions and so the experts’ judgements may differ from one to other. Observation is a skill that all people naturally have, nevertheless some notice different things, and some are more observant generally than others [207]. Thus, these differences may have affected the results of the evaluations, in that most of the experts’ assessments matched the system’s, but some did not. It remains a matter of debate as to whether it is a failing of a system not to match such human variation, or should we regard this as a way in which such a system is actually better than human experts – by regarding such variation as being due to 'human fallibility'. However, in conclusion, and regardless of these considerations, we can say that the human and system scores were not the same just in relation to success1, and so H3.2 is 9/10 supported.

Overall, the human experts found that making their own assessments of the students' activities in the virtual world was somewhat difficult (H 3.1) and that the system matched the human experts closely on most measures relating to task performance (H 3.2). These results, therefore, with some reservations, do support the main hypothesis (H3): *The Observe Portal provides assessments that are very similar to human-expert assessments; these Observe Portal assessments are produced using less effort overall.* These results, in turn, support the *MIVO* research framework and its proof-of-concept prototype.

Hypothesis 4

Hypothesis 4 specified that the users would prefer the *Observe Portal* assessment results as compared to traditional teachers' assessments. The strategy used to evaluate users' preference was to ask learner and expert participants who experienced both assessment approaches in phase 2 (Cond 2.2) to express their preferences.

Based on the students' feedback, a general preference for the *Observe Portal* system emerged from the students' feedback, but this differed significantly from the neutral midpoint judgment on only three of the eight relevant quantitative PA items. Thus, notably, the majority of participants (87%) thought that the *Observe Portal* system demonstrated significant advantages as compared to expert observations as regards assessing students' performance in the virtual world. Specifically, most of the students found the system useful and that it provided more information about the students' social collaborative skills and also their task related success. In terms of assessing interaction levels, however, only a little over half of the students (57%) considered that the system's assessments were better than the teachers' (this result not to be statistically significant). In addition, 67% of the participants believed that the assessments made by the system were in general more accurate and useful than the teacher's assessments (this was also found not to be significant).

Turning now to the experts' preferred approach, the analysis of the experts' PA questionnaire responses showed that on the two relevant closed response items they very significantly agreed that using the *Observe Portal* assessment system had significant advantages over the traditional methods for assessing students. Likewise, all the experts reported that they preferred to use the *Observe Portal* system for assessing collaborative students operating within the virtual world – as opposed to the traditional method.

The experts' qualitative comments revealed their reasons for choosing the system for use in assessment (as opposed to simply making their own assessments). Firstly, some of these participants reported that they found the system was more accurate in terms of collecting learning evidence, especially when more than two students were being assessed. Secondly, others reported that using *Observe Portal* was much easier than assessing the students manually and saved time and effort, again this was especially true where more than two at a time had to be assessed. While the first point matches what some students said, of course, the saving of effort and time was a key factor only to the experts/teachers. Technology can greatly improve educational processes and supports teachers by providing evidence of students' progress and reducing the amount of time required to assess learners [208].

Some experts did also make comments partly echoing what some of the abovementioned students said, that the system assessment might be more relevant to certain types of learning; indeed that it could improve the learning and assessment which could be achieved in immersive environments. According to Gobert [101], educators can encounter serious issues when attempting to assess learning within 3D environments. There is a significant lack of theoretical foundations which can be discovered in the literature regarding learning assessment and assessment approaches that are relevant to virtual environments. Thus, this thesis aimed to overcome this limitation, exploit the affordances of 3D virtual worlds, and investigate the ways in which students can be assessed in such spaces, in order to enhance their (these environments') learning effectiveness.

In addition to the previously detailed PA results, there were significant differences between the experts and the students as regards the PA items. The experts accepted the *Observe Portal* assessment significantly more strongly than the students did. This is a very interesting result - that human experts/teachers preferred the use of the system for assessment more than their students did. This could be due to at least two factors. First, the experts, because of their background knowledge of computing, presumably had a far deeper understanding of how the system worked than most of the students did. Hence, they might have been able to see its merits more clearly, including in terms of issues such as how it might be arriving at its assessments; thus, they were less likely (than the students) to judge it by the output it produced as feedback to the students. Students, on the other hand, were presumably judging the system primarily just on the basis of the feedback it produced for them individually. Second, the priorities of the teachers were different. While students were interested in the content of the feedback, teachers, as we can see from the remarks cited earlier, were interested in the ease of use and speed of the system and the ways in which it could save them effort.

Finally, we found that, on two parallel questionnaire items, the experts emerged as significantly more positive than the students about the merits of the expert system assessment. It can be concluded that the findings concerning the experts' and students' preferred approach support the *Observe Portal* in H4, that *Students and experts prefer the Observe Portal's assessment feedback over and above that yielded from human experts.* However, the support for this from the experts was considerably stronger than that from the students.

Hypothesis 5

Finally, hypothesis 5 focused on assessing the students' and experts' acceptance of the *Observe Portal* system, as defined by the core variables of the TAM model (6.1). This was evaluated in terms of two dimensions: the student acceptance and expert acceptance. Both were measured via questionnaire items targeting perceived usefulness, perceived ease of use and the intention to use the expert system when available.

The data was purely quantitative and both students and experts made significantly positive responses to all the items in all the categories. This was except for one item under ease of use – as responded to by the experts. Furthermore, the results from the students' questionnaire implied that ease of use was not affected by any of the student background variables (computing expertise, virtual worlds experience, intelligent environment knowledge, programming experience). In summary, the results derived from both the expert and the student participants' responses clearly revealed strong acceptance of the use of the *Observe Portal* assessment system (H 5).

8.3.Applications

The MIVO framework and its models, as discussed in this thesis, can be applied in any other learning environment that requires students to collaborate together to solve problems. However, some of the following should be taken into account when applying the research framework and prototype in such a way:

1- The scalability of the virtual environment has been tested in [52], and hence it has been shown to work successfully with more than 10 students, however it is been found that the fewer the group members, the better the group performance and the learning outcomes that these groups produce [52]. Thus, the research evaluation experiments were limited to the use of groups of two or four students, and the experiments were not applied to larger groups because it was felt that the students would get confused if asked to rate more than three other students during the time in which they worked together. Although it is possible for many students to use the *Observe Portal* at the same time, the rating method can be scaled up to groups of four students.

2- The outcomes of the current research can be applied to any educational institution and can thus readily be applied to my sponsor university (Umm AlQura University) in Saudi Arabia. The current research platform (*Observe Portal*) is able to help teach students about the topic of smart houses, in computer science classes, and can help to assess any necessary group collaborative work.

The following section discusses the limitations inherent in the research platform that should be considered when applying it.

8.4.Limitations

There were some limitations in the research framework and the prototype developed according to it; they resulted from the restricted time available for carrying out the research. These limitations were:

1- The first design limitation was that the *Observe Portal* (Chapter 4) did not permit teachers to customise the learning assessments by choosing which OLens lenses they wanted to apply within the learning sessions. In addition, there were no facilities which allowed teachers to introduce additional new lenses and to specify the rules for such lenses.

2- Another limitation, perhaps, was in the design of the classified chat window classification buttons. Some users claimed that the chat buttons were awkward to use and unnecessary. However, the chat interface was designed in this way in order to enable the classification of student sentences and then to assist the system in judging the students' collaboration. The automatic classification of such "chats" would have been a research project in itself.

3- The 3-points rating (low-middle-high) limited the students in terms of their ability to rate their peers. Indeed, some students suggested widening the rating scale to one of five points instead to allow them to both criticise and praise their peers more accurately. However, widening the scale might well have overloaded the students' thinking processes, since, as it was, they were already being expected to undertake many tasks at the same time. In addition, the main purpose of implementing the rating system was to provide the necessary information to the system so that it could give better assessments than it otherwise would have been able to. Further rating gradations were unnecessary in terms of this requirement.

4- The prototype was limited in its assessment abilities because it only provided the students with quantitative assessment, using charts, and produced no qualitative feedback. Such feedback would have been useful for the students in terms of learning more about their performance.

5- MIVO was utilised here for a specific learning context - how to program a smart house using simple rules made available via a 3D virtual world; this limited context was employed in order to demonstrate the validity of the overall vision of the framework. Applying it in another learning context would be beneficial in relation to the general research area.

8.5. Chapter Summary

This chapter has discussed the findings of the experiments undertaken to evaluate the assessment system, and their wider implications for the research area. Overall, the value of the *MIVO* framework and the components of this which were employed in the proof-of-concept prototype was supported in terms of it being a suitable computational framework for supporting the assessment of collaborative learning in VWs. In addition, the empirical results support the effectiveness of the *Observe Portal* assessment system in terms of it collecting the learning evidence from collaborative distance-learners working in a virtual world environment and providing effective feedback to those learners.

Looking Table 7-1 we can see that most of our hypotheses received some support and that this support was particularly strong in the case of H5 - acceptance based on the TAM model. Nevertheless, some valuable points were made by participants in the open response data which prompt the examination of a number of improvements which could be made; these will be considered in the final chapter.

The next and final chapter concludes this thesis with an account of the implications of the findings for the further development of our proposed expert assessment system. It also concludes with a consideration of the remaining challenges in the field and suggestions for future research work in this area.

Chapter 9

9. Concluding Remarks

"It always seems impossible until it's done."

– Nelson Mandela

The motivation of this research was to create a novel computational framework that enhances the observing and assessing of collaborative learning in immersive environments. It focuses more on providing methods for identifying and classifying learning evidence and assessing group working then on mapping all these elements to an appropriate learning design. In order to achieve this, the study includes a detailed literature review concerning virtual learning environments and learning assessment and the integrating of these with the use of fuzzy logic and multi-agent approaches – in order to enhance the assessment of learners (Chapter 2). In line with this literature, the present research proposes an innovative computational framework called the Mixed Intelligent Virtual Observation (*MIVO*) framework which includes the Virtual Observation Lenses (*OLens*) and the Mixed Agents model (*MixAgent*) (Chapter 3); this framework can be used to support the assessment of collaborative students' learning within virtual worlds. The *MIVO* framework, along with the *OLens* and *MixAgent* models, has been implemented here within a proof-of-concept prototype called the *Observe Portal* system (Chapter 4). This prototype utilised fuzzy logic methods to imitate expert-based evaluations; the integration of the fuzzy logic system into the *Observe Portal* is described in Chapter Five. The research also employed collaborative learning activities for the evaluation of the effectiveness of the models (Chapter 6), reporting on these evaluations and the analyses of them in Chapter 7. The outcomes of these evaluations are discussed in Chapter 8. Finally, this final chapter (Chapter 9) includes a summary of the thesis and explains the possible future directions which this research work could take - and the attendant challenges.

9.1.Summary of Achievements

The research hypotheses were defined in Chapter 1, section 1.3. To underline the themes of the research hypotheses, this thesis describes empirical user-based experiments involving both students and experts who all participate in a number of educational sessions within a virtual world. Each session used a group of two students to imitate collaboration between learners and one computer science expert to imitate classroom teacher observation. Also, these experiments employed a functional prototype (the *Observe Portal*) which was implemented as a proof-of-concept of the *MIVO* framework. The *Observe Portal* prototype, as described in Chapter 4, applied the *MixAgent* model – which involves natural and software agents both playing a central role in the gathering of learning evidence in real-time to assess student performance. Furthermore, the system utilised the *OLens* model which uses a fuzzy learning system (created for this research). The *OLens* model comprises four lenses: event detection, learning interactions, student success and performance outcome. To evaluate students via each lens, data from the *MixAgent* model was processed employing a fuzzy reasoning approach as a means to combine the data generated from the agents; learners' performance was assessed using the data representing the inferred learning outcomes obtained from the collaborative activities involving individual students and groups (Chapter 5). In consequence, the *MixAgent* and *OLens* models were applied to real-time collaborative learning activities within the *Observe Portal*, and the results from the observation lenses were presented to the users via dashboard interfaces and video recordings.

The learning activities in the immersive environment operate using the educational model relating to hands-on laboratories and are based on the educational paradigms associated with these. These activities were designed to allow students to collaborate and work together to control virtual smart home appliances, as explained in Chapter 6. The experiments also involved experts who were asked to evaluate students manually and observe their (the students') work; subsequently, the system's assessment was compared with these experts' assessments. The results from the user experiments illustrated that the use of the *MIVO* computational framework supports the proficient collection of learning evidence from distance-learning collaborative students for assessments by enabling the integration of software and natural agents to enhance such assessments.

In addition, the statistical results showed that the learners had, in general, positive experiences of the research prototype. The findings derived from the natural agents' perceptions evidenced that the learners had a generally positive attitude to working as natural agents and to evaluating the other student(s) through the rating tool; indeed, the rating logs showed that learners rated each other several times while working on the tasks. With respect to the peer ratings, a three-point scale is the most practicable. However, we would add a further pedagogical dimension by offering a space for students to make their own, open, qualitative comments and to offer advice to their peers on how to improve their (peers') performance. It would not be obligatory to complete this, but the facility would be available for those who wanted to use it; this is as suggested in the users' comments. In addition to the rating tool, the classified chat window which offers several chat-classification options via its buttons was considered easy to use and students strongly agreed that they achieved good communication within their groups. Also, the analysis of the chat logs indicated that most of the participants categorised their sentences correctly via the chat buttons. All these results support the thesis hypothesis 1: *Users express positive attitudes towards their roles as human agents when performing distance-learning tasks in the virtual world.*

This research also assessed the effectiveness of the *MIVO* framework by measuring whether the *Observe Portal* system provides distance collaborative learners with valued assessment feedback in terms of helping them to understand their weaknesses and strengths. To do so, the research evaluates the students' assessment experience and the experts' reflection on the assessment. Generally, the evaluation of the students' assessment experience concluded that the majority of the students had positive experiences and agreed the assessment had a significant impact on their ability to understand their performance. Furthermore, the experts' reflections on the assessments made were encouraging, and these experts agreed that the assessments yielded from the system were reliable, valid and useful for helping students to improve their performance. These findings supported the effectiveness of our *MIVO* framework in supporting proficient assessment and they also supported hypothesis 2, *that the Observe Portal system provides collaborative distance learners with assessment feedback, and users report positive experiences of such assessment feedback.*

An important goal for the virtual observation framework and prototype was that it should improve the effectiveness of the assessing of collaborative students as compared to experts' traditional forms of observation, as proposed. This aspect was assessed by obtaining the expert assessment experience measurement and comparing an evaluation of the system's observations with the expert observations. The results showed that although some experts stated that assessing students effectively, manually, was an easy task, many of them could not complete the assessment sheets, especially as regards assessing the levels of collaborative skill exhibited. In contrast, the system was able to present a complete set of user assessment results back to the users. The experts found that assessing individual students operating in *Observe Portal* was easy when they only had to take into account the activities of just two students; however, they also stated that observing more than two students interacting could be a very difficult task, tracking many users and evaluating them at the same time was confusing to them. They also realised that it was very time-consuming and inconvenient to spend time simply watching the learners operate in the virtual world. Furthermore, a comparison between the system's assessments and the experts' assessments showed that there were no significant differences between these on any of the measures except one measure (the success of task 1). However, a longer training period for the participants would have been useful with regard to the evaluation experiments - so that the students might have understood how to use the system better. This might help to bring the results of the expert evaluations closer to those of the system evaluations, especially in regard to the first tasks. Overall, the system and the expert assessments were closely matched apart from one measure. These results therefore (with some minor reservations) support the main hypothesis 3, *the Observe Portal provides assessments that are very similar to human-expert assessments; these Observe Portal assessments are produced using less effort overall.*

The experiments also evaluated the users' preferences in terms of the assessment approach. The strategy as regards this was to ask both learner and expert participants who experienced both assessment approaches - in the second phase of the experiment (phase 2 (condition 2.2), as described in Chapter 6) - to express their preferences. The subsequent analysis showed that the majority of the participants believed that the *Observe Portal* system had a significant advantage over traditional methods (teacher observation) in relation to assessing students’ performance in the virtual world. Such findings evidenced that in many respects the participants much preferred the system’s to the teachers’ assessments. However, in addition, the results highlighted the interesting finding that the human experts preferred the system’s assessments more than the students did. Overall, it can be concluded that the findings derived from the experts’ and the students’ expressed preferences in terms of approach hypothesis 4- *students and experts prefer the Observe Portal’s assessment feedback over and above that yielded from human experts*. All these results give positive support to the computational framework proposed here and the prototype used to implement it.

Finally, this research studied the importance of assessing the users’ acceptance when implementing a project such as the *Observe Portal* system. This (the user acceptance) was evaluated by assessing both the students’ levels of acceptance and the experts’ levels of acceptance. In summary, most of the student and expert participants agreed that the system was easy to use, useful and indeed that they would use it where it was available. Participants positively accepted the use the *Observe Portal* as a tool for assessing students and groups’ performance. These findings advocate for hypothesis 5, *that students and experts express their acceptance of using the Observe Portal assessment system*.

It is important to be able to measure the performance of individual students because it is only by doing so that one can determine whether a student has achieved the desired learning objectives. Such an approach is obviously also highly valuable to teachers in relation to reviewing their learners’ work and then helping them to enhance their learning activities within the VW based on their (the students) performance. In addition, the feedback generated by the system helped to show to the students themselves their own weakest areas so that they could work at them and improve their overall performance.

9.2. Contributions

The first, main contribution of this research is the proposing of a novel computational framework that synthesizes and integrates learning theories and computational models related to the observing and assessing of collaborative learning in VWs. This innovative framework contains within it two observation models: the Observation Lenses model (*OLens*) and the Mixed Agents model (*MixAgent*) (*Chapter 3*). *OLens* is a novel observation model for identifying and structuring evidence of student learning in 3D virtual environments as part of an e-learning assessment process. *MixAgent* is a computational model that integrates software and natural agents in order to implement a mechanism for collecting learners' data and inferring from these data to identify learning evidence and to assess the learning achieved by groups and individuals.

There is extensive coverage in the empirical literature of the merits of appraising students in real world classrooms; however, there is a lack of research concerning the observing and assessing of students who are operating within virtual worlds. Thus, this computational framework overcomes this limitation, exploits the affordances of 3D virtual worlds and allowed for an investigation into the ways in which students can be assessed, in VWs - in order to enhance their learning effectiveness. In particular, it dealt with the challenges of gathering learning evidence and analysing learning events to measure the quality and quantity of learning outcomes, and it addressed the limitations involved with the collecting of learning evidence by technology and the collecting of such evidence by human experts. Such a mechanism can be utilised as the basis of assessing learning in immersive environments.

The second contribution of this thesis is the construction of a proof-of-concept prototype (*Observe Portal*) which implements the proposed framework. This prototype employed the virtual observation components, assessing students' performance in real-time from different perspectives based on a pedagogical framework (*Chapter 4 and Chapter 5*). The assessments involved evaluating collaborative learners as groups and as individuals; this included the evaluation of learning interactions, learning success and collaborative skills.

Finally, the thesis contributed to the research area by presenting empirical research findings from the evaluation of the prototype which demonstrated the effectiveness of the approaches and the models used in relation to collaborative learning activities in 3D virtual environments, and compared them (favourably) to the equivalent traditional approaches *(Chapter 6 and Chapter 7*).

The experimental findings have demonstrated an expansion of the affordances of assessment in 3D VWs. These have advanced our knowledge of virtual assessment by showing that it is possible to apply existing educational frameworks to immersive environments and to harness the power of human observation allied with technologically-based observation to support the gathering of high quality learning evidence, in quantity, from collaborative learning activities; this was in order to provide valuable feedback so that educationalists can understand the students' weaknesses and strengths from a number of differing perspectives and then improve their (the students') performance.

Moreover, this research also delivered the following secondary (in terms of being less important) contributions:

1- The design of assessment interfaces driven by learners' performance (to assess students taking part in collaborative tasks).

2- The creation of hierarchical fuzzy logic systems which emulate human reasoning as expressed in 3D virtual environments.

3- The encoding of a collaborative problem-solving skills taxonomy [14] and applying this in a 3D virtual environment to assess student collaborative skills.

4- The introduction of instruments for measuring the effectiveness of the natural agent tools; these instruments measure, specifically, the perception of the chat communication (COMM) subsystem and the perception of the natural agent rating procedure (NA).

5- The use of instruments for measuring the experts' assessment experience (EXP) and the experts' reflections on the effectiveness of the assessment methods (REF).

9.3. Future Work

1- The evaluation of *MIVO* was carried out over a relatively short period of time. Undertaking a larger scale (in terms of numbers of learners) and longer duration evaluation might yield significant results, demonstrating further the model's suitability (or otherwise) for improving learning and the assessment of learning in immersive environments.

2- Producing an instructor interface that enables teachers to configure the assessment lenses and properties, define learning activities, decide assessment characteristics and choose the type of assessment feedback required potentially represents an area for future research.

3- A future study could generalise the MIVO framework and its models to work with other learning activities and apply them in other types of teaching environments, such as mixed reality environments; this represents an open research topic. Additional challenges might emerge if these models were to be applied to other settings and contexts.

4- A further research direction would be to apply natural language processing (NLP) approaches to analyse the users' conversations in the chat window and so do away with the need to have multiple buttons to classify students' dialogues. NLP has been used in many studies in order to process and analyse large amounts of human language data from a variety of different sources such as social media and games [209-211]. Applying NLP to the students' chat in order to classify student conversations would enable users to concentrate more on their own work rather than having to give attention to the question of which chat buttons they must use to classify their sentences. Thus, AI-based methods could help in gathering more learning evidence from the collaborative learners.

5- It is important to consider assessing other attributes and skills relating to collaborative work in distributed learning environments. Assessing other learning skills, not just the collaborative skills, and so add more assessment lenses to the *OLens* model to evaluate students from different points of view, would enhance the learning process offered and support the educators to evaluate the learners' work in more depth and according to additional perspectives; this is another possible future research direction.

6- The Corona Virus Pandemic (COVID-19) has demonstrated the importance that technologies such as our virtual learning environment may have in the future, for both teaching and assessment. In many conceivable circumstances, students may not be able to physically attend places of learning and in such circumstances many educators will use online video calling applications for teaching. However, most educators and students have claimed that video calls/conferences are not as effective as physical classes because of the loss of attention and enthusiasm which is endemic with this type of online learning. The application of 3D environments provides more interesting scenarios whereby students can learn. However, developing immersive systems such as virtual reality (VR) or augmented reality (AR) environments that can provide a wide range learning content and activities and still support the accurate assessment of learners will require a great deal of development and programming. A future study could usefully look at the possibility of flexible and easy creation of immersive environments by educators, customising systems for their own learning purposes, and how to make building a new 3D lesson as easy as producing a set of presentation slides.

Appendix A: Experiment Instruments

A.1. Student Preliminary Survey

Constructs:

• General information (GI)
• Computers use (CU)
• Programming experience (PE)
• Video games experience (VGE)
• Virtual worlds experience (VWE)
• Intelligent environments knowledge (IE)

Table A.1: Students Preliminary Survey

Code	Question	Response scale
GI -1	Full Name	- [open ended]
GI -2	Email	- [open ended]
GI -3	Gender	- Male / - Female
GI -4	Age	- [open ended]
GI -5	Nationality	- [open ended]
GI -6	Level of English	- Elementary proficiency - Limited working proficiency - Professional working - Full professional proficiency - Native or bilingual proficiency
GI -7	Level of studies	- First-year undergraduate - Second-year undergraduate - Third-year undergraduate - Postgraduate(Master/PhD)
GI -8	Subject of studies	- [open ended]
CU -1	Do you own a personal computer?(laptop desktop)	- Yes - No
CU -2	[If PRE-2 answer = YES] Which are the main uses you give to your personal computer?	- Leisure - Studies - Social interactions - Paid work - Other
CU -3	What do you consider your computing expertise?	- Beginner - Intermediat - Expert
PE-1	How experienced are you in programming?	- Beginner - Intermediat - Expert
VGE -1	How often do you play video games per week?	- Not at all - Once or twice per week - 4-5 times per week - Every day

VGE -2	If you play video games please name the ones you use	- [open ended]
VWE-1	Are you familiar with virtual worlds?	- Yes - No
VWE-2	[IF VW-1 answer = YES] How often do you use virtual worlds?	- Not at all - Once or twice per week - 4-5 times per week - Every day
VWE-3	Please select the virtual worlds that you have used or heard of	- OpenSim - Second Life - RealXtend - Open Wonderland - Meshmoon - IMVU - Club Penguin - Habbo - Other Option
IE-1	Are you familiar with smart houses/intelligent spaces?	- Yes - No
IE-2	Have you ever been involved in doing practical activities/ assignments in a computer engineering lab?	- Yes - No
IE-3	Have you used or heard of technology to make your house "smart"?	- Yes - No
GW-1	Do you like to work in groups?	- Yes - No
GW-2	Why did you choose this answer in Q-22?	- [Open-Ended]
GW-3	Rate your collaborative skill level	- I have High collaborative skill - I have Middle collaborative skill - I have Low collaborative skill

A.2. Experts Preliminary Survey

Constructs:

• General information (GI)
• Experience in using educational technology (ET)
• Student Assessment (SA)
• Personal innovativeness (PI)
• Virtual worlds experience (VWE)
• Intelligent Environments knowledge (IE)

Table A.2: Experts preliminary survey

Code	Question	Response scale
GI -1	Full Name	- [open ended]
GI -2	Email	- [open ended]
GI -3	Gender	- Male/ - Female
GI -4	Age	- [open ended]
GI -5	Nationality	- [open ended]
GI -6	Level of English	- Beginner - Intermediate - Expert
GI-7	Level of education	- Master's Degree - First-year PhD - Second-year PhD - Third-year PhD - Fourth-year PhD - PhD Degree
GI-8	Major	- [open ended]
GI-9	What subjects do you teach?	- [open ended]
GI-10	Teaching Experience	- Less than one year - 1-5 years - 6-10 years - 11-15 years - 16-20 years - 20+ years
GI-11	Have you taught/observed students in computer engineering labs?	- Yes - No
SA-1	How do you usually assess students' learning?	- Tests - Assignments - Group Projects - Individual projects - Observation - All the above - Other (please specify [ ])
SA-2	Do you assign group projects and group activities to students?	- Yes - No
SA-3	[If SA2 answer=yes] How do you assess the groups?	- Self-evaluation - Peer evaluation - Final product evaluation - Teacher evaluation - All the previous

		- Other (please specify [ ])
ET-1	Do you use technology (software or hardware) in your classes or modules?	- Yes - No
ET-2	Do you use educational software that helps in teaching?	- Yes - No
ET-3	[IF ET-1 OR ET-2 answer = YES] Please write the name(s) of the software you have used?	- [open ended]
PI-1	I like using new technologies in teaching.	- Strongly agree - Agree - Disagree - Strongly Disagree
PI-2	I can use and understand new technologies quite easily.	- Strongly agree - Agree - Disagree - Strongly Disagree
VWE-1	Are you familiar with virtual worlds?	- Yes - No
VWE-2	[IF VWE-1 answer = YES] How often do you use virtual worlds?	- Not at all - Once or twice per week - 4-5 times per week - Every day
VWE-3	Have you used virtual worlds for teaching?	- Yes - No
VWE-4	Please select the virtual worlds that you have used.	- OpenSim - Second Life - RealXtend - Open Wonderland - Meshmoon - IMVU - Club Penguin - Habbo - Other Option
IE-1	Are you familiar with smart houses/intelligent spaces?	- Yes - No
IE-2	Have you used or heard of technology to make your house "smart"?	- Yes - No

A.3. Participants Information Sheet

Certification

I, (Samah Felemban) certify that the details of this project have been fully explained and described and my contact details have been provided to the participants for their replies, communications or inquiries.

The purpose of this study

The research is about applying a computational framework to enhance the effectiveness of assessing and observing collaborative students in 3D virtual worlds compared to traditional classroom observation. Also, it provides distance learners with effective feedback to better understand the students' and groups' weaknesses and strengths. This research evaluates the effectiveness of the used assessment approaches through measuring learners and expert's experiences when using the 3D virtual environments and their acceptance of the assessment results.

Do I have to take part?

Your participation is voluntary. I would like you to consent to participate in this study as I believe that you can make an important contribution to the research. If you do not wish to participate, you do not have to do anything in response to this request. Asking you to take part in the research because I believe you can provide important information to the research evaluation that I am undertaking.

What will I do if I take part?

If you are an adult and happy to participate in the research, I will ask you to read this information sheet and sign the consent form. When we receive this you will be asked to enrol yourself in simple learning tasks (writing rules for smart home sensors) either remotely or locally. When you finish all the tasks, you will be asked to fill in a survey. Then, the researcher will contact you to discuss your participation in the next experiment if needed.

What are the possible benefits of taking part?

Whilst there may be no personal benefits to your participation in this study, the information you provide can contribute to the future development of e-learning when using up-to-date immersive technologies.

Will my taking part in the study be kept confidential?

All information you provide to us will be kept confidential. The participant data and survey data will be stored separately on a secure database. Only the researcher will have access to it. Providing your personal information is optional.

What will happen to the results of the research study?

All information provided by you will be stored anonymously on a computer with analysis of the information obtained undertaken by the researcher based at the University of Essex. The results from this analysis will be available in one or more of the following sources; the researcher PhD thesis, scientific papers in peer-reviewed academic journals, presentations at conferences or local seminars.

For more information, please contact Samah Felemban (ssyfel@essex.ac.uk).

CONSENT FORM

Title of the Project: Towards Recognising Learning Evidence in Collaborative Virtual Environments

Researcher: Samah Felemban

Notice: If you are an adult (aged 18 and more), please sign this constant form and continue filling the following surveys. If not, please don't participate in this research experiment.

	Please check the appropriate boxes	YES	NO
1.	**I confirm that I** have read and understood the Information Sheet dated for the above study		
2.	I understand that my participation is voluntary and that I am free to withdraw from the project at any time without giving any reason and without penalty.		
3.	I agree to take part in the project. Taking part in the project will include being interviewed and recorded (video).		
4.	I understand that my words (anonymised) may be quoted in publications, reports, web pages, and other research outputs.		
5.	I understand that the collected survey data and users' data provided will be securely stored and accessible only by the researcher involved in the study and that confidentiality will be maintained.		
6.	I understand that the analysis of the data collected in this study will be used as appropriate and for publication of findings, in which case data will remain completely anonymous.		
7.	I agree the researcher may contact me.		
8.	I agree to take part in the above study.		

Participant Name: **Date:** **Signature:** **Email:**

Researcher Name: **Date:** **Researcher Signature:**

A.4. Experts Manual Sheets

A.4.1 Social collaborative problem-solving skill sheet

Element	Indicator	Low	Middle	High
Participation				
Action	Activity within environment	No or very little activity	Activity in familiar contexts	Activity in familiar and unfamiliar contexts
Interaction	Interacting with, prompting and responding to the contributions of others	Acknowledges communication directly or indirectly	Responds to cues in communication	Initiates and promotes interaction or activity
Task completion/ perseverance	Undertaking and completing a task or part of a task individually	Maintains presence only	Identifies and attempts the task	Perseveres in task as indicated by repeated attempts or multiple strategies
Perspective taking				
Adaptive responsiveness	Ignoring, accepting or adapting contributions of others	Contributions or prompts from others are taken into account	Contributions or prompts of others are adapted and incorporated	Contributions or prompts of others are used to suggest possible solution paths
Audience awareness (Mutual modelling)	Awareness of how to adapt behaviour to increase suitability for others	Contributions are not tailored to participants	Contributions are modified for recipient understanding in the light of deliberate feedback	Contributions are tailored to recipients based on interpretation of recipients' understanding
Social regulation				
Negotiation	Achieving a resolution or reaching compromise	Comments on differences	Attempts to reach a common understanding	Achieves resolution of differences
Self evaluation (Metamemory)	Recognising own strengths and weaknesses	Notes own performance	Comments on own performance in terms of appropriateness or adequacy	Infers a level of capability based on own performance
Transactive memory	Recognising strengths and weaknesses of others	Notes performance of others	Comments on performance of others in terms of appropriateness or adequacy	Comments on expertise available based on performance history
Responsibility initiative	Assuming responsibility for ensuring parts of task are completed by the group	Undertakes activities largely independently of others	Completes activities and reports to others	Assumes group responsibility as indicated by use of first person plural

A.4.2 Learning Interactions Sheet

Learning Interactions				Rate Individual Interaction
Student Name	0 - 5 min	5 - 10 min	10 - 15 min	
				High - Middle - Low
				High - Middle - Low

Group Learning Interactions				Rate Group Interaction
Student Name	0 - 5 min	5 - 10 min	10 - 15 min	
				High - Middle - Low
				High - Middle - Low

A.4.2 Task Success Sheet

Task Success				**Rate Individual Success**
Student Name	**0 - 5 min**	**5 - 10 min**	**10 - 15 min**	
				High - Middle - Low
				High - Middle - Low

Group Task Success				**Rate Group Success**
Student Name	**0 - 5 min**	**5 - 10 min**	**10 - 15 min**	
				High - Middle - Low
				High - Middle - Low

A.5. Student Post-Questionnaires

A.5.1. Student post-questionnaire – Phase 1 (the physical classroom)

Constructs:

- Perceive collaboration and communication (COMM)
- Assessment Experience: Quantity and timing of feedback (QTF), Quality of feedback (QF), Doing with the feedback (UF).
- User interface (GUI)

Table A.5.1: Student post questionnaire – Phase 1 (the physical classroom)

Code	Question	Response scale			
COMM1	I found difficulties when communicating with the other student(s)	Strongly Agree	Agree	Disagree	Strongly Disagree
COMM2	It was comfortable to communicate with the other student(s)	Strongly Agree	Agree	Disagree	Strongly Disagree
COMM3	Explain the reasons why it was comfortable (or not) to communicate with the other student(s)	[open ended]			
COMM4	How would you rate your experience of collaborating with students in the same location?	Very Poor	Poor	Good	Very Good
COMM5	Please provide any extra comment you have on your experience working with the other student(s) in the experiment.	[open ended]			
GUI1	Using the programming interface while working with the other student(s) was EASY	Strongly Agree	Agree	Disagree	Strongly Disagree
GUI 2	Using the programming interface while working with the other student(s) was FUN	Strongly Agree	Agree	Disagree	Strongly Disagree
GUI 3	Using the programming interface while working with the other student(s) was USEFUL	Strongly Agree	Agree	Disagree	Strongly Disagree
GUI 4	Using the programming interface while working with the other student(s) was INTERESTING	Strongly Agree	Agree	Disagree	Strongly Disagree
GUI 5	Using the programming interface while working with the other student(s) was DIFFICULT	Strongly Agree	Agree	Disagree	Strongly Disagree
GUI 6	Using the programming interface while working with the other student(s) was ANNOYING	Strongly Agree	Agree	Disagree	Strongly Disagree
GUI 7	Using the programming interface while working with the other student(s) was BORING	Strongly Agree	Agree	Disagree	Strongly Disagree
GUI 8	Explain the reasons why it was comfortable (or not) using the	[open ended]			

	programming interface with other student(s)				
GUI 9	Do you have any additional comments about the overall experience when rating others in the virtual world?	[open ended]			
QTF1	On this activity, I get plenty of feedback from the system on how I did.	Strongly Agree	Agree	Disagree	Strongly Disagree
QTF2	The feedback comes back very quickly.	Strongly Agree	Agree	Disagree	Strongly Disagree
QTF3	There is hardly any feedback on my performance when I finish.	Strongly Agree	Agree	Disagree	Strongly Disagree
QTF4	I would learn more if I received more feedback.	Strongly Agree	Agree	Disagree	Strongly Disagree
QTF5	Whatever assessment I get comes too late to be useful.	Strongly Agree	Agree	Disagree	Strongly Disagree
QF1	The expert assessment mostly tells me how well I am doing in relation to others	Strongly Agree	Agree	Disagree	Strongly Disagree
QF2	The assessment from the expert is useful to understand the individual and the group INTERACTION LEVEL in the virtual world	Strongly Agree	Agree	Disagree	Strongly Disagree
QF3	The assessment from the expert is useful to understand the individual and the group SUCCESS LEVEL in the virtual world	Strongly Agree	Agree	Disagree	Strongly Disagree
QF4	The assessment from the expert is useful to understand the individual and the group COLLABORATIVE SKILLS LEVEL	Strongly Agree	Agree	Disagree	Strongly Disagree
QF5	The assessment from the expert helps me to understand things better	Strongly Agree	Agree	Disagree	Strongly Disagree
QF6	The assessment from the expert shows me how to do better next time	Strongly Agree	Agree	Disagree	Strongly Disagree
QF7	Once I have read the assessment, I understand what I did	Strongly Agree	Agree	Disagree	Strongly Disagree
QF8	I don’t understand some of the expert assessment .	Strongly Agree	Agree	Disagree	Strongly Disagree
QF9	I understand what the assessment is saying.	Strongly Agree	Agree	Disagree	Strongly Disagree
UF1	I use the assessment feedback to go back over what I have done in the learning activity.	Strongly Agree	Agree	Disagree	Strongly Disagree
UF2	I think the assessment will help me with any subsequent activity.	Strongly Agree	Agree	Disagree	Strongly Disagree
UF3	I do not use the assessment for revising.	Strongly Agree	Agree	Disagree	Strongly Disagree
UF4	I tend to only read the marks.	Strongly Agree	Agree	Disagree	Strongly Disagree

A.5.2 Student Post Questionnaire – Phase 2 (Condition 2-1 (virtual classroom with system observation))

Constructs:

- Perceive collaboration via chatting tool (COMM)
- Perceive rating as natural agent (NA)
- Assessment experience: quantity and timing of feedback (QTF), quality of feedback (QF), doing with the feedback (UF)
- Student acceptance of the system (SA): perceived usefulness (PU), perceived ease of use (PEOU), intension to use (IU)

Table A.5.2: Student Post Questionnaire – Phase 2 (Condition 2-1)

Code	Question	Response scale			
Section 1: Natural Agents (COMM & NA)					
COMM1	I found difficulties when communicating with the other student(s) via the multi-buttons chat window	Strongly Agree	Agree	Disagree	Strongly Disagree
COMM2	It was comfortable to communicate with the other student(s) through the virtual interface (i.e. using the chat window with classifying buttons)	Strongly Agree	Agree	Disagree	Strongly Disagree
COMM3	Explain the reasons why it was comfortable (or not) to communicate with the other student(s) through the chat window?	[open ended]			
COMM4	How would you rate your experience of collaborating with students in other location(s) using the classified chat?	Very Poor	Poor	Good	Very Good
COMM5	Please provide any extra comment you have on your experience working with the other student(s) in the virtual world.	[open ended]			
NA1	Using the rating tools to evaluate the other student(s) while working or at the end were EASY	Strongly Agree	Agree	Disagree	Strongly Disagree
NA2	Using the rating tools to evaluate the other student(s) while working or at the end were FUN	Strongly Agree	Agree	Disagree	Strongly Disagree
NA3	Using the rating tools to evaluate the other student(s) while working or at the end were USEFUL	Strongly Agree	Agree	Disagree	Strongly Disagree
NA4	Using the rating tools to evaluate the other student(s) while working or at the end were INTERESTING	Strongly Agree	Agree	Disagree	Strongly Disagree
NA5	Using the rating tools to evaluate the other student(s) while working or at the end were DIFFICULT	Strongly Agree	Agree	Disagree	Strongly Disagree
NA6	Using the rating tools to evaluate the other student(s) while working or at the end were ANNOYING	Strongly Agree	Agree	Disagree	Strongly Disagree

NA7	Using the rating tools to evaluate the other student(s) while working or at the end were BORING	Strongly Agree	Agree	Disagree	Strongly Disagree
NA8	Explain the reasons why it was comfortable (or not) to rate the other student(s) through the rating tools	[open ended]			
NA9	Do you have any additional comments about the overall experience when rating others in the virtual world?	[open ended]			
Section 2: Assessment Experience					
QTF1	On this activity, I get plenty of feedback from the system on how I did.	Strongly Agree	Agree	Disagree	Strongly Disagree
QTF2	The system assessment comes back very quickly.	Strongly Agree	Agree	Disagree	Strongly Disagree
QTF3	There is hardly any feedback on my performance when I finish.	Strongly Agree	Agree	Disagree	Strongly Disagree
QTF4	I would learn more if I received more feedback from the system.	Strongly Agree	Agree	Disagree	Strongly Disagree
QTF5	Whatever assessment I get comes too late to be useful.	Strongly Agree	Agree	Disagree	Strongly Disagree
QF1	The system assessment mostly tells me how well I am doing in relation to others	Strongly Agree	Agree	Disagree	Strongly Disagree
QF2	The assessment from the system is useful to understand the individual and the group INTERACTION LEVEL in the virtual world	Strongly Agree	Agree	Disagree	Strongly Disagree
QF3	The assessment from the system is useful to understand the individual and the group SUCCESS LEVEL in the virtual world	Strongly Agree	Agree	Disagree	Strongly Disagree
QF4	The assessment from the system is useful to understand the individual and the group COLLABORATIVE SKILLS LEVEL	Strongly Agree	Agree	Disagree	Strongly Disagree
QF5	The assessment from the system helps me to understand things better.	Strongly Agree	Agree	Disagree	Strongly Disagree
QF6	The assessment from the system shows me how to do better next time.	Strongly Agree	Agree	Disagree	Strongly Disagree
QF7	Once I have read the assessment from the system, I understand what I did.	Strongly Agree	Agree	Disagree	Strongly Disagree
QF8	I don’t understand some of the system assessment.	Strongly Agree	Agree	Disagree	Strongly Disagree
QF9	I understand what the assessment is saying.	Strongly Agree	Agree	Disagree	Strongly Disagree
UF1	I use the assessment feedback to go back over what I have done in the learning activity.	Strongly Agree	Agree	Disagree	Strongly Disagree
UF2	I think the system assessment will help me with any subsequent activity.	Strongly Agree	Agree	Disagree	Strongly Disagree
UF3	The system assessment prompts me to go back over video material recorded earlier to understand the scores.	Strongly Agree	Agree	Disagree	Strongly Disagree

UF4	I tend to only read the marks from the system without watching the recorded video.	Strongly Agree	Agree	Disagree	Strongly Disagree
Section 3: Student acceptance of the system (SA)					
PU1	The use of *Observe Portal* is useful for assessment in 3D-VW collaborative activities	Strongly Agree	Agree	Disagree	Strongly Disagree
PU2	The use of the *Observe Portal* system is not suitable for assessing students performance.	Strongly Agree	Agree	Disagree	Strongly Disagree
PU3	The use of the *Observe Portal* system allows me to get a deeper understanding of the individual and the group performance.	Strongly Agree	Agree	Disagree	Strongly Disagree
PU4	I find that *Observe Portal* is useful in assessing students interactions, success, and collaborative skills.	Strongly Agree	Agree	Disagree	Strongly Disagree
PU5	I don't see that *Observe Portal* makes any difference in assessing students interactions, success, and collaborative skills.	Strongly Agree	Agree	Disagree	Strongly Disagree
PEOU1	I find the *Observe Portal* is easy to use in virtual world collaborative activities.	Strongly Agree	Agree	Disagree	Strongly Disagree
PEOU2	The feedback obtained from the system is clear and understandable.	Strongly Agree	Agree	Disagree	Strongly Disagree
PEOU3	It is difficult to use the *Observe Portal* assessment system.	Strongly Agree	Agree	Disagree	Strongly Disagree
PEOU4	Students assessment through *Observe Portal* is easy	Strongly Agree	Agree	Disagree	Strongly Disagree
IU1	I would use *Observe Portal* to assess my performance.	Strongly Agree	Agree	Disagree	Strongly Disagree
IU2	I would use *Observe Portal* to understand my performance.	Strongly Agree	Agree	Disagree	Strongly Disagree
IU3	Assume that I had access to the *Observe Portal*. I intend to use it to understand the individual and the group performance.	Strongly Agree	Agree	Disagree	Strongly Disagree
X	Do you have any additional comments on your overall experience?	[open ended]			

A.5.3 Student post questionnaire – Phase 2 (Condition 2-2 (virtual classroom with system and expert observation))

Constructs:

- Perceive collaboration via chatting tool (COMM)
- Perceive rating as natural agent (NA)
- Assessment experience: quantity and timing of feedback (QTF), quality of feedback (QF), doing with the feedback (UF).
- Student preferred approach (PA)
- Student acceptance of the system (SA): perceived usefulness (PU), perceived ease of ese (PEOU), intension to use (IU).

Table A.5.3: Student post questionnaire – Phase 2 (Condition 2-2)

Code	Question	Response scale			
Section 1: Natural Agents (COMM & NA)					
COMM1	I found difficulties when communicating with the other student(s) via the multi-buttons chat window	Strongly Agree	Agree	Disagree	Strongly Disagree
COMM2	It was comfortable to communicate with the other student(s) through the virtual interface (i.e. using the chat window with classifying buttons)	Strongly Agree	Agree	Disagree	Strongly Disagree
COMM3	Explain the reasons why it was comfortable (or not) to communicate with the other student(s) through the chat window?	[open ended]			
COMM4	How would you rate your experience of collaborating with students in other location(s) using the classified chat?	Very Poor	Poor	Good	Very Good
COMM5	Please provide any extra comment you have on your experience working with the other student(s) in the virtual world.	[open ended]			
NA1	Using the rating tools to evaluate the other student(s) while working or at the end were EASY.	Strongly Agree	Agree	Disagree	Strongly Disagree
NA2	Using the rating tools to evaluate the other student(s) while working or at the end were FUN.	Strongly Agree	Agree	Disagree	Strongly Disagree
NA3	Using the rating tools to evaluate the other student(s) while working or at the end were USEFUL.	Strongly Agree	Agree	Disagree	Strongly Disagree
NA4	Using the rating tools to evaluate the other student(s) while working or at the end were INTERESTING.	Strongly Agree	Agree	Disagree	Strongly Disagree
NA5	Using the rating tools to evaluate the other student(s) while working or at the end were DIFFICULT.	Strongly Agree	Agree	Disagree	Strongly Disagree
NA6	Using the rating tools to evaluate the other student(s) while working or at the end were ANNOYING	Strongly Agree	Agree	Disagree	Strongly Disagree

NA7	Using the rating tools to evaluate the other student(s) while working or at the end were BORING	Strongly Agree	Agree	Disagree	Strongly Disagree
NA8	Explain the reasons why it was comfortable (or not) to rate the other student(s) through the rating tools	[open ended]			
NA9	Do you have any additional comments of the overall experience when rating others in the virtual world?	[open ended]			
	Section 2: Assessment Experience				
QTF1	On this activity, I get plenty of feedback from the system on how I did.	Strongly Agree	Agree	Disagree	Strongly Disagree
QTF2	The system assessment comes back very quickly.	Strongly Agree	Agree	Disagree	Strongly Disagree
QTF3	There is hardly any feedback on my performance when I finish.	Strongly Agree	Agree	Disagree	Strongly Disagree
QTF4	I would learn more if I received more feedback from the system.	Strongly Agree	Agree	Disagree	Strongly Disagree
QTF5	Whatever assessment I get comes too late to be useful.	Strongly Agree	Agree	Disagree	Strongly Disagree
QF1	The system assessment mostly tells me how well I am doing in relation to others	Strongly Agree	Agree	Disagree	Strongly Disagree
QF2	The assessment from the system is useful to understand the individual and the group INTERACTION LEVEL in the virtual world.	Strongly Agree	Agree	Disagree	Strongly Disagree
QF3	The assessment from the system is useful to understand the individual and the group SUCCESS LEVEL in the virtual world.	Strongly Agree	Agree	Disagree	Strongly Disagree
QF4	The assessment from the system is useful to understand the individual and the group COLLABORATIVE SKILLS LEVEL.	Strongly Agree	Agree	Disagree	Strongly Disagree
QF5	The assessment from the system helps me to understand things better	Strongly Agree	Agree	Disagree	Strongly Disagree
QF6	The assessment from the system shows me how to do better next time.	Strongly Agree	Agree	Disagree	Strongly Disagree
QF7	Once I have read the assessment from the system, I understand what I did	Strongly Agree	Agree	Disagree	Strongly Disagree
QF8	I don’t understand some of the system assessment.	Strongly Agree	Agree	Disagree	Strongly Disagree
QF9	I understand what the assessment is saying.	Strongly Agree	Agree	Disagree	Strongly Disagree
UF1	I use the assessment feedback to go back over what I have done in the learning activity.	Strongly Agree	Agree	Disagree	Strongly Disagree
UF2	I think the system assessment will help me with any subsequent activity.	Strongly Agree	Agree	Disagree	Strongly Disagree
UF3	The system assessment prompts me to go back over video material recorded earlier to understand the scores.	Strongly Agree	Agree	Disagree	Strongly Disagree

UF4	I tend to only read the marks from the system without watching the recorded video.	Strongly Agree	Agree	Disagree	Strongly Disagree
Section 3: Student acceptance of the system (SA)					
PU1	The use of ObservPortal is useful for assessment in 3D-VW collaborative activities	Strongly Agree	Agree	Disagree	Strongly Disagree
PU2	The use of the *Observe Portal* system is not suitable for assessing students' performance.	Strongly Agree	Agree	Disagree	Strongly Disagree
PU3	The use of the *Observe Portal* system allows me to get a deeper understanding of the individual and the group performance.	Strongly Agree	Agree	Disagree	Strongly Disagree
PU4	I find that *Observe Portal* is useful in assessing students interactions, success, and collaborative skills.	Strongly Agree	Agree	Disagree	Strongly Disagree
PU5	I don't see that *Observe Portal* makes any difference in assessing students' interactions, success and collaborative skills.	Strongly Agree	Agree	Disagree	Strongly Disagree
PEOU1	I find the *Observe Portal* is easy to use in virtual world collaborative activities.	Strongly Agree	Agree	Disagree	Strongly Disagree
PEOU2	The feedback obtained from the system is clear and understandable.	Strongly Agree	Agree	Disagree	Strongly Disagree
PEOU3	It is difficult to use the *Observe Portal* assessment system.	Strongly Agree	Agree	Disagree	Strongly Disagree
PEOU4	Students assessment through *Observe Portal* is easy.	Strongly Agree	Agree	Disagree	Strongly Disagree
IU1	I would use *Observe Portal* to assess my performance.	Strongly Agree	Agree	Disagree	Strongly Disagree
IU2	I would use *Observe Portal* to understand my performance.	Strongly Agree	Agree	Disagree	Strongly Disagree
IU3	Assume that I had access to the *Observe Portal*. I intend to use it to understand the individual and the group performance.	Strongly Agree	Agree	Disagree	Strongly Disagree
Section 4: *Observe Portal* & Teacher Assessment					
PA1	I think the use of the *Observe Portal* system has a significant advantage over traditional methods (teacher observation) for assessing students' performance in the virtual world	Strongly Agree	Agree	Disagree	Strongly Disagree
PA2	Which method provides more information about the group and the students' performance?	Teacher Assessment		*Observe Portal* System Assessment	
PA3	Which assessment method provides USEFUL information about the group and the students' performance?	Teacher Assessment		*Observe Portal* System Assessment	

PA4	Which method assesses better the individual and the group INTERACTION level?	Teacher Assessment	*Observe Portal* System Assessment
PA5	Which method assesses better the individual and the group SUCCESS level?	Teacher Assessment	*Observe Portal* System Assessment
PA6	Which method assesses better the individual and the group COLLABORATIVE SKILLS level?	Teacher Assessment	*Observe Portal* System Assessment
PA7	Which approach would you prefer to use for assessing the group and individual performance?	Teacher Assessment	*Observe Portal* System Assessment
PA8	Based on the previous question (A7), explain the reason for your choice?	[open ended]	
PA9	I believe that the assessment provided by the .... is more accurate.	Teacher Assessment	*Observe Portal* System Assessment
PA10	Based on the previous question (PA9), explain the reason for your choice?	[open ended]	
X	Do you have any additional comments on your overall experience?	[open ended]	

A.6. Expert Post-Questionnaires

A.6.1 Expert Post-Questionnaire – Phase 1 (Physical Classroom)

Construct:

- Expert experience (EXP)
- Expert Reflection about system assessment (REF)

Table A.6.1: Expert Post-Questionnaire – Phase 1

Code	Question	Response scale			
EXP1	It is EASY to observe students' performance using the manual sheets.	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP2	It is DIFFICULT to observe students' performance using the manual sheets.	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP3	It is INCONVENIENT to observe students' performance using the manual sheets	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP4	It is TIME-CONSUMING to observe students' performance using the manual sheets	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP5	Explain the reasons why it was difficult (or not) observing and assessing the students' performance using the manual sheets	[Open Ended]			
EXP6	If there were more than two students, I'll observe and assess them easily.	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP7	Based on the previous question (EXP6), explain the reason for your choice?	[Open Ended]			
EXP8	At some point, I got lost while I'm observing the students.	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP9	Collecting learning evidence from collaborative activities was difficult.	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP10	Assessing the groups' and the individuals' INTERACTIONS was easy.	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP11	Assessing the groups' and individuals' TASK SUCCESS was difficult.	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP12	Assessing the COLLABORATIVE SOCIAL SKILLS of the groups and individuals was easy.	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP13	Do you have any additional comments about the overall experience using the manual sheets for the assessment?	Strongly Disagree	Disagree	Agree	Strongly Agree
REF1	Learners receive useful feedback from me after the activity.	Strongly Disagree	Disagree	Agree	Strongly Agree
REF2	The marking of the learners' performance from me is helpful.	Strongly Disagree	Disagree	Agree	Strongly Agree
REF3	The provided assessment can help learners to improve their performance.	Strongly Disagree	Disagree	Agree	Strongly Agree

REF4	I think that the assessment is reliable.	Strongly Disagree	Disagree	Agree	Strongly Agree
REF5	Why do you think the assessment is reliable or not?	[Open Ended]			
REF6	I think that the assessment is valid.	Strongly Disagree	Disagree	Agree	Strongly Agree
REF7	Why do you think the assessment is valid or not?	[Open Ended]			

A.6.2 Expert Post Questionnaire

Constructs:

- Expert perceive usefulness (E-PU)
- Expert perceive ease of use (E-PEOU)
- Expert perceive intention to use (E-IU)
- Expert experience (EXP)
- Expert reflection about system assessment (REF)
- Expert preferred approach (E-PA)

Table A.6.2: Expert Post Questionnaire (Phase 2)

Code	Question	Response scale			
E.PU1	The Observ Portal is useful for assessment in 3D-VW collaborative activities.	Strongly Disagree	Disagree	Agree	Strongly Agree
E.PU2	The use of the *Observe Portal* system allows me to get a better understanding of the individuals' and groups' performance.	Strongly Disagree	Disagree	Agree	Strongly Agree
E.PU3	I think the *Observe Portal* system is not suitable for assessing students' performance.	Strongly Disagree	Disagree	Agree	Strongly Agree
E.PU4	I find that *Observe Portal* is useful for assessing students' interactions, success and collaborative skills.	Strongly Disagree	Disagree	Agree	Strongly Agree
E.PU5	I don't see that *Observe Portal* makes any difference in assessing students interactions, success and collaborative skills.	Strongly Disagree	Disagree	Agree	Strongly Agree
E.PEOU1	I find the *Observe Portal* is easy to use in virtual worlds.	Strongly Disagree	Disagree	Agree	Strongly Agree
E.PEOU2	Using the *Observe Portal* system is clear and understandable.	Strongly Disagree	Disagree	Agree	Strongly Agree
E.PEOU3	It is difficult to use the *Observe Portal* assessment system.	Strongly Disagree	Disagree	Agree	Strongly Agree
E.PEOU4	Assessment through the *Observe Portal* system is easy.	Strongly Disagree	Disagree	Agree	Strongly Agree
E.IU1	I would use *Observe Portal* to assess students performance in collaborative activities.	Strongly Disagree	Disagree	Agree	Strongly Agree
E.IU2	I would use the system for assessing students interactions, success and collaborative skills.	Strongly Disagree	Disagree	Agree	Strongly Agree
E.IU3	Assume that I had access to the *Observe Portal*, I intend to use it to assess student performance.	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP1	It is EASY to observe students' performance in the virtual environment using the manual sheets.	Strongly Disagree	Disagree	Agree	Strongly Agree

EXP2	It is DIFFICULT to observe students' performance in the virtual environment using the manual sheets	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP3	It is INCONVENIENT to observe students' performance in the virtual environment using the manual sheets.	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP4	It is TIME-CONSUMING to observe students' performance in the virtual environment using the manual sheets	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP5	Explain the reasons why it was difficult (or not) observing and assessing the students' performance using the manual sheets	[Open Ended]			
EXP6	If there were more than two students, I'll observe and assess them easily.	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP7	Based on the previous question (EXP6**)**, explain the reason for your choice?	[Open Ended]			
EXP8	At some point, I got lost while I'm observing the students.	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP9	Collecting learning evidence from collaborative activities in VW was difficult.	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP10	Assessing the groups' and the individuals' INTERACTIONS was easy	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP11	Assessing the groups' and individuals' TASK SUCCESS was difficult.	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP12	Assessing the COLLABORATIVE SOCIAL SKILLS of the groups and individuals was easy.	Strongly Disagree	Disagree	Agree	Strongly Agree
EXP13	Do you have any additional comments about the overall experience using the manual sheets for the assessment?	Strongly Disagree	Disagree	Agree	Strongly Agree
REF1	Learners understand the purpose of the system assessment.	Strongly Disagree	Disagree	Agree	Strongly Agree
REF2	Learners receive useful feedback from the system after the activity.	Strongly Disagree	Disagree	Agree	Strongly Agree
REF3	The marking of the learners' performance from the system is helpful.	Strongly Disagree	Disagree	Agree	Strongly Agree
REF4	The provided system assessment can help learners to improve their performance.	Strongly Disagree	Disagree	Agree	Strongly Agree
REF5	I think that the *Observe Portal* assessment is reliable.	Strongly Disagree	Disagree	Agree	Strongly Agree
REF6	Why do you think the assessment is reliable or not?	Strongly Disagree	Disagree	Agree	Strongly Agree

REF7	I think that the *Observe Portal* assessment is valid.	Strongly Disagree	Disagree	Agree	Strongly Agree
REF8	Why do you think the assessment is valid or not?	[Open Ended]			
REF9	Peer evaluation is a good approach to assess the quality of student performance in VWs.	Strongly Disagree	Disagree	Agree	Strongly Agree
REF10	Using the rating tool to assess the other group member(s) is an appropriate method in the collaborative activity.	Strongly Disagree	Disagree	Agree	Strongly Agree
REF11	Based on the previous question (REF10), explain the reason for your choice?	[Open Ended]			
E-PA1	I think the use of the *Observe Portal* system has a significant advantage over traditional methods (teacher observation) for assessing students' performance.	Strongly Disagree	Disagree	Agree	Strongly Agree
E-PA2	Which approach would you prefer to use for assessing collaborative students in the virtual world?	Traditional methods		The *Observe Portal* system	
E-PA3	Based on your previous (E-PA2), explain the reason for your choice?	[Open Ended]			
E-PA4	Could you give us your view of the *Observe Portal* system?	[Open Ended]			
E-PA5	What aspects of the *Observe Portal* do you think helps most in assessing students' learning in the virtual world?	[Open Ended]			

Appendix B: Statistical Tables

B.1. Kolmogorov-Smirnov Test for Student Constructs

Table B.1: Kolmogorov-Smirnov Test for Student Constructs

Question	Statistic (D)
COMM1	.321
COMM2	.234
COMM4	.361
NA1	.308
NA2	.321
NA3	.321
NA4	.255
NA5	.321
NA6	.325
QTF1	.295
QTF2	.263
QTF3	.289
QTF4	.253
QTF5	.254
QF1	.393
QF2	.327
QF3	.346
QF4	.297
QF5	.346
QF6	.329
QF7	.312
QF8	.308
QF9	.293

DF1	.287
DF2	.259
DF3	.250
DF4	.280
PU1	.418
PU2	.294
PU3	.374
PU4	.376
PU5	.293
PU6	.393
PEU1	.329
PEU2	.361
PEU3	.465
PEU4	.365
IU1	.379
IU2	.361
IU3	.301
PA1	.346
PA2	.374
PA3	.411
PA4	.355
PA5	.447
PA6	.482
PA7	.337
PA9	.411

B.2. Kolmogorov-Smirnov Test for Experts Constructs

Table B.2: Kolmogorov-Smirnov Test for Expert Constructs

Question	Statistic (D)
E.PU1	.433
E.PU2	.272
E.PU3	.433
E.PU4	.482
E.PU5	.329
E.PU6	.272
E.PEU1	.461
E.PEU2	.482
E.PEU3	.274
E.PEU4	.381
E.IU1	.284
E.IU2	.284
E.IU3	.272
REF1	.200
REF2	.277
REF3	.277
REF4	.277
REF5	.240
REF7	.267
REF9	.267
REF11	.240

B.3. Perception of natural agent rating (NA)

Table B.3.1: NA1/NA5 composite DIFFICULT-EASY values

	EASY (NA1)		DIFFICULT (-1) (NA5)			
	Std No.	%	Std No.	%	Total	%
Strongly agree	23	34%	34	50%	57	42%
Agree	38	56%	26	38%	64	47%
Disagree	6	9%	6	9%	12	9%
Strongly disagree	1	1%	2	3%	3	2%
	68	100%	68	100%	136	100%

Table B.3.2: NA2/NA6 composite FUN-ANNOYING values

	FUN (NA2)		ANNOYING (-1) (NA6)			
	Std No.	%	Std No.	%	Total	%
Strongly agree	27	40%	27	40%	54	40%
Agree	26	38%	26	38%	52	38%
Disagree	11	16%	11	16%	22	16%
Strongly disagree	4	6%	4	6%	8	6%
	68	100%	68	100%	136	100%

Table B.3.3: NA4/NA7 composite INTERESTING-BORING values

	INTERESTING (NA4)		BORING (-1) (NA7)			
	Std No.	%	Std No.	%	Total	%
Strongly agree	24	35%	22	32%	46	34%
Agree	36	53%	35	51%	71	52%
Disagree	5	7%	10	15%	15	11%
Strongly disagree	3	4%	1	1%	4	3%
	68	100%	68	100%	136	100%

B.4. Student perception of chat communication (COMM)

Table B.4.1: COMM1/COMM2 DIFFICULT-EASY values after reversal

	Difficult (-1) (COMM1)		Easy (COMM2)			
	Std No.	**%**	**Std No.**	**%**	**Total**	**%**
Strongly agree	29	43%	26	38%	55	40%
Agree	31	46%	31	46%	62	46%
Disagree	8	12%	8	12%	16	12%
Strongly disagree	0	0%	3	4%	3	2%
	68	100%	68	100%	136	100%

B.5. Student assessment experience (AEQ)

Table B.5.1: QTF1/QTF3/QTF4 composite values Quantity of Feedback

	QTF1		QTF3 (-1)		QTF4 (-1)			
	Frequency	**%**	**Frequency**	**%**	**Frequency**	**%**	**Total**	**%**
Strongly Disagree	2	2.9	0	0	4	5.9	6	3%
Disagree	7	10.3	3	4.4	26	38.2	36	18%
Agree	39	57.4	35	51.5	26	38.2	100	49%
Strongly Agree	20	29.4	30	44.1	12	17.6	74	36%
Total	68	100.0	68	100.0	68	100.0	204	100%

Table B.5.2: QTF2/QTF5 composite values Timing of Feedback

	QTF2		QTF5(-1)			
	Frequency	**%**	**Frequency**	**%**	**Total**	**%**
Strongly Disagree	1	1.5%	1	1.5%	2	1%
Disagree	3	4.4%	10	14.7%	13	10%
Agree	36	52.9%	33	48.5%	69	51%
Strongly Agree	28	41.2%	24	35.3%	52	38%
Total	68	100%	68	100%	136	100%

Table B.5.3: QF1/QF5/QF6/QF7/QF8/QF9 total composite values

	QF1		QF5		QF6		QF7		QF8 (-1)		QF9		Total	
	Freq	**%**	**Freq**	**%**	**Freq**	**%**	**Freq**	**%**	**Freq**	**%**	**Freq**	**%**	**Total**	**%**
Strongly Disagree	12.0	18	1.0	1	2.0	3	3.0	4	1.0	1	1.0	1	20	4.20%
Disagree	45.0	66	7.0	10	9.0	13	8.0	12	7.0	10	11.0	16	87	18.28%
Agree	11.0	16	37.0	54	31.0	46	34.0	50	38.0	56	31.0	46	182	38.24%
Strongly Agree	68.0	100	23.0	34	26.0	38	23.0	34	22.0	32	25.0	37	187	39.29%
Total	**68**	**100**	**68**	**100**	**68**	**100**	**68**	**100**	**68**	**100**	**68**	**100**	**476**	**175%**

B.6. Experts Experience (EXP)

Table B.6.1: EXP1-r/EXP2/EXP6-r/EXP8/EXP9/EXP10-r/EXP11/EXP12-r composite values

	EXP1-r		EXP2		EXP6-r		EXP8		EXP9	
	Expert No.	%	Expert No.	%	Expert No.	%	Expert No.	%	Expert No.	%
Strongly Disagree	7	70%	6	60%	1	10%	1	10%	0	0%
Disagree	0	0%	0	0%	0	0%	4	40%	4	40%
Agree	3	30%	4	40%	7	70%	0	0%	0	0%
Strongly Agree	0	0%	0	0%	2	20%	5	50%	6	60%
Total	10	100%	10	100%	10	100%	10	100%	10	100%

EXP10-r		EXP11		EXP12-r			
Expert No.	%	Expert No.	%	Expert No.	%	Total	%
5	50%	1	10%	3	30%	24	30%
1	10%	4	40%	1	10%	14	18%
4	40%	0	0%	6	60%	24	30%
0	0%	5	50%	0	0%	18	23%
10	100%	10	100%	10	100%	80	100%

B.7. Users Acceptance (TAM Variables)

Table B.7.1: PU1/PU2/PU3 composite values

	PU1		PU2 (-1)		PU3			
	Freq	%	Freq	%	Freq	%	**Freq Total**	**%**
Strongly Disagree	1	1.5	2	2.9	1	1.5	**4**	**2%**
Disagree	5	7.4	14	20.6	4	5.9	**23**	**11%**
Agree	45	66.2	35	51.5	46	67.6	**126**	**62%**
Strongly Agree	17	25.0	17	25.0	17	25.0	**51**	**25%**
Total	68	100.0	68	100	68	100	**204**	**1.00**

Table B.7.2: PU4/PU5 composite value

	PU4		PU5 (-1)			
	Freq	%	Freq	%	**Freq Total**	**%**
Strongly Disagree	4	5.9	3	4.4	**7**	**5%**
Disagree	3	4.4	8	11.8	**11**	**8%**
Agree	43	63.2	35	51.5	**78**	**57%**
Strongly Agree	18	26.5	22	32.4	**40**	**29%**
Total	68	100.0	68	100.0	**136**	**100%**

Table B.7.3: Composite values of PEOU items

	PEOU1		PEOU2		PEOU3(-1)		PEOU4			
	Freq	**%**	**Freq**	**%**	**Freq**	**%**	**Freq**	**%**	**Total Freq**	**%**
Strongly Disagree	2	2.9	2	2.9	3	4.4	1	1.5	8	3%
Disagree	3	4.4	5	7.4	3	4.4	2	2.9	13	5%
Agree	40	58.8	35	51.5	40	58.8	41	60.3	156	57%
Strongly Agree	23	33.8	26	38.2	22	32.4	24	35.3	95	35%
Total	68	100.0	68	100.0	68	100.0	68	100.0	272	100%

Table B.7.4: Composite value for IU items

	IU1		IU2		IU3			
	Freq	%	Freq	%	Freq	%	**Total Freq**	**%**
Strongly Disagree	3	4.4	2	2.9	2	2.9	**7**	**3%**
Disagree	8	11.8	5	7.4	7	10.3	**20**	**10%**
Agree	33	48.5	36	52.9	36	52.9	**105**	**51%**
Strongly Agree	24	35.3	25	36.8	23	33.8	**72**	**35%**
Total	68	100.0	68	100	68	100.0	**204**	**100%**

Table B.7.6: E.PU1/E.PU2/E.PU3 composite values

	E.PU1		E.PU2		E.PU3(-1)			
	Freq	%	Freq	%	Freq	%	**Freq Total**	**%**
Strongly Disagree	0	0	0	0	0	0	**0**	**0%**
Disagree	0	0	0	0	0	0	**0**	**0%**
Agree	3	30%	2	20%	7	70%	**12**	**40%**
Strongly Agree	7	70%	8	80%	3	30%	**18**	**60%**
Total	10	100%	10	100%	10	100%	**30**	**100%**

Table B.7.7: E.PU4/E.PU5 composite value

	E.PU4		E.PU5 (-1)		Total	
	Freq	%	Freq	%	**Freq Total**	**%**
Strongly Disagree	0	0	0	0	**0**	**0%**
Disagree	0	0	1	10%	**1**	**5%**
Agree	5	50%	5	50%	**10**	**50%**
Strongly Agree	5	50%	4	40%	**9**	**45%**
Total	10	100%	10	100%	**20**	**100%**

Table B.7.8: Composite values of E.PEOU items

	E.PEOU1		E.PEOU2		E.PEOU3(-1)		E.PEOU4			
	Freq	**%**	**Freq**	**%**	**Freq**	**%**	**Freq**	**%**	**Total Freq**	**%**
Strongly Disagree	1	10%	0	0	2	20%	0	0	3	8%
Disagree	0	0	0	0	0	0	0	0	0	0%
Agree	1	10%	2	20%	3	30%	6	60%	12	30%
Strongly Agree	8	80%	8	80%	5	50%	4	40%	25	63%
Total	10	100%	10	100%	10	100%	10	100%	40	100%

Table B.7.9: Composite value for E.IU items

	E.IU1		E.IU2		E.IU3			
	Freq	%	Freq	%	Freq	%	**Total Freq**	**%**
Strongly Disagree	0	0	0	0	0	0	**0**	**0%**
Disagree	1	10%	1	10%	1	10%	**3**	**10%**
Agree	2	20%	2	20%	5	50%	**9**	**30%**
Strongly Agree	7	70%	7	70%	4	40%	**18**	**60%**
Total	10	100%	10	100%	10	100%	**30**	**100%**

References

[1] M. D. Childress and R. Braswell, "Using massively multiplayer online role-playing games for online learning," *Distance Education*, vol. 27, pp. 187-196, 2006.

[2] M. Alrashidi, K. Almohammadi, M. Gardner, and V. Callaghan, "Making the invisible visible: real-time feedback for embedded computing learning activity using pedagogical virtual machine with augmented reality," in *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, 2017, pp. 339-355.

[3] B. Dalgarno and M. J. Lee, "What are the learning affordances of 3-D virtual environments?," *British Journal of Educational Technology*, vol. 41, pp. 10-32, 2010.

[4] M. R. Gardner and W. W. Sheaffer, "Systems to support co-creative collaboration in mixed-reality environments," in *Virtual, Augmented, and Mixed Realities in Education*, D. Liu, C. Dede, R. Huang, and J. Richards, Eds., ed Singapore: Springer Singapore, 2017, pp. 157-178.

[5] S. Felemban, "Distributed pedagogical virtual machine (d-pvm)," presented at the The Immersive Learning Research Network Conference (iLRN 2015), 2015.

[6] A. Pena-Rios, V. Callaghan, M. Gardner, and M. J. Alhaddad, "Remote mixed reality collaborative laboratory activities: learning activities within the InterReality Portal," presented at the Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 03, 2012.

[7] P. C. Blumenfeld, R. W. Marx, E. Soloway, and J. Krajcik, "Learning with peers: from small group cooperation to collaborative communities," *Educational Researcher*, vol. 25, pp. 37-40, 1996.

[8] G. Wells, *Action, talk, and text: Learning and teaching through inquiry* vol. 16: Teachers College Press, 2001.

[9] D. Schallert, J. Reed, M. Kim, A. Beth, Y. Chen, M. Yang, *et al.*, "Online learning or learning on the line: Do students learn anything of value in a CMD," in *meeting of the national reading conference, San Antonio, TX*, 2004.

[10] Y.-H. V. Chiang and D. L. Schallert, "Instructional design meets politeness issues in virtual worlds," *Immersive Environments, Augmented Realities, and Virtual Worlds: Assessing Future Trends in Education*, p. 123, 2012.

[11] O. De Troyer, F. Kleinermann, and A. Ewais, "Enhancing virtual reality learning environments with adaptivity: lessons learned," in *Symposium of the Austrian HCI and Usability Engineering Group*, 2010, pp. 244-265.

[12] T. A. Mikropoulos and A. Natsis, "Educational virtual environments: A ten-year review of empirical research (1999–2009)," *Computers & Education*, vol. 56, pp. 769-780, 2011.

[13] M. Gardner and J. Elliott, "The Immersive Education Laboratory: understanding affordances, structuring experiences, and creating constructivist, collaborative processes, in mixed-reality smart environments," *EAI Endorsed Transactions on Future Intelligent Educational Environments*, vol. 14, p. e6, 2014.

[14] F. Hesse, E. Care, J. Buder, K. Sassenberg, and P. Griffin, "A framework for teachable collaborative problem solving skills," in *Assessment and Teaching of 21st Century Skills*, ed: Springer, 2015, pp. 37-56.

[15] D. H. Schunk, *Learning Theories: An Educational Perspective*. New York: Macmillan, 1991.

[16] P. A. Ertmer and T. J. Newby, "Behaviorism, Cognitivism, Constructivism: Comparing Critical Features From an Instructional Design Perspective," *Performance Improvement Quarterly*, vol. 26, pp. 43-71, 2013.

[17] N. Saeed and S. Sinnappan, *Effects of Media Richness on User Acceptance of Web 2.0 Technologies in Higher Education*: INTECH Open Access Publisher, 2009.

[18] T. Mayes and S. De Freitas, "Review of e-learning theories, frameworks and models," *JISC E-Learning Models Desk Study*, 2004.

[19] B. Dalgarno, "The potential of 3D virtual learning environments: A constructivist analysis," *Electronic Journal of Instructional Science and Technology*, vol. 5, pp. 3-6, 2002.

[20] S. Wheeler, P. Yeomans, and D. Wheeler, "The good, the bad and the wiki: Evaluating student-generated content for collaborative learning," *British Journal of Educational Technology*, vol. 39, pp. 987-995, 2008.

[21] A. A. Gokhale, "Collaborative learning enhances critical thinking," *Journal of Technology Education*, vol. 7, 1995.

[22] J. Bruner, "Vygotsky: A historical and conceptual perspective," *Culture, communication, and cognition: Vygotskian perspectives*, pp. 21-34, 1985.

[23] A. W. Woolley, I. Aggarwal, and T. W. Malone, "Collective intelligence and group performance," *Current Directions in Psychological Science*, vol. 24, pp. 420-424, 2015.

[24] V. Callaghan, "The maker fables," in *Intelligent Environments (Workshops)*, 2013, pp. 306-313.

[25] Hajj Hackathon. (2018). Available: https://hajjhackathon.devpost.com/

[26] T. A. Angelo, "Reassessing (and defining) assessment," *AAHE BULLETIN*, vol. 48, pp. 7-9, 1995.

[27] L. Suskie, *Assessing student learning: A common sense guide*. San Francisco: Jossey-Bass, 2009.

[28] L. Suskie, "What is good assessment?," in *Assessing Student Learning: A Common Sense Guide*, Second Edition ed San Francisco: Jossey-Bass, 2009.

[29] N. M. Webb, "Group collaboration in assessment: Multiple objectives, processes, and outcomes," *Educational Evaluation and Policy Analysis*, vol. 17, pp. 239-261, 1995.

[30] L. S. Vygotsky, *Mind In Society: The Development of Higher Psychological Processes*: Harvard university press, 1980.

[31] D. Boud, R. Cohen, and J. Sampson, "Peer learning and assessment," *Assessment & Evaluation in Higher Education*, vol. 24, pp. 413-426, 1999.

[32] R. Watkins, "Groupwork and assessment: The handbook for economics lecturers. Economics Network," ed, 2004.

[33] University Assessment Framework [Online]. (2010). Available: https://metranet.londonmet.ac.uk/fms/MRSite/psd/hr/capd/Assessment%20Framework/University%20Assessment%20Framework%20Oct%202010.pdf

[34] C. Eberly. (2017). How can I assess group work?. Available: https://www.cmu.edu/teaching/designteach/design/instructionalstrategies/groupprojects/assess.html

[35] M. Binkley, O. Erstad, J. Herman, S. Raizen, and M. Ripley, "Defining 21st century skills and assessments. ," *Assessment and Teaching of 21st Century Skills*, 2010.

[36] F. L. Oswald, N. Schmitt, B. H. Kim, L. J. Ramsay, and M. A. Gillespie, "Developing a biodata measure and situational judgment inventory as predictors of college student performance," *Journal of Applied Psychology,* vol. 89, p. 187, 2004.

[37] M. M. Shultz and S. Zedeck, "Predicting lawyer effectiveness: broadening the basis for law school admission decisions," *Law & Social Inquiry,* vol. 36, pp. 620-661, 2011.

[38] P. C. Kyllonen, "Measurement of 21st century skills within the Common Core State Standards," in *Invitational Research Symposium on Technology Enhanced Assessments. May*, 2012, pp. 7-8.

[39] B. S. Connelly and D. S. Ones, "An other perspective on personality: meta-analytic integration of observers' accuracy and predictive validity," *Psychological Bulletin,* vol. 136, p. 1092, 2010.

[40] G. S. Maxwell, *Teacher Observation in Student Assessment*: Queensland School Curriculum Council, 2001.

[41] K. Morrison, L. Manion, and L. Cohen, "Observation," in *Research Methods in Education,* ed: Routledge, 2007, pp. 396-413.

[42] K. Burke, "What are observation checklists?," in *How to Assess Authentic Learning*, ed: Corwin Press, 2009.

[43] W. Xing, R. Wadholm, E. Petakovic, and S. P. Goggins, "Group learning assessment: developing a theory-informed analytics," *Educational Technology & Society,* vol. 18, pp. 110-128, 2015.

[44] G. D. Borich, *Observation skills for effective teaching*: Routledge, 2016.

[45] M. Davis, G. Hunter, L. Thalaal, V. Tran Ba, and A. Wooding-Olajorin, "Developing “Smart” Tutorial Tools to Assist Students Learn Calculus, Taking Account of Their Changing Preferred Approaches to Learning," in *Intelligent Environments 2019: Workshop Proceedings of the 15th International Conference on Intelligent Environments*, 2019, p. 227.

[46] M. Davis and G. Hunter, "CalculEng–An On-Line Tutorial Tool to Assist the Teaching and Learning of Calculus," in *The 18th SEFI Mathematics Working Group seminar on Mathematics in Engineering Education*, 2016, p. 82.

[47] G. Hunter, D. Livingstone, P. Neve, and G. Alsop, "Learn programming++: The design, implementation and deployment of an intelligent environment for the teaching and learning of computer programming," in *2013 9th International Conference on Intelligent Environments*, 2013, pp. 129-136.

[48] M. Stiles, "Effective learning and the virtual learning environment," in *EUNIS 2000: Towards Virtual Universities: Proceedings of the European University Information System 2000 Conference held at INFOSYSTEM 2000*, 2000, pp. 171-180.

[49] P. Dillenbourg, D. Schneider, and P. Synteta, "Virtual learning environments," in *3rd Hellenic Conference "Information & Communication Technologies in Education"*, Rhodes, Greece, 2002, pp. 3-18.

[50] K. M. Stanney and K. S. Hale, *Handbook of Virtual Environments: Design, Implementation, and Applications*: CRC Press, 2014.

[51] C. Dede, "Immersive interfaces for engagement and learning," *Science,* vol. 323, pp. 66-69, January 2, 2009.

[52] A. Pena-Rios, "Exploring Mixed reality in distributed collaborative learning environments," Doctor of Philosophy Ph.D. Thesis, School of Computer Science and Electronic Engineering, University of Essex, 2016.

[53] A. Alzahrani, "Towards the development and understanding of collaborative mixed-reality learning spaces," Doctor of Philosophy Ph.D. Thesis, School of Computer Science and Electronic Engineering, University of Essex, 2017.

[54] M. Alrashidi, "Making the invisible visible in constructionist learning tasks: an explanation framework based on a Pedagogical Virtual Machine (PVM)," Doctor of Philosophy Ph.D., School of Computer Science and Electronic Engineering University of Essex, 2017.

[55] E. Jambi, M. Gardner, and V. Callaghan, "A Generalized Pedagogical Framework for Creating Mixed-Mode Role-Play in Multi-User Virtual Environments," in *International Conference on Immersive Learning*, 2019, pp. 158-171.

[56] E. Jambi, M. Gardner, and V. Callaghan, "Supporting mixed-mode role-play activities in a virtual environment," in *2017 9th Computer Science and Electronic Engineering (CEEC)*, 2017, pp. 49-54.

[57] E. Longford, M. Gardner, and V. Callaghan, "Social organisation and cooperative learning: identification and categorisation of groups and sub-Groups in non-cooperative games," in *International Conference on Immersive Learning*, 2019, pp. 131-143.

[58] H. Wu, S. W. Lee, H. Chang, and J. Liang, "Current status, opportunities and challenges of augmented reality in education," *Computers & Education,* vol. 62, pp. 41-49, 2013.

[59] P. Milgram and A. F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE TRANSACTIONS on Information and System,* vol. E77-D, pp. 1321-1329, 1994.

[60] Y. Uematsu and H. Saito, "Vision-based augmented reality applications," in *Computer Vision*, X. Zhihui, Ed., ed: InTech, 2008.

[61] H. Tamura, H. Yamamoto, and A. Katayama, "Mixed reality: Future dreams seen at the border between real and virtual worlds," *Computer Graphics and Applications, IEEE,* vol. 21, pp. 64-70, 2001.

[62] A. Blum. (2018). *The Multiple Uses of Augmented Reality in Education*. Available: https://www.emergingedtech.com/2018/08/multiple-uses-of-augmented-reality-in-education/

[63] M. B. Ibáñez, Á. Di Serio, D. Villarán, and C. Delgado Kloos, "Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness," *Computers & Education,* vol. 71, pp. 1-13, 2// 2014.

[64] M. Alrashidi, V. Callaghan, and M. Gardner, "ViewPoint: an augmented reality tool for viewing and understanding deep technology," in *Cloud-of-Things (CoT'13)*, Athens, Greece, 2013.

[65] J. M. Andujar, A. Mejías, and M. A. Marquez, "Augmented reality for the improvement of remote laboratories: an augmented remote laboratory," *Education, IEEE Transactions on,* vol. 54, pp. 492-500, 2011.

[66] K. Squire and E. Klopfer, "Augmented reality simulations on handheld computers," *Journal of the Learning Sciences,* vol. 16, pp. 371-413, 2007.

[67] T. Arvanitis, A. Petrou, J. Knight, S. Savas, S. Sotiriou, M. Gargalakos*, et al.*, "Human factors and qualitative pedagogical evaluation of a mobile augmented reality system for science education used by learners with physical disabilities," *Personal and Ubiquitous Computing,* vol. 13, pp. 243-250, 2009.

[68] L. Kerawalla, R. Luckin, S. Seljeflot, and A. Woolard, ""Making it real: exploring the potential of augmented reality for teaching primary school science," *Virtual Reality,* vol. 10, pp. 163-174, 2006.

[69] B. E. Shelton and N. R. Hedley, "Exploring a cognitive basis for learning spatial relationships with augmented reality," *Technology, Instruction, Cognition and Learning,* vol. 1, p. 323, 2004.

[70] C. Dede, T. Grotzer, A. Kamarainen, S. Metcalf, and M. Tutwiler, "EcoMobile: Blending virtual and augmented realities for learning ecosystems science and complex causality," *Journal of Immersive Education,* 2012.

[71] R. A. Ruddle, "The effect of translational and rotational body-based information on navigation," in *Human walking in virtual environments*, ed: Springer, 2013, pp. 99-112.

[72] G. C. Burdea and P. Coiffet, *Virtual reality technology*: John Wiley & Sons, 2003.

[73] S. Wei and Z. Wen-qi, "Virtual Reality technology in modern medicine," in *Audio Language and Image Processing (ICALIP), 2010 International Conference on*, 2010, pp. 557-561.

[74] A. Uribe-Quevedo, D. Rojas, and B. Kapralos, "Customization of a low-end haptic device to add rotational DOF for virtual cardiac auscultation training," in *Information, Intelligence, Systems & Applications (IISA), 2016 7th International Conference on*, 2016, pp. 1-6.

[75] P. Paiva, L. Machado, A. Valen, T. Batista, and R. Moraes, "SimCEC: a collaborative vr-based simulator for surgical teamwork education," *Comput. Entertain.,* vol. 16, pp. 1-26, 2018.

[76] S. Wesley. (2018). *Augmented and Virtual Reality: The Future of Learning Experiences*. Available: https://virtualspeech.com/blog/augmented-virtual-reality-future-of-learning-experience

[77] M. Fetscherin and C. Lattemann, "User acceptance of virtual worlds," *Journal of Electronic Commerce Research,* vol. 9, pp. 231-242, 2008.

[78] M. Alvarez, "Second Life and school: The use of virtual worlds in high school education," *Texas, USA,* 2006.

[79] A. Kim. (2018). *Radiology education gets Second Life in virtual world*. Available: https://www.auntminnie.com/index.aspx?sec=ser&sub=def&pag=dis&ItemID=120417

[80] M. Ebner and A. Holzinger, "Successful implementation of user-centered game based learning in higher education: An example from civil engineering," *Computers & education,* vol. 49, pp. 873-890, 2007.

[81] M. Rymaszewski, *Second life: The official guide*: John Wiley & Sons, 2007.

[82] J. Seng and E. Edirisinghe, "Teaching computer science using Second Life as a learning environment," *Proceedings of the Australasian Society for Computers in Learning in Tertiary Education (ascilite), Singapore*, 2007.

[83] J. A. Betz, "Computer games: Increase learning in an interactive multidisciplinary environment," *Journal of Educational Technology Systems,* vol. 24, pp. 195-205, 1995.

[84] C. Depradine, "A role-playing virtual world for web-based application courses," *Computers & Education,* vol. 49, pp. 1081-1096, 2007.

[85] A. P. Schouten, B. van den Hooff, and F. Feldberg, "Real decisions in virtual worlds: team collaboration and decision making in 3d virtual worlds," in *ICIS,* 2010, p. 18.

[86] T. Coffman and M. B. Klinger, "Utilizing virtual worlds in education: The implications for practice," *International Journal of Social Sciences,* vol. 2, pp. 29-33, 2007.

[87] I. Duncan, A. Miller, and S. Jiang, "A taxonomy of virtual worlds usage in education," *British Journal of Educational Technology,* vol. 43, pp. 949-964, 2012.

[88] D. C. Frezzo, J. T. Behrens, R. J. Mislevy, P. West, and K. E. DiCerbo, "Psychometric and evidentiary approaches to simulation assessment in packet tracer software," in *Networking and Services, 2009. ICNS '09. Fifth International Conference on*, 2009, pp. 555-560.

[89] D. Kerr and G. K. Chung, "Identifying key features of student performance in educational video games and simulations through cluster analysis," *JEDM-Journal of Educational Data Mining,* vol. 4, pp. 144-182, 2012.

[90] B. Csapó, J. Ainley, R. Bennett, T. Latour, and N. Law, "Technological issues for computer-based assessment," in *Assessment and Teaching of 21st Century Skills*, P. Griffin, B. McGaw, and E. Care, Eds., ed: Springer Netherlands, 2012, pp. 143-230.

[91] A. Merceron and K. Yacef, "Mining student data captured from a web-based tutoring tool: Initial exploration and results," *Journal of Interactive Learning Research,* vol. 15, pp. 319-346, 2004.

[92] R. J. Mislevy, R. G. Almond, and J. F. Lukas, "A brief introduction to evidence-centered design," *ETS Research Report Series,* vol. 2003, pp. i-29, 2003.

[93] G. Chung, E. Baker, T. Vendlinski, R. Buschang, G. Delacruz, J. Michiuye*, et al.*, "Testing instructional design variations in a prototype math game," presented at the In Current Perspectives from Three National R&D Centers Focused on Game-based Learning: Issues in Learning, Instruction, Assessment, and Game Design., Denver, CO, 2010.

[94] T. Reiners, S. Gregory, and H. Dreher, "Educational assessment in virtual world environments," in *ATN Assessment Conference 2011: Meeting the Challenges*, 2011, pp. 132-140.

[95] P. R. Bloomfield and D. Livingstone, "Immersive learning and assessment with quizHUD," *Computing and Information System Journal,* vol. 13, pp. 20-26, 2009.

[96] D. Livingstone and J. Kemp, "Integrating web-based and 3D learning environments: Second Life meets Moodle," *Next Generation Technology-Enhanced Learning,* vol. 8, 2008.

[97] K. Thompson and L. Markauskaite, "Identifying Group Processes and Affect in Learners: A Holistic Approach to," *Cases on the Assessment of Scenario and Game-Based Virtual Worlds in Higher Education,* p. 175, 2014.

[98] C. D. Schunn and J. R. Anderson, "The generality/specificity of expertise in scientific reasoning," *Cognitive Science,* vol. 23, pp. 337-370, 1999.

[99] A. Bernardini and C. Conati, "Discovering and recognizing student interaction patterns in exploratory learning environments," in *Intelligent Tutoring Systems*, 2010, pp. 125-134.

[100] S. H. Tesfazgi, "Survey on behavioral observation methods in virtual environments," *research assignment, Delft Univ. of Tech,* 2003.

[101] J. D. Gobert, M. A. Sao Pedro, R. S. Baker, E. Toto, and O. Montalvo, "Leveraging educational data mining for real-time performance assessment of scientific inquiry skills within microworlds," *JEDM-Journal of Educational Data Mining,* vol. 4, pp. 111-143, 2012.

[102] R. J. Mislevy and M. Riconscente, "Evidence-centered assessment design," *Handbook of test development,* pp. 61-90, 2006.

[103] V. J. Shute, L. Rieber, and R. Van Eck, "Games... and... learning," *Trends and issues in instructional design and technology,* vol. 3, 2011.

[104] R. J. Mislevy, "Evidence-centered design for simulation-based assessment," *Military medicine,* vol. 178, pp. 107-114, 2013.

[105] V. J. Shute, "Stealth assessment in computer-based games to support learning," *Computer games and instruction,* vol. 55, pp. 503-524, 2011.
[106] M. Minsky, *Society of Mind*: Simon and Schuster, 1988.
[107] G. Weiss, *Multiagent Systems: A Modern Approach To Distributed Artificial Intelligence*: MIT press, 1999.
[108] M. Wooldridge and N. R. Jennings, "Intelligent agents: theory and practice," *Knowledge Engineering Review,* vol. 10, pp. 115-152, 1995.
[109] E. H. Durfee and V. R. Lesser, "Negotiating task decomposition and allocation using partial global planning," *Distributed Artificial Intelligence,* vol. 2, pp. 229-244, 1989.
[110] N. Kamila and R. Swain, "Role based architecture for complex agent," *American Journal of Intelligent Systems,* vol. 3, pp. 20-27, 2013.
[111] J. A. Sánchez, "A taxonomy of agents," *Rapport technique, ICT-Universidad de las Américas-Puebla, México,* 1997.
[112] T. Selker, "COACH: a teaching agent that learns," *Commun. ACM,* vol. 37, pp. 92-99, 1994.
[113] J. C. Schlimmer and L. A. Hermens, "Software agents: Completing patterns and constructing user interfaces," *JAIR,* vol. 1, pp. 61-89, 1993.
[114] H. Kautz, B. Selman, and A. Milewski, "Agent amplified communication," in *AAAI/IAAI, Vol. 1*, 1996, pp. 3-9.
[115] M. Ikeda, S. Go, and R. Mizoguchi, "Opportunistic group formation," in *Artificial Intelligence and Education, Proceedings of AIED,* 1997, pp. 167-174.
[116] T. Supnithi, A. Inaba, M. Ikeda, J. i. Toyoda, and R. Mizoguchi, "Learning goal ontology supported by learning theories for opportunistic group formation," *Proc. of AIED99,* pp. 67-74, 1999.
[117] A. Alzahrani, "Exploring adjustable autonomy in online tutoring systems," Doctor of Philosophy Ph.D. Thesis, School of Computer Science and Electronic Engineering, University of Essex, 2017.
[118] R. S. Yadav and V. P. Singh, "Modeling academic performance evaluation using soft computing techniques: A fuzzy logic approach," *International Journal on Computer Science and Engineering,* vol. 3, pp. 676-686, 2011.
[119] P. Albertos and A. Sala, "Fuzzy expert control systems: Knowledge base validation," vol. 6, ed, 2002.
[120] C. M. Bishop, *Pattern recognition and machine learning*: springer, 2006.
[121] F. Mereani, "Investigating the detection of stored scripting attacks using machine learning," City, University of London, 2021.
[122] F. Mereani and J. Howe, "Preventing Cross-Site Scripting Attacks by Combining Classifiers," 2018.
[123] H. R. Berenji, "Treatment of uncertainty in artificial intelligence," *Machine intelligence and autonomy aerospace systems,* vol. 115, pp. 233-247, 1988.
[124] D. Dubois and H. Prade, "What are fuzzy rules and how to use them," *Fuzzy sets and systems,* vol. 84, pp. 169-185, 1996.
[125] B. Al-Najjar and I. Alsyouf, "Selecting the most efficient maintenance approach using fuzzy multiple criteria decision making," *International journal of production economics,* vol. 84, pp. 85-100, 2003.
[126] P. Baraldi, L. Podofillini, L. Mkrtchyan, E. Zio, and V. N. Dang, "Comparing the treatment of uncertainty in Bayesian networks and fuzzy expert systems used for a human reliability analysis application," *Reliability Engineering & System Safety,* vol. 138, pp. 176-193, 2015.

[127] J. M. Mendel, "Fuzzy logic systems for engineering: a tutorial," *Proceedings of the IEEE,* vol. 83, pp. 345-377, 1995.

[128] C. Wang, "A study of membership functions on mamdani-type fuzzy inference system for industrial decision-making," 2015.

[129] P. Singhala, D. Shah, and B. Patel, "Temperature control using fuzzy logic," *arXiv preprint arXiv:1402.3654,* 2014.

[130] I. Y. Subbotin and M. G. Voskoglou, "A triangular fuzzy model for assessing critical thinking skills," *International Journal of Applications of Fuzzy Sets and Artificial intelligence,* vol. 4, pp. 173-186, 2014.

[131] A. Alam, S. Ullah, and N. Ali, "The effect of learning-based adaptivity on students performance in 3d-virtual learning environments," *IEEE Access,* vol. 6, pp. 3400-3407, 2018.

[132] A. Alam, S. Ullah, M. Burqi, A. ULLAH, and N. Ali, "Evaluating students performance in adaptive 3D-virtual learning environments using fuzzy logic," *Sindh University Research Journal-SURJ (Science Series),* vol. 48, 2016.

[133] Rasim, Y. Rosmansyah, A. Langi, and Munir, "Selection of learning materials based on students' behaviors in 3DMUVLE," *Telkomnika,* vol. 16, pp. 2127-2136, 2018.

[134] C. Luis, P. Adriana, and G. Alfredo, "Towards an Automated Model to Evaluate Collaboration Through Non-Verbal Interaction in Collaborative Virtual Environments," in *Intelligent Systems: Concepts, Methodologies, Tools, and Applications*, ed Hershey, PA, USA: IGI Global, 2018, pp. 1570-1586.

[135] K. Chrysafiadi and M. Virvou, "Evaluating the integration of fuzzy logic into the student model of a web-based learning environment," *Expert Systems with Applications,* vol. 39, pp. 13127-13134, 2012.

[136] S. Felemban, M. Gardner, and V. Callaghan, "Virtual observation lenses for assessing online collaborative learning environments," presented at the Immersive Learning Research Network (iLRN 2016), Santa Barbra, USA, 2016.

[137] S. Felemban, M. Gardner, V. Callaghan, and A. Pena-Rios, "Mixed Agents Virtual Observation Lenses for Immersive Learning Environments," *Journal of Universal Computer Science,* vol. 24, pp. 171-191, 2018.

[138] B. S. Bloom and D. R. Krathwohl, "Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain," 1956.

[139] J. T. Mayes and C. J. Fowler, "Learning technology and usability: a framework for understanding courseware," *Interacting with computers,* vol. 11, pp. 485-497, 1999.

[140] I. G. Consortium. (2003). *IMS learning design best practice and implementation guide* Available: https://www.imsglobal.org/learningdesign/ldv1p0/imsld_bestv1p0.html

[141] J. Venn, *Assessing Students With Special Needs*: Prentice Hall, 2006.

[142] S. Felemban, M. Gardner, and V. Callaghan, " An Event Detection Approach for Identifying Learning Evidence in Collaborative Virtual Environments," in *2016 8th Computer Science and Electronic Engineering Conference (CEEC),* 2016.

[143] S. Felemban, M. Gardner, and V. Callaghan, "Towards Recognising Learning Evidence in Collaborative Virtual Environments: A Mixed Agents Approach," *Computers,* vol. 6, 2017.

[144] M. Hadzic, E. Chang, T. Dillon, J. Kacprzyk, and P. Wongthongtham, *Ontology-Based Multi-Agent Systems* vol. 219: Springer, 2009.

[145] S. Felemban, M. Gardner, V. Callaghan, and A. Pena-Rios, "Towards observing and assessing collaborative learning activities in immersive environments," presented at the

Immersive Learning Research Network: Third International Conference, iLRN 2017 Proceedings, Coimbra, Portugal, 2017.

[146] M. B. Ibáñez, R. M. Crespo, and C. D. Kloos, "Assessment of knowledge and competencies in 3D virtual worlds: A proposal," in *Key Competencies in the Knowledge Society*, ed: Springer, 2010, pp. 165-176.

[147] D. W. Johnson, *Cooperation in the classroom*: American Psychological Association, 1991.

[148] R. Bartle, "Hearts, clubs, diamonds, spades: Players who suit MUDs," *Journal of MUD research*, vol. 1, p. 19, 1996.

[149] J. Lugrin, M. E. Latoschik, M. Habel, D. Roth, C. Seufert, and S. Grafe, "Breaking bad behaviors: A new tool for learning classroom management using virtual reality," *Frontiers in ICT*, vol. 3, p. 26, 2016.

[150] G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education," *EDUCAUSE review*, vol. 46, p. 30, 2011.

[151] J. Fiaidhi, "The Next Step for Learning Analytics," *IT Professional*, vol. 16, pp. 4-8, 2014.

[152] W. Greller and H. Drachsler, "Translating learning into numbers: A generic framework for learning analytics," *Journal of Educational Technology & Society*, vol. 15, p. 42, 2012.

[153] M. Worsley, "Multimodal learning analytics: enabling the future of learning through multimodal data analysis and interfaces," in *Proceedings of the 14th ACM international conference on Multimodal interaction*, 2012, pp. 353-356.

[154] E. F. Olivares, P. Albert, J. van Helvert, and M. Gardner, "Designing a learning analytics application to improve learner success in interactions based on multimodal dialogue systems," in *International Conference on Immersive Learning*, 2016, pp. 171-179.

[155] J. A. Larusson and B. White, *Learning analytics: From research to practice* vol. 13: Springer, 2014.

[156] N. Mogharreban and L. Dilalla, "Comparison of defuzzification techniques for analysis of non-interval data," in *NAFIPS 2006-2006 Annual Meeting of the North American Fuzzy Information Processing Society*, 2006, pp. 257-260.

[157] N. Padhy, *Artificial Intelligence and Intelligent Systems:* Oxford University Press, 2005.

[158] C.-C. Lee, "Fuzzy logic in control systems: fuzzy logic controller. II," *IEEE Transactions on systems, man, and cybernetics*, vol. 20, pp. 419-435, 1990.

[159] P. De Byl, *Holistic game development with unity: An all-in-one guide to implementing game mechanics, art, design and programming*: AK Peters/CRC Press, 2017.

[160] R. Trötschel, J. Hüffmeier, D. D. Loschelder, K. Schwartz, and P. M. Gollwitzer, "Perspective taking as a means to overcome motivational barriers in negotiations: When putting oneself into the opponent's shoes helps to walk toward agreements," *Journal of personality and social psychology*, vol. 101, p. 771, 2011.

[161] R. S. Peterson and K. J. Behfar, "Leadership as group regulation," in *The psychology of leadership*, ed: Psychology Press, 2004, pp. 157-178.

[162] G. Joughin, "Assessment, learning and judgement in higher education: A critical review," in *Assessment, Learning and Judgement in Higher Education*, G. Joughin, Ed., ed Dordrecht: Springer Netherlands, 2009, pp. 1-15.

[163] P. Ramsden, *Learning to Teach in Higher Education*: Routledge, 2003.

[164] P. Black and D. Wiliam, "Assessment and classroom learning," *Assessment in Education: principles, policy & practice*, vol. 5, pp. 7-74, 1998.

[165] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS quarterly*, pp. 319-340, 1989.

[166] V. Venkatesh, M. G. Morris, and P. L. Ackerman, "A longitudinal field investigation of gender differences in individual technology adoption decision-making processes," *Organizational behavior and human decision processes,* vol. 83, pp. 33-60, 2000.

[167] I. Ajzen, "The theory of planned behavior," *Organizational behavior and human decision processes,* vol. 50, pp. 179-211, 1991.

[168] J. L. Hale, B. J. Householder, and K. L. Greene, "The theory of reasoned action," *The persuasion handbook: Developments in theory and practice,* vol. 14, pp. 259-286, 2002.

[169] E. Rogers, "Diffusion of innovations . Delran," *NJ: Simon & Schuster. Schneider, L.(1971). Dialectic in sociology. American Sociological Review,* vol. 36, p. 667678, 2003.

[170] V. Venkatesh, J. Y. Thong, and X. Xu, "Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology," *MIS quarterly,* vol. 36, pp. 157-178, 2012.

[171] J. Brooke, "SUS-A quick and dirty usability scale," *Usability evaluation in industry,* vol. 189, pp. 4-7, 1996.

[172] R. Saade, F. Nebebe, and W. Tan, "Viability of the" technology acceptance model" in multimedia learning environments: a comparative study," *Interdisciplinary Journal of E-Learning and Learning Objects,* vol. 3, pp. 175-184, 2007.

[173] M. Igbaria and J. Iivari, "The effects of self-efficacy on computer usage," *Omega,* vol. 23, pp. 587-605, 1995.

[174] R. Agarwal and J. Prasad, "Are individual differences germane to the acceptance of new information technologies?," *Decision sciences,* vol. 30, pp. 361-391, 1999.

[175] Y. Lee, K. A. Kozar, and K. R. Larsen, "The technology acceptance model: Past, present, and future," *Communications of the Association for information systems,* vol. 12, p. 50, 2003.

[176] L. Halawi and R. McCarthy, "Measuring faculty perceptions of blackboard using the technology acceptance model," *Issues in Information Systems,* vol. 8, p. 160, 2007.

[177] D. Gefen and M. Keil, "The impact of developer responsiveness on perceptions of usefulness and ease of use: an extension of the technology acceptance model," *ACM SIGMIS Database: the DATABASE for Advances in Information Systems,* vol. 29, pp. 35-49, 1998.

[178] Y. Huang, S. J. Backman, K. F. Backman, and D. Moore, "Exploring user acceptance of 3D virtual worlds in travel and tourism marketing," *Tourism Management,* vol. 36, pp. 490-501, 2013.

[179] S. Altarteer and V. Charissis, "Technology acceptance model for 3d virtual reality system in luxury brands online stores," *IEEE Access,* vol. 7, pp. 64053-64062, 2019.

[180] C. Lorenzo, L. Lezcano, and S. S. Alonso, "Language Learning in Educational Virtual Worlds-a TAM Based Assessment," *J. UCS,* vol. 19, pp. 1615-1637, 2013.

[181] C. M. Rasimah, A. Ahmad, and H. B. Zaman, "Evaluation of user acceptance of mixed reality technology," *Australasian Journal of Educational Technology,* vol. 27, 2011.

[182] M. B. Ibáñez, J. J. García Rueda, D. Morillo, and C. Delgado Kloos, "Creating test questions for 3D collaborative virtual worlds: The WorldOfQuestions authoring environment," 2012.

[183] V. Venkatesh, "Creation of favorable user perceptions: Exploring the role of intrinsic motivation," *MIS quarterly,* pp. 239-260, 1999.

[184] V. Venkatesh and H. Bala, "Technology acceptance model 3 and a research agenda on interventions," *Decision Sciences,* vol. 39, pp. 273-315, 2008.

[185] R. Saadé and B. Bahli, "The impact of cognitive absorption on perceived usefulness and perceived ease of use in on-line learning: an extension of the technology acceptance model," *Information & management*, vol. 42, pp. 317-327, 2005.
[186] V. Venkatesh and F. D. Davis, "A theoretical extension of the technology acceptance model: Four longitudinal field studies," *Management science*, vol. 46, pp. 186-204, 2000.
[187] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "User acceptance of computer technology: a comparison of two theoretical models," *Management science*, vol. 35, pp. 982-1003, 1989.
[188] C. M. Jackson, S. Chow, and R. A. Leitch, "Toward an understanding of the behavioral intention to use an information system," *Decision sciences*, vol. 28, pp. 357-389, 1997.
[189] B. H. Sheppard, J. Hartwick, and P. R. Warshaw, "The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research," *Journal of consumer research*, vol. 15, pp. 325-343, 1988.
[190] D. Hounsell, V. McCune, J. Hounsell, and J. Litjens, "The quality of guidance and feedback to students," *Higher Education Research & Development*, vol. 27, pp. 55-67, 2008.
[191] D. J. Nicol and D. Macfarlane-Dick, "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice," *Studies in higher education*, vol. 31, pp. 199-218, 2006.
[192] G. Gibbs, C. Simpson, and R. Macdonald, "Improving student learning through changing assessment–a conceptual and practical framework," in *European Association for Research into Learning and Instruction Conference, Padova, Italy*, 2003.
[193] G. Gibbs and C. Simpson, "Measuring the response of students to assessment: the Assessment Experience Questionnaire," in *11th Improving Student Learning Symposium*, 2003, pp. 1-12.
[194] G. Gibbs, "How assessment frames student learning," *Innovative assessment in higher education*, vol. 23, 2006.
[195] J. Nunnally, "Psychometric methods," ed: New York: McGraw-Hill, 1978.
[196] J. F. Hair, W. C. Black, B. J. Babin, R. E. Anderson, and R. L. Tatham, "Multivariate data analysis (Vol. 6)," ed: Upper Saddle River, NJ: Pearson Prentice Hall, 2006.
[197] N. Heckert and J. Filliben, "Kolmogorov-Smirnov Goodness-of-Fit Test," in *NIST/SEMATECH e-Handbook of Statistical Methods*, ed: National Institute of Standards and Technology, 2003.
[198] H. Boone and D. Boone, "Analyzing likert data," *Journal of Extension*, vol. 50, pp. 1-5, 2012.
[199] R. L. Ott and M. T. Longnecker, *An Introduction to Statistical Methods and Data Analysis*: Nelson Education, 2015.
[200] G. QingKe, H. Dan, W. Zhao, and S. Kan, "Effects of positively and negatively worded items in personality measurement," *Acta Psychologica Sinica*, vol. 38, pp. 626-632, 2006.
[201] S. Bostock, "Student peer assessment," *Learning Technology*, vol. 5, 2000.
[202] J. Margolis. (2019). *IBM's debating computer suggests human brains are nothing special*. Available: https://www.ft.com/content/36c669c0-791b-11e8-af48-190d103e32a4
[203] Quora. (2016). *How powerful is the human brain compared to a computer?* Available: https://www.forbes.com/sites/quora/2016/03/02/how-powerful-is-the-human-brain-compared-to-a-computer/#27a88709628e

[204] R. Lu and L. Bol, "A comparison of anonymous versus identifiable e-peer review on college student writing performance and the extent of critical feedback," *Journal of Interactive Online Learning,* vol. 6, 2007.

[205] A. Darejeh and D. Singh, "A review on user interface design principles to increase software usability for users with less computer literacy," *Journal of computer science,* vol. 9, p. 1443, 2013.

[206] D. W. Schanzenbach, "Limitations of experiments in education research," *Education Finance and Policy,* vol. 7, pp. 219-232, 2012.

[207] N. Cowie, "Observation," in *Qualitative Research in Applied Linguistics: A Practical Introduction*, J. Heigham and R. A. Croker, Eds., ed London: Palgrave Macmillan UK, 2009, pp. 165-181.

[208] J. King, J. South, and K. Stevens, "Reimagining the role of technology in education: 2017 national education technology plan update," ed: U.S. Department of Education, Office of Educational Technology, 2017, p. 87.

[209] J. Chamberlain, "Harnessing collective intelligence on social networks," University of Essex, 2015.

[210] J. Chamberlain, U. Kruschwitz, K. Fort, and C. Cieri, "Games and gamification for natural language processing," 2018.

[211] J. Chamberlain, R. Bartle, U. Kruschwitz, C. Madge, and M. Poesio, "Metrics of games-with-a-purpose for NLP applications," 2017.