Flexible Assessment in Immersive Environments

Development of an Externalised and Automated Assessment Approach in the Exemplary Domain of STEM Education

Diploma Thesis
at
Graz University of Technology

submitted by

Joachim Maderer

Supervisor:
Univ.-Doz. Dipl.-Ing. Dr. techn. Christian Gütl
Graz University of Technology

Co-Supervisor:
Dr. techn. Mohammad AL-Smadi, MSc
Graz University of Technology
Tallinn University, Estonia

Institute for Information Systems and Computer Media (IICM)
Graz University of Technology
A-8010 Graz, Austria

Graz, October 2013

Flexibles Assessment in Immersiven Umgebungen

Entwicklung eines externen und automatisierten Ansatzes für Assessment im exemplarischen Bereich der MINT Bildung

Diplomarbeit
an der
Technischen Universität Graz

eingereicht von

Joachim Maderer

Betreuer:
Univ.-Doz. Dipl.-Ing. Dr. techn. Christian Gütl
Technische Universität Graz

Mitbetreuer:
Dr. techn. Mohammad AL-Smadi, MSc
Technische Universität Graz
Tallinn University, Estonia

Institut für Informationssysteme und Neue Medien (IICM)
Technische Universität, Graz
A-8010 Graz, Österreich

Graz, Oktober 2013

Abstract

Learning and teaching of natural science and similar subjects is generally not considered a simple matter. Studies do not only indicate socio-cultural problems but also that instructional methods and contents are often not authentic and not always in line with the lived-in world of students. Constructivist learning theories seem particularly appropriate for science education but require special resources in terms of time and equipment. Besides, also formative assessment and feedback are a part of each learning process; moreover, learning outcomes are increasingly formulated as competencies.

In recent years, 3-D computer graphics and the Internet have become available over a wide area and reasonable for the greater part of the population. Based on these technologies so called immersive environments – such as 3-D computer games, simulations and virtual online worlds – offer several aspects which could support constructivist approaches to learning. Nevertheless, it becomes apparent that also in simulations, formative assessment and feedback represent a necessary requirement to facilitate learning. Beyond that, assessment mechanisms and feedback are also an immanent aspect for the success of computer games. Due to the complexity and increasing choice of e-learning systems it is hardly possible for teachers to integrate meaningful, that means didactically relevant, assessment options into 3-D environments and simulations on their own. The analysis of available literature about e-assessment regarding immersive environments indicates that current approaches are either less flexible, require advanced knowledge about platform-dependent scripting technologies or focus too much on traditional assessment strategies such as multiple-choice tests and similar concepts.

Based on these issues, this thesis proposes a flexible assessment framework for assessment in different immersive environments. Particularly the assessment of behaviour, the reusability for different application scenarios as well as a platform-independent externalisation of the assessment design are significant requirements. Beginning with the enhancement and generalisation of a first approach in game-based environments, a prototype in the realm of virtual worlds has been developed and tested in the context of a simulated physics experiment. The approach is focusing on a semantic-enabled concept that provides generic event tracking and feedback facilities within the immersive environment, whereas a web service is responsible to interpret the user behaviour and return immediate feedback messages. In addition, a first demonstration has been conducted for the purpose of an expert evaluation which received positive feedback for the greater part and provided useful feedback for further improvements.

Kurzfassung

Das Lernen und Lehren von Naturwissenschaften und ähnlichen Gebieten gilt allgemein nicht als einfach. Studien zeigen nicht nur gesellschaftliche Probleme sondern auch, dass Lehrmethoden und Inhalte oft nicht authentisch sind und nicht mit der Lebenswelt von Studierenden und SchülerInnen im Einklang stehen. Besonders konstruktivistische Lerntheorien scheinen für naturwissenschaftlichen Unterricht besonders geeignet zu sein, benötigen aber spezielle zeitliche und materielle Ressourcen. Daneben sind auch eine lernbegleitende Leistungsfestellung (formative Assessment) und Feedback Bestandteil eines jeden Lernprozesses; außerdem werden Lernziele zunehmend in Form von Kompetenzen formuliert.

In den letzten Jahren sind 3-D Computergrafik und Internet flächendeckend und günstig für den Großteil der Bevölkerung verfügbar geworden. Basierend auf diesen Technologien bieten sogenannte Immersive Environments – darunter 3-D Computerspiele, Simulationen und virtuelle Online-Welten – einige Aspekte die konstruktivistische Lernmethoden unterstützen können. Jedoch zeigt sich, dass eine lernbegleitende Leistungsfestellung und Feedback auch in Simulationen ein notwendiges Kriterium sind um einen Lernerfolg zu erzielen. Darüber hinaus sind Beurteilungsmechanismen und Feedback auch immanent für den Erfolg von Computerspielen. Wegen der Komplexität und steigenden Auswahl an E-Learning Systemen ist es kaum möglich, dass eine Lehrkraft sinnvolle, d.h. didaktisch relevante Beurteilungssysteme, in 3-D Umgebungen und Simulationen selbstständig integriert. Die Analyse verfügbarer Literatur bezüglich E-Assessment in immersiven Umgebungen deutet darauf hin, dass aktuelle Ansätze entweder wenig flexibel sind, fortgeschrittenes Wissen über plattformabhängige Skriptsprachen (Technologien) erfordern oder den Fokus zu sehr auf traditionelle Strategien zur Leistungsfeststellung legen, z.B. Multiple-Choice-Fragen und ähnliche Konzepte.

Basierend auf diesen Problemen wird in dieser Diplomarbeit ein flexibles Framework für E-Assessment in unterschiedlichen immersiven Umgebungen vorgeschlagen. Besonders die Beurteilung der Handlungsweise von Lernenden, die Wiederverwendbarkeit für verschiedene Anwendungsszenarien, sowie eine plattformunabhängige Externalisierung des Entwurfes der Leistungsfeststellung stellen signifikante Anforderungen dar. Ausgehend von einer Erweiterung und Verallgemeinerung eines ersten Ansatzes in spielbasierten Umgebungen, wurde ein Prototyp im Bereich von virtuellen Welten entwickelt und im Kontext eines simulierten Physikexperiments getestet. Der Ansatz stützt sich auf ein semantisch orientiertes Konzept, welches generische Ereignisaufzeichnung und Feedbackfunktionen bereitstellt, wohingegen ein Webdienst dafür verantwortlich ist, das Benutzerverhalten zu interpretieren und sofortiges Feedback zu retournieren. Darüber hinaus wurde eine erste Demonstration im Rahmen einer Expertenevaluierung durchgeführt, welche größtenteils positives Feedback erhielt und nützliches Feedback für weitere Verbesserungen eingebracht hat.

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Ort	Datum	Unterschrift

Acknowledgements

First of all, I want to express gratitude to my family for their emotional, moral and financial support over the years. Deep appreciation and thanks go to my girlfriend Sabrina who has supported me all the time, including proofreading this thesis, as well as several other texts. I also want to thank her entire family for their support and for letting me feel welcome all the time.

Further acknowledgments are directed to Dr. Christian Gütl, not only for supervising this thesis, but also for the oppertunity to participate in an interesting and current field of research; also to Dr. Mohammad AL-Smadi, for being initial mentor during the ALICE project and co-supervisor of this thesis.

Furthermore, I also want to thank Dominik Kowald and Gregor Dumphart, who have been appreciated friends since secondary school and reliable group members throughout several courses at the university.

Finally, I am also grateful to Johanna Pirka for her recommendations on tools and tips about Open Wonderland, and particularly to Dr. Gudrun Wesiak for her additional support.

Graz, October 2013 Joachim Maderer

"Get over the idea that only children should spend their time in study. Be a student so long as you still have something to learn, and this will mean all your life."

Henry L. Doherty

Contents

List of Figures

List of Tables

1 Introduction

Teaching and learning has always been a complicated matter. Particularly STEM fields such as science and mathematics are considered the requisite know-how of modern society, but education fails on a large scale to be interesting and successful among students as well as society in general. If nothing else, the lack of a practical, realistic and true-to-life context in science education is made responsible for this plight. (cf. Osborne & Dillon, 2008; Bybee, 2010) It is clear that it is not always easy to provide all the means necessary for a perfect learning environment. Nevertheless, the increasing support for information and communication technology (ICT) within the last two decades has also opened new options for learning. Immersive environments such as 3-D virtual worlds represent a promising new technology in the realm of e-learning to facilitate, for instance, physics experiments within an authentic context (Machet, Lowe, & Gütl, 2012).

1.1 Motivation

Assessment is a crucial component of each planned instruction. But it is no longer that assessment is only considered as a means of grading. Formative assessment provides important feedback about the learning progress for students and teachers. Besides, assessments are increasingly authentic and focus on realistic situations rather than rigid test scenarios. (Suskie, 2010) Nevertheless, it is not a simple matter to provide authentic e-assessment tasks, especially when it should be based on automated solutions. But with the increased usage of technology, also assessment tasks require an accurate computer-based implementation. That means also that it should be flexible so that it can be used with different learning environments, further enable teachers and students to use all aspects of the assessment process with ease. (AL-Smadi & Gütl, 2008) Since the usage of immersive environments is a rather new field of e-learning there has been less research about proper assessment solutions.

The aim of this thesis is to research the theoretical concepts behind immersive education and examine related work regarding assessment approaches in immersive learning environments. Based on these findings this thesis proposes a flexible assessment framework and a proof of concept in the domain of physics education. The virtual world platform Open Wonderland is used to implement a virtual experiment that is supported through an external assessment system. Thus, feedback is immediately provided for the learner based on his or her actions. The first showcases are further presented to a group of experts in the context of a small expert evaluation.

1.2 Outline

In general this thesis consists of a theoretical part – summarising all important background information – and a practical part. The practical part will document the development of the prototype and its exemplary application, as well as the results of a small expert evaluation.

The summary of the theoretical background starts in *Chapter 2* with the introduction to learning theories and the challenges involved with STEM education. Further, all relevant aspects of a learning process will be outlined. *Chapter 3* is supposed to make the reader familiar with the idea of immersive environments and to classify computer terminology such as virtual reality and virtual worlds. Also several important software platforms for the development of virtual worlds will be introduced. Based on that, different forms of immersive education, including digital games-based learning, simulations and virtual worlds are discussed. This will be supported by state of the art examples covering all three categories. *Chapter 4* is finally concerned with the application and integration of assessment into immersive learning scenarios. This will cover general aspects on e-assessment, as well as new theoretical approaches and frameworks. Based on an extensive summary of recent work, related to assessment in immersive environments, issues and motivation for this thesis will be emphasised.

With *Chapter 5* the practical part of this thesis begins. Besides the general project idea, the requirements and conceptual architecture of a flexible assessment framework for immersive environments will be introduced. Based on this framework a feasible set of prototypes is selected and designed, as well as the layout for an integrated proof of concept. *Chapter 6* explains the development of the prototypes, including a special assessment module for the virtual world platform Open Wonderland that is connected with an already existing assessment system. Beyond that, the development of an assessment-compliant simulation of a simple pendulum is documented, as well as necessary improvements regarding the aforementioned external assessment system. *Chapter 7* describes the setup of three assessment showcases based on the implemented prototypes. In addition, the results and feedback of a small expert evaluation are reported.

Finally, *Chapter 8* discusses important lessons learned from the theoretical, technical and user perspective; whereas *Chapter 9* concludes this work with a final summary and outlook for future work.

2 STEM Education and Learning

The importance of science education, as well as the related disciplines of technology, engineering and mathematics, usually referred to as STEM, is to a great extent motivated in economic considerations on a national level. Countries require staying competitive on the global market, and technology and engineering are major industrial sectors, supported by progress in science and mathematics. (cf. Bybee, 2010) Based on these concerns, but also due to the technological evolvement of society in general, Bybee emphasises the importance for integrated STEM education in the context of the United State education system from elementary schools onwards.

According to Osborne and Dillon (2008), several European countries express similar concerns. Students are hardly motivated and engaged with science, and the number of students in higher STEM disciplines is decreasing. Although problems in teaching are present and considered seriously, there are social and cultural aspects as well. As Sjøberg and Schreiner (2010) showed in a report of the *ROSE project* ("Research on Science Education"), there are several significant findings and correlations among a wide range of European and non-European countries on different development stages. For instance, it seems the general interest in science decreases in better developed countries. Nevertheless, they also comment that this does not have to mean that students from poor countries are per se more interested in science, but might be anyway motivated due to prospects for a better life. Further, statistical analysis indicates several gender differences. Regarding the contents, girls are hardly interested in traditional technical aspects, but prefer topics about health, the human body and the natural environment.

The aim of this Chapter is to discuss the major problems of science-related education and introduce the most prominent solution approaches. The first section is supposed to make the reader acquainted with the most important terms in education. Based on that, modern considerations in science education, including inquiry-based learning and competences will be discussed. Finally, computer-supported approaches are introduced.

2.1 General Aspects of Formal Learning Settings

It seems important at this point to examine the basic terms and concepts commonly inherent in formal learning settings. These are fundamental terms for the remaining work and also related to important issues discussed in this thesis. Therefore, this section covers learning theories, instructional design and didactics, and most importantly assessment and feedback.

2.1.1 Learning Theories

The basis of each planned instruction for learning is an underpinning concept that explains what learning is and how it works. Literature reveals numerous theories, however, the following groups of theories are of most importance (Carlile & Jordan, 2005):

- *Behaviourism* is based on empirical deliberations, thus only considering external stimuli of the environment and observable responses of the learner. In that sense, learning is defined as a lasting change of behaviour. Therefore, two basic strategies, *classical conditioning* and *behaviour modification* are known. The latter is more important for education. Desired behaviour is *reinforced* through rewards whereas unwanted behaviour is ignored – thus replaced by desired behaviour – or explicitly weakened through punishment. However, reward is generally considered more effective than punishment; (Carlile & Jordan, 2005; Woollard, 2010)
- In contrast to behaviourism, *Cognitivism* is focused on thinking processes. It was influenced by progress in the fields of human perception, attention and memory retention as well as approaches in the sector of artificial intelligence. Based on that, the proper organisation of information, preparation of well-structured learning materials and usage of different media are crucial aspects; whereas the teacher is also supposed to be in full charge of the learning process. Furthermore, the adoption and development of general schemas for critical thinking is an immanent aspect; (Carlile & Jordan, 2005)
- *Constructivism* is more different in that external information, thus objective knowledge, is not the key aspect of learning, but how individuals interpret information based on their own experiences and foreknowledge. In this sense, learners construct knowledge for themselves and maintain a personal model of reality with specific perspectives and understanding. This model is formed with every new experience and situation. Different learners might construct different conclusions after concrete experiences, as the previous experiences will not be the same for each learner. However, at this point constructivism separates into two different subtypes: The so called *radical constructivism* defines

learning and knowledge construction as an entire individual process; whereas *social constructivism*, basically following the same principles, suggests that knowledge is always constructed through interaction with others, including cultural background and context. This could also be related to philosophical considerations about the way science – as a community driven process – works. Under these deliberations teachers are supposed to be not more than facilitators of learning. (Carlile & Jordan, 2005; Pritchard & Woollard, 2010)

2.1.2 Instructional Design and Didactics

According to Zierer and Seel (2012), didactics is defined by Dolch (1967) as *"the science of learning and teaching in general"*. It features aspects of learning theories, as well as instructional design; the latter being *"considered as the American way of planning and organizing instruction"*. Both concepts developed quite independently. What is known as *General Didactics* has more importance in the context of school education, whereas instructional design has a broader application including school education, but also higher education, military training, etc. Looking closer into the differences, several didactical approaches are focused on contents rather than teaching methods and differ therefore from Dolch's definition. In addition, also 'Bildung', a humanistic concept particularly associated with German education is more inherent in such approaches. However, there are some didactical approaches, such as systematic didactics or learner-oriented didactics which are rather comparable to instructional design. Nevertheless, instructional design has a more general scope. Zierer and Seel define it as *"the entire process of instructional planning and implementation"*, which includes even entire education systems. Besides that, the assessment respectively assessability of learning outcomes seems also a concern in several models of both concepts.

2.1.3 Assessment and Feedback

Assessment is usually understood as the evaluation of students' learning outcomes in order to give grades. While this is essentially a part of assessment, it also reflects a much older understanding. Modern approaches consider assessment an integral part of the learning process. Therefore it must be differentiated between assessment that evaluates the final outcome of learning (referred to as *summative assessment*) and such assessment that supports the learning process itself (*formative assessment*). The latter can be focused on the progress of individual students, which includes feedback for improvement, but also on the quality of instructional design. The assessment of the instructional methods of a teacher is related to the field of action research, which is concerned with the practical improvement of concrete situations. (Suskie, 2010; see also Shermis & DiVesta, 2011)

Beside this, modern assessment approaches are generally better in line with the desired learning outcomes and focus on thinking and skill development, rather than pure reproduction of knowledge. Nevertheless, students still complain that formal course assessment does often focus on knowledge that has to be learned by heart and is not related to realistic problems that would occur at the workplace. This demotivates intensive learning based on understanding and encourages shallow learning to pass the test. This happens also if separate courses are not well-aligned within the curriculum, thus not providing a red thread throughout the entire learning afford. Alternate forms of assessment, and among that authentic assessment activities, involve students with realistic problems. Nevertheless, there are still possible issues, for instance, if assessments focus only on the solvation of a problem instead of formulating the problem itself. (Suskie, 2010; Sambell, McDowell, & Montgomery, 2012)

According to Nicol and Macfarlane-Dick (2006), feedback is further a crucial element of formative assessment that links desired outcomes of learning activities with the current performance of the learners. Each learner produces internal feedback to make such comparisons, which is related to self-regulation. However, in order to develop such abilities of self-regulation, external feedback is also an essential component. External feedback influences the motivation and feelings of the learners. It further interferes with the current conceptions of the students and requires them to actively construct a meaning out of the feedback, which means it is not just a quality message that can be passively absorbed. Nevertheless, following recommendations for good feedback practices, a remarkable aspect is that feedback should be provided at a time where students can improve their work, not afterwards when the learning activity has already finished. It is also better to provide the feedback through dialogues and discussions, for instance with teachers and other students.

2.2 Challenges in Science Education

According to Osborne and Dillon (2008), there is already a mismatch of what science education in schools is intended for. Some interest groups believe that it is the first stage to higher education in science. However, school should not be seen as a recruitment facility. Under this premise it is more important to give young students the opportunity of a comprehensive and open-minded education, in order to understand the basic principles of the world they live in and develop a general appreciation for the importance, usefulness, but also limits of science. This is not less important for students who eventually pursue a career in science later on. Therefore, contents should focus on topics which are aligned with state of the art progress of science, as inherent in society and presented in media. Students have to see a connection between science and their lived-in world. Based on that, several recommendations are given for more successful science education. This includes approaches such as hands-on experiments, inquiry-based learning, curriculum adjustments to motivate female students, collaborative assignments, but also changes in assessment strategies.

2.2.1 Inquiry-based Learning and Constructivism

According to Savery (2006) *inquiry-based learning (IBL)* is a *"student-centered, active learning approach focused on questioning, critical thinking, and problem solving"*. It is usually associated with hands-on activities in science education. Colburn (2000) has more closely discussed the role of *inquiry* and *constructivism* in the context of science education. According to him, a major problem is that students already have their own explanations of phenomena, which are most often based on *misconceptions*. Because of that, traditional instruction will not easily change these assumptions. Therefore, Colburn discussed six aspects for improved teaching and learning from the perspective of constructivism:

1. *hands-on activities* are considered useful but only if implemented as open-ended activities, whereas students are supposed to answer new questions based on their foreknowledge. This could reveal conceptual problems and increase the acceptance for different explanations;
2. *cooperative learning* involves students into discussions with peers. This can also reveal conceptual problems based on challenging questions which students ask each other and it is more likely that students accept their misconception under these circumstances;
3. *open-ended questions* should be used to ask students about their ideas and concepts. But students require sufficient *time* to think about the question before they can respond;
4. *demonstrations* should be prepared to result in unexpected outcomes or students should be asked to predict the outcome of an experiment;
5. *lectures* and *textbooks* are not an adequate teaching material before students are aware of their conceptual problems. Furthermore, such content should never be used isolated;
6. *assessment* is further required to be changed. It should consist of short conceptual questions as well as authentic assessment as part of inquiry activities, including laboratory experiments and problem-solving.

Nevertheless, there are also concerns that IBL and similar approaches might fail to improve learning due to the minimal guidance involved with such approaches (see Kirschner, Sweller, & Clark, 2006).

2.2.2 Competences

In order to improve science education, it is suggested to change assessment paradigms and focus also on competencies which the students should acquire (Osborne & Dillon, 2008). But what are competencies? In comparison to knowledge and skills, the term competence appears rather complicated to describe. According to Winterton, Delamare-Le Deist, and Stringfellow (2006), it is not possible to give a unified definition based on literature due to its different interpretation and application. However, it has been commonly identified as a concept that expresses abilities that incorporate knowledge and skills as requirements but are more profound and also related to social skills and attitudes. It is further stated that competences are always individual and related to context.

National educational models of EU member states have recently been changed from content-oriented (assessment) approaches to learning outcomes, which are often formulated as competencies. (Winterton et al., 2006) Due to the demand of EU members for a reference framework that provides standardisation for the exchange of educational outcomes the *European Qualification Framework (EQF)* has been developed. (European Commission, 2008) The framework – which is based on the dimensions of knowledge, skills and competences – understands these three terms as follows:

> ”knowledge“ means the outcome of the assimilation of information through learning. Knowledge is the body of facts, principles, theories and practices that is related to a field of work or study. In the context of the European Qualifications Framework, knowledge is described as theoretical and/or factual;
> ”skills“ means the ability to apply knowledge and use know-how to complete tasks and solve problems. In the context of the European Qualifications Framework, skills are described as cognitive (involving the use of logical, intuitive and creative thinking) or practical (involving manual dexterity and the use of methods, materials, tools and instruments);
> ”competence“ means the proven ability to use knowledge, skills and personal, social and/ or methodological abilities, in work or study situations and in professional and personal development. In the context of the European Qualifications Framework, competence is described in terms of responsibility and autonomy. (European Commission, 2008)

2.3 Computer-supported Science Education

According to Rogers (2008) computers can support activities either dedicated to construction or information presentation and retrieval. For instance, measurement data might be recorded through special USB interfaces, evaluated and visualised. Video capturing enables motion tracking of objects in video sequences. The automated and fast capabilities of computers to record and compute data allow students to work with large data sets and increases the time they can focus on important aspects; also providing information in real time. Simulations can be used to provide virtual experiments which are possibly too dangerous or expensive, but also too difficult and special for usual classroom settings. Nevertheless, pedagogical aspects such as clear learning objectives and corrective actions are important for a successful application.

In addition to these 'traditional' approaches there are new ideas in computer-supported education going even one step further. According to Machet et al. (2012) there is an increasing interest in so called immersive 3-D virtual worlds. Particularly for learning strategies based on constructivism these environments are considered beneficial in several ways. Besides reporting about first promising approaches to science education in 3-D virtual worlds, Machet et al. propose further the usage of 3-D immersive virtual worlds for conducting virtual and remote laboratory experiments. They emphasise that real experiments might be limited to a less authentic context and that 3-D virtual worlds feature the technical means to provide any kind of realistic context. This could enable students to better transfer theoretical models and concepts to different situations and examples of application.

2.4 Summary

This Chapter has given a brief overview of the economic, socio-cultural and educational challenges of STEM education. Students lack motivation and engagement with science and the teaching methods involved with science education are too much focused on knowledge and not well aligned with students daily experiences.

Most important learning theories have been introduced, including behaviourism, cognitivism, as well as radical and social constructivism. Didactics and instructional design are both concerned with the preparation of learning settings but have also several differences. Assessment – which is an integral aspect of instructional approaches – has often only been considered for final examination and is increasingly important to provide feedback to students and instructors during learning activities. Feedback itself stands in important relation to the development of self-regulation abilities of learners and should therefore be provided in time to allow students improving their work.

In order to improve science education it is suggested to focus on authentic learning experiences and facilitate approaches such as hands-on experiments and inquiry-based learning. This can support the eradication of misconceptions if properly applied. Besides, also assessment should be authentic and focus on competencies. Competencies describe profound abilities to apply knowledge and skills in complex situations, which is nowadays of significant importance for education and qualification.

Finally, some basic ideas have been discussed of how computer technology might support traditional classroom activities. Especially virtual experiments have relevance in the context of this thesis. Beyond that, the usage of immersive 3-D virtual worlds is considered promising to improve science education. The next Chapter will therefore define and discuss *immersive environments* more precisely as well as their application in the context of education with focus on science, respectively physics.

3 Immersive Environments and Education

There appears to be an intensive use of the term *immersive environment* in related literature lately. Howland (1999) has determined *immersiveness* as an important term for the creation of computer games. Based on the multiple definitions of ‘immerse’ in Webster’s Dictionary he explained that a computer game should *“engross and absorb”* a player. Thus, immersiveness can be considered a quality measurement for such an environment to achieve this goal. Broader definitions include not only pure digital environments but also real world settings, such as theme parks and museums, which are supported by digital media installations as well. In such environments different technologies but also narrative and interactive elements contribute to immerse users. (Miller, 2008)

Nevertheless, particularly for *digital immersive environments*, as Abbasi and Baroudi (2012) stated, the domain of application is not restricted to entertainment, but includes also educational and business applications. Their simplified expression of immersion is *“that users feel like they are part of the simulated world”*. The Immersive Education Initiative[1] for instance explains that *“Immersive Education is designed to immerse and engage students in the same way that today’s best video game grab and keep the attention of the player”*, but this includes also computer simulations and other technologies (Immersive Education Initiative, n.d.).

The aim of this Chapter is dedicated to the research progress, effects and benefits of immersive education environments. The first sections will more closely look into the psychological background of immersive environments and define some technical terms. This includes the explanation of well-known terms like virtual reality, virtual worlds and virtual environments in more detail. While most of these notions sound similar – and have even been used interchangeable – it is important to clearly classify these terms (cf. Schroeder, 2008). Furthermore, the educational benefits of 3-D learning environments will be discussed, including concepts such as digital game-based learning, simulations and virtual worlds. Finally, two state of the art examples including all three concepts are introduced.

[1] http://europe.immersiveeducation.org/ (retrieved on Mai 2, 2013)

3.1 Psychological Background

The effects behind immersive environments can be linked to psychological research and terms. According to Wirth et al. (2007) the term *presence* was first introduced to explain 'immersion' from a theoretical perspective. But conceptual problems revealed that more fundamental psychological concepts such as *attention* and *involvement* should be considered. Based on that several subtypes emerged with *spatial presence* being the most appropriate concept to reflect the meaning in the context of immersive environments. Spatial presence is a newer term for what is known as *telepresence*, but represents – as Wirth et al. argue – a broader concept that is applicable to different types of media, besides virtual reality also television or traditional books. It describes a psychological state that is perceived as the *"conviction of being located in a mediated environment"*. This state is considered a response to the immersiveness of an environment. However, it is assumed that the state is either engaged or not, there is no continuum.

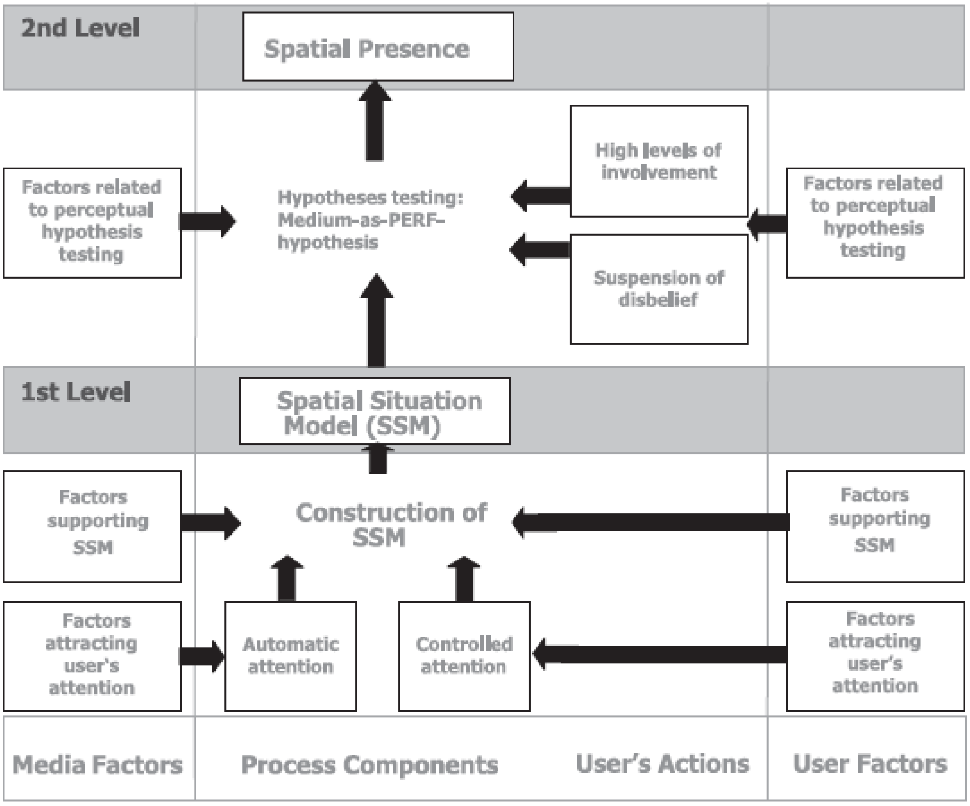


Figure 3.1: Two-level model to explain the formation of spatial presence (Wirth et al., 2007)

In order to describe the *"formation of spatial presence"*, Wirth et al. have further proposed a process model that consists of two levels (see Figure 3.1). Without going into details, the model is interpreted as follows: First of all, the *attention* of the user has to be caught. This happens automatically – based on media qualities – and is called *automatic attention*; or through the content, which the user might find interesting, thus called *controlled attention*. Second, based on the provided information a decision is made whether the received stimulation describes a spatial situation which leads eventually to the construction of a *spatial situation model (SSM)*. If that succeeds the second level of the model becomes relevant. Here the mind evaluates the role of the user in relation to the SSM and the real world. If the user experiences a feeling of *involvement* and the surrounding environment factors become insignificant or can be ignored (*suspension of disbelief*), the psychological state of spatial presence is considered engaged.

According to Weibel and Wissmath (2011) the mental state of *flow* is another important psychological concept that contributes to immersive environments such as computer games. It roughly refers to the *"immersion into an activity"* and was originally suggested by Csikszentmihalyi (1975). The concept of flow applies to different kind of activities. It describes a situation of strong involvement where a person can draw intrinsic satisfaction and pleasure from an activity, whereas the process itself is often more important than the result. Nevertheless, this has only a chance to happen if a person gets enough challenge and a clear objective. But the skills a person possesses must outweigh the degree of difficulty so that there is a realistic chance of success. This includes also the necessity of immediate feedback. Someone experiencing this mental state is deeply concentrated and feels in full control of the situation. Typical side effects include a different perception of time and a lack of self-consciousness. While flow was originally suggested and studied in the context of traditional recreational activities like sports but also different working situations, it has recently been transferred and analysed to the field of human-computer interaction and computer games. (Weibel & Wissmath, 2011; Nakamura & Csikszentmihalyi, 2002)

Weibel and Wissmath have studied the effects and relation between spatial presence and flow. Based on three empirical studies they confirm that both concepts are different despite a positive correlation that exists between them. They consider flow the dominant factor of immersion for computer games rather than spatial presence. But it is possible that spatial presence supports flow as a kind of precondition because it takes the attention away from the surrounding environment. Nevertheless, both factors will depend on the actual type of game or environment as well as personal conditions, such as motivation and tendencies towards immersive media.

3.2 Virtual Reality

Although virtual reality is a popular term its usage and understanding among average users and mass media does not necessarily match the definition of researchers and experts (cf. Sherman & Craig, 2002). Furthermore, also Miller (2008) states that there is a difference between what a lot of people call virtual reality, although featuring only some aspects, and *true* virtual reality solutions which are often limited to research, military and industrial applications.

According to Vince (2004), as well as Sherman and Craig (2002), the term virtual reality can basically be considered an analogue to the model concept of virtual images in optics. In general, virtual refers to something which is not actually existent but appears and behaves as it was. While an image in a mirror – a virtual image in fact – gives the impression of an identical reality that is present behind that mirror, virtual reality is supposed to represent a similar real impression for a simulated environment that only exists digitally. That means that several senses of the user have to be stimulated in order to achieve this. More precisely, Sherman and Craig determine four key elements which are necessary for a virtual reality experience:

- *Virtual world*: it represents the actual content which can be any kind of imaginary space. In the context of a computer simulation this refers to a database of related object descriptions and rules. It does not depend on a particular presentation technology;
- *Immersion*: particularly physical immersion (cf. spatial presence in the previous section) is important, as the idea of virtual reality as an access technology is not concerned with the content of the virtual world, and therefore narrative aspects, in the first place;
- *Sensory feedback*: special hardware is needed to stimulate the senses of the participants in order to physically immerse them. While the visual representation is considered the most important factor in most cases, it is directly followed by rich audio representation and sometimes also haptic feedback. Less often even balance, taste and smell are simulated through special devices. A tracking mechanism to follow the movements of a user is considered an important requirement;
- *Interaction*: this refers to the ability to navigate within the virtual world but also to the manipulation of virtual objects. More advanced collaborative environments also feature interaction among multiple users. This requires usually additional virtual objects to provide a presentation for each other, referred to as avatar.

Based on these elements Sherman and Craig (2002) construct a concrete definition for virtual reality, including a broad domain of virtual reality technologies:

> a medium composed of interactive computer simulations that sense the participant's position and actions and replace or augment the feedback to one or more senses, giving the feeling of being mentally immersed or present in the simulation (a virtual world).

A question rather complicated to be answered is whether a common computer game that uses 3-D graphics can be considered as virtual reality or not. This definition and further explanations of Sherman and Craig seems to insist on at least the ability to track the user's position by adding additional hardware to a modern computer system. Nevertheless, Vince (2004) distinguishes between immersive and non-immersive virtual reality systems. He states that early definitions have only included systems that feature head-mounted displays (HMD) but includes modern 3-D applications such as CAD systems and computer games as part of low immersive virtual reality technology.

3.3 Virtual Worlds and Virtual Environments

Sherman and Craig used the term *virtual world* in a rather generic way to describe all types of virtual (3-D) places that could be accessed through virtual reality technology. But there is a more particular meaning behind virtual worlds and virtual environments.

Bartle (2003) described virtual worlds as (imaginary) *self-contained* environments on different scopes (i.e. not necessarily a planet) which are realised through a simulation running on a single computer or a network of such. There is less difference up to this point, however, several more conventions usually apply to virtual worlds: The most important properties are a *shared* access among users in *real time* and *persistence*. The latter means that the world must not cease to exist when all users leave the world. On the contrary, it is possible that the environment will continue to develop. The users interact with the entire virtual world through a personal *character* and all possible actions and changes are determined by *automated rules*, referred to as physics.

Virtual worlds itself have emerged from early *multi-user dungeons (MUD)*. While first implementations were merely text-based games, graphical MUDs continued their development and used the term *persistent worlds* meanwhile. With the increasing number of users, graphical MUDs in the context of games are currently known as *massively multiplayer online role-playing games (MMPORGS)*. Although different fields of application have emerged from a pure usage in entertainment, it is still common to use game terminology to describe different aspects when working with virtual worlds (e.g. players instead of users). (Bartle, 2003)

The remaining sections will discuss a newer – eventually more accurate – definition of virtual worlds, outline the general architecture of virtual worlds and introduce a selection of important toolkits respectively platforms to create virtual worlds.

3.3.1 Modern Definitions

Considering the interdisciplinary interest on virtual worlds, Bell (2008) has expressed the need for a distinct definition because researchers and other professionals have used the term with somewhat different meanings. Based on a combination of important previous definitions, Bell gives a combined, short and solid definition for virtual worlds: *"A synchronous, persistent network of people, represented as avatars, facilitated by networked computers"*. Each part expresses a distinct requirement (Bell, 2008):

- *Synchronous*: expresses the requirement for real time communication. It is necessary for a real interaction among users and supports the perception of an environment that is comparable to real life;
- *Persistent*: the environment continuous to exist and evolve independently of a specific user being connected or not;
- *Network of people*: emphasises the possibilities for communication and social groups. This is the most important aspects for Bell, as he would otherwise consider virtual worlds as *"an empty data warehouse"*;
- *Represented as avatars*: an avatar acts as a surrogate entity and represents the user in a way that is significant different than a profile page of a traditional social network (e.g. Facebook). All actions conducted by a user are realised through the avatar and the avatar is addressed as if it was a person;
- *Facilitated by networked computers*: this term adds the requirement for computer and network technology. Otherwise a virtual world could (theoretically) also be realised with common utilities, although still following the remaining terms of the definition. Nevertheless the fast processing capabilities of networked computers are a significant feature of virtual worlds.

With this definition, it appears, Bell has excluded all possible ambiguities in relation to similar technologies and settings; although Bartle had already (at least implicitly) addressed most of these concerns, except for the strong emphasis on social aspects (cf. Bartle, 2003). In addition, Schroeder (2008) indicates that the difference between *virtual environment* and *virtual world* could be found in the aspect of socialising. While online games could also support social interaction their primary purpose is entertainment. But he does not define a strict separation between virtual reality and virtual environment. This matches with the statements of Vince (2004) that virtual environments could be considered as a general synonym for systems which are similar to virtual reality but are not supposed to provide a perfect duplicate of reality; rather imaginary worlds.

3.3.2 Architecture

According to all definitions that have been presented in the last subsection, all environments of such kind will require quite a similar architecture. First, there have to be one or more *servers* (or *clusters of servers*) which are responsible to continually manage instances of persistent virtual worlds. Each instance will at least include databases to store objects and entities the world consists of as well as an avatar database that associates users with in-world characters. Server clusters provide load balancing for client connections but share the same world and avatar databases. The *user database* is a central component. It is particularly important in the realm of commercial usage where players might own a central account but have different avatars belonging to different virtual world instances. Thus, it is usually a separated high-end hardware and software system. *Front ends* are different pieces of software which mediate between different types of clients and virtual world components. Clients might connect from full-strength client software – running on a common computer system such as a PC or gaming console – or with a less feature-rich browser interface. Even connections from mobile devices are possible, however, rather not for actual playing but to access current information about in-world events. Figure 3.2 shows a conceptual diagram of a virtual world architecture. (Bartle, 2003)

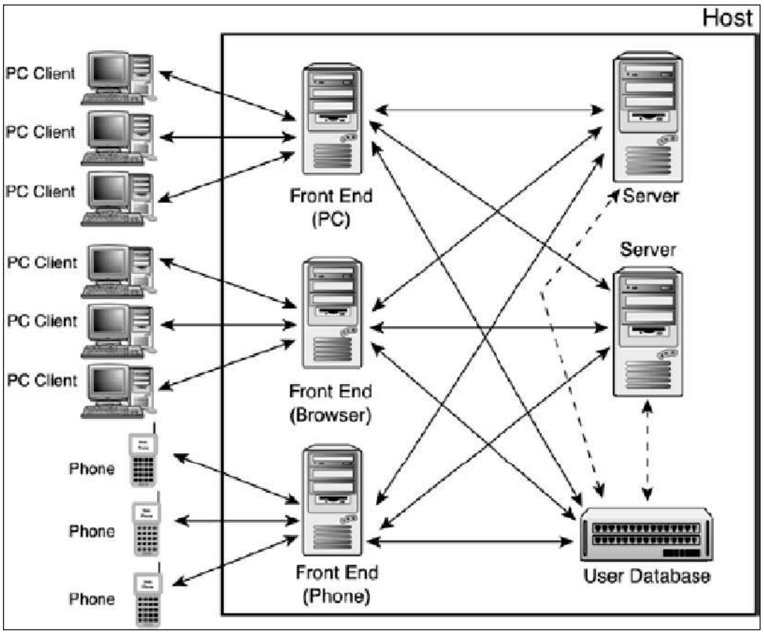


Figure 3.2: Typical virtual world architecture (Bartle, 2003)

3.3.3 Platforms

Several software platforms exist for creating and hosting immersive 3-D virtual worlds. This section will give an overview of *Second Life*, *Open Simulator* as well as *Open Wonderland* and mention a few further platforms.

Second Life

Based on public discussion and media resonance, Second Life is probably the most popular representative for a virtual world platform. Its technical description can be shortened to that it basically fulfils all the definitions given so far. As it is based on a client-server infrastructure, users have to use client software in order to connect to the virtual world. The client can be downloaded from the Second Life website[2] for free. However, several virtual objects and land available within Second Life have to be purchased with virtual money, called Linden Dollar. It is possible to obtain Linden Dollar for the exchange of real money or by earning it from transactions conducted within the virtual world. Beyond that it is also possible for users to construct and design new items. Nevertheless, uploading certain assets necessary for such items — e.g. images and textures — will also include fees. (White, 2007)

Despite its technical characteristics it is hard to define what Second Life is actually supposed to be. Among different standpoints, Second Life is hardly a game. It does not feature typical artificial systems like non-player characters (NPC) and there are no such achievements to gain that would be expected from a computer game. Users rather refer to it as a 3-D chat and community where they can express themselves. However, the evolution within this virtual world has grown to an extent where mostly everything that could be imagined in real life can be found in Second Life. This includes museums, any kind of sports activity, as well as virtual shopping; simply almost anything that could be named. (see White, 2007)

Open Simulator

In contrast to Second Life, Open Simulator[3] is an open source platform for creating and hosting virtual worlds, as well as virtual environments. While it features several communication protocols, it also supports the Second Life protocol. Thus, it is possible to use the Second Life client software to connect to virtual worlds hosted by Open Simulator servers. The server component itself is programmed in C# and can be either hosted on Windows through .NET Framework, as well as on Unix compatible

[2] http://www.secondlife.com
[3] http://www.opensimulator.org

systems by relying on Mono as runtime environment. The current development state supports all features necessary for a virtual world to be operated, including even features like inventories, friends lists, combat actions and virtual money. In general, Open Simulator is supposed to provide a multiple-purpose application server for different 3-D applications. Beyond that, the creators claim that Open Simulator should work as the very foundation for a future 3-D web; as such, providing a minimal extensible server, most likely as what is the Apache webserver currently to the traditional web. ("OpenSimulator Website", 2013)

Open Wonderland

Developed in Java and based on open standards, Open Wonderland[4] is also an open source platform for creating virtual worlds. First initiated by Sun Microsystems but then continued as an open source project it is still in an early development stage. The project focuses on real-time collaboration capabilities. Therefore technologies necessary for rich audio support, sharing of standard desktop applications within the 3-D world, as well as embedded Java applications are major aspects. However, Open Wonderland is highly extensible and even most of the core features are implemented as modules. A sophisticated security model has been integrated in order to provide reliable authentication in terms of business applications and for educational usage. Also Open Wonderland is considered a step towards a 3-D web that will share similarities with the World Wide Web. Hence, being based on numerous decoupled servers one can visit through a universal browser that loads assets like 3-D models and related program code. (Kaplan & Yankelovich, 2011; Open Wonderland Foundation, 2012)

Other Platforms

This overview could only introduce a few platforms to create immersive 3-D virtual worlds. There are many other (open source) platforms available. These include, for instance, *realXTend* (http://realxtend.org/), *Sirikata* (http://www.sirikata.com) and *Open Cobalt* (http://www.opencobalt.org/).

3.4 3-D Learning Environments

Dalgarno and Lee (2010) have presented a systematic model that explains the pedagogical benefits of immersive learning environments. Instead of separating between single-user environments and virtual worlds they use the term *3-D virtual learning*

[4] http://www.openwonderland.org

environments to combine games, simulations as well as virtual worlds. Based on a pool of characteristics that either belong to the categories *representational fidelity* or *learner interaction*, three quality factors for immersion are described. These are *construction of identity* within the environment, *sense of presence* and *co-presence*. The latter actually describes that others are present as well, expressing the collaborative aspects. Based on these influence factors they derive the following affordances – the potential pedagogical benefits – such environments can offer:

- *Spatial knowledge representation*: 3-D graphics allow an accurate representation of real phenomena with less limitation and can therefore support the development of conceptual understanding. This includes also the option to implement dangerous or unfeasible tasks, e.g. because of high expenses;
- *Experiential learning*: it means that learners can easily experience situations which can hardly be provided under real circumstances such as the exploration of atoms or invisible forces;
- *Engagement*: it refers to the intrinsic motivation a learner can draw from personal involvement. This includes game-based approaches as well as narrative aspects, further also the mental state of *flow* (cf. section 3.1).
- *Contextual learning*: there is evidence that learning can succeed better if it is facilitated in the same situational context where the acquired knowledge and skills are needed. Thus, 3-D environments can provide this kind of context;
- *Collaborative learning*: this refers to social constructivist learning approaches. In contrast to other collaboration and communication tools, 3-D environments provide enhanced options for teamwork and synchronous interaction on shared resources.

Nevertheless, Dalgarno and Lee also critically emphasise that empirical studies will be required in the future as a lot of publications only report in a qualitative manor. The following subsections will discuss game-based learning, simulations and virtual worlds in more details.

3.4.1 Digital Game-based Learning

The usage of games for learning is not a new idea. It is indicated that already 3000 BC games were considered as pedagogical tools in China. In addition, also traditional games and simulations had already been approved on their value for formal education within the second half of the last century. (Akilli, 2007) But research on educational computer games did only become a serious academic matter within the last ten years (Chang, Wang, Lin, & Yang, 2009). The abilities of computer games to motivate and engage players – quite similar to comparable media such as good movies – motivates the usage of computer games as a medium in general to transfer educational content. But there are also commercial entertainment games that support common educational or pedagogical objectives, such as skills training, knowledge about history or moral conflicts. Nevertheless, the identification of successful concepts alone does not guarantee a successful application on dedicated educational games. It is indeed a significant challenge that educational and gaming goals do not hinder each other. (Becker, 2007; Egenfeldt-Nielsen, 2006)

Serious Games

According to Djaouti, Alvarez, Jessel, and Rampnoux (2011) computer games are lately called *serious games* if their major purpose is not dedicated to entertainment, but target fields such as education and training. This term implies a (possibly intended) contradiction. But the earliest games were in fact developed within a serious context, whether it was a scientific project or for training purpose. Another aspect might be the definition of a game itself. Akilli (2007) gives an extended definition of games as *"a competitive activity that is creative and enjoyable in its essence, which is bounded by certain rules and requires certain skills"*. This competitive aspect alone, which is also immanent in sports, might add a certain seriousness (cf. Djaouti et al., 2011). Nevertheless, although the emergence of the term serious games is more complicated and older, Djaouti et al. (2011) believe it was emphasised within the last years in order to deliberately distinguish between computer games for entertainment – which have become the default, sometimes negative, association to this term – and such games intended for (but not only) education purpose.

In addition, also specific educational games – so called *Edutainment* – have been produced by several companies for quite a while, even since the earlier times of home computer devices. But such games have been criticised due to shortcomings in intrinsically motivating the player and to facilitate learning beyond simple training. The main issue is that behaviourism was often the immanent learning theory only focusing on reinforcement and rewards. (Egenfeldt-Nielsen, 2006)

Instructional Design

Becker (2007) has conducted an intense literature study and has compared different instructional design and learning theories against the design and methods of successful commercial computer games. The central point is that such games rather not provide a new approach of learning, as they do already incorporate sophisticated theories. The latter is understandable as commercial games need to address intrinsic motivation and adapt to different consumers in order to make profit. Beside many other aspects, mechanisms for *guidance*, *assessment* and *feedback* have been found an integral part of successful games.

A recent framework for game-based learning was proposed by van Staalduinen and de Freitas (2011). The framework is essentially based on a combination of several previous theories and incorporates aspects of learning, instruction and assessment, but also a proper alignment between them. Therefore an important step was to identify different elements of computer games. These elements have been assigned to the categories of a *four-dimensional framework* for game-based learning (see De Freitas & Oliver, 2006). The categories are *learner specifics*, *pedagogy*, *representation* and *context* (see Figure 3.3).

Learner Specifics Challenge Conflict Progress	**Pedagogy** Adaptation Assessment/Feedback Debriefing/Evaluation Instructions/Help/Hints Safety
Representation Action-Domain Link Control Interaction (Equipment) Interaction (Interpersonal) Interaction (Social) Location Problem-Learner Link Representation Sensory Stimuli	**Context** Fantasy Goals/Objectives Language/Communication Mystery Pieces or Players Player Composition Rules Theme

Figure 3.3: Assignment of game elements to the four-dimensional framework (van Staalduinen & de Freitas, 2011)

The actual framework is depicted in Figure 3.4. It consists of the three columns *learning*, *instruction* and *assessment* which should be considered in that order. Besides *learning content*, *learning objects* and *player goals* are deliberately separated as they can divert. The instructional design consists of a loop of *user behaviour*, *player feedback*, *user engagement*, and *user learning*. Feedback is considered a crucial element to foster engagement and promote learning. For assessment they suggests either *debriefing* or *system feedback*, possibly in the form of scores. The framework should support the development of new education games but can also be used to analyse previous works. (van Staalduinen & de Freitas, 2011)

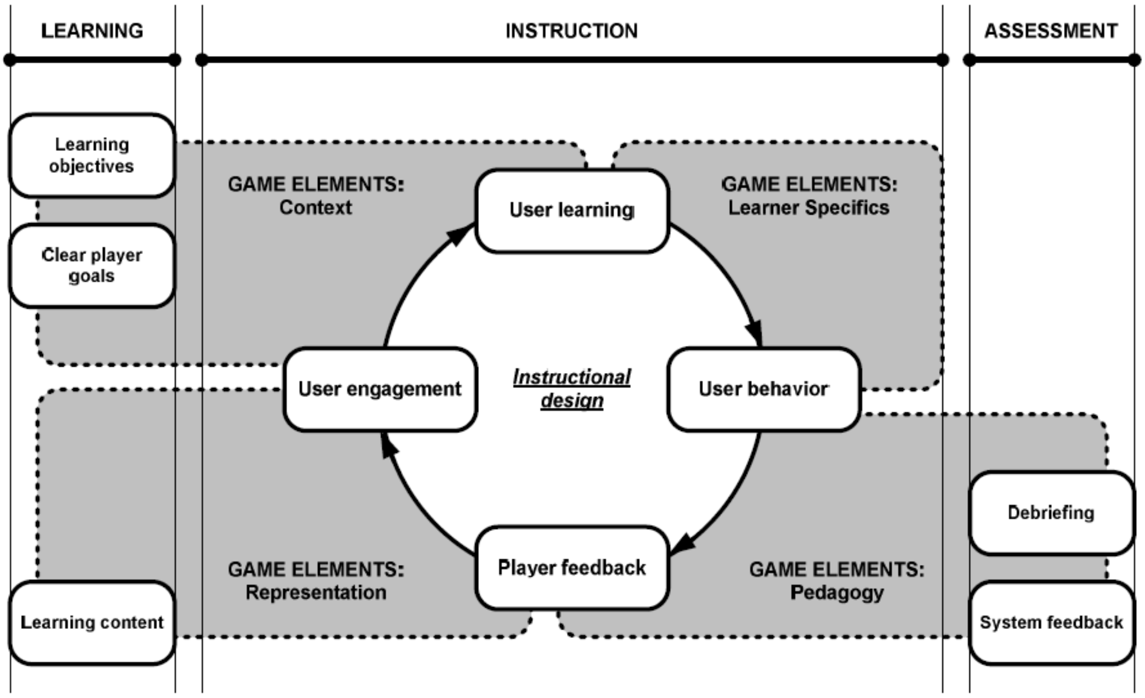


Figure 3.4: Framework for game-based learning (van Staalduinen & de Freitas, 2011)

Effectiveness and Efficiency

If there is a real practical benefit from game-based learning seems not yet conclusive. Egenfeldt-Nielsen (2006) remained rather sceptical and stated: *"It can certainly be said that video games facilitate learning, but the evidence for saying any more than this is weak"*. The majority of studies between the years 1981 and 2005 reported about increased motivation and engagement of students. However, a lot of studies have only collected data on the impression and opinions of the participants, rather than conducting a serious empirical study that would compare the learning outcomes against control groups (Egenfeldt-Nielsen, 2006; Akilli, 2007). Based on similar criticisms, Blunt (2009) reports on three successful studies conducted over a period of two years, which support the general effectiveness of digital game-based learning. Control groups of students who visited courses in the field of business, economics and management (involved several hundred students) have been additionally trained with computer games. Their results were in all three cases significantly better. But Blunt also expresses the need for subsequent studies. Nevertheless, despite if it is working or not there remains also the question if the benefits outweigh all the necessary technological and user-oriented affords in the end (Egenfeldt-Nielsen, 2006).

Tsai, Yu, and Hsiao (2012) have further reported on the influence of *learning* *motivation*, *learning ability* as well as *playing motivation* and *playing skill* regarding the learning outcome of educational games. Therefore, they have created a role-playing game to teach the concepts of electrical power and energy. The primary outcome is that playing skills influence the learning progress under all circumstances. That means the participants did only reach the highest possible results when they had also good skills in playing computer games, next to general learning motivation and also developed learning ability.

3.4.2 Simulations

Simulations are a different concept than games. But the difference is often not clearly formulated and both concepts also melt together in a broad area of application (cf. Prensky, 2001). This section should give a definition of simulations, highlight their usage for science education and report about the effectiveness and certain design considerations regarding a proper application.

Definition and Usage in Science Education

Based on a re-evaluation from literature, Sauvé, Renaud, Kaufman, and Marquis (2007) determine simulations as *"simplified, dynamic and precise representation of reality defined as a system"*. Thus, it is clear that many computer games feature several characteristics of simulations. For instance, the implementation of basic physical laws in most 3-D computer games but especially the development of sports or business games has a high demand on creating realistic models of real world processes. (cf. Prensky, 2001) Nevertheless, according to Prensky accurate representations of reality alone – which may also include realistic tasks – can become boring for the user. Thus, the combination of a sophisticated computer simulation with game-based concepts could result in a more attractive situation for learning. This separate class of *simulation games* are therefore particularly interesting for learning, which may explains why many authors usually mention both concepts consistently together.

Nevertheless, particularly science education seems to rely on what Gredler (2004) calls *symbolic simulations*; which she further divides into *system simulations* and *laboratory-research simulations.* The latter might be as well called virtual experiment. According to Rutten, van Joolingen, and van der Veen (2012) computer simulations in science education can be used to improve the conceptual understanding of students. The pedagogical and instructional advantages are manifold. Without the need to prepare and supervise experiments teachers can focus on the students' needs. Whereas students can investigate phenomena in real-time or different speeds, experiment with different simulation parameters and make use of different kinds of presentation. Beyond that, computer simulations can be used to support *scientific discovery learning (SDL)* where students slip into the role of researchers and can experience and entire research cycle starting with the formulation of a research question up to the experimental confirmation of an own hypothesis.

Effectiveness and Design Considerations

In contrast to game-based learning, there appears to be more conclusive evidence available for learning with computer simulations. Rutten et al. (2012) conducted an extensive literature review regarding the usage of computer simulations in science education. This review still leaves 48 empirical studies within the last decade after filtering inappropriate material. It was focused on the effectiveness of computer simulations in general, but also on the impact of representation and visualisation, the role of instructional support and the best approach to use computer simulations in relation to a classroom teaching settings. The most important aspects are briefly summarised in the following paragraphs.

With the exception of one indifferent outcome, most studies dedicated to determine the effectiveness of computer simulations indicate significant improvement of understanding or task and exam performance. But students have problems to draw connections from their observations and those who prefer research-oriented approaches have increased benefits in gaining new knowledge from computer simulations. Another study could also demonstrate an increased satisfaction of students, although no cognitive improvements were found. Beyond that, the usage of laboratory simulations, respectively virtual experiments, has been approved to prepare students for real experimentation. These preparation exercises lead to an improved theoretical background, knowledge about laboratory work, conceptual understanding, as well as a better performance during a real laboratory tasks. Especially students with shortcomings were found to benefit most from such virtual laboratory exercises. It is even evident that there is no difference between the learning outcomes of traditional and simulated experiments. But this also depends on the learning material as a subsequent study demonstrated. In addition, written instruction has been found an important requirement and students should be supported in formulating hypothesis when learning with computer simulations. (Rutten et al., 2012)

Regarding the influence factors, it was found that that 3-D graphics improve the conceptual understanding, but hardly with students who have low spatial abilities. However, highly immersive virtual reality technology – such as stereoscopic goggles – does not remarkably improve the learning outcomes. But other immersive factors, such as involvement into the activity as well as students being personally addressed, contribute significantly. From the instructional perspective, guidance has been found important in order to foster learning. Though, it is important to find a right balance as too much guidance interrupts the exploratory affordances of simulations. Nevertheless, it was found that students require appropriate information not only prior to a task but also during its completion. Finally, collaboration of students has also a significant positive effect on the learning outcomes. (Rutten et al., 2012)

3.4.3 Virtual Worlds as Learning Environments

A more recent approach on immersive education has been established through the introduction of large-scale 3-D virtual worlds like Second Life. While such virtual environments can also represent games – as discussed in section 3.3.1 – and therefore be an advanced platform for hosting game-based learning, this section will focus on the usage of virtual worlds as communication and collaboration platforms for the enrichment of online learning activities.

Motivation and Benefits

Dickey (2003) belongs to the earlier researchers who have recognised the pedagogical potential of 3-D virtual worlds, as she claims less literature was available at that time. She highlights the combination of immersive 3-D graphics as well as the communication capabilities of chat systems and earlier multi-user environments. Both of these aspects have been found to support constructivist learning approaches. In addition, the representation through avatars could lower the threshold of students to ask questions. Based on a case study, a participating teacher commends the immediate feedback that allows the adaption of teaching strategies. Students have also been observed to make use of the different perspectives to improve their understanding as well as helping each other.

According to Gütl (2011) 3-D virtual worlds have great potential to represent the next level of development in the context of e-learning. The immense success of recent online games such as *World of Warcraft*, as well as the extensive online world of *Second Life* confirms that the necessary technology has evolved from early research projects. They have recently reached the average user and are now a serious matter of business and economy. Traditional e-learning systems cause significant trade-offs such as isolating the user from others because of limited means of communication and collaborative activities. 3-D virtual worlds, on the other hand, promise an approach towards learning settings which are comparable to real world activities. Nevertheless, they can still incorporate additional benefits only possible in virtual environments. These include collaboration among learners in distant locations as well as the general benefits of simulated environments (dangerous and expensive experiments, hidden phenomena).

Gütl has further reported about two case studies. The first focused on the need of distance education, a particular issue in the Australian system for higher education considering the distribution of the continent's population. The second was more specifically focused on physics education, whereas the implementation of a physics simulation in Open Wonderland was considered as a supplement to traditional classroom teaching. Motivational aspects included the realisation of otherwise impossible activities and representations but also to facilitate social interaction and collaboration among students and teachers. Both studies reported about positive attitudes and experiences among participants. But it was also emphasised that further research is required due to technical and usability issues.

Design Considerations

Schmeil, Eppler, and de Freitas (2012) emphasise the need for a formal and structured approach to implement collaborative activities in 3-D environments. They criticise that current applications do not exploit the real potential of virtual worlds. For instance, just placing collaborative tasks – e.g. simple text documents or presentations – inside a 3-D environment will not necessarily add any additional value compared to more simplistic solutions. Consequently, the *Avatar-based Collaboration (ABC) Framework* is proposed, which is based on collaboration patterns that are described through language-theoretical aspects (semiotics). The framework consists of the following layers:

- *Infrastructure*: refers to the items and actions available in-world. Virtual objects are differentiated between *static*, *automated* and *interactive objects*. The latter are supposed to have instructional relevance. Based on these objects, higher level actions such as communication, navigation and interaction with objects is considered. This layer reflects the syntactic level of a collaboration pattern;
- *Dramaturgy*: this layer adds semantic meaning to the syntactic elements. It determines the relations between users, the location where interaction happens and also the dimension of time (when and in which sequences do actions occur). It further takes into account appropriate actions based on rules and social context;
- *Context* and *Goal*: refer to the application domain of the scenario, as well as the concrete objective for the scenario. This layer is defined broader and includes not only educational goals, but also collaborative work and playing activities. In terms of learning, the goals are defined in accordance with the levels of *Bloom's Taxonomy*.

3.5 Examples

Literature reveals several research scenarios and examples for digital game-based learning, simulations and the usage of 3-D virtual worlds as learning environments. Due to the limited scope, two works have been selected for a brief overview; both featuring dedicated science simulations. The first is Ludwig, a recent high-fidelity educational computer game and research project. The second is the implementation of real-world university course for learning physics within Open Wonderland.

3.5.1 Ludwig

Ludwig[5] is an Austrian educational game, teaching about renewable energy and featuring state of the art graphics, audio and game-play. It focuses on the learning theory of constructivism with a strong alignment between game and learning goals. Nevertheless, also aspects of behaviourism and cognitivism are present in some game elements. Students can explore the concepts of renewable energy within an authentic context. To this end, the game is supported by a catching story and students have to solve problems in order to proceed. The story is centred on a robot called Ludwig who is dispatched from an alien world to find new sources of energy and eventually crashes on a future earth. But the planet has already been abandoned because all resources have been exhausted. Several items and phenomenon can be tested in virtual laboratories and used by the player to support Ludwig on his mission. The development was based on multiple iterative cycles including professionals from the field of computer game development, didactical experts, as well as psychologists and pedagogues. But also feedback from students was an important aspect, especially when it comes to gameplay and the quality of game design. Didactical experts from the University of Graz designed learning content in accordance with the curriculum for physics. Therefore, concept maps have been used to model and present the knowledge on renewably energy in form of an in-game knowledge base. (Pfeiffer, 2012; OVOS, 2012)

The game was tested in several lower secondary school classes and students were also asked to provide feedback. Most students reported about great fun playing Ludwig, which would not be the case when playing conventional learning games. A scientific evaluation was conducted as well to obtain information on the effectiveness on the game. Although learning effects could be determined it was also emphasised that it is important to discuss the game contents with the students and connect them to physical processes of the real world. The game was also awarded as best “Serious Game 2011” in Germany and has created interest among other countries, including France and the United States. (Pfeiffer, 2012)

[5] http://www.playludwig.com/ (retrieved on Mai 2, 2013); demo version available online

Figure 3.5: In-game impression of Ludwig (screenshot taken from the demo version)

3.5.2 TEALsim in Open Wonderland

Pirker, Berger, Gütl, Belcher, and Bailey (2012) have reported on porting the Java-based simulation environment *TEALsim* to Open Wonderland (OWL), as well as its application within a virtual learning scenario. The TEALsim tool is usually supposed to support the *technology-enhanced active learning (TEAL)* approach that is used at Massachusetts Institute of Technology (MIT) for training students in elementary physics. Since this learning method requires special classrooms and a lot of equipment, it is too expensive for many universities to adopt this approach. Therefore, the motivation was to provide a virtual scenario within OWL that enables students to communicate and collaborate on a comparable basis. While the TEALSim software itself does already incorporate 3-D simulations, the control panels and display features were decoupled, adapted and directly placed into the 3-D interaction area of OWL. Also other elements which are usually part of the real world course have been realised with equivalent facilities, e.g. introductory videos and written information instead of teacher instruction. Figure 3.6 shows the falling coil simulation, placed inside OWL.

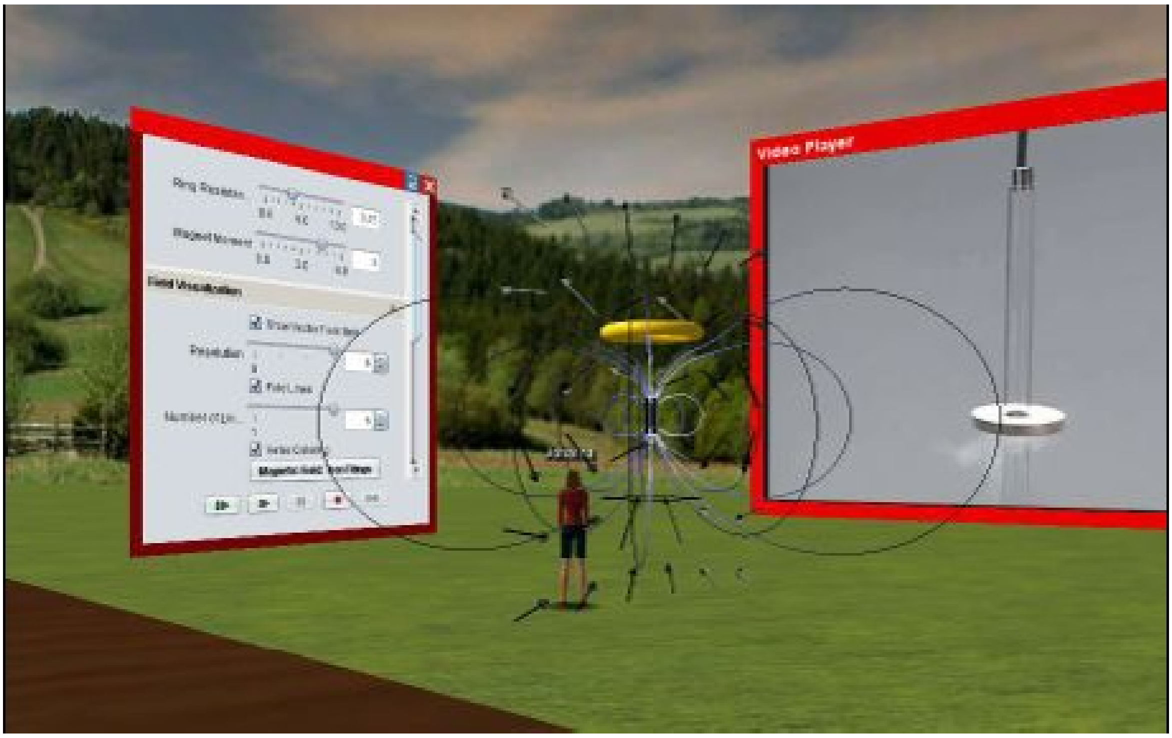


Figure 3.6: The falling coil simulation, placed inside OWL (Pirker, Berger, Gütl, Belcher, & Bailey, 2012)

3.6 Summary

Immersive environments are compelling digital and semi-digital scenarios based on psychological concepts such as spatial presence and flow in terms of computer games. Virtual reality is technically a medium that provides access to a computer simulated world through intensive stimulation of users' senses. But there are also people who call simple 3-D computer games and similar environments at least low immersive versions of virtual reality. Besides, virtual worlds are interactive online communities which are persistent and where users must be represented as avatars. Virtual environments – with less focus on social interaction – could be considered a subset of virtual worlds. Nevertheless, it is still possible that all these terms are confused in different literature. Several platforms have been introduced that allow the hosting of virtual worlds and environments. The most popular but also a commercial one is Second Life; open source projects include Open Simulator and Open Wonderland.

The usage of 3-D learning environments reveals five potential pedagogical benefits. These are *spatial knowledge representation*, *experiential learning*, *engagement*, *contextual learning* and *collaborative learning*. The engagement and motivation students experience when playing computer games motivates the usage of such immersive environments as medium to communicate educational content. The term serious game is now popular to discriminate from usual entertainment computer games. However, also the latter category can feature pedagogical benefits. The success of computer games is related to appropriate instructional design, whereas assessment, feedback and guidance contribute significantly. The framework of van Staalduinen and de Freitas (2011), which relies on a four-dimensional classification of game elements, considers feedback as a crucial aspect to design appropriate educational games. Nevertheless, at this time it appears not conclusive if serious games can really support learning effectively and efficiently in all situations. Simulations are a different concept but are still important for several computer games. Their effectiveness for science education is already significant in certain aspects but still needs further research. Applications include the improvement of conceptual understanding but also the support of scientific discovery learning. Especially virtual experiments have been found to prepare students for real laboratory exercises. Another approach to immersive education is provided through 3-D virtual worlds. These collaborative environments are supposed to provide an authentic and constructivistic learning approach within a social context. But this requires an appropriate orchestration of the environments' infrastructure and possible interactions to align it with learning objectives and define how students can collaborate in order to draw a real benefit from the social aspects of virtual worlds.

The computer game Ludwig has combined game-based learning with strong narrative elements and virtual laboratory exercises to teach about renewably energy. The work of Pirker et al. (2012) used the TEALsim simulation framework within Open Wonderland to provide a collaborative learning setting for an elementary physics course. The primary motivation was to provide a comparable replacement for real world settings where those would have been too expensive. Based on the importance of feedback and guidance in games and simulations as well as the general importance of assessment as part of learning, the next Chapter will briefly discuss e-assessment in general and review how assessment measurements have been explicitly integrated with immersive education environments.

4 Assessment in Immersive Learning Environments

The previous Chapter has shown that assessment and feedback are immanent concepts in game-based environments. Furthermore, also dedicated computer simulations of physical phenomena require appropriate support and guidance systems. E-assessment is obviously the computer-supported equivalent to traditional assessment activities. Considering the general importance of assessment and timely feedback for the learning process, it is clear that an increasing usage of e-learning technologies requires a proper integration of e-assessment approaches into digital learning resources. Based on the authentic affordances of modern technologies such as immersive environments, also assessment should be computer-based and integrated in order to provide authentic and individual assessment conditions. (cf. AL-Smadi & Gütl, 2008; Wesiak, Al-Smadi, & Gütl, 2012; Ridgway, McCusker, & Pead, 2004)

The aim of this Chapter is to give a general introduction to e-assessment and explain how it can be embedded with complex learning resources. Beyond that, a theoretical framework for feedback in serious games and similar domains is introduced. The remaining parts will then report about related work on assessment in immersive environments. Finally, necessary work and issues will be identified and discussed.

4.1 E-Assessment

The application of the term e-assessment is rather broad and most people might not be aware of the actual scope. *Computer-based assessment (CBA)* requires computers for the actual test situation and can also include feedback for the students. If students do not interact with computers themselves, there are still further tasks involved in the assessment process – such as the creation of reports or the analysis of results – that can be conducted with the support of computer systems. These activities are typically subsumed as *Computer-assisted assessment (CAA)*, although both categories are sometimes confused. The most prominent example for the latter is possibly the automated correction of paper-based multiple-choice questions through optical mark readers. (Charman & Elmes, 1998) Nevertheless, there seems to be less literature dealing with e-assessment in the context of school education. This might be the case because institutions such as universities had at least historically a greater practical need for e-assessment than schools because of their different scope. (cf. Charman & Elmes, 1998; Ridgway et al., 2004)

The purpose of this section is to motivate the usage of e-assessment and give an overview of its current status as well as the most important types and strategies for e-assessment applications. Beyond that, a new theoretical framework is introduced which promises a sound selection and application of e-assessment methods also including complex resources such as immersive environments.

4.1.1 Motivation and Current Status

The main motivation for e-assessment came less surprising out of the need to cope with the increasing amount of students, which is a problem of academic education in the first place. Thus, electronic solutions such as automated multiple-choice question are a perfect way to save time when conducting necessary summative assessments. Nevertheless, it is further emphasised that one should also consider pedagogical benefits which are rather manifold. E-assessment is certainly better structured and treats students therefore fair and equitable. But beside other advantages, its particular benefits are seen in the potential to provide immediate and consistent feedback for students and that is increasingly important regarding formative assessment concerns. (Charman & Elmes, 1998)

Besides that, Ridgway et al. (2004) emphasises the potential and necessity for e-assessment also in the context school education. Information and communication technology (ICT) does increasingly facilitate different learning tasks. But when it comes to the assessment process students must still accept paper-based tests. In general, Ridgway et al. describe several advantages of e-assessment. E-assessment systems can adapt to meet individual deficits of learners or can provide interactive content such as simulations or large data sets. This should enable more sophisticated and authentic assessment situations than possible under traditional assessment conditions and to address modern educational goals such as critical thinking and problem-solving but also communication and collaboration skills.

Nevertheless, e-assessment also earned criticism because it does not live up to these expectations. Instead of focusing on the additional benefits it just implements the same kind of assessment already used in traditional settings and is particularly limited due to the desire of automated evaluation. (AL-Smadi & Gütl, 2008) The e-assessment introduction of Cook and Jenkins (2010) still reports basically only about simple techniques such as multiple-choice questions, gap texts or numeric answers, which have been explained in (Crisp, 2007); although mentioning other computer-assisted approaches such as e-portfolios – an electronic documentation and collection of students' project achievements (Ridgway et al., 2004). Nevertheless, Crisp (2007) also subsumes some early approaches on assessment integrated into simulations and similar interactive domains.

4.1.2 Types and Strategies of E-Assessment

The most important classification of assessment is certainly given through its overall purpose in the learning process. Crisp (2007) aligns traditional assessment concerns (as partially discussed in section 2.1.3) with e-assessment and separates between three types which surround an enclosed learning process:

- *Diagnostic Assessment* is conducted prior to the actual learning process. It is supposed to collect information about the knowledge and skills level of the students and to adapt the learning process so that students' needs are satisfied. There are practically no consequences for the students (*low stake assessment*) and it is not necessary that students authenticate with the e-assessment system.
- *Formative assessment* informs teachers and students about the learning progress during learning activities. It is considered the most important type of assessment as students can practice and receive individual feedback and the learning process can still be adapted. It is still not a requirement to authenticate students and there are little to average consequences involved.
- *Summative assessment* requires the highest security measurements and online-based assessment is therefore more complicated to implement. It is mostly used as a final assessment activity in order to conclude the learning process, evaluate the students and give grades. Thus, the benefit for learning itself is rather low.

According to AL-Smadi and Gütl (2008) typical e-assessment systems can be differentiated between *fixed response* and *free response* systems. The former means that students can only choose among a limited number of provided answers; whereas the latter allows for an arbitrary input. Kowald (2012) has further looked up several strategies of e-assessment in literature. There are *automated assessment* approaches – such as fixed response questions (e.g. single-choice) but even first approaches towards the evaluation of free text answers processed by a computer system (see Gütl, 2008). Besides, there are several strategies that require an active participation of students or teachers for the evaluation process (Kowald, 2012):

- *Teachers* and *tutor assessment* puts the instructors in the exclusive position of correcting and grading the students;
- *Self-assessment* lets the student assess his or her own work, whereas reflection and increased engagement contributes to learning because students can find their weaknesses;

- *Peer assessment* enables students to assess each other. The requirement to formulate proper feedback does not only improve learning but is also an option to relieve teachers from some of their work;
- *Collaborative assessment* allows students to assess their own work but leaves grading in the domain of the teachers; thus, appropriate as an improved form of summative assessment;
- *Group assessment* can be combined with some of the approaches (peer assessment, teacher assessment) to target not only individuals but an entire group as a whole.

The combination of almost all these assessment strategies has already been realised in the context of collaborative writing in non-immersive learning environments (see Kowald, 2012; AL-Smadi, Höfler, & Gütl, 2011).

4.1.3 Integrated Model for E-Assessment

Based on the increasing requirements and complexity of e-learning resources, Wesiak et al. (2012) have developed a theoretical framework for the selection of such *complex learning resources* combined with appropriate integrated assessment methods. This *Integrated Model for e-Assessment (IMA)* consists of a general model to choose suitable complex learning resources, as well as a submodel that provides a structured approach towards an appropriate assessment method. It should guarantee an efficient and effective e-learning setting, taking into account all related aspects such as preconditions specific to individuals (possible adaptivity), the knowledge domain, didactical considerations, and most important: the defined learning outcomes and the associated assessment strategy to proof them correctly. The selection process is based on the following steps:

1. *Learning objectives* are selected based on the desired outcomes of the course and in accordance with didactical objectives. The latter might include competences and metacognitive skills.
2. The appropriate *complex learning resource* is selected, based on the requirements of the learning objectives. Typical examples include collaborative tasks, simulations and serious games.
3. The appropriate *assessment method* is selected by applying the actual *assessment model* (see Figure 4.1). It covers all implementation-relevant aspects, such as *assessment type* (diagnostic, formative, summative) and *assessment strategy* (e.g. assessment through individuals or automated), as well as the actual *assessment method* (fixed response, or activity tracking, etc.) and finally also feedback in terms of type, source and frequency.

4. Finally, the setup is *evaluated* (methods and procedures are assessed) and *validated* (is the setup appropriate to reflect the learner status?).

The entire process can be repeated until the desired outcomes can be met with the selected learning resources and assessment methods. The model was evaluated by five experts. Based on two relevant issues, regarding abstractness and the missing of practical experiences, the model has been improved to the current version and applied on practical examples (see Wesiak et al., 2012)

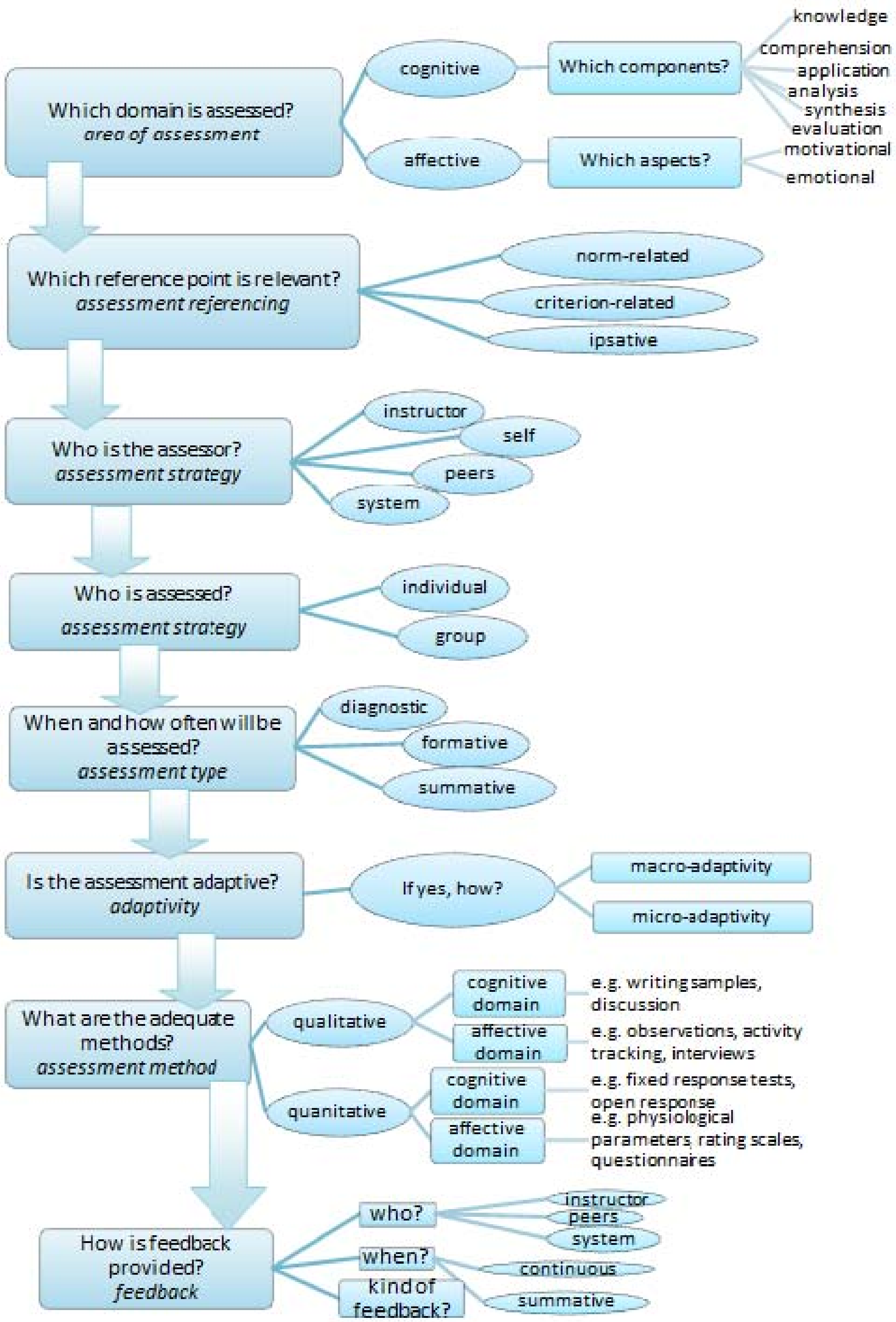


Figure 4.1: Assessment model (submodel) (Wesiak et al., 2012)

4.2 Feedback in Serious Games

Dunwell, De Freitas, and Jarvis (2011) have further proposed a four-dimensional framework to describe how feedback should be implemented in serious games, but also related abstract domains. Two aspects that have been researched more closely regard the timing (frequency) of how feedback should be delivered, as well as the contents of the feedback message. The approach is based on the classification of Carl Rogers (1951), including the feedback categories *evaluative*, *interpretive*, *supportive*, *probing* and *understanding*. The classification of Rogers has been applied to the context of serious games in the following way:

- *Evaluative*: the feedback is presented as a simple score;
- *Interpretive*: an explanation is added to describe what was wrong;
- *Supportive*: score is supported by an explanation, of what needs to be improved;
- *Probing*: feedback requires user response to learn about the difficulties of the player in order to dynamically adapt the scenario;
- *Understanding*: causal explanation of fundamental error, based on previous response.

In addition, a new type called *evolutionary* feedback was added. It should describe the immanent feedback of games, based on changes in the environment that could fall under the category of consequences being caused, observed and understood by players themselves; e.g. the death of a virtual patient. *Evaluative* and *interpretive* feedback can be implemented easily. However, all other types of feedback content (evolutionary feedback excluded) require approaches towards artificial intelligence. The complete model (see Figure 4.2) is then based on *type* and *frequency*, as well as *content* (essential or desirable task, based on the selected learning objectives) and *format* (text, image, voice or simulated outcome). Nevertheless, this model is as well placed within the context of pedagogical, learner-specific and external considerations. (Dunwell, De Freitas, & Jarvis, 2011)

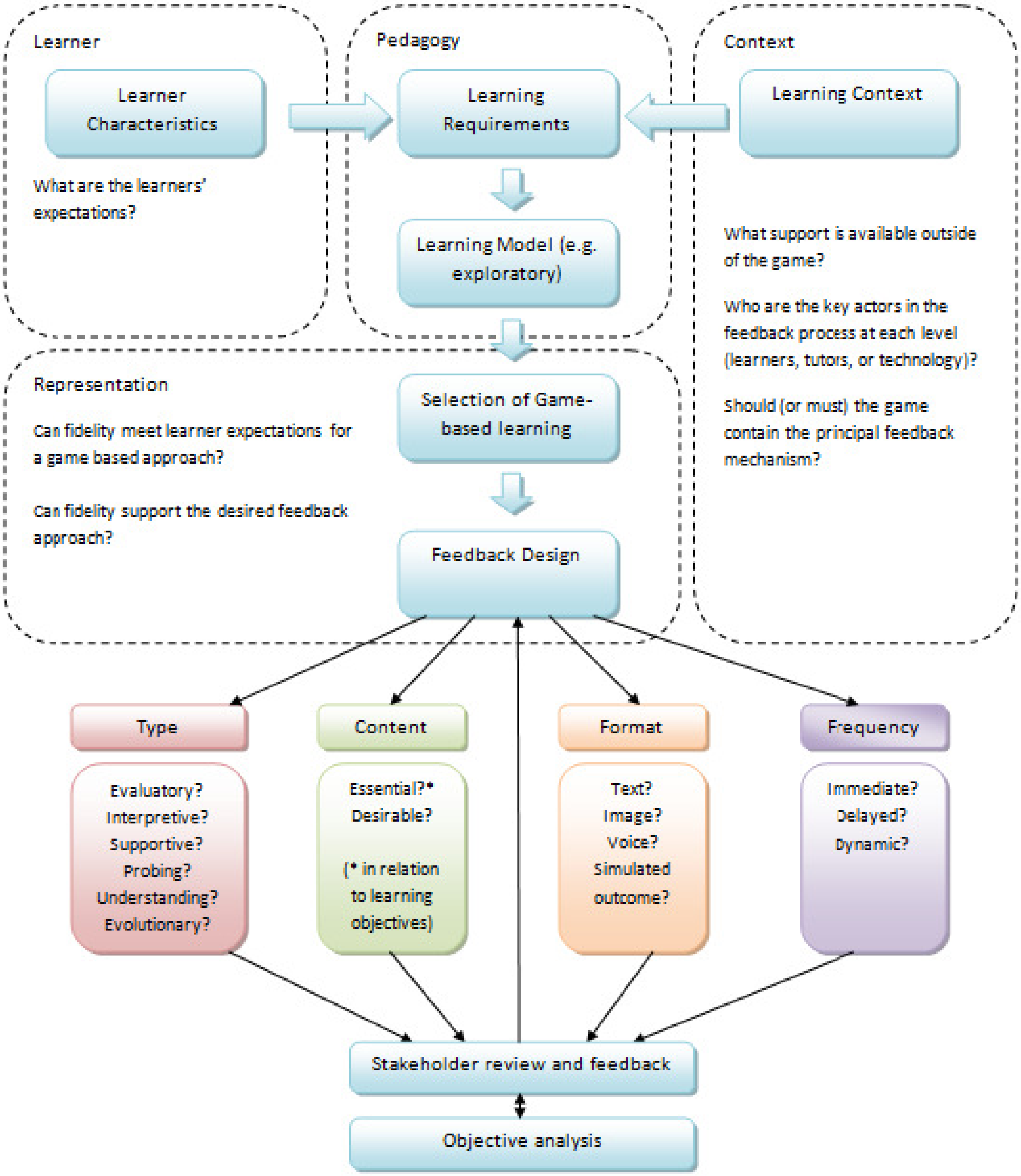


Figure 4.2: Four-dimensional approach to feedback in serious games (Dunwell, De Freitas, & Jarvis, 2011)

4.3 Related Work

Digital game-based learning that is based on commercial and proprietary computer games cannot easily implement high-level assessment features and special feedback mechanisms. But serious games and 3-D virtual learning environments – which are specifically developed for educational purpose – are well concerned with embedding assessment and feedback for formative and summative purpose. The extensive literature review covered in this section should summarise the current progress and challenges of integrated assessment approaches.

4.3.1 Integrated Assessment in Storytelling Games

Martínez-Ortiz, Moreno-Ger, Sierra, and Fernández-Manjón (2006) have reported on the integration of computer games as enclosed learning units into learning management systems (LMS). They have developed a game engine for storytelling games (adventures) that features an embedded assessment mechanism. A game is described by an XML document and related assets, such as images and sound files. The XML file defines a set of scenes, each featuring a background image, as well as exit areas that define transitions to other scenes. Items and characters are placed inside those scenes. They are considered the major medium for communicating the learning content of the game. The player is supposed to interact with the items and communicate with the characters through predefined dialogues. The game flow is then realised by adding flags (variables) that describe the state of the game and conditions which have to be met in order to activate items or to continue at certain exit areas.

Based on the game state, assessment rules can be defined separately by learning designers. This assessment rules are supposed to generate assessment notifications which could be send to the LMS while playing the game. Assessment notifications are discriminated by their type (e.g. achievements) and importance. The results are later compiled to a report and retrievable through the LMS. Although it is stated that the instructors have always access to the assessment notification – even during game play – there is no indication that students get supportive information whilst playing based on the assessment rule. (Martínez-Ortiz et al., 2006)

In subsequent works, e-Game – meanwhile known as e-Adventure – has been used to develop adaptive storytelling games. Based on a questionnaire, that is conducted before the actual game starts, the LMS can gather information on the status of the player and allow the entrance to the game at different stages. There have further been approaches to couple information about current knowledge (and also learning style) directly with the LMS profile of the learner and enforce adaptivity in-game. (see Moreno-Ger, Burgos, Sierra, & Manjón, 2007; Moreno Ger, Sancho Thomas, Martínez Ortiz, Sierra, & Fernández Manjón, 2007)

4.3.2 Competence-based Assessment in Computer Games

The work of Shute, Ventura, Bauer, and Zapata-Rivera (2009) was concerned with assessing complex behaviours based on hidden player observation (*"stealth assessment"*) in order to decide on the increase of competences. They have further stressed the possibilities for stealth assessment and formative assessment for feedback and dynamic adaption within immersive computer games.

The approach is based on the evidence-centred assessment design. It is supposed to assess competences rather than factual knowledge. Therefore, starting from desired outcomes as incorporated in a competence model, an evidence model is constructed that consists of a student model, as well as a set of rules that define the necessary behaviour to update this model. The student model reflects a concrete status of a single student in relation to the competence model. Finally, tasks are designed accordingly to provoke the required behaviour the evidence model depends on. The relations between the models is shown in Figure 4.3. An important fact about the evidence-based approach is that competences cannot be directly observed, as they are rather abstract concepts. That means, in order to make a decision about the mastery of certain competences, the approach involves a probabilistic evaluation method facilitated through Bayes networks. Thus, competences are reflected through conditional probabilities of their subordinate indicator variables. (Shute et al., 2009)

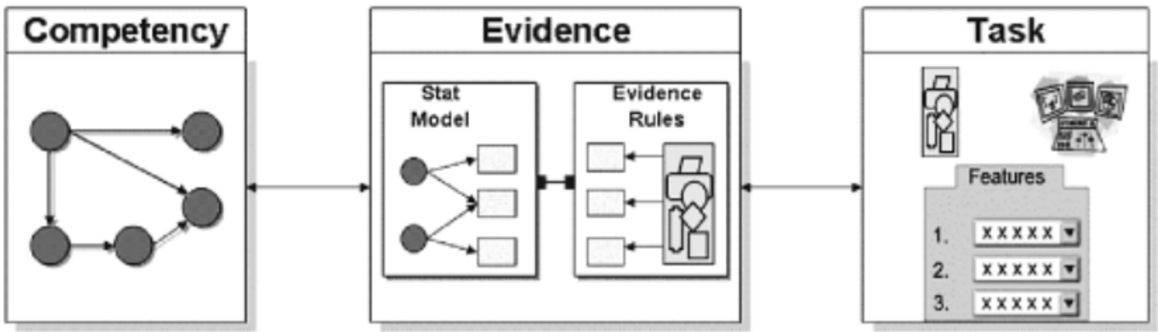


Figure 4.3: Models involved in evidence-centred design for assessment (Shute et al., 2009)

The evidence-based approach had already been realised by Bauer et al. (in 2003) in the context of a simulation for training network skills. This included embedded stealth assessment as well as accurate feedback for the students. The approach was ported to an example conducted within the immersive commercial computer game *The Elder Scrolls IV: Oblivion*. The goal was to measure creative problem solving based on the evaluation of different approaches to get across a river. Swimming was hardly an option due to the dangerous fish residing in the river. Nevertheless, considering its nature as a fantasy game, the environment enabled several fancy solutions, e.g. freezing the water. The actual evaluation was based on a log file provided by the game. (Shute et al., 2009)

Not completely different is the work of Kickmeier-Rust and Albert (2010) which also focused on a probabilistic estimation of competencies. But they concentrated on active feedback delivery through the game interface, however, using only interventions that sustain the immersive characteristics of the gaming experiences. Thus, the approach was named *micro-adaptivity*. It is based on a theoretical framework that is called *Competence-based Knowledge Space Theory (CbKST)* which is concerned with the connection between observable behaviour and the competencies associated with a specific knowledge domain. In order to provide feedback, player behaviour is automatically analysed, whereas different actions cumulatively contribute to the measurement of probabilities for the absence or existence of certain competencies. Possible feedback interventions include scenarios such as concrete feedback, hints or motivational statements, as well as dialogues between NPCs and the learner. (see also Kickmeier-Rust, Steiner, & Albert, 2009)

The approach was tested in the context of physics education, more precisely optics, as part of the *ELEKTRA* project. Implemented as a typical static computer game, the assessment system operates via different components. The state of game items is represented through a simplified, discrete position system (position category) that guarantees a finite game state. Interaction with the game leads to changes regarding the competence probabilities which further feed a reasoning logic. The logic provides recommendations for adaption in accordance with learning objectives and other aspects. This feedback is converted to a specific representation within the gaming environment based on the current context. The assessment system is based on an ontology, storing the required information and relationships between the different components. The game itself was realised as a 3-D adventure whereas feedback is manifested in the character of Galileo Galilei, functioning in the implicit role of a teacher. Finally, the requirement for generalising the approach and extend it to different application domains has been stated. (Kickmeier-Rust, Steiner, & Albert, 2009; Kickmeier-Rust & Albert, 2010)

4.3.3 Achievement-based Assessment and Feedback in Serious Games

This assessment approach, proposed by Dunwell, Petridis, et al. (2012), has been implemented in the context of a civil defence exercise regarding the evacuation of a school building on the off chance that an earthquake occurs. It is based on the enhancement of diverse content items during game development. This includes triggering specific actions such as taking a schoolbag or calling a lift (see Figure 4.4) but also the combination of several other conditions. For instance, the location of the player in certain areas can influence the correctness of specific actions, e.g. crawling, instead of walking, when passing an area full of smoke. Based on these rules, evaluative and interpretive feedback is provided. The environment is based on a non-linear setting that allows learners to explore correct and incorrect behaviour. Moreover, their conceptual framework also suggests that data of such serious games might be recorded and used for different assessment engines.

Figure 4.4: Different content items triggering events for achievement-based assessment (Dunwell, Petridis, et al., 2012)

The approach has further been externalised in order to provide information about player actions and attributes such as location to an external service. This service can evaluate the game events and generate feedback (see AL-Smadi, Wesiak, & Gütl, 2012; Dunwell, Petridis, et al., 2012).

4.3.4 Assessment of Competences in 3-D Virtual Worlds

Ibáñez, Crespo, and Kloos (2010) have investigated the potential of 3-D virtual worlds for the assessment of competencies from a theoretical perspective. Based on the *European Qualification Framework (EQF)* – as well as the *Secretary's Commission on Achieving Necessary Skills (SCANS)*, a United States framework – the ability to assess knowledge, skills and competencies has been examined. While knowledge assessment, and also skills assessment, could actually be based on similar techniques as in 2-D learning settings, competencies have been a missing or not well developed type of assessment. The latter holds also true for traditional e-learning environments.

Knowledge and simple skills assessment can follow a reoccurring scheme, based on a way to input information, possible responses and feedback. Starting from simple text-based interfaces, through images and other media, it is suggested that question dialogues and feedback can be provided through a NPC, even by pointing at certain 3-D objects. Beyond that it is mentioned that interactions of learners can easily be observed by an assessment engine, including mouse and keyboard interaction, player movement and manipulations regarding the environment or certain 3-D objects. Further, as part of social learning, fellow students could also provide feedback. The assessment of competencies is rooted in the fact, the virtual worlds can provide an extensive but problem domain focused environment that allows for complex interaction and tasks – involving a rich set of skills from knowledge acquisition to accurate application. They further proposed a case study, placed in the context of history teaching (industrial revolution), which they claim had been in development as of the publication of the proposal. On the competency level it should focus on students' collaboration. (Ibáñez et al., 2010)

4.3.5 Further Assessment Approaches in 3-D Environments

Crisp, Hillier, and Joarder (2010) did a review on current assessment approaches in virtual 3-D worlds. They tested *Sloodle*, a project that links Second Life (SL) with *Moodle*, a popular learning management systems (LMS). It includes features to access Moodle resources, such as question items, load presentations and display them within SL, as well as facilitate the distribution of learning materials across students. There is further a system that allows for students to display their status on learning tasks inside Moodle. A sample scenario was created by the authors to evaluate the quality of this approach. They reported about general usefulness, also emphasising that teachers can easily design learning settings – including assessment – without knowledge about scripting languages. The reuse of existing Moodle content is pointed out as well. Another assessment tool that has been investigated is *QuizHUD* (Quiz Heads-Up Display). It can also be used with SL but is not integrated with an external LMS. QuizHUD associates items or parts of items with response actions as part of tasks such as multiple-choice questions. Here, too, scripting is not required from teachers. From the users' perspective, QuizHUD is once activated and available for the user whilst exploring the world. Finally, they were conducting own experimentations on scripting primitive objects and chatbots to implement assessment strategies. Their conclusion was that these approaches generally have potential, but further research is necessary to provide simple interfaces to enable sophisticated assessment to be designed by teachers with less programming skills. Crisp (2012) reported further about a first solution approach dealing with this issue. An editor for Moodle allows to configure simple chat dialogues associated with events on prims (primitive SL objects).

Arroyo et al. (2010) reported about the integration of the *Question and Test Interoperability (QTI)* standard with Open Wonderland. This standard covers the exchange of question items and results, such as multiple-choice questions and other similar formats. While an external QTI interpretation system (web service) was used to load XML files – containing material following the QTI specification – an additional XML definition was developed to determine interaction methods for responses in 3-D virtual worlds, as this is not covered by QTI itself. This sustains the flexibility already provided by QTI. They have implemented response systems such as spatial zones to be entered or 3-D objects to be clicked in order to give answers. Feedback and additional information is also displayed on the HUD to guide the learner, as further questions can be spread around the scene.

4.4 Identified Issues

Chapter 2 has already pointed out that assessment – especially formative aspects – is essential for learning. Nevertheless some aspects of assessment and especially feedback are inherent in all kinds of virtual 3-D environment, as part of their medial preconditions – a game without basic response would not be considered a functional game. However, the related work has shown that there is a great interest on assessment from a pedagogical and instructional perspective. Also high-level feedback in immersive environments such as serious games is considered an important aspect. Nevertheless, based on the related work several issues have been identified as outlined in this section. Parts of this section are based on (Maderer, Gütl, & AL-Smadi, 2013).

4.4.1 Common Issues

The following more general issues have been extracted from the related work:

1. *Interoperability and flexibility*: according to AL-Smadi, Gütl, and Helic (2009) e-assessment systems should be flexible and following standards in order to cope with the amount of LMS available on the market. This includes specifically the proposal of an service-oriented approach to provide (among other requirements) *interoperable* and *reusable* assessment technologies among different application settings. Furthermore, flexible techniques should also increase the control of the

instructor to provide assessment mechanisms after the development of a serious game has finished (AL-Smadi, Wesiak, & Gütl, 2012).

Although the integration of existing e-learning standards, such as QTI, as well as connecting existing LMS, has been shown in the context of 3-D virtual worlds (cf. Arroyo et al., 2010; Crisp et al., 2010), it is still limited to simple assessment formats and basic content. Several related work presented in this section has been based on a prototypical approach that did either rely on an unspecified, probably platform dependent log file (Shute et al., 2009), or scripted, respectively programmed the entire assessment tools based on a specific selection of platforms such as native game-engines (Kickmeier-Rust & Albert, 2010; Dunwell, Petridis, et al., 2012) as well as Second Life (SL), or a combination of SL with Moodle to react on rather simple user actions (Crisp et al., 2010; Crisp, 2012). Thus, assessment strategies focusing on advanced interaction scenarios in immersive environments will also require an increasingly flexible and interoperable interface standard in general. The externalisation of the achievement-based assessment approach in game-based learning (see AL-Smadi, Wesiak, & Gütl, 2012) is a first step towards such a flexibility. It is the direct basis of this work which will be further explained in the next Chapter;

2. *Usability for assessment designers*: as pointed out by Crisp et al. (2010) achieving (sophisticated) assessment through extensive in-world scripting is not feasible for most instructors. While it can certainly not be expected from teachers to be expert programmers, it can only be less expected to be expert on several different virtual world platforms. As it has already been shown in Chapter 3, virtual worlds follow a complex architecture involving client and server aspects to be considered. This emphasises the need for appropriate editing tools to fulfil the requirement of e-assessment systems being *accessible* (AL-Smadi, Gütl, & Helic, 2009). Editing question items, dialogues and simple search tasks for immersive environments is already a feature in LMS such as Moodle (cf. Crisp, 2012). The challenge is now to provide the same level of usability also for more sophisticated and complex assessment scenarios in immersive 3-D virtual worlds and game-based environments.

4.4.2 Lack of Behavioural and Authentic Assessment in 3-D Virtual Worlds

Placing the assessment items within an authentic virtual world does not automatically make the assessment itself more authentic. Using only methods such as multiple-choice questions and similar fixed response questions, but even chatbots, appears rather limited for domains such as science education. For complex physics simulations and scientific tasks it will hardly suffice; especially regarding continuous formative assessment and immediate feedback in the context of complicated experimentation tasks. (cf. Crisp, 2012)

The first examples in the domain of game-based learning are certainly more authentic. Shute et al. (2009) had reported on the usage of evidence-based design for *stealth assessment* to automatically estimate competency levels based on sequences of actions provided by an immersive game, but had also mentioned the potential on feedback and adaption. Even more advanced is the realisation of actual pedagogical intervention in the context of the ELEKTRA project (Kickmeier-Rust & Albert, 2010). But literature has not yet revealed assessment approaches for 3-D virtual worlds – such as Open Wonderland or Second Life – that would automatically evaluate complex player behaviour and provide immediate feedback. Related work rather indicates that pedagogically relevant intervention is built upon key interactions. Real behavioural assessment – understood as evaluating complex sequences of actions, as well as environmental consequences and relations – appears to be missing (cf. Crisp, 2012). However, Ibáñez et al. (2010) emphasised that recording all kind of users actions within virtual worlds is not a problem, and should therefore be used to assess competences. But it is not clear at this point if this statement refers only to the general interception and suitable preparation of user actions, e.g. for human interpretation, or a systematic and automatic evaluation of entire sequences of events.

Nevertheless, classroom teaching (also practical university courses) is based on teacher observation, having a great perspective of the overall situation, including the possibility for giving feedback and guidance. The level of feedback will also be based on the concrete foreknowledge and skills of the group or individual learner. Thus, it can be concluded that transferring teaching lessons into 3-D virtual worlds will also require appropriate arrangements in-world. Feedback in real exploratory settings is not based on students answering multiple-choice questions. There is maybe the possibility for intermediate questions of the teacher to guide the learner towards the right direction, especially after an obvious error was made or questions are asked from students. However, error detection will be based on series of actions and environmental observations before intervention starts. Based on these deliberations, the lack and requirement for automated behavioural assessment techniques has been identified. That such e-assessment system should be automated is aligned with the requirement for reducing workload but also due to the permanent availability of e-learning systems. Otherwise it would be necessary that teachers and tutors are available online and in-world all the time. That feedback itself is important for the learning process, has already been stated several times.

4.5 Summary

E-assessment ranges from supportive activities in traditional assessment settings to entirely computer-based assessment approaches. The motivation for e-assessment is broad. Starting from simple logistical reasons due to the large numbers of students, also many pedagogical benefits arise from e-assessment. It is fairer for students and formative assessment can be better supported because of the immediate feedback that is technically possible. Particularly modern educational objectives are believed to be better evaluated through extensive computer-based activities. But there is also criticism regarding the current implementations and it does not seem that the average teacher has the chance to use much more than simple question formats. Despite the separation into diagnostic, formative and summative assessment several automated and non-automated assessment strategies exists. These include not only teacher-based assessment but also self or peer assessment. The theoretical framework of Wesiak et al. (2012) is supposed to provide an appropriate application of assessment strategies in accordance with learning objectives and complex learning resources. Whereas Dunwell, De Freitas, and Jarvis (2011) has more closely examined feedback strategies for serious games and similar environments.

Recent literature has shown different approaches to assessment in digital game-based learning environments, as well as 3-D virtual worlds. Some approaches are concerned with the assessment of competences, such as the evidence-based design approach or the general proposal of how competencies can be assessed in learning environments based on 3-D virtual worlds. Further projects have demonstrated the integration of e-assessment standards (QTI) into learning settings placed within 3-D virtual worlds or the enhancement of learning settings with scripted assessment items. Storytelling games have already been completely integrated with LMS.

Several issues have been identified within the related work. Current approaches are not flexible enough, support only integration of existing LMS materials (questions items or traditional 2-D content) and do – as far as 3-D virtual worlds are concerned – not provide automated assessment features that would rely on analysis of complex player behaviour. Although some promising approaches have been shown in the context of computer games, no porting of the concept to 3-D virtual worlds has been reported. The next Chapter introduces the practical part of this thesis and promotes a flexible assessment approach based on the discussed issues.

5 Requirements and Design of a Flexible Assessment System

The literature review has shown that STEM education is a challenging field. But computer-supported approaches, such as immersive environments are promising tools to improve education in general. These include computer games, simulations and 3-D virtual worlds. Several pedagogical benefits are based on the idea that students construct knowledge in a social context. Nevertheless, assessment and feedback are not only important for learning but have also been found integral part for the success of computer games. Furthermore, guidance and supportive information is not less important in exploratory simulation settings. The related work on assessment in immersive environments indicates several outstanding issues. Current approaches are less flexible, hardly feasible to be used by teachers and consider only simple actions or fixed response question formats. Behavioural approaches to assess competencies are promising but less developed at this time, particularly in 3-D virtual worlds.

Based on these issues, the aim of this Chapter is to define requirements for a flexible assessment approach and propose a general flexible assessment framework. Furthermore, a reduced and feasible set of actual prototype components is designed and applied to a test scenario in the domain of physics education.

It is important to note that the first findings about the practical part of this thesis have already been published (Maderer, Gütl, & AL-Smadi, 2013) and further improvements have been submitted for publication (Maderer & Gütl, in press). But the author of this thesis was responsible for the design and implementation of the entire technical parts. Some parts of the paper, as well as certain images will be used in the remaining work. These are cited as necessary, as well as if something was not contributed alone.

5.1 Main Idea and Objectives

The initial idea for this work emerged from requirements of the *ALICE project*[1], an abbreviation for *Adaptive Learning via Intuitive/Interactive Collaborative and Emotional*

[1] http://www.aliceproject.eu/

systems. The project was motivated by several outstanding issues in contemporary e-learning software, namely, the lack of interaction, challenge, empowerment and social identity. Therefore the project was devoted to the improvement of learning activities, including encouragement for collaborative learning, the usage of simulations and serious games as complex learning objects, as well as new forms of assessment. The latter should not only be able to evaluate the outcomes of complex learning resources, but also provide information to improve activities by updating several models, such as the learner model that represents cognitive state and preferences of learners. (ALICE, n.d.) Parts of work package *"WP5: new forms of assessment"* have been implemented by the AEMT[2] group at Graz University of Technology. Towards the final months the author joint as a project fellow and was responsible for the externalised assessment approach applied to the civil defence game that has already been mentioned in the related work section (see AL-Smadi, Wesiak, & Gütl, 2012; see further AL-Smadi, Gütl, Dunwell, & Caballe, 2012). That means further ideas presented here are based on this initial work and include knowledge at first hand.

Based on the first findings, this thesis proposes an improved and generalised conceptual approach for the integration of flexible assessment in different immersive environments, including game-based environments as well as 3-D virtual worlds. The practical part is following two main objectives:

1. Enhancement of a 3-D virtual world platform – in this case Open Wonderland – in order to provide the necessary means to analyse the behaviour of user actions relevant for formative and summative assessment;
2. Implementation of a virtual physics experiment in accordance with the assessment approach to extend the domain of application into science education and examine the assessment approach with real-time simulations and laboratory tasks.

The prototypes are further intended to examine the possibilities of immediate learner feedback, as well as to evaluate its general applicability for competency-based learning objectives. For that reason, the prototype is demonstrated to a group of experts in order to evaluate the approach and collect feedback for further improvements.

5.2 Flexible Assessment Framework

This section is supposed to subsume all important requirements for a flexible assessment framework that focuses on the analysis of complex player behaviour. Based on these requirements a general solution approach and conceptual architecture is presented. This section is an extended version of Maderer, Gütl, and AL-Smadi (2013).

[2] http://www.iicm.tugraz.at/about/Homepages/cguetl/team

5.2.1 Requirements

The requirements presented in this subsection are based on the first findings reported in Maderer, Gütl, and AL-Smadi (2013) but further extended and refined in accordance with findings and suggestions of the related work section of this work (cf. section 4.3) as well as Chapter 4 in general. These are general requirements relevant for a flexible assessment approach and are not supposed to reflect the actual outcome of the prototypes development.

Functional Requirements

The following four groups of functional requirements have been declared:

- *Interoperability of Computer Systems*
  - Support for different immersive environments (game engines, 3-D virtual world platforms, etc.)
  - Interchangeable assessment system ("assessment logic")
  - Integration of external information systems, including learning management systems (LMS)
- *Flexibility of Assessment Methods*
  - Support for different assessment paradigms, application domains and knowledge domains
  - Configurable for different learning objectives
  - Useable for different assessment types (formative and summative aspects)
- *Adaptivity*
  - Consideration of individual or group-level learner preferences
  - Import and export (update) of learner status (e.g. knowledge and skills level, task achievements or competency models)
- *Feedback and Guidance*
  - support for different feedback mechanisms based on the available resources of a specific platform (e.g. different content formats, including text or audio).
  - Support for different guiding systems (for instance NPC)
  - Identification, navigation to and manipulation of in-world objects to enable guidance systems for direct interference.

Non-Functional Requirements

There are two general non-functional requirements that must be supported by a flexible assessment framework:

- *Usability concerns*: It must be possible for instructors to design assessment tasks (e.g. assessment rules) as simple as possible. More precisely – in accordance with the actual assessment system used – the skills required to design assessment tasks must not exceed the logical affordances of basic script programing in terms of complexity. Especially any requirements regarding platform-dependent expertise in terms of software development (scripting) must be completely avoided.
- *Performance*: The entire evaluation and feedback cycle must comply with real-time measurements. Hence, it must be possible for learners to get immediate feedback and interact with dialogue based feedback systems dynamically.

5.2.2 Solution Approach

The main idea is a combination of several considerations. Binding assessment rules to game states is simple for a finite state machine as used in adventure games (cf. Martínez-Ortiz et al., 2006; Kickmeier-Rust & Albert, 2010) but turns out to be practically impossible for an open-ended environment. Thus, the first approach for serious games focused on domain specific measurement variables that describe the current environmental situation of a single player as well as the actions the player performs. For instance, if a player is currently present in a critical area and performs a pedagogically wrong action the external assessment system can raise an appropriate feedback message. But also adding additional close objects with its own properties has already been considered as an option in the early prototype. Nevertheless, the knowledge about how these events have been assembled in detail is limited, as the civil defence game has only been delivered closed source to this work group.

Besides, software agents are a popular concept also used in 3-D computer games for several years. These autonomous systems are usually bound to a NPC or similar items and use different levels of artificial intelligence approaches to react on sensory information from the environment. The quality of sensory information and internal logic can range from the simplest conditional expressions up to complex decision systems. (see Nareyek, 2001) There are also recommendations about semantically self-descriptive virtual objects in 3-D virtual environments (see Schmeil et al., 2012; Tutenel, Bidarra, Smelik, & Kraker, 2008).

Considering the required flexibility between several platforms and similar application domains as well as the objective to support complex behaviour analysis it seems reasonable to focus on a semantic-enabled approach that suits the needs of software agents on different levels. Based on that an external system can observe the situation and interfere without the need to be aware of details about the target platform. Based on the original usage in the game-based context, and in compliance with these considerations, two types of useful information have been identified for a flexible assessment framework and need to be available for external systems:

- *Environmental state*: involves all properties of virtual objects, as well as properties of spatial sections (e.g. weather conditions, or the sound of an alarm) that appear within the virtual environment (cf. Tutenel et al., 2008). It is important to note, that only such properties are of interest, which have a meaning for human players, and could be (at least indirectly) observed by them. Otherwise, the information would not be useful to provide meaningful and comprehensible assessment and feedback for the player. Under this term it is further understood that also avatars represent environment-relevant objects, especially for other players that might be involved in collaboration and observe the avatars of each other.
- *Player actions*: describe the user interactions players perform through their avatars on virtual objects or with their avatars. In contrast to environmental conditions, actions define either a single event located in time or the transition between environmental conditions. That means, a player action either stands alone or is combined with the update of an environment condition at the same time. The latter is especially interesting in that it provides a significant difference between the actual observation of the action, or the observation of the persistent outcome, once any observer comes into range of perception, after an action occurred.

The sum of all environmental states which are within the spatial range of perception of a certain player, define the situational context for that player. The assessment of actions a player performs is than evaluated in regard of that situational context. That means player behaviour is only considered wrong or right in terms of the available information a player could have had, not information that arises somewhere else within the virtual environment.

5.2.3 Conceptual Architecture

Based on the first approach in game-based learning an improved and generalised conceptual architecture is supposed to fulfil the requirements stated in the previous sections. The concept is basically built upon a three-tier architecture (see Figure 5.1), which differentiates between *immersive environment*, *middleware* and *background systems*.

The architecture consists of the following components:

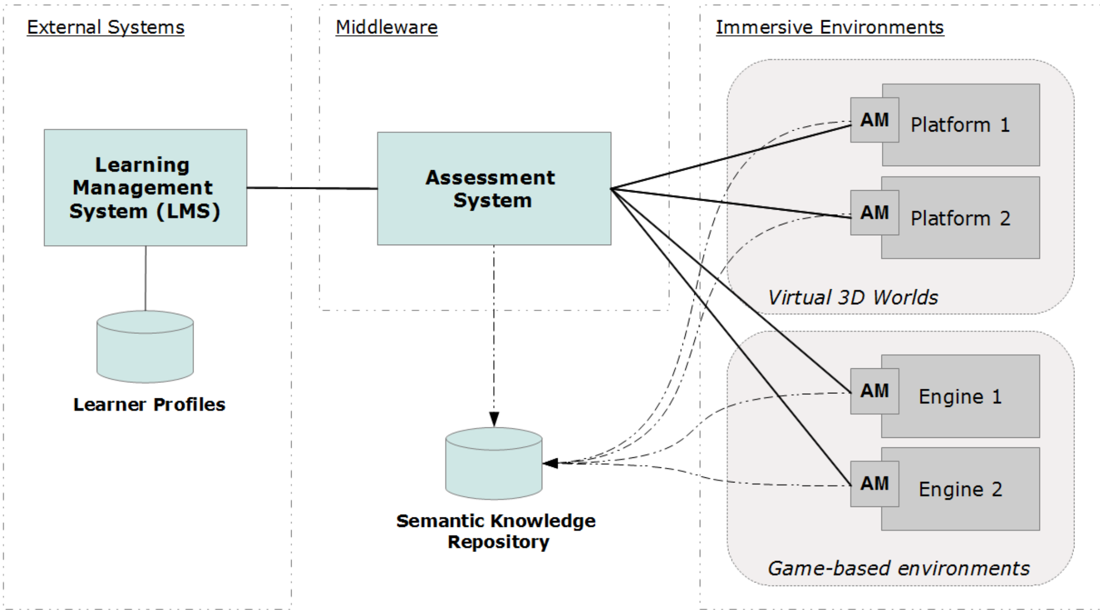


Figure 5.1: Conceptual architecture of the generalised flexible assessment approach (Maderer, Gütl, & AL-Smadi, 2013)

- *3-D virtual world* and *game-based environment* refer to possible types of immersive environments, each represented through a couple of different platforms;
- *Assessment system*: an assessment system implements the necessary logic to evaluate the behaviour of users based on the given environmental context (conditions of objects and places). The actual evaluation approach is unspecified. Nevertheless, the communication between immersive environment and assessment system must follow a specified protocol that includes the reception of events from the immersive environment and the delivery of feedback messages back to the immersive environment.;
- *Semantic Knowledge Repository*: represents an information contract between the assessment system and immersive environments. This information should provide metadata to describe places, objects, conditions of objects and possible user interactions. The repository should further provide a library of reusable descriptions throughout different knowledge domains, including a common set of basic items required for most 3-D virtual environments. On an (optimistic) outlook it can be understood as a central database offered through a web service that allows access to standardised and reusable semantic annotations that can be applied to 3-D elements; (cf. Tutenel et al., 2008)
- *Assessment module (AM)*: a software component (e.g. a module or plug-in) has to

be implemented once for a specific platform, in order to provide mechanisms that allow to track information about the state of the environment and user behaviour. In addition, the module should also feature a set of support functions to play feedback and guiding information which is generated through the assessment system. What kind of feedback can actually be used will depend on the available resources of the platform. Preferably, text-based feedback messages – possibly communicated through an NPC – as well as pre-defined dialogues between the assessment system and the player should be supported. Beyond that – based on the specifics of the platform – design time tools may also provide the necessary infrastructure for world designers to annotate places and objects on a semantic level.

The assessment module is further responsible to pre-process events occurring in-world such that information is useful and detailed enough to support the remote evaluation process but does not overstrain the network connection;

- *Learning management system (LMS)*: The LMS essentially represents the most important external system. Through the LMS the assessment system should be able to access the learner profile which enables the adaption of feedback in terms of details, form or frequency based on preferences and foreknowledge;
- *Learner profile*: The learner profile stores information about individual learners and provides data about task achievements, knowledge and skill levels, or competency models. The assessment system is supposed to update this data based on the assessment outcomes.

The supposed workflow for this approach is that world builders and developers prepare a semantically-enhanced environment, whereas instructors design or configure assessment rules in the context of the external assessment system. The approach focuses only on conditions and user actions. This should guarantee that also more concrete resources such as fixed-response questions and dialogues can be framed within the approach. Nevertheless, due to the complexity it is clear that the entire framework cannot be implemented in a single step. Especially the research and development for standards regarding the semantic knowledge repository will require further attention, as well as the connection to learner profiles via learning management systems. Based on its central importance an assessment module for Open Wonderland has been chosen as the next logical development step.

5.3 Technical Overview of Open Wonderland

The virtual world platform Open Wonderland (OWL) has already been introduced in Chapter 3. OWL was chosen as target platform for a few reasons: First, there are already several research projects focusing around learning in virtual environments; which includes also projects of this institution. Second, it is open source and features an extensive API for customisation, which would have been more complicated with commercial systems like Second Life. Based on these considerations it was standing to reason to take this platform for a first prototype development of an assessment module. This section will explain the most important technical aspects of OWL, including its communication architecture and all crucial core technologies that have been used, as these technologies also influence the development of further modules.

5.3.1 General Architecture

OWL is solely based on Java technology and relies on several open standards and libraries. An overview of the architecture is depicted in Figure 5.2. On the server side, four independent server components provide the infrastructure of OWL. A *web server* is responsible to host the management interface that allows for the control and configuration of the remaining services and offers a single sign-on mechanism required for all other services. Besides that, it also works as a repository for all installed modules, as well as other assets – including 3-D models, textures, and other resources. The other services use appropriate protocols, according to their specific task. The *Darkstar server* is the primary component responsible for the actual in-memory representation of the 3-D space and provides the necessary means for synchronous communication of object state and interaction between clients and server. Thus, it is the minimal required component beside the web server in order to operate a virtual world instance. The *shared application server (SAS)* is restricted to Linux and Solaris systems, as it provides collaboration on conventional desktop applications that run on the *X Window System*. The SAS can collect the graphics output and forward it to all participating clients, as well as deliver input events that have been conducted in-world to control the application that is running in the background. Finally, the *voice bridge* enables real-time audio communication. Nevertheless, it is possible to provide additional custom service nodes. The *Wonderland Client* loads all required resources from the server-side module and asset repositories. *MT Game* engine and the underlying *JMonkeyEngine* are responsible for the rendering and calculation cycles, supported by a layer of core services and a communications API. (Kaplan & Yankelovich, 2011)

The following subsections will more closely explain the Darkstar server, MTGame and JMonkey Engine, as well as the most important OWL terms and concepts.

Darkstar Server

Project Darkstar (PDS) does not longer officially exist, but has been released as a new open source project that is now known as *RedDwarf Server (RDS)* (cf. Kaplan &

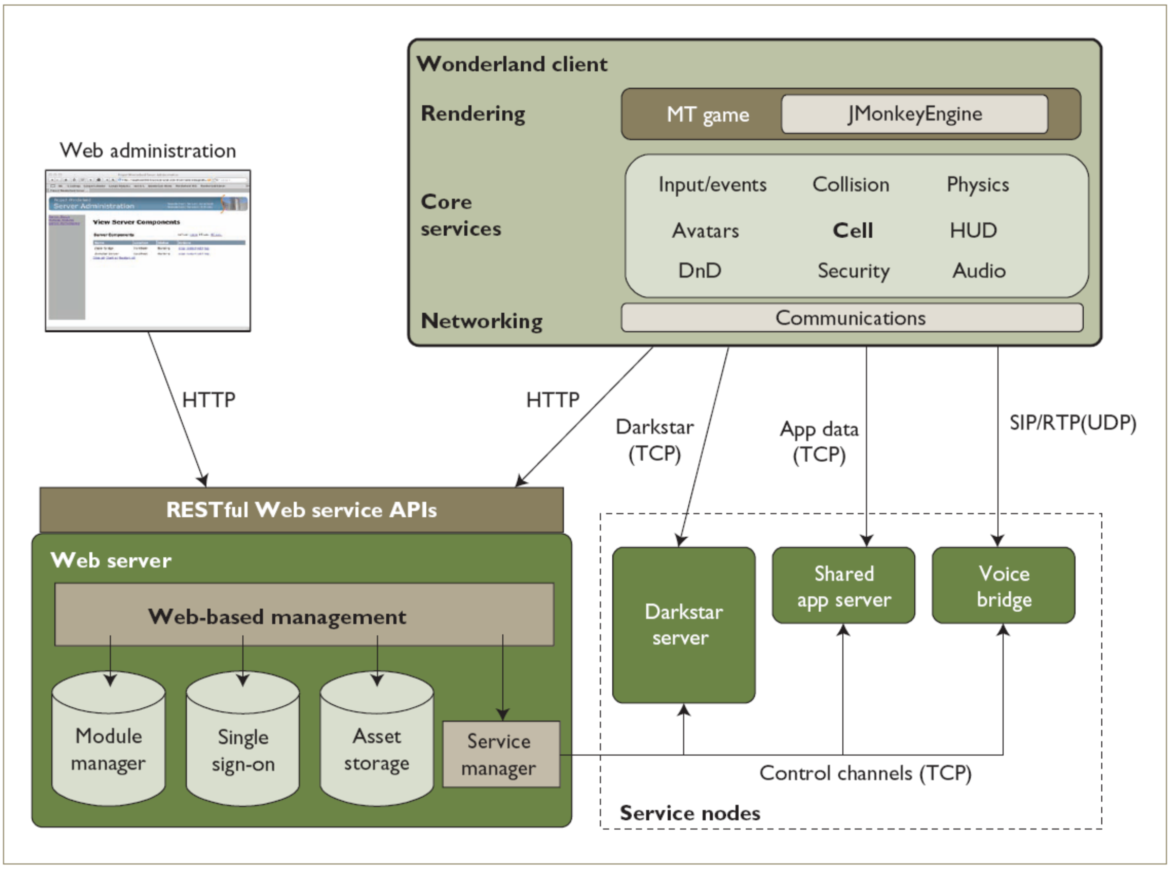


Figure 5.2: Open Wonderland architecture (Kaplan & Yankelovich, 2011)

Yankelovich, 2011; RedDwarf Contributors, 2010). Unfortunately, PDS is still inherent in all documentation surrounding OWL, but as no enclosed documentation could be found about PDS anymore, the following technical information is taken from the documents available on the RDS website[3]. Nevertheless, the reaming work will continue to reference it as *Project Darkstar* and *Darkstar server*.

Project Darkstar provides essentially the backbone of OWL. It addresses the issues of virtual worlds regarding persistence and multi-user access and aims to ensure scalable and robust software design. The Darkstar server offers an event-driven programming model that allows application developers to design software in a single-threaded manor. Nevertheless, the server is capable of parallel execution and keeps track of concurrency issues. This is achieved by an integrated architecture that relies on the following key concepts (RedDwarf Contributors, 2010):

- *Tasks*: a task is an object that represents the smallest transactional unit of programming logic. That means a task only effects the environment after it

[3] http://sourceforge.net/apps/trac/reddwarf/

has successfully finished; otherwise, all changes are reverted. This includes even network communication, which is not release before the task has finished. Consequently, this requires tasks to be finished very quickly. Tasks are either scheduled implicitly or via the task manager, which allows the application developer to start periodic or delayed tasks;

- *Managed objects* and *managed references*: these two concepts support both, the thread-safe execution of concurrent tasks, as well as the persistence mechanism. A managed object marks the smallest chunk of an object graph that can be stored or retrieved by the system. Managed references connect different managed objects. Basically, tasks access objects through managed references for read access. If a task requires manipulating an object, it must be accessed for update in order to get exclusive write permission on a copy. This copy is reintegrated after the transaction has been committed. Typical examples for managed objects include entities such as 3-D objects and player items, as well as avatars but also any other kind of internal support structure required for the application. Beyond that, also scheduled tasks are persisted and can be activated after a server crash and reboot;
- *Managers*: certain code cannot be executed in the simplified context of a task that is only allowed to run for several milliseconds (100 ms is the default value). Therefore, the Darkstar API provides several manager classes which give access to external services, such as database interfaces and other IO systems, but also the task manager itself. Tasks deploy requests and receive responses asynchronously. Therefore, such services can run significantly longer as conventional tasks;
- *Channels*: provide a communication interface to send messages between clients and servers. The project is not limited to Java clients, in fact several implementations exist for different programming languages ("RedDwarf Website", 2013).

JMonkeyEngine and MTGame

The *jMonkeyEngine (jME)* is a typical 3-D scene graph toolkit with many features. However, it is only single-threaded and is not supported through an appropriate processing model that would facilitate a simple implementation of real-time applications. Based on that *MTGame* was developed as a *"multi-threaded game engine"* around jME to compensate these issues. In comparison to other game engines, MTGame introduces multi-threading to separate computation among 3-D entities, rather than on the problem domain, such as physics, rendering and artificial intelligence. This approach is supposed to promise increased scalability in future. Also here, all primary synchronisation issues are implemented transparent for the developer. A component model is supposed to make 3-D entities easily extendable, in some cases even without the need for additional programming. (Twilleager, n.d.)

In order to create an animation, a *processor component* has to be implemented and attached to an entity. Basically, MTGame divides the process into a general computation phase – that could be easily computed concurrently among different threads – and the actual rendering phase that happens only on a single renderer thread. Thus, the rendering methods on all objects' renderers are called sequentially. It is further recommended that the computation of animations depends on passed time intervals instead of amount of frames, as the latter is not guaranteed to be constant. (Twilleager, n.d.)

5.3.2 Important Wonderland Concepts and Terminology

This subsection describes the most important terms and concepts used in the development of OWL features. The remainder of the work will refer to them frequently. This section as well as the remaining information acquired about OWL is based on several spread resources, including Kaplan and Yankelovich (2011) and the OWL website (Open Wonderland Foundation, 2012), more precisely different tutorials provided in "Open Wonderland Documentation Wiki" (n.d.), as well as personal source code studies (see "Wonderland: Subversion Repository", n.d.; "Wonderland Modules: Subversion Repository", n.d.) and Java source code documentation ("Open Wonderland: JavaDoc", n.d.).

Cells and Cell Messages

A cell is essentially a 3-D object. It has a representation on the clients as well as on the server. The server-side version is supposed to represent the synchronised and persistent state of a cell, whereas client cells are mostly responsible for interaction, rendering, animation and detailed physics calculation. Therefore the server-side object implements the `ManagedObject` interface of Darkstar by extending the base class `CellMO`. Information between clients and server is facilitated through several communication concepts: If configuring a cell, the server state is loaded directly into the client editor tools in order to make changes. If a cell gets required to be considered on a client, the server will sent the current client state – almost identical to the server state – to the client. Dynamic changes are communicated through `CellMessage` classes. Cell messages are an extended version of Darkstar messages, but are already routed towards specific cells. In more generic terms, a cell could be interpreted as a spatial communication node between clients and server – comparable to a cellular phone network – capable of displaying any kind of graphics within the 3-D space but also additional items such as *heads-up display (HUD)* components (2-D windows) on the surface of the OWL client window.

Components

This concept is known as capability to users. A component extends the functionality of a cell through composition instead of inheritance. Cells are usually slim containers and even the core functionalities – such as cell positioning and message channels – have been transferred into components. However, such components have been registered as default components and will be added each time a cell object is created on the server or client. Components actually share the same paradigm with cells. They can consist of server-side managed objects and client-side objects, feature server and client state and communicate through cell messages.

Plugins

Plugins are used to provide general customisation to the Wonderland server or client. Such plugins are automatically activated and executed when the server starts up or a specific client connects to a Wonderland server. It can be used to initialise any kind of modification in the context of Darkstar and Wonderland, including the setup of new menu items on the client or the registration of default components.

Module

A module is an administrative software packaging unit. Although realised through a Java Archive (JAR) it does not directly contain Java classes. The archive features again a separate JAR for client and server code. Based on convention, Java code for OWL modules is divided into three Java packages ('namespaces') called `client`, `server` and `common`. The built process yields two archives which either contain client or server classes and attaches the common classes automatically to both of them. The common package contains all classes that have to be shared between client and server such as message objects. Beyond that also artwork and dependent libraries can be stored in the relevant locations. Dependencies to other Wonderland modules are managed through a versioning system and the related requirements are stored in an XML file within the module.

5.4 Assessment Module

Based on the technical overview of OWL this section explains the technical considerations and the conceptual design of an assessment module. Most paragraphs of this section have already been published in Maderer, Gütl, and AL-Smadi (2013) and are borrowed almost word for word with explicit consent of the co-authors.

5.4.1 Feature Analysis and Design Considerations

Prior to the actual development, an analysis of the current state of OWL regarding the aspects of semantic event tracking for assessment purpose was conducted. This relates to the interception of user actions, available information on cells (3-D objects) and the detection of location changes of the avatar. This information has particularly been determined through personal examination and extensive source code studies (refer to the previous section for all available and used sources).

User Interaction

On the client-side, *user actions* are implemented through different mechanisms, including context menus, control panels (also available in-world) and direct processing of mouse and keyboard events within the 3-D space. While basic mouse and keyboard events are less interesting due to the raw information they provide, actions revealed by context menus and control panels were examined more closely. It has been found impossible to derive information that can be used to automatically create semantic events, because there is no guarantee that actions will expose any information beyond graphical or textual representation required for the user interface. Especially textual representations depend on the current language setting of the client; therefore it does not render a valid source for independent information. As a considerable amount of actions have to be shared with other clients, cell messages are finally sent to the server to request those changes. Although these messages can be intercepted, there is still the problem that messages are simple serializable and derived Java objects, which do not feature a common interface to expose semantically usable information.

Cell Status

Available *information of cells* is rather limited. Besides a name there are no attributes available that describe the purpose of the 3-D object. As cells are implemented as derived classes and stored in different modules, all information and behaviour is encapsulated. Consequently, there is no common way to determine which kind of user interaction with an object is supported. This concerns also the status of an object, whereas in that case status refers to something that could be understood and observed by a human being, e.g. the colour of a traffic light or something similar.

Avatar Location

Regarding the current location of an avatar, it is not complicated to obtain the corresponding 3-D coordinates. However, the usage of exact coordinates is not desirable, as the assessment system should be able to work with different scenarios, thus not relying on coordinates, but rather on abstract descriptions of location on a semantic level. Furthermore, there is no predefined mechanism for tagging spatial areas. OWL features enable the concept of proximity listeners, which can define 3-D spatial areas (called bounding volumes) that trigger events whenever an avatar enters or leaves such an area.

5.4.2 Conceptual Architecture

As the previous section highlighted, retrieving semantically useful information from OWL without further arrangements appears to be impossible. Nevertheless, a few features have been identified that can support the development of an appropriate tracking mechanism. Therefore this section will introduce a conceptual architecture for an assessment module that includes tracking capabilities and facilities to provide feedback and guidance to the learners. Although there is no dedicated external design tool available for OWL, the assessment module is supposed to cover the following two phases:

- During *design time* world builders are not only supposed to layout the world but also provide additional annotations that support the assessment approach. This includes the creation of spatial sections that can be tagged with information describing the location of the avatar in abstract terms. Beyond that, attributes can be applied to simple (static or animated) objects in order to describe them semantically. Interactive objects on the other hand are already supposed to provide all means necessary for interaction and state tracking internally. Nevertheless, for both kinds of objects distances can be configured to describe proximity in discrete terms. In this phase – which is activated through a privileged command – all annotation items get visible to the world builders.
- In the *runtime* phase the assessment module is supposed to track events that consist of user interactions and environmental changes. These events will be forwarded to an external assessment system which will evaluate the behaviour based on its internal logic and return feedback messages and guidance actions to the OWL server. From there, feedback can be forwarded to clients included in the assessment plan.

Due to the architecture of OWL itself, the assessment module (see Figure 5.3) is divided into aspects that belong to either the client context or the server context. The conceptual architecture has the following components:

1. *Client 3-D object* and *shared 3-D object* refer to a typical pair of client-side cell and server-side managed cell objects in OWL. While the cell object on the

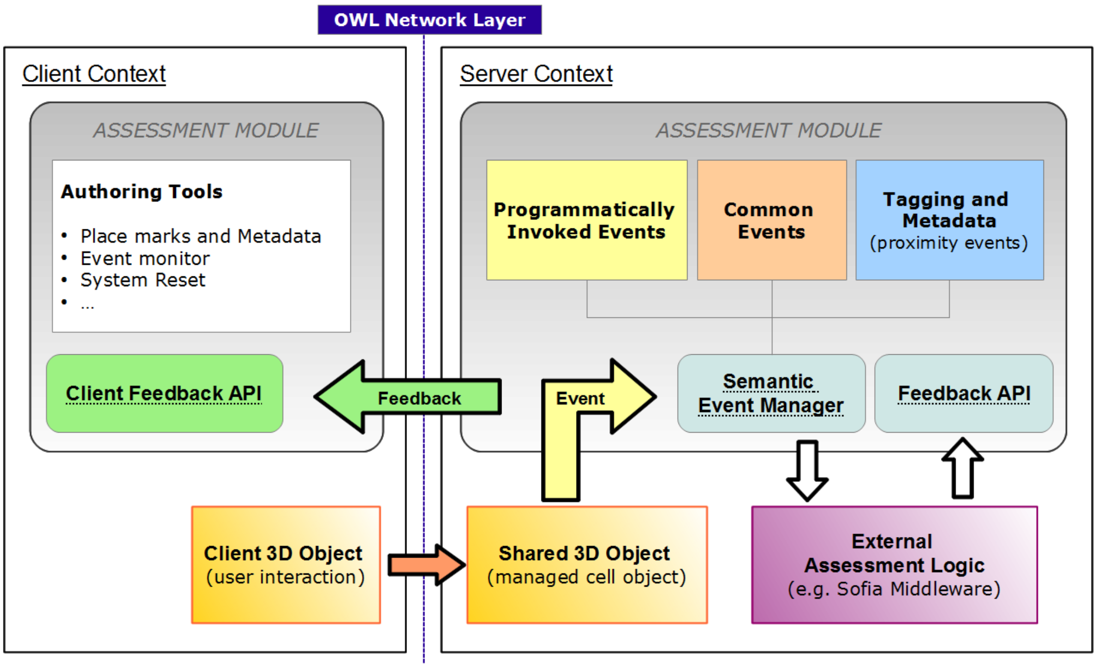


Figure 5.3: Conceptual architecture of the assessment module for Open Wonderland (Maderer, Gütl, & AL-Smadi, 2013)

client-side is responsible for user interaction and presentation, relevant actions that affect all participating clients are sent to the server through cell messages. The managed cell object on the server-side is supposed to construct semantic events and sent them to the assessment module for further processing;

2. The *semantic event manager* is responsible for collecting semantic events created within the virtual world and forwarding the information to all registered listeners. The following mechanisms are involved in creating semantic events:

   a) *Tagging and metadata components* are used to annotate spatial sections with place marks and add additional information to certain cells (3-D objects), including proximity zones – discrete definition of distance from an avatar to the cell. This can be used to detect behaviour such as users entering or leaving rooms; or approaching certain cells. Both actions are accomplished by world designers during design time. In the runtime phase, semantic events are created automatically based on the location of the avatar;

   b) *Programmatically invoked events* are created by embedded functionality that is added to existing and newly created modules (usually server-side managed cells) by software developers. This kind of events is used to describe interaction with a cell object – as well as to report on its state – e.g. the amplitude of a pendulum swing. Here again, a state is always considered to be information that is directly observable and understandable by a human player;

c) *Common events* include very basic OWL actions like start and stop of avatar movements or gestures, which are again created automatically. However, there is no particular design time feature that exists for such events. Therefore, a global configuration is supposed to provide mappings to yield proper semantic information for such built-in events;

3. The *external assessment interface* is supposed to register with the semantic event manager and communicate with the external assessment system. The assessment interface uses the server-side feedback API to create feedback content items based on the results received from the external assessment system;
4. The *feedback API* is a server-side interface to create and send feedback content to specified clients based on the participating users. Which users are involved in certain feedback messages is determined by the external assessment system. The structural representation of the feedback content is based on the four-dimensional approach for feedback in serious games (Dunwell, De Freitas, & Jarvis, 2011). However, while pedagogical aspects are decided in the context of the assessment system, the module is responsible for presentation. The current version is capable of displaying text messages only;
5. The *client feedback API* is a collection of utility functions to play feedback content items on the OWL client and to activate different guiding mechanisms;
6. The *authoring tools* on the client-side are used by content designers to create place marks for annotating spatial areas and attach metadata to existing cells. To support debugging, an event monitor and a system reset function are included as well.

Based on the presented architecture the assessment module requires at least a connection to an external assessment system and a programmatically enhanced cell that offers certain user interaction and status information. The next section will introduce and briefly describe the *SOFIA evaluation engine*, the first prototype for an assessment logic that has been used for the proof of concept in serious games. It is not completely compatible with the newest approach but should suffice for a couple of showcases.

5.5 SOFIA Evaluation Engine

This prototype is conceptually named as part of the *Service-Oriented Flexible and Interoperable e-Assessment* approach as mentioned in AL-Smadi, Wesiak, and Gütl

(2012). Based on the general requirements of the ALICE project, this assessment system has been designed and implemented by the author of this thesis but before the official start of the thesis (see also AL-Smadi, Gütl, Dunwell, & Caballe, 2012). Thus, this section is located between a technical overview as well as subsequent documentation of details that have not been published until this point. However, the information is important for the implementation of actual showcases regarding a physics experiment, as well as for lessons learned.

5.5.1 Conceptual Architecture

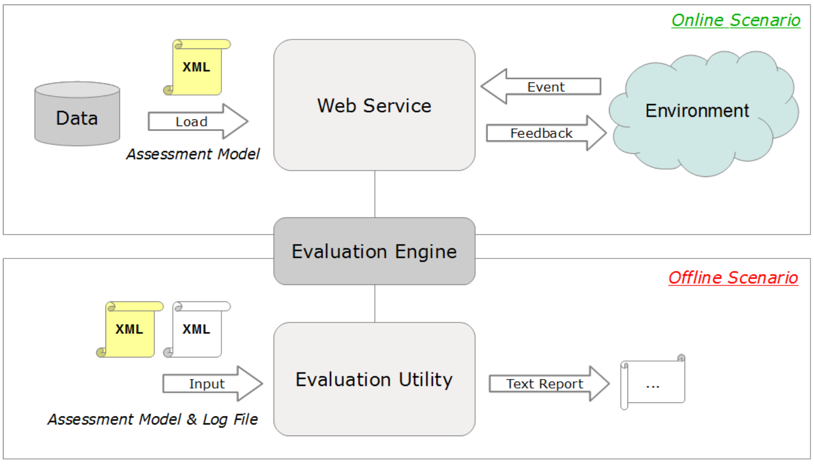


Figure 5.4: Conceptual Architecture of the SOFIA evaluation system.

The assessment system has been designed to support formative and summative assessment scenarios. The conceptual architecture is depicted in in Figure 5.4. The actual *evaluation engine* is supposed to be a standalone library that can either be used with a web service or as standalone utility application. The web service is primarily supposed to support online evaluation scenarios, where events from the (immersive) environment are immediately sent to the web service. Based on the assessment type the response may contain immediate feedback messages. However, it is also possible to upload an entire log file to the web service and receive a collection of all raised feedback messages. An *assessment model* is represented through an XML file, providing the necessary predefined rules and feedback information. The incoming events – represented through the log file or provided one at a time – are iteratively evaluated against the assessment model. In both cases the assessment model is identified through a task identifier and loaded from a *data repository*. The offline scenario uses a standalone desktop application (*evaluation utility*) that requires an assessment model as well as a log file as input and generates a plaintext report. Although theoretically considered, the first draft has not yet included an actual connection to a LMS.

5.5.2 Assessment Model

This subsection explains the structure of the XML format that describes a specific assessment model. It was in parts inspired by the work of Martínez-Ortiz et al. (2006), however, with the mentioned shortcomings (cf. section 5.2.2). The evaluation process is considered to be based on context-dependent measurement rules. These rules are referred to as behaviour patterns.

Behaviour Pattern

A behaviour pattern is a sequence of conditions or groups of conditions, which are supposed to match against incoming updates of environmental conditions and player actions. The concept is basically borrowed from regular expression libraries. However, as information within the environment changes through time, the process is designed as iterative approach that is automatically suspended when there is no incoming data. But based on the time that passes between information updates, it is also possible to abort a specific match process, once the given time frame is exceeded. The different elements a behaviour pattern consists of are called fragments. The different categories of fragments are explained in the following subsections.

Matching Expressions

Matching expressions can either represent containers or leave nodes. That means an expression either evaluates a concrete condition or it depends on the evaluation of its child nodes combined with further conditions. The possible evaluation results of a fragment are success, fail or suspend. The latter refers to a situation that cannot be decided at the given time. For instance, if a complete test cycle of the current node has already reached the beginning again, the match process will not finish based on the current information. The following fragments are available:

- `match-action`: tests if an action with the given name is available in the current evaluation cycle. If that is the case, an inner condition can further check on the action parameters;
- `match-property-changed`: evaluates successfully if the specified property name has been updated in the current cycle;
- `match-condition`: evaluates the given condition based on all information available within the perception context.
- `match-anything-else`: evaluates always successfully;

- `match-group`: matches successfully if all child fragments match successfully. If one of the child fragments suspends also this fragment will be suspended. It is further possible to define the minimum and maximum times this fragment should repeatedly match, before continuing with the next fragment. If no upper limit is defined, the match process will automatically favour the next fragment following the group fragment once it evaluates to true. That way it is possible to define a 'non-greedy wildcard' search pattern in combination with the `match-anything-else` fragment.

Consequences and Feedback

A `consequence` tag is a special fragment that is also embedded into the behaviour pattern among the other matching expressions. But instead of checking for specific conditions a consequences block can be used to execute internal procedures as well as raising feedback messages for the learner. Internal procedures include the storage and retrieval of measurement variables as well as conditional expressions and a set of logical and mathematical constructs (`compare`, `and`, `subtract`). The latter are used inside other fragments too. Finally, an integrated stopwatch mechanism (cf. Moreno-Ger, Blesius, Currier, Sierra, & Fernández-Manjón, 2008) can be used to evaluate performance criteria for a behaviour pattern.

5.6 Integrated Proof of Concept

The final design step is dedicated to the definition of a test scenario to apply the flexible assessment framework and the assessment module for OWL to a concrete physics simulation. This includes the implementation of a *simple pendulum* cell, as well as a *stopwatch* to take measurements with the pendulum. The example has been selected due to its simple instructional nature as well as the time-critical real-time movements involved. Measuring the periodic time and calculating the frequency is considered a possible task in this setting.

The contextual situation of the integrated proof of concept is illustrated in Figure 5.5. *World builders and developers* prepare the setting and implement assessment-compliant cells. These include the *simple pendulum*, a *stopwatch* as well as a *submission form* to check the calculated frequency. Furthermore, *tools and static objects* are added to provide an authentic environment and install all required learning materials (e.g. whiteboard or written instruction). Based on the metadata capabilities of the *assessment module* the spatial area that encloses the experiment workplace is annotated to define location in semantic terms. In addition, the web service is configured together with a proper task name. The instructional designer creates an XML-based *assessment*

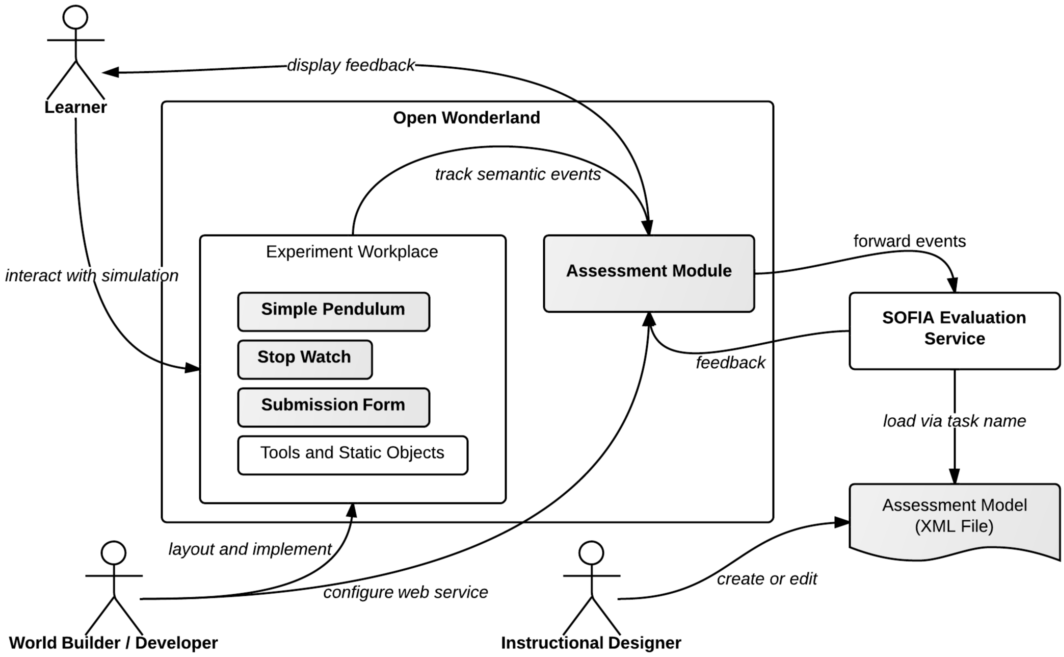


Figure 5.5: Schematic overview of the integrated proof of concept.

model for the *SOFIA evaluation engine*. When the student enters the world and triggers the first event an assessment context will be created based on the configured URL and task name which loads the proper assessment model. While the student is interacting with the scenario the assessment module is tracking and forwarding events to the evaluation engine. Incoming feedback is processed and displayed to the learner.

5.7 Summary

Based on literature findings and first results of the ALICE project, requirements for a flexible assessment framework have been defined. Considering aspects such as semantically self-descriptive 3-D environments and approaches in the development of software agents, an improved and generalised conceptual architecture to support behavioural assessment in different immersive environments and application domains has been defined. Immersive environments are supposed to implement an assessment module that provides a semantically-enabled tracking mechanism and an infrastructure to play feedback items such as simple text-based messages but also more complicated formats. An external assessment system includes the actual assessment logic and should be connected to external systems such as learning management systems. Individual settings and learning progress can be used to adapt the assessment plan and formative assessment results can be further used to update the learner profile. Communication between immersive environments and assessment systems is based on a contract described through a semantic knowledge repository.

Because of the enormous extend of such an approach the main objectives for this work have been limited. The implementation of an assessment module for Open Wonderland (OWL) is considered the most important step in order to examine if such an approach is feasible within a complex 3-D virtual world architecture. OWL has been chosen due to its extensibility and familiarity to the remaining research group. It is based on a client-server architecture; whereas the Darkstar (RedDwarf) gaming server represents the most important system component, providing transactional execution among clients and a persistent environment. The most important concept within OWL is a cell, representing an independent spatial communication node used to create 3-D objects which are shared among clients.

The assessment module for OWL is designed based on a feature analysis. Due to the limits regarding available information on objects and possible user interactions a threefold approach for event tracking has been introduced: Common events include basic cell-independent operations such as moving or using gestures. Tagging and metadata events are designed to track the avatar location in abstract terms and report about perceptual and operational ranges of 3-D objects. Programmatically-invoked events are necessary for interactive, self-descriptive 3-D objects. Semantic events are collected on the server and delivered to an external assessment system. Incoming feedback items are processed and distributed to the relevant clients.

The SOFIA evaluation engine is a context-dependent rule-based assessment engine designed to detect behaviour patterns consistent of conditional fragments and user actions. Consequences include the control of measurement variables and generation of feedback entries. The engine has been provided as a web service and standalone evaluation utility loading the assessment rules from an XML-based assessment model. The standalone utility requires and additional log file.

Finally, an integrated proof of concept has been outlined. Students are supposed to control a simple pendulum simulation, as well as a stopwatch and a submission form to determine the periodic time and frequency of the pendulum. The assessment module is used together with the SOFIA evaluation engine to provide immediate feedback. Based on this concept the next Chapter will explain several important implementation details. These include the general OWL assessment module together with an interface to the SOFIA evaluation engine as well as the pendulum simulation and its necessary tools.

6 Implementation of the Prototypes

The last Chapter has defined a flexible and generalised approach for assessment in immersive environments. For a proof of concept more concrete design specifications have been given for a reusable assessment module for Open Wonderland (OWL). An example scenario has been outlined that combines the assessment module with the already existing SOFIA evaluation service. This includes further the assessment-compliant simulation of a simple pendulum and its necessary tools.

Based on these specifications, the aim of this Chapter is to describe and document all important aspects regarding the implementation of the prototypes. This includes the implemented features of the assessment module and also the integration with the SOFIA evaluation service. Furthermore, the development of the simple pendulum and its additional tools is described. Finally, necessary improvements of the SOFIA evaluation service that have been implemented during this project will be summarised as well.

6.1 Assessment Module

The assessment module has been implemented as an OWL module. In accordance with the conceptual architecture the main aspects have been implemented server-side. Besides the structural representation of semantic events, this section will therefore focus on detailed decisions about the event tracking mechanisms. Further aspects include the feedback mechanism and client-side design time tools. Beyond that, the integration with the external SOFIA web service is a critical aspect and described in more detail.

6.1.1 Semantic Event Structure

The structure of semantic events as represented in the context of the assessment module is a refined version of the event objects that have been defined as part of the interface provided by the SOFIA evaluation service. It is included in the *common* package of the assessment module to be also available on the client-side for debug reasons. The objects and relations are depicted in the class diagram in Figure 6.1. A `SemanticEvent` instance is characterised by a time code that represents the exact time the event occurred, usually up to the level of milliseconds. The event can further contain a set of `ObjectUpdate` instances, which refer to the changed state of a single entity within the 3-D virtual world. The object update can describe three different scenarios:

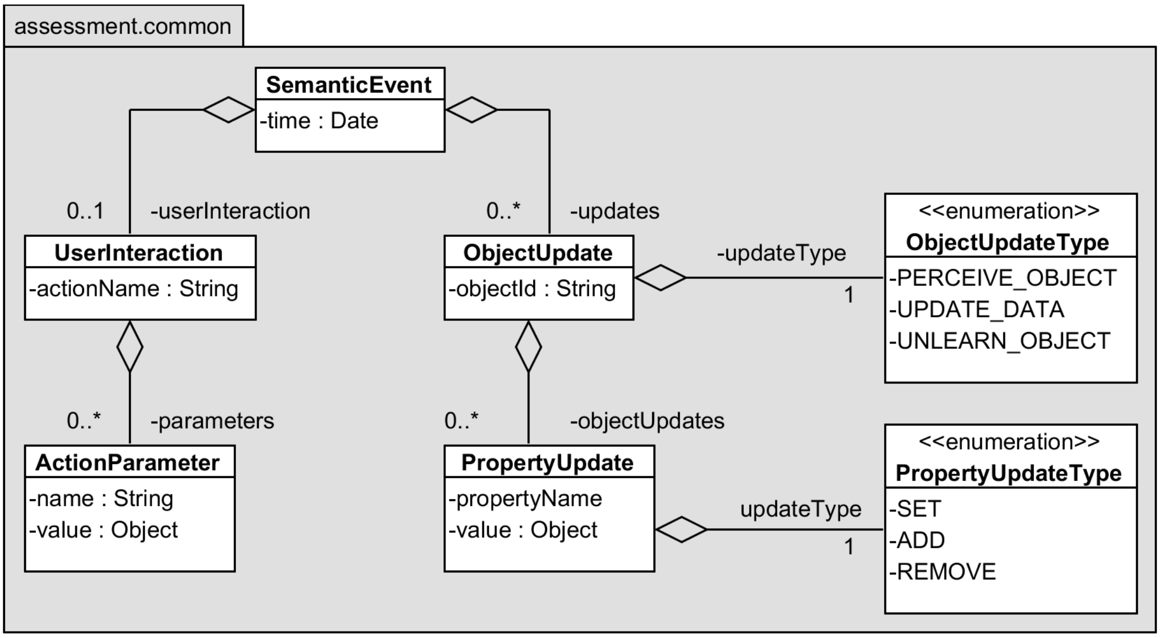


Figure 6.1: Class diagram illustrating the structure of semantic events

1. `PERCEIVE_OBJECT`: an object (cell) has come into existence for one or more users. This should either happen when a single user approaches an object with his or her avatar, or when the object has just been created, which would possibly affect the perception of multiple users at the same time being present within the proximity range of the newly created object. Therefore a set of usernames is also included in the object update;

2. `UPDATE_OBJECT`: informs the assessment system that an object has changed its state, respectively reporting the most recent dynamic information about the object. No usernames are required to be included in this update; the external assessment system should already be aware of which users are currently observing this object. Nevertheless, it is still open to further design decisions if the assessment module should already pre-process these filtering operations; but the information is anyway redundant;

3. `UNLEARN_OBJECT`: the object cannot longer be considered to be observed by the given usernames. The assessment system should remove the existence of the object from the evaluation context of the specified users. In the reverse direction, this should either happen if a specific avatar leaves the proximity of an object, or

an object gets deleted from the world, thus involving all users currently being within the proximity range of the deleted object;

With the exception of the `UNLEARN_OBJECT` type, all object updates are supposed to include a set of `PropertyUpdate` objects. These property updates are not an absolute measure of state, but report about what has changed. That means an update must not always include all properties of the object.

Beside the object updates, exactly one `UserInteraction` object can be added, indicating that changes in possible one or more objects have been induced through an actual user operation. Nevertheless, a user interaction could also describe an unrelated action, such as moving the avatar or using gestures. This interaction object features the major action name, information on which user has performed the action, as well as additional parameters narrowing the meaning of the action. The latter could also include the object identifiers of the operated objects if some exist. Nevertheless, object identifiers are not really provided by means of high level object identification. They should support possible implicit information linkage within the external assessment system, as well as provide an anchor for triggering invasive feedback actions, thus manipulating objects remotely from the assessment system. Finally, as the state of the avatar should also be reflected through an object that is identified through the username, the assessment system also has information about which users could have observed the interaction.

6.1.2 Triggering Mechanisms for Semantic Events

Based on the data format presented in the last subsection the question is how the semantic events should be created and filled with proper information. This subsection explains concerns about the amount of data being processed (also discussed in Maderer & Gütl, in press), how the event creation strategies defined in the conceptual architecture have been implemented, and how semantic events are actually constructed before they get finally deposed at the semantic event manager.

Data Volume Strategies

During the implementation phase of the prototype it became clear that especially complex simulations can include highly fluctuating values, such as distances or velocity, changing rapidly. Nevertheless, it is possible that some of these object properties may be important for the assessment process. While this could possibly be neglected for a few major objects, communicating internal simulation propagation of many objects in the scene and throughout the world simultaneously would rather stress the assessment system unnecessarily. This holds especially true when considering that the assessment system is designed as an external support service that cannot play the same role as, for instance, the OWL Darkstar server. Besides, it would also result in large log files.

To address this problem an adequate event creation strategy has been developed. The first part of this strategy consists of differentiating the importance, respectively timely relevance, of data an object can provide into three cumulative levels of information:

1. *Dynamic*: data on this level is usually part of the simulation model of an object and changes extremely frequent. It is only considered important whenever a user interaction occurs within the context of perception of the object in question;
2. *Changable*: properties on this level are subject to persistent changes that occur regularly within the 3-D virtual world. This either includes autonomous changes that can for instance happen based on random events or as a side effect of user interactions being performed on an object;
3. *Full*: this level involves the entire set of properties an object has to feature, also including static information that identifies the purpose, attributes and possible operations of an object. As the name implies, static information will not change during normal operations – only during world editing – thus it can be assumed that this level of information needs only be included during declarative operations. This is whenever an object is created, edited on the metadata level, or a user approaches the object (see `PERCEIVE_OBJECT` in the last subsection).

The second part of the strategy involves an `EnvironmentManager` class that is supposed to keep track of users being present within the perception range of semantically enhanced objects. The following section explains how this is implemented and combined with the event creation strategies of the conceptual architecture.

Source of Events

In accordance with the event creation strategies of the conceptual architecture, several facilities have been implemented that provide the relevant information. It is important to mention that not all aspects which are described as part of the semantic event structure could be implemented due to time limits of this work. Especially the support for triggering events related to the creation, static update and removal of cells, as well as recognition of perceptual ranges between different avatars might not work as expected due to the implementation details and limits of the underlying OWL components that have been used to develop the prototype.

To support the *tagging and metadata* concept, two features have been embedded into the assessment module:

1. The `SpatialPlacemarkCellMO` class represents the server-side implementation of a new cell type that is usually concealed from the average user. It defines a spatial area that could be tagged with attributes which describe the location of this area in abstract terms. Proximity listeners are used to check whether an avatar *enters* or *leaves* this section. Both operations are communicated to the environment manager which will record the change and eventually prepare proper semantic events in order to communicate the avatars' abstract location to the assessment system based on the tagging information provided on the spatial area.
2. The `MetadataComponentMO` class is a custom OWL component – the behavioural extension of a cell – that gets registered among the default components. That means it is automatically assigned to every cell. This component is supposed to define proximity spheres for the *perceptual* range, as well as the *operational* range of an object. One would argue that it is be better to only observe the avatars' surroundings to check on approaching and diverging cells. While that is true, proximity listeners in OWL have been designed to check on view cells – that means avatars – entering a proximity range and not vice versa. Thus, it is necessary to install the proximity listeners on each cell. Nevertheless, for the sake of the prototype, it has been decided to rely on existing infrastructure rather than fundamentally improving these concepts at this time.

For *programmatically-invoked events*, i.e. events required for interactive objects, event creation is based on two concepts: First, each interactive object is supposed to implement the `SemanticStateProvider` interface in order to provide information about its static attributes, as well as to report about *changeable* or *dynamic* object state. Second, whenever an event happens – random occurrence or user interaction – the metadata component (already present for each cell) is referenced and used to inform about an autonomous state update or to depose a user interaction object. Up from this point, event compilation is managed by the assessment module itself.

Common events have not been implemented as it was not an urgent requirement regarding the showcases included in this work. In fact there are many basic events of OWL that could be considered for this broad group of events (e.g. world editing or avatar movement). But the basic idea to do so has been theoretically examined and is straight forward: similar to the *event recorder* module that is included in the OWL module repository (see "Wonderland Modules: Subversion Repository", n.d.), it would be possible to intercept specific well-known cell messages, map these to a proper semantic event object and forward it to the environment or semantic event manager.

Event Compilation Process

Considering the discussed infrastructure and data volume strategies, this following paragraphs explain how a semantic event gets actually compiled based on different user interactions.

If the user interaction is associated with a cell, all *changeable* – that means also *dynamic* – information of the cell is added to the semantic event object. In a subsequent operation, all cells that are currently in the context of this interaction, i.e. all 3-D objects within the perceptual range of the user that caused the interaction, are considered as well. However, these cells are only requested to attach their *dynamic* state to the operation, as the *changeable* data level is not assumed to have changed. In other words, the semantic event represents a snapshot of the dynamic user-centred situation that was given at the time the user performed the operation.

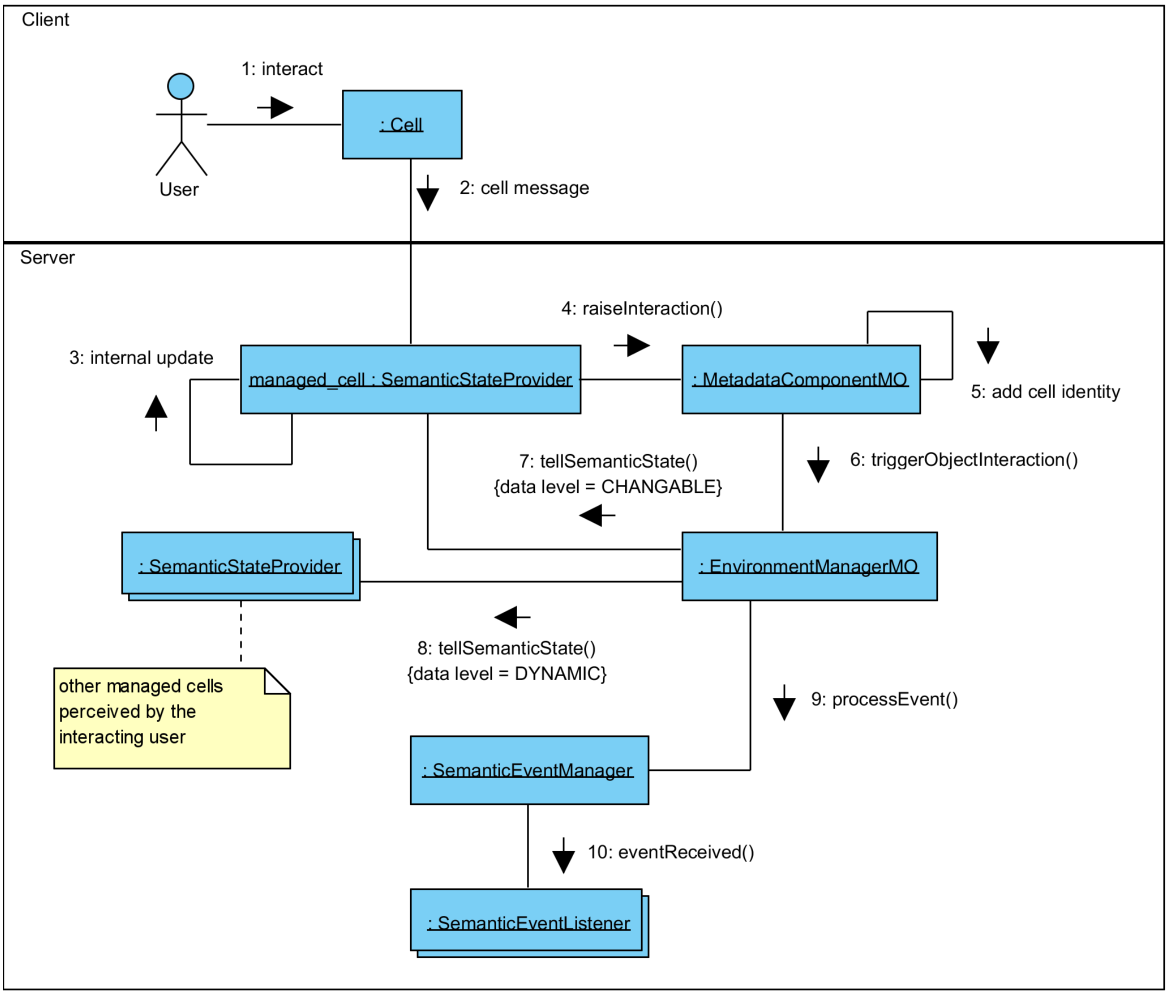


Figure 6.2: Communication diagram illustrating the compilation of an interaction-based semantic event

The complete process is exemplarily illustrated as a communication diagram in Figure 6.2. Starting from a client-side input operation, the information is transmitted to the server through a cell message, where the internal status of the server-side managed cell is changed and the attached metadata component is invoked with a user interaction object. The metadata component adds then the ID of the cell and forwards the interaction object to the environment manager. The environment manager holds lookup tables that associate users with perceived objects and vice versa. Based on this data it is possible to gather the dynamic information. After both, changeable and dynamic state information, have been collected, the semantic event is delivered to the semantic event manager where further processing will be dispatched to arbitrary event consumers.

For events which are not based on a user interaction, the semantic event will only contain the changeable state of the related cell; respectively the full state of the cell if a proximity enter event on the perceptual range of a cell was the source of the event. Finally, leaving and entering spatial sections represents a special case. They are communicated as user interactions along with property changes on the object that is named after the username; however they do not include dynamic information about the local environment. The latter was not found significantly important for such events; at least at this time.

6.1.3 Feedback Mechanism

Fairly similar to the semantic event structure, feedback items have been implemented that represent any kind of feedback that should be injected into the assessment module; the design also derived from the SOFIA evaluation engine. The relevant objects are depicted in the class diagram in Figure 6.3. The top-level object is an assessment feedback class that can contain one or more feedback content items, as well as the username that is supposed to identify the addressee of the feedback. If no user is specified, the message is delivered to all clients. It is already clear that future improvements will require more sophisticated target discriminations, e.g. group or location based mechanisms. Nevertheless, regarding the feedback content items itself, only text-based feedback has been developed at this time. A `TextFeedbackContent` item consists of the actual text that should be displayed on the users screen, as well as a colour code that emphasises the general category of message, including success messages, hints (warnings) and errors.

The feedback is then further forwarded to the OWL client through a message object via an additional Darkstar communication channel (`AssessmentConnection` class). On the client-side the feedback messages are displayed through a HUD component – not to be confused with cell components – on the southern area of the screen. Size and colour is automatically adjusted. The message can be dismissed by the user through a click on the message area. However, unconfirmed messages do not lock the screen or input mechanisms, thus letting the users continue his or her work unhindered. The messages

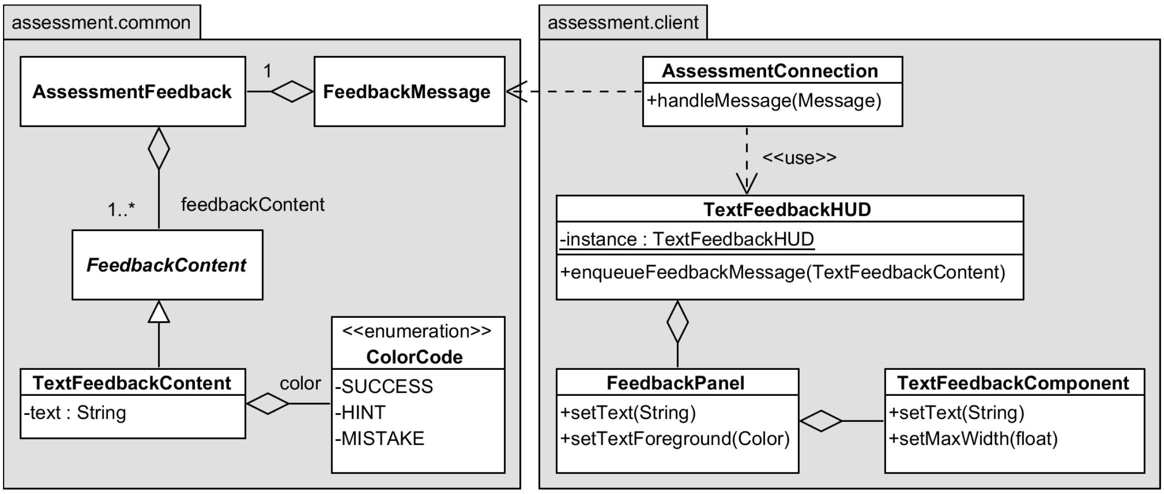


Figure 6.3: Class diagram illustrating the feedback structure and client-side display mechanism for text-based feedback

will further disappear automatically after some time. Subsequent messages will be queued until the user confirms any previous messages. The `TextFeedbackComponent` is a special Java Swing component that dynamically adjusts the required display height based on the length of the feedback text.

6.1.4 Design Time and Client Tools

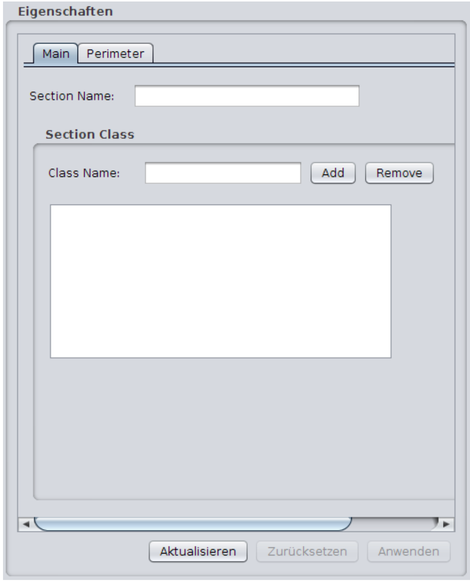
The assessment module installs a new *'Assessment'* submenu into the *'Tools'* menu of the OWL client. The toggle switch *'Enable editing ...'* activates the design time tools, thus allowing the user on the current client to see all spatial place marks that have been added to the world. In addition, a toolbar that appears at the bottom of the main window offers quick access to add new spatial place marks. Metadata components are registered as default capability for each cell and will be added automatically. That means they can neither be added nor removed by hand, but settings can be adjusted.

Spatial Place Marks

For the world designers to quickly distinguish between spatial annotations, place marks are highlighted through semi-transparent shapes (see Figure 6.4a). The centre point is depicted as a small pyramid, labelled with the set of attributes that describe the location on an abstract level.

Through the object properties window (see Figure 6.4b), it is now possible to access the configuration of the spatial place mark and adjust the list of section attributes, as

(a) 3-D representation of spatial place marks during design time



(b) Spatial place mark properties sheet

Figure 6.4: Representation and settings of the spatial place mark feature

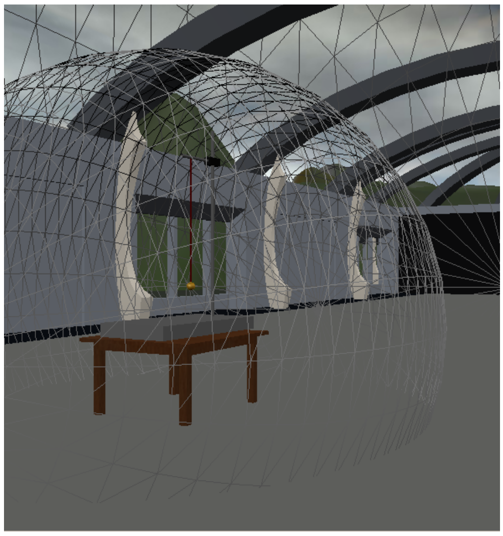
well as the perimeter. The perimeter defines the spatial extend of the place mark. It is currently defined as a box where width, height, and depth can be adjusted. Apart from this, spatial place marks can be nested at will. Through this nesting it is possible to provide more concrete information from level to level. For example, the entire area could be annotated as school or university, whereas single rooms could be tagged as laboratory.

Metadata Component

The metadata component is responsible to classify 3-D objects and to define proximity ranges for perceptual and operational distances. Which properties can be adjusted by the designer is determined by the type of the object. If the object is interactive – that means user interactions and semantic state provision has been considered during the development phase – then the object is already considered semantically self-descriptive. In that case (see Figure 6.5a) the checkbox *'Enable semantic data'* is ticked (true); although it is disabled as the status cannot be changed. Beyond that, also attributes that describe the object type are not supposed to be modified. However, if the cell is a legacy item or just a conventional static or animated 3-D entity then its semantic description is not activated by default. But all options are allowed to be edited. Nevertheless, both types of cells allow the adjustment of proximity distances, operation range – from which distance is an avatar considered to be interacting with or investigating the item –

and perceptual range – from which distance should the object be considered existent for the assessment plan.

(a) Metadata properties sheet



(b) Highlight of *perceptual* and *operational* ranges while the metadata properties sheet is enabled for editing

Figure 6.5: Representation and options of the metadata component (capability)

Another feature of the metadata component is the visualisation of the proximity distances (see Figure 6.5b). This behaviour has been inspired by the *EZScript*[1] module, which also relies on proximity listeners and displays the boundaries similar to this (see also "Wonderland Modules: Subversion Repository", n.d.).

Control Console

Further included in the assessment menu is the *'Show Console'* command. It opens an additional window (see Figure 6.6) where it is possible to observe all semantic events which are generated within the current virtual world. Usually events are not forwarded to clients; however by activating the checkbox *"Activate global event log"* the client will be registered to receive all semantic event that have been delivered to the semantic event manager on the server.

The purpose for this is obviously to provide an important debug feature at this early stage. Nevertheless, the events will not only show up on the console but are

[1] http://openwonderland.org/module-warehouse/module-warehouse/doc_details/275-ezscript?cat=add_ons&Itemid=123 (retrieved on January 5, 2013)

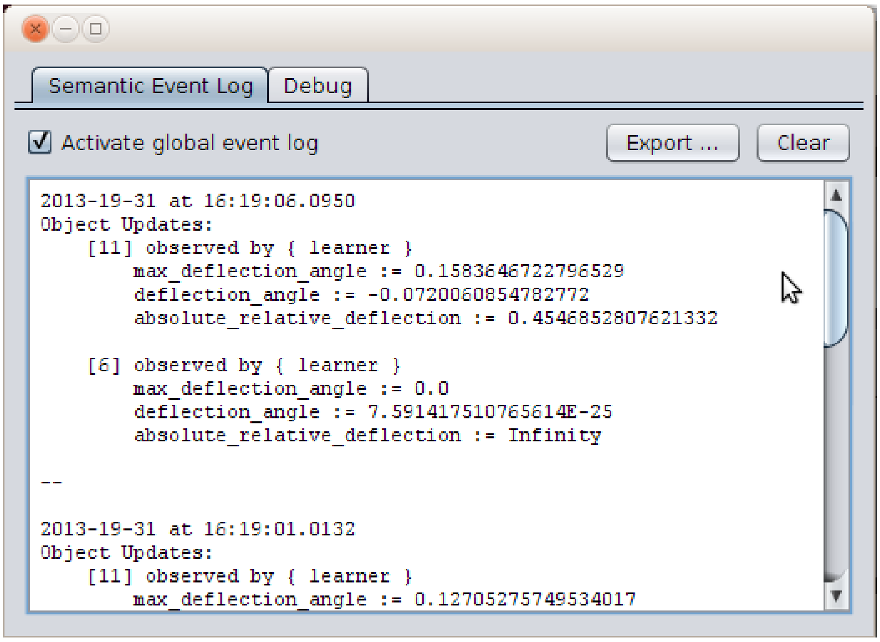


Figure 6.6: Semantic Event Log

further collected in the background. It is then possible to export the events in form of an XML based log file that is compatible with the newest improvements of the semantic event structure. Consequently, scenarios can be tested offline when developing in the context of the assessment system or when testing new assessment rules. Clearing the text frame will also flush the collected events in the background.

6.1.5 Integration of the SOFIA Evaluation Service

Up to this point, the assessment module operates only as a collector for semantic events and offers a set of functions to support the delivery of feedback information to the player. Although it is supposed to be a framework to attach any kind of assessment logic, an interface to communicate with the SOFIA assessment service has directly been embedded into the assessment module itself. The following subsections will give details on the server-side and client-side implementations.

Explanation and Design Considerations

The external assessment service is represented as a cell (3-D object) within the virtual world and called assessment controller. Although not implemented in the prototype, this cell is also supposed to provide a spatial area that defines the boundaries of the assessment task. As OWL worlds might be used to host several experiments and task, it would be necessary in production environments to provide several independent assessment contexts. In such a scenario, the assessment controller would only receive events which are located inside the boundaries of the assessment controller.

The primary communication interface has been implemented within the server context of OWL. Several considerations have been involved with this design decision: On the one hand, the architecture of the assessment module already suggests that semantic events of all users are available on the server-side. On the other hand, the SOFIA evaluation service has also conceptually been designed with multi-user assessment in mind. Thus, a single connection to the evaluation service is sufficient to communicate the user interactions and environmental changes for all participants involved in an assessment situation. This makes it further easier to dispatch feedback information that addresses an entire group.

Server-Side Implementation

The server-side counterpart for the assessment controller (`AssessmentControllerMO` class) implements the `SemanticEventListener` interface of the assessment module. Once it is activated, it registers with the `SemanticEventManager` in order to collect all semantic events that have been raised by the assessment module. The entire situation, including the following paragraphs, is illustrated in the class diagram in Figure 6.7.

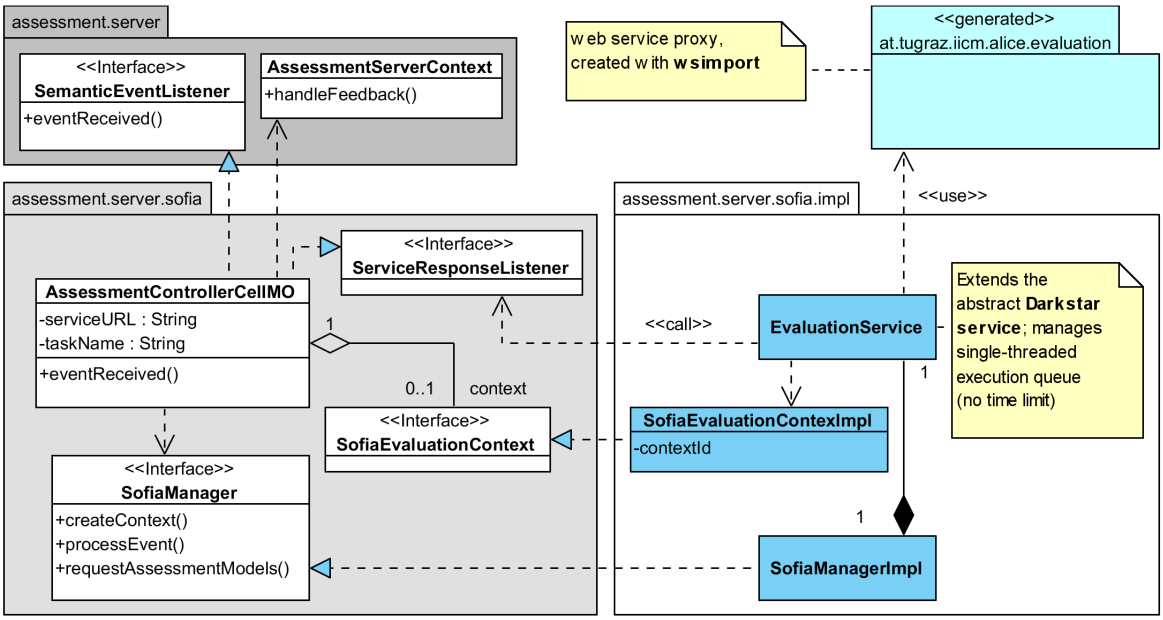


Figure 6.7: Classes required for the integration of the SOFIA evaluation service (simplified presentation)

The most critical part of the integration process is related to the server-side transaction model of OWL (Darkstar). As each transaction is only allowed to run for some milliseconds, it was not possible to call the external web service directly from within the context of a transaction. The exact similar situation applies to the *Twitter Viewer Module*[2] which uses the Twitter web service API to obtain information from that social network (see "Wonderland Modules: Subversion Repository", n.d.). As the source code is available among the other OWL modules, the approach has been adapted to solve the current problem. It was therefore necessary to provide a dedicated Darkstar service (`EvaluationService`) that could handle long-term operations asynchronously. In accordance with what seems to be a dependency-injection model used by Darkstar, the service is not directly operated, but accessed via a manager interface (`SofiaManager`). It's actual or default implementation is configured through the *ant* script used to build the assessment module from source code.

Using the evaluation service to assess user behaviour involves now two steps: First, it is necessary to create a specific evaluation context. Based on a given task name that is internally associated with a certain assessment model – i.e. the actual assessment rules used for evaluation – a token is returned that is provided in further communication with the system. In accordance with the manager concept, the OWL implementation for the web service sustains the abstraction by only exposing the empty `SofiaEvaluationContext` interface to the assessment system, while hiding the real implementation that holds the token. Second, after the context has been created, all semantic events which are received from the assessment module are forwarded to the `SofiaManager` together with the evaluation context. The `ServiceResponseListener` interface is used to receive responses from the service, including possible feedback which will be injected into the assessment system by invoking the `AssessmentServerContext` class.

Finally, the items in the class diagram have been separated into three different colours in order to distinguish between core features of the assessment system, the classes related to the assessment controller approach, as well as the concrete implementation that depends on the web service client proxy (package), generated with the web service import tool (*wsimport*) that is deployed with the Java development kit.

To better understand the approach, the dynamic situation has been depicted in a sequence diagram (see Figure 6.8). After receiving an event, the assessment controller forwards the event via the manager classes to the evaluation service. The evaluation service, which controls an own thread, immediately schedules a new request object into a single-threaded processing queue and returns within a time limit of the magnitude of 100 milliseconds to avoid breaking the Darkstar transaction model. Once the scheduled request is fetched from the queue it can run as long as necessary to perform the external

[2] http://openwonderland.org/module-warehouse/module-warehouse/doc_details/252-twitter-viewer-module?cat=add_ons&Itemid=123 (retrieved on February 5, 2013)

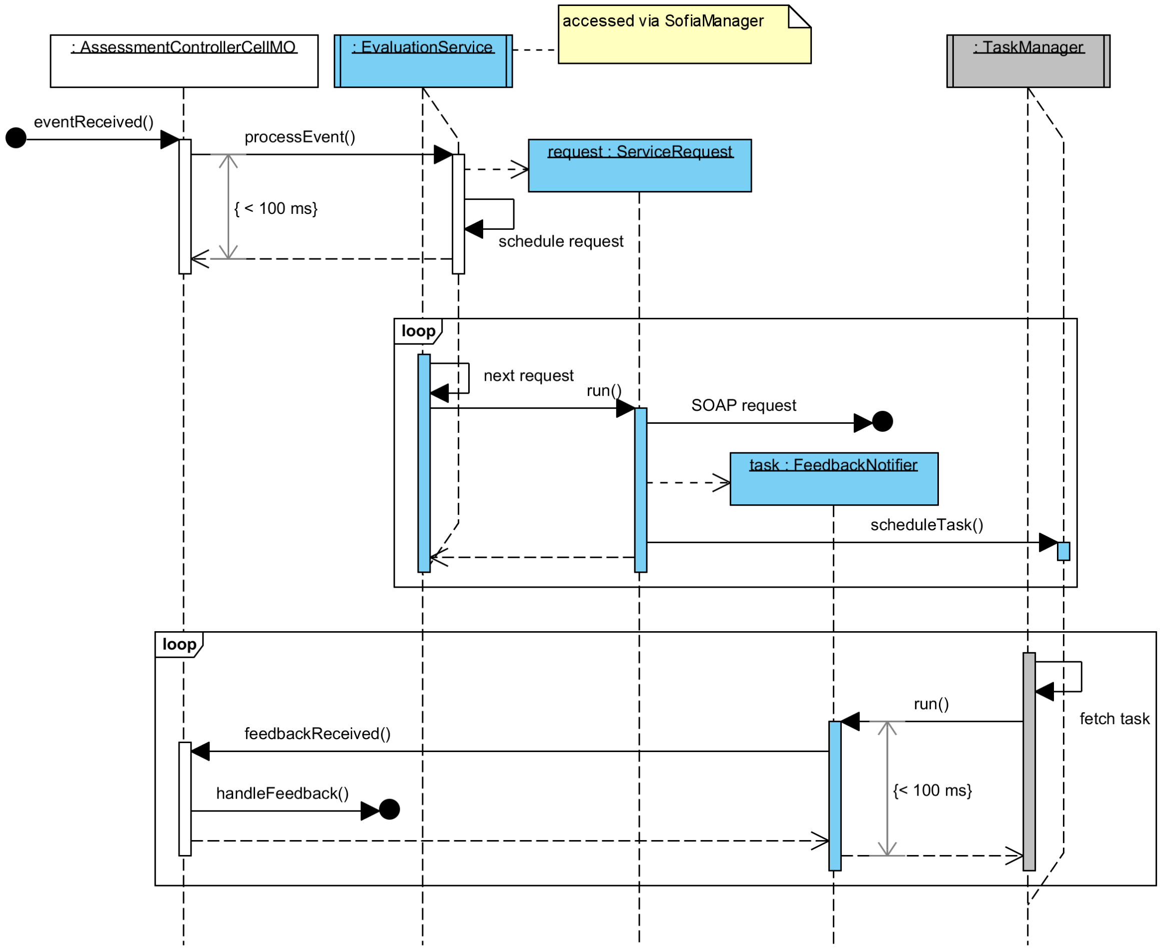
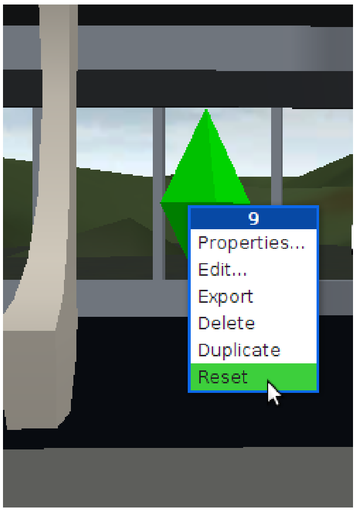


Figure 6.8: Sequence diagram describing the process of forwarding a semantic event and receiving immediate feedback messages

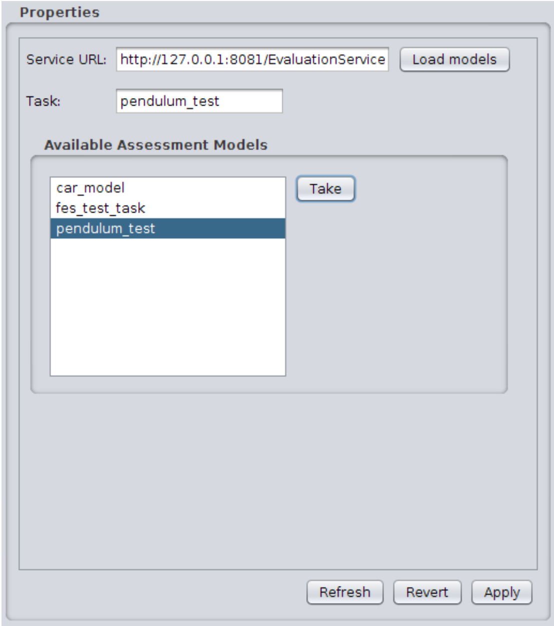
web service request (SOAP). When the request has finished, a notification task is created and scheduled to be run within the transactional context (task manager) again, where it eventually calls the feedback response method of the assessment controller.

Client-Side Representation and Configuration

In order to connect the virtual world instance with the *SOFIA evaluation service*, a new *'Assessment Controller'* has to be inserted. This is achieved through the *'Insert Object ...'* dialogue of the OWL client. The corresponding cell is represented through a green hovering pyramid symbol (see Figure 6.9a). This symbol is also considered a design time feature. However, at this time it is not hidden to provide an easier access to the support functions, such as forcing the reset of the external assessment system.



(a) 3-D representation and context menu



(b) Web service configuration

Figure 6.9: Assessment controller user interface

The properties sheet of the assessment controller (Figure 6.9b)allows the world builders, respectively the assessment designers, to configure the location of the web service as well as the task name. In order to quickly switch between different assessment models (equivalent to task names) available through the specified web service, a list can be requested after the URL has been entered (*'Load Models'*). The *'Take'* button will allow to copy the selected model name into the *task name* field.

6.2 Assessment-compliant Simulation of a Simple Pendulum

The central component for an assessment example in the domain of STEM education is a 3-D object that is used for experimentation. Therefore, a new OWL cell has found to be necessary that provides an authentic simulation of a physical process and is capable of communicating state and user interaction as required by the assessment module architecture. A simple pendulum has been chosen as an exemplary object.

6.2.1 System Overview

The pendulum cell has been implemented as a new OWL module, featuring client and server aspects, as well as a set of commonly shared classes required for communication. Figure 6.10 gives a schematic overview the situation.

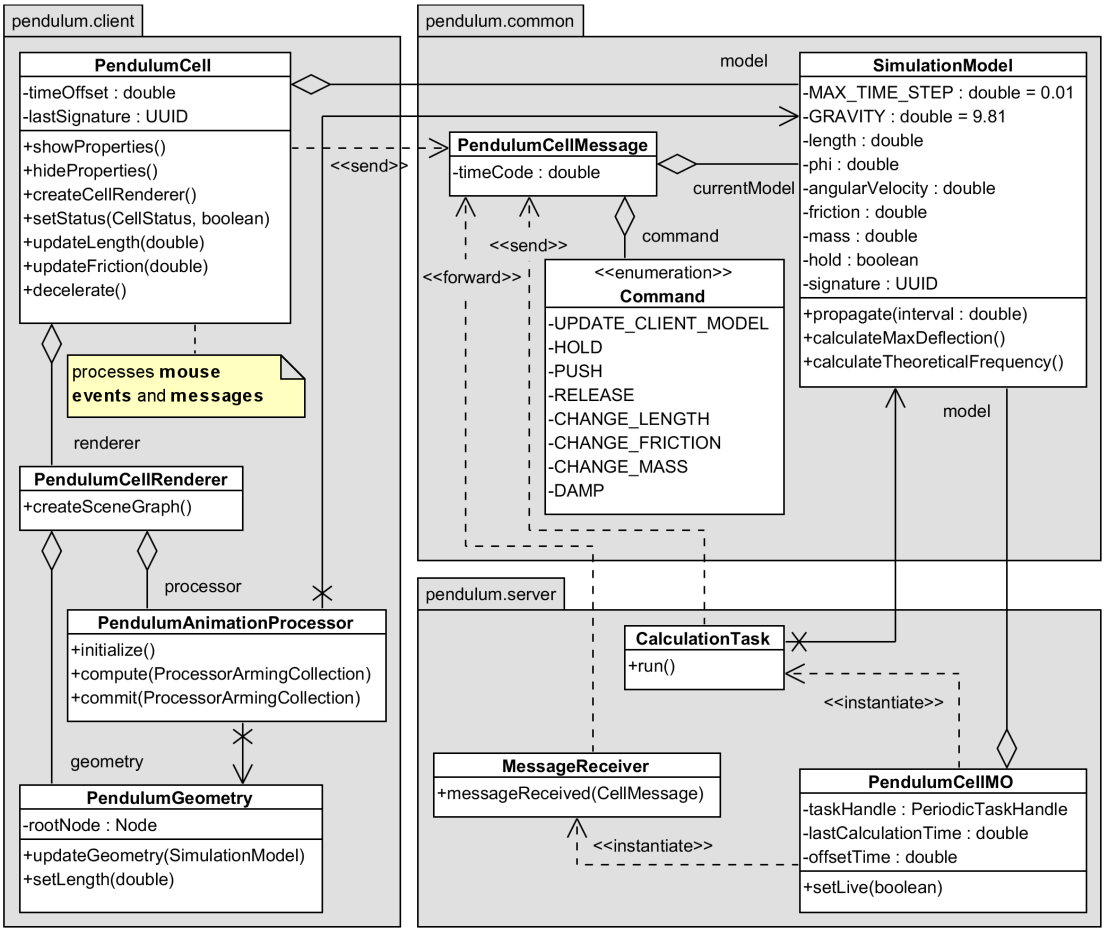


Figure 6.10: Class diagram of the pendulum module including the most important classes and operations

The central component of the pendulum simulation is the `SimulationModel` class that is shared between client and server. The simulation model is a numerical representation of the equations of motion involved in a simple pendulum animation. The physical state of the pendulum is continuously calculated by both, the clients and server, based on the elapsed time. On the client-side, the calculation is coupled with the rendering process. The `PendulumAnimationProcessor` class is a *MTGame* processor component that is associated with the client's frame rate and continuously propagates the local simulation model. In its commit-phase the associated `PendulumCellRenderer` class that holds a 3-D model of the pendulum cell (`PendulumGeometry`) is updated with the current state of the simulation model. In contrast to this, the server-side version of the simulation model is propagated through a periodically scheduled Darkstar task (`CalculationTask`); intervals approximately 100 ms. Whenever the propagation process is started (also on the client), the update of the model is performed in constant small time intervals, starting from the end of the last propagation cycle to the current time. The updated simulation model is then sent to all connected clients in order to prevent small deviations over time. In this sense, the server can be considered the simulation master. Nevertheless, the situation changes once a user takes direct control over the moving parts of the pendulum. In that case, the server will just forward the updated versions of the responsible client's simulation model to all other connected clients. After the client stops the direct manipulation, the server becomes the simulation master again. The entire communication between client and server is facilitated through instances of the `PendulumCellMessage` class, featuring several possible commands and a copy of the current simulation model.

Simulation Model

The physical and mathematical knowledge behind the simulation model is based on the related Wikipedia article as well as a general physics textbook (see "Pendulum (mathematics)", 2013; Tipler & Mosca, 2004).

A simple pendulum – as the names states – is a physically simplified model of real world conditions. In such a model, the rod or thread that connects the mounting with the mass is considered weightless and damping (e.g. air resistance) as well as mechanical friction is omitted as well. Such a pendulum is in general not a harmonic oscillator; that means the periodic time is not independent of the maximum deflection angle. The problem is that the resilience force depends on trigonometric relations which are not linear. But under the assumption that for small angles $sin(x) \approx x$, the analytical solution is rather easy to find and allows the treatment as a harmonic oscillator. With these simplifications the theoretical frequency of a pendulum is given as:

$$f \approx \frac{1}{2\pi}\sqrt{\frac{g}{l}} \tag{6.1}$$

Nevertheless, in order to provide a more realistic behaviour for the experiment, it has anyway been decided to include velocity-dependent damping and implement the pendulum based on numeric integration. That means it is possible to reach conditions that would not be included if just treating the pendulum as harmonic oscillator with damping. The mechanical situation is depicted in Figure 6.11. The general differential equation for such a system is given as follows:

$$ml\frac{d^2\phi}{dt^2} + cl\frac{d\phi}{dt} + mg\sin\phi = 0 \tag{6.2}$$

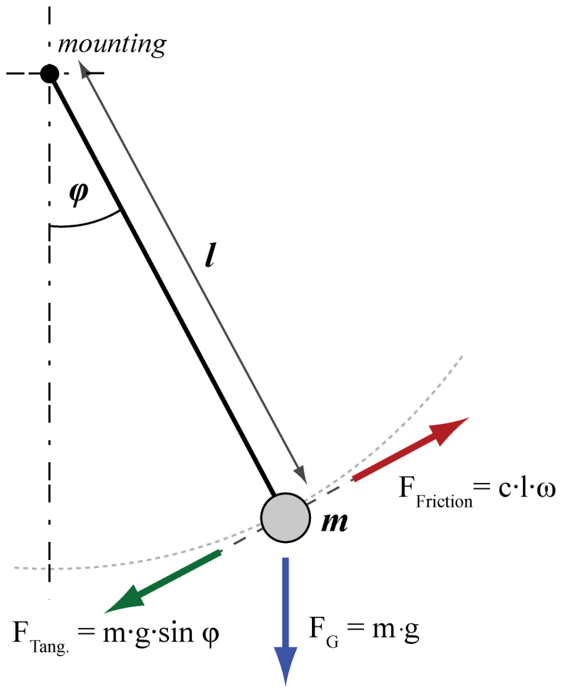


Figure 6.11: Mechanical situation of a simple pendulum with additional velocity-dependent damping (own drawing) (based on Tipler & Mosca, 2004)

This equation includes the following parameters used in the simulation model: l is the length of the thread (interpreted as radius); m the swinging point mass; and c is a coefficient that determines the velocity-dependent frictional force; symbol g denotes the constant of gravitation. The internal state of the simulation model is then described through current deflection angle ϕ and angular velocity ω (the first time derivative of ϕ). The accelerating force $F_T(\phi, \omega)$ on the pendulum is the tangential component of the gravitational force, inducing a torsional momentum as a function of the current deflection angle combined with the opposite directed velocity-dependent damping. Thus, the resulting function for the angular acceleration is given as:

$$\alpha(\phi, \omega) = -\frac{g}{l} \cdot \sin\phi - \frac{c}{m} \cdot \omega \tag{6.3}$$

The trivial idea was to solve the equation of motion for this system through simple numeric integration. Based on the current time step Δt one could propagate the deflection angle based on the current angular velocity; whereas the angular velocity is updated based on the acceleration function $\alpha(\phi, \omega)$. Unfortunately, this approach, which is also known as *Euler's forward method* (cf. Sandvik, 2012), turns out to accumulate significant errors and leads to a divergence of the system's energy, thus causing undesired behaviour of the pendulum. Therefore, according to Sandvik (2012), the *velocity Verlet algorithm* has been chosen among a few comparable methods. It is a stable algorithm for solving simple differential equations. As it guarantees a timely symmetry between forward and backward propagations, computational errors can be neglected and the energy is conserved for periodic movements without damping. Nevertheless, the usage of velocity-dependent damping requires additional adjustments and a more complicated velocity approximation which is further explained in Sandvik (2012). Based on the formulas provided by Sandvik, the propagation steps in terms of the pendulum – by using angular notation – have finally been implemented as follows:

$$\phi_{n+1} = \phi_n + \Delta t \cdot \omega_n + \frac{1}{2}(\Delta t)^2 \cdot \alpha(\phi_n, \omega_n) \tag{6.4}$$

$$\tilde{\omega}_{n+1} = \omega_n + \frac{1}{2m}\Delta t \cdot [\alpha(\phi_n, \omega_n) + \alpha(\phi_{n+1}, \omega_n + \Delta t \cdot \alpha(\phi_n, \omega_n))] \tag{6.5}$$

$$\omega_{n+1} = \omega_n + \frac{1}{2}\Delta t \cdot [\alpha(\phi_n, \omega_n) + \alpha(\phi_{n+1}, \tilde{\omega}_{n+1})] \tag{6.6}$$

The intermediate term $\tilde{\omega}_{n+1}$ is an additional velocity approximation to reach a better convergence as the force at ϕ_{n+1} also depends on the velocity at this time which is not known in advance (cf. Sandvik, 2012). It can further be seen that the mass influences only the damping. The simplified formula for the theoretical frequency of a simple pendulum (without damping) is still provided as part of the simulation model. The value is supposed to serve as a theoretical reference for the accuracy of measurement activities in the context of assessment showcases. Furthermore, also kinetic and potential energies are considered to estimate the current peak amplitude of the pendulum.

Synchronisation Process

The synchronisation between client and server was an important but complicated aspect (also discussed in Maderer & Gütl, in press). Particularly for this example, but for dynamic physics simulations in general, the assessment measurements might be based on time-critical features. However, this leads to a problem. All clients and the server should always be completely synchronised. On the one hand, the semantic event generation is server-based, which requires the server to maintain the exactly same simulation status as the client. On the other hand, all participating clients should also share the same simulation state as the assessment process may apply to an entire group. However, just sending the simulation model continuously to the clients is not enough.

First test runs revealed coarse transitions whenever the client received the newest state of the simulation model from the server.

The first important aspect has to be considered regarding the actual network communication. As a message takes several milliseconds reaching its destination, it is important to incorporate this delay when copying and continuing the simulation model at the destination, whether it is a client or the server. But it cannot be assumed that client and server clocks operate synchronously. Therefore, a support service was necessary to synchronise the clocks between client and server. This has been implemented as a simplified roundtrip message exchange; whereas the server clock is always considered going right. Once the client starts up, the current time is requested from the server. The client further adds the exact sent time to the message; whereas the subsequent server response will contain both the server time as well as the original request time of the client. Based on that, the client calculates and includes the difference between client and server time as well as an additional time offset by cutting the total time of the request cycle in half. This all happens under the assumption that both communication directions have nearly constant and equal delay throughout the process. (see, "Network Time Protocol", 2013; see also, Mills, 1991) Based on the assumption that the time is then sufficiently synchronised between the two systems, each cell message carries a timestamp identifying the last moment the simulation model was propagated. Incoming models are immediately propagated to the current time before being integrated into further rendering or calculation cycles.

In addition, whenever the client takes over to become the simulation master – that is when a user is moving the pendulum manually – it is not guaranteed that no other user would do so meanwhile. Thus, blocking incoming model updates under any circumstances is not a solution. In order to detect if a specific client has really seamlessly taken the absolute control of the pendulum one must wait for a full roundtrip of messages between client and server to ensure no other client has come first. Therefore, the simulation model was extended by a random UUID signature field. Based on this signature an exclusion mechanism expects the following cases:

1. If no last signature is known to the client or the last signature is identical to the incoming signature, the server – or any other client – is assumed to be simulation master and the new model state is accepted;

2. If the incoming signature is identical with the last one, but the local simulation model has generated a new signature, the incoming model is declined. Thus, the client is in a pending state to become simulation master;

3. Finally, if the incoming signature is identical to the local signature, the clients claim is confirmed; otherwise the exclusive control is immediately released.

In accordance with all these considerations, the first system experiments appear to be working almost as expected. However, they have not been tested with real collaborative interaction so far.

6.2.2 Interface to the Assessment Module

Up to this point, it has been ensured that the simple pendulum provides an authentic simulation as well as a smooth synchronisation of the distributed simulation models. The final step is to connect the managed cell on the server with the tracking system of the assessment module. Thus, the `PendulumCellMO` class implements the `SemanticStateProvider` interface of the assessment module (see Figure 6.12) in order to report about *dynamic* simulation data (`deflection_angle` and `max_deflection_angle`) as well as *changeable* properties (`length`, `mass`, `friction` and `theoretical_frequency`). Furthermore, a few static attributes are supposed to describe the 3-D object in semantic-enabled terms.

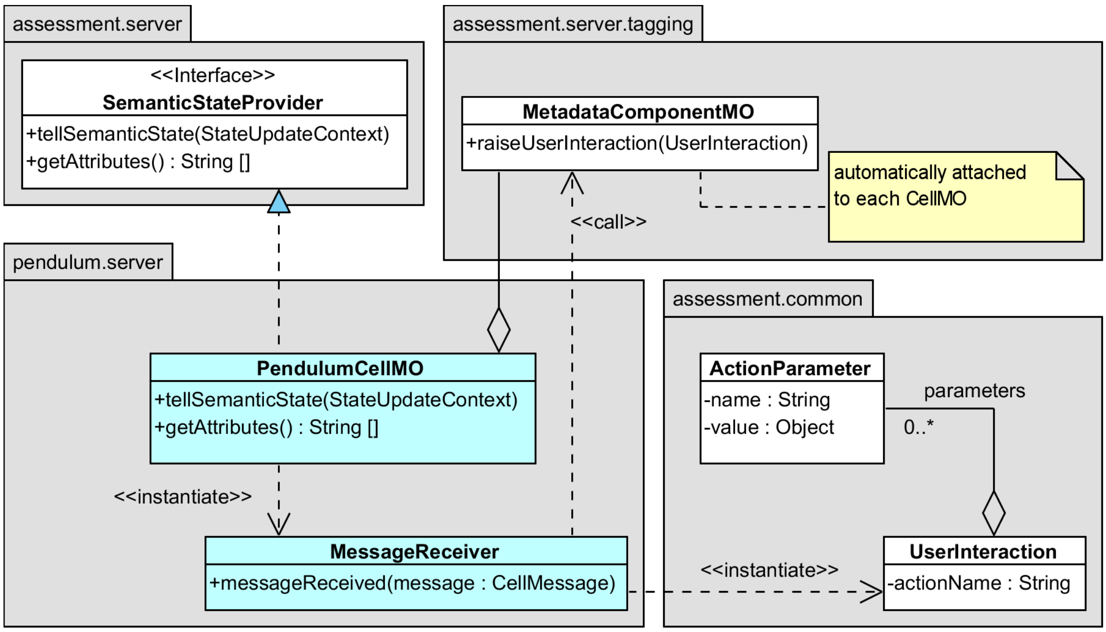
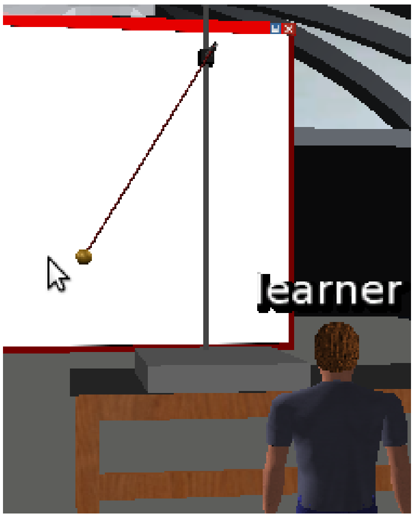


Figure 6.12: Connection between server-side pendulum cell and assessment module

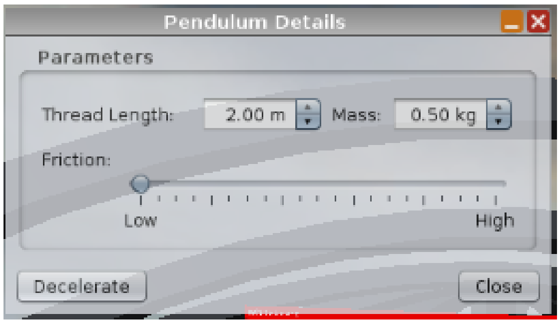
The nested `MessageReceiver` instance of the managed cell object is further responsible to construct `UserInteraction` objects based on incoming client messages and raise user interactions through the attached `MetadataComponentMO` object. This leads to the construction of the appropriate semantic events. Supported player actions include *hold* and *release* of the pendulum mass with the mouse pointer, as well as the change of *length*, *mass* and *friction*.

6.2.3 User Interface

The client-side representation of the pendulum is depicted in Figure 6.13a. The mass can directly be moved through the mouse pointer and it is either possible to release the pendulum from standing position or during movement (interpreted as push). By double-clicking the platform of the pendulum or using the context menu item *'Pendulum details ...'* a HUD panel (see Figure 6.13b) with additional options is opened. This panel allows the adjustment of simulation parameters, including the length of the thread, the mass and a rather unspecified level of friction (the velocity-dependent damping).



(a) 3-D direct motion control



(b) Pendulum Details

Figure 6.13: User interface of the simple pendulum simulation.

6.2.4 Tools for an Exemplary Assignment

The pendulum simulation represents without any doubt the most important component for an applied example of the assessment system. However, it was also necessary to find a concrete task for learners in order to conduct a useful experiment with the pendulum. But for this – in order to make measurements – at least a *stopwatch* was found an important additional tool; which needs also being compatible with the assessment approach. Due to the limited time, an additional cell has been created – directly within the context of the same OWL module – that offers a simple environment for an assignment (also discussed in Maderer & Gütl, in press). This assignment is represented as box – labelled with a question mark (Figure 6.14a) – that opens a HUD panel once it is clicked (Figure 6.14b). This panel includes not only a simple stopwatch, but also a submit form to enter a frequency value and finish the task.

(a) 3-D representation of the 'assignment box' (b) Stopwatch and submission HUD panel

Figure 6.14: Representation of the assignment to determine the periodic time and frequency of the pendulum.

Technically, the assignment box represents an independent cell within OWL. That means it is not in any way coupled with the pendulum simulation. The only relation is given through theoretical (semantic) considerations, as well as in the context of an external assessment system. Thus, the server-side managed cell of the assignment box does also implement the `SemanticStateProvider` interface and is consequently capable of reporting user interactions and state updates for the assessment module.

6.3 Improvements of the SOFIA Evaluation Service

The SOFIA evaluation service has been used as external assessment system to realise assessment showcases in the context of OWL. Due to external circumstances, it was necessary to ensure that the service is operational on Linux systems. Besides, a few smaller additions and changes have been made to the original system.

6.3.1 Porting to Mono

First of all, as the evaluation engine has originally been developed under the original *Microsoft .NET Framework* it was important to ensure that it also functions with the *Mono* runtime environment. That is particularly the case because a Windows server was no longer available. Although Mono is commonly supposed to be mostly compatible on the binary level with applications that have been written on .NET Framework this is not always the case (cf. Mono, n.d.). In the given situation slightly different behaviour among the XML serializer libraries made some modifications necessary. In fact it has turned out that defining a fluent XML format and mapping it with the default automated serialization solutions to an object hierarchy is rather problematic. Nevertheless, without going into detail, with some adjustments and workarounds it was possible to get the *SOFIA evaluation engine* operational with Mono. It can only be recommended to use better alternatives for futher developments.

Regarding the web service, by using XSP – a basic webserver system that is able to execute ASP.NET web applications – it is no problem to host the web service application in the context of Apache or as a standalone application. Like many Linux applications Mono and XSP can be rather easily compiled from its source code. That makes it possible to compile the environment against a specific target path within a Linux user account and install arbitrary versions in parallel. Thus, the entire environment can be hosted from a less privileged user account. (see Mono, n.d.)

6.3.2 Implemented Features

This subsection should briefly describe the new features that have been implemented to support the proof of concept in OWL:

- For the sake of the physics experiments it was found useful to extend the existing logical *compare* operation of the *assessment model* by a new type of comparison. This *'EqualWithDelta'* type is supposed to compare two values within a given tolerance. This tolerance is either specified as an *absolute* or *relative* value. The latter calculates the difference between the two values and uses the first value as point of reference. This behaviour should simplify the formulation of assessment rules to evaluate the measurement performance of learners during experimentations.
- On the level of the web service a new operation has been added to provide a list of available assessment modules – file names are currently interpreted as task names – to make it easier to switch between different assessment models from within the immersive environment.
- Based on an intermediate advice, colours to indicate different types of content for text-based feedback have been added. This change effected the feedback statement for the assessment model, as well as the web service API. The codes are supposed to differentiate between success (or positive) messages, recommendations or hints for improvements, as well as errors.
- The evaluation engine has further been modified to not remember internal state variables in the context of a single behaviour pattern but throughout the evaluation

context. Thus, it is possible to adapt the evaluation and feedback process based on previously matched patterns. This would classify the engine as a *triggering agent* (cf. Nareyek, 2001).

6.4 Summary

This Chapter has explained importing aspects of the implementation and improvements of all prototypes involved with the proof of concept. It included implementation details about the assessment module for OWL, the development of an assessment-compliant pendulum simulation, as well as some additional features for the SOFIA evaluation engine.

The assessment module itself is responsible for tracking semantically-enabled information about the movements and interactions of the users' avatars and incorporates a mechanism to distribute feedback messages across the involved clients. Spatial place marks and a metadata component – added to each cell – have been implemented to specify location in an abstract semantic way and to provide identification, as well as discrete distance definitions for the perception of cells. The proximity definitions of the metadata component are further required to decide on which information is relevant for an accurate overview of the user's perceptual context. Therefore, relevance of data has been separated into three levels – full, changeable and dynamic. When a user interaction on a 3-D objects is raised, dynamic state of each interactive and assessment-compliant cell in the perceptual context of the user is taken into account, including the relevant changeable data of the triggering object itself. Beyond that, an interface to the web service of the SOFIA evaluation engine has also been integrated to the assessment module.

The simulation of the simple pendulum is based on numeric integration of the equations of motions to provide a rather authentic experience. Time-critical aspects for a smooth and exact synchronisation between clients and server had to be considered in order to provide a seamless user experience as well as appropriate data for the external assessment system. This included the implementation of a basic time synchronisation mechanism, necessary to incorporate the message delays when sending updated versions of the simulation model to the clients or server. To prevent race conditions when a client takes over the dynamic simulation, a signature-based roundtrip check was implemented when replacing the local state with the incoming simulation model. In addition, an exemplary assignment cell has been developed, featuring a stopwatch, as well as a submit form a calculated frequency value.

Finally, the SOFIA evaluation engine has been ported to Mono and a couple of new features have been implemented. This includes a tolerance-based comparison function, colour codes for text-based feedback, and storage of internal states across different assessment rules.

The next Chapter deals with the user perspective of the implemented prototypes. This includes the available user interfaces, a set of exemplary showcases, as well as the results of an expert evaluation based on the showcases.

7 Showcases and Expert Evaluation

Based on the implemented prototypes three assessment showcases have been designed. These showcases were further used to demonstrate the approach to a couple of experts. The aim of this Chapter is to present the showcases and describe the intention, setup and assessment rules behind them. Furthermore, the methodology and results of the expert evaluation are discussed.

7.1 Assessment Showcases

The purpose of this section is to document three showcases that have been set up in OWL based on the implemented prototypes. Several items have been placed in order to mimic a laboratory setting. The SOFIA evaluation engine has been used to define assessment rules in order to give immediate feedback based on the player actions performed within the virtual world. The first two showcases have already been reported (earlier development state) in Maderer, Gütl, and AL-Smadi (2013); whereas the third showcase is also introduced in Maderer and Gütl (in press).

7.1.1 Approaching the Experiment

This first showcase provides a very basic demonstration of the spatial place mark feature. The laboratory setup has been wrapped up with a spatial place mark, tagged as `experiment-workplace`. Whenever an avatar enters this area its set of location attributes is unified with the tags associated with the spatial place mark. A user interaction called `EnterSection` is communicated along with the modified attributes as well as a parameter containing the new attributes.

The instructional idea behind this showcase arises from practical observations. Learners that have already been conducted the same or similar experiments know that they should read the instructions first. Nevertheless, a newcomer may not know where to find these instructions or does not know what to do. The feedback message (see Figure 7.1) guides the learner to the presentation wall where the PDF document contains all necessary information. In a more sophisticated scenario the assessment rule would consolidate the learner profile via an LMS in order to prevent the feedback message if not necessary for more experienced learners.

The XML assessment rule (see Listing 7.1) for the external web service has been constructed to react on two conditions: First, an internal state variable stores the information if this feedback rule has already been executed since the beginning of the assessment scenario. If that was not the case, the assessment rule further matches against the action of entering a new spatial section where the new attributes must contain the `experiment-workplace` tag. After a successful evaluation, a feedback message is generated and the internal state is updated in order to prevent further messages of the same kind.

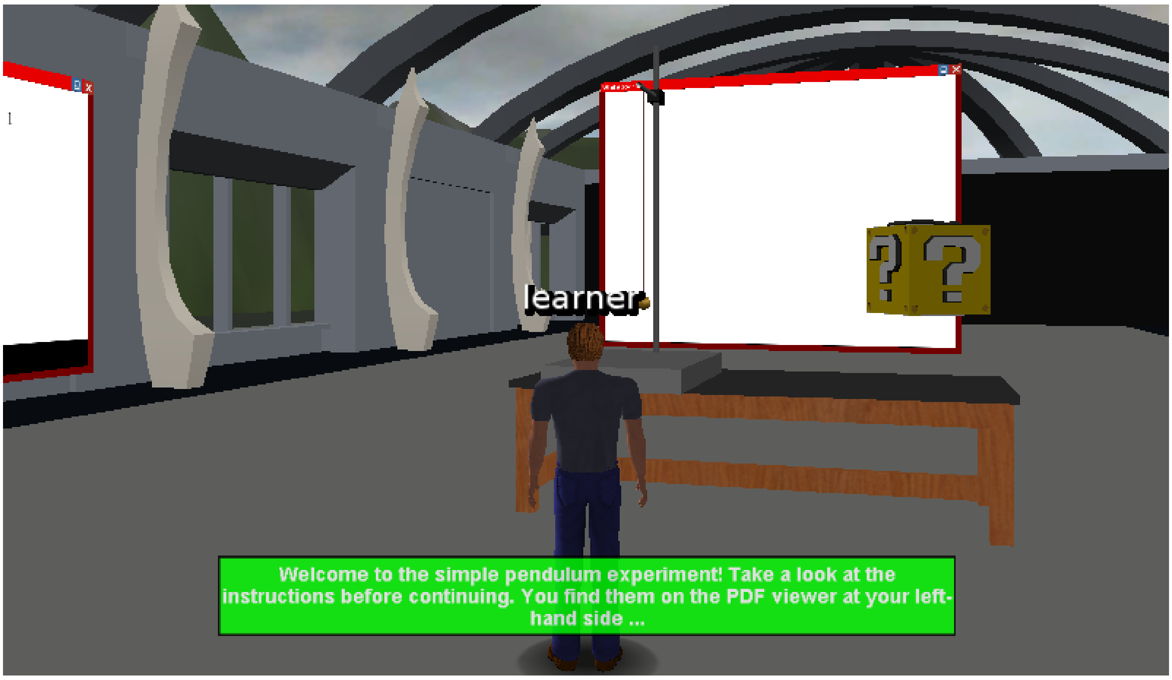


Figure 7.1: An avatar entering the experiment area for the first time receives advice on reading the PDF

```
<behaviour-pattern>
    <match-condition>
        <compare type="Equals">
            <get property-name="welcome_message_done" />
            <value type="Boolean">false</value>
        </compare>
    </match-condition>
    <match-action name="EnterSection">
        <compare type="Contains">
            <get property-name="location-entered" />
            <value>experiment-workplace</value>
        </compare>
```

```
    </match-action>
    <consequences>
        <set property-name="welcome-message-done">
            <value type="Boolean">true</value>
        </set>
        <feedback>
            <text>
                Welcome to the experiment 'simple pendulum'.
                Please refer to the presentation wall.
                It provides you with the necessary theory and
                task descriptions.
            </text>
        </feedback>
    </consequences>
</behaviour-pattern>
```

Listing 7.1: Assessment rule for guiding the learner to the presentation wall

7.1.2 Exceeding Recommended Deflection

The second showcase relies on the state of the pendulum cell itself as well as its possible interactions. After the player has pushed, or deflected and released the pendulum, an appropriate user interaction is generated and the resulting semantic event contains the current state of the pendulum. This includes also its peak amplitude (`maximum_deflection_angle`) derived from the current mechanical energy of the system.

From the instructional perspective the second showcase is concerned with the prevention of systematic errors in measurements. In a later showcase the learner will be required to determine the (theoretically idealised) frequency of the pendulum. This can only be correct if the pendulum still operates in the range of a harmonic oscillator which is valid to be assumed for small deflection angles but not for larger ones. In reality there is a non-linear proportion between periodic time and peak amplitude (cf. subsection 6.2.1, simulation model). If the described behaviour happens an appropriate feedback message (see Figure 7.2) warns the player about the potential mistake.

In this case the assessment rule is based on a match against the `release` action, as well as a subsequent comparison between the defined threshold of approximately $\phi \approx 35° \approx 0{,}52\,\mathrm{rad}$ and the actual peak amplitude. If the entire rule evaluates successfully, a feedback is created that warns about the deflection being too high.

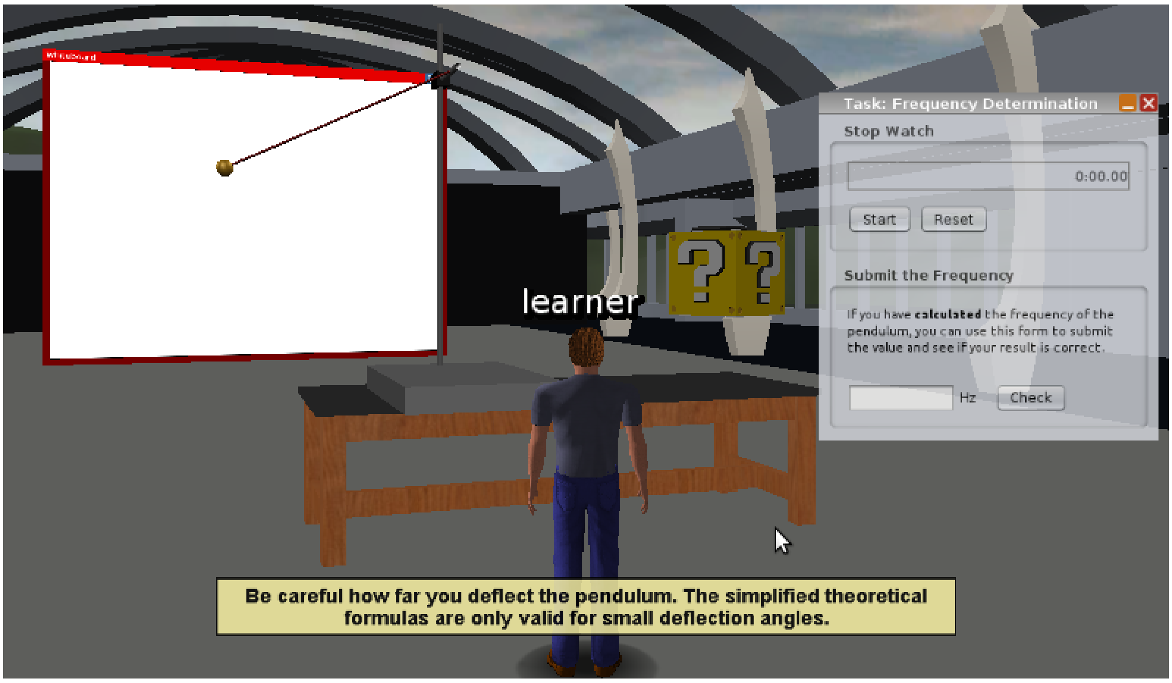


Figure 7.2: The avatar has induced the pendulum too intensively. The advice is warning that this could be too much for accurate measurements.

7.1.3 Frequency Measurement

From the technical perspective this showcase relies on two independent objects that are available in the situational context of the avatar. On the one hand, interaction and state of the pendulum contribute to the assessment process. On the other hand, the assignment panel that contains the stop watch as well as the submit form, which represent the important key actions in this scenario.

This showcase requires the conceptual understanding and skills of the learner. The task – as given by the assignment panel, respectively a possible statement within the guidance PDF – requires the learner to determine the frequency of the pendulum. This is achieved by taking the time of at least one full period and calculating the frequency through the reciprocal value of the determined periodic time.

The assessment rule for this showcase is more complicated. It separates into two behaviour patterns. The first one is responsible to evaluate the quality of measuring the periodic time of the pendulum. This is attained by the comparison of the deflection angles at the times the start and stop buttons of the stop watch were operated. The result is stored internally, with the option to communicate an intermediate message for less experienced learners – exemplarily realised as a duplicated and slightly modified assessment model at this time. The second pattern is triggered by the submit button,

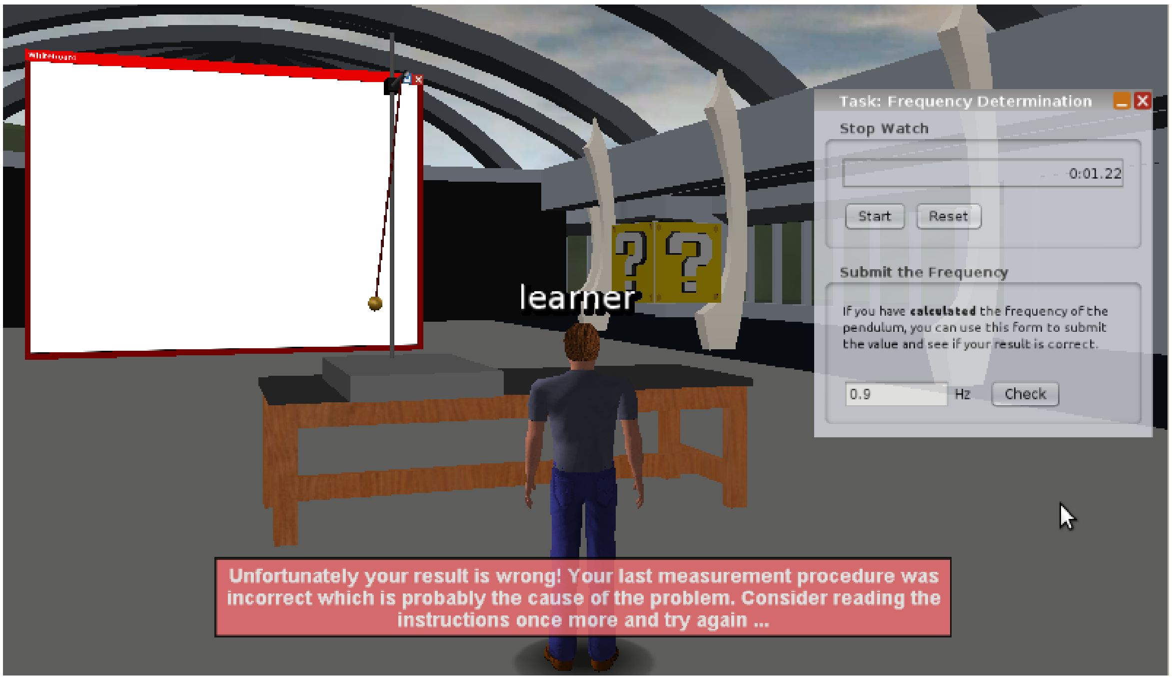


Figure 7.3: Avatar has activated the assignment task and has successfully stopped the time in order to calculate the frequency.

communicating the changes of the frequency input field. Depending on the quality of the previous measurement, wrong results are differentiated between problems regarding either the measurement activity (see Figure 7.3) or calculation, based on the internal state of the assessment engine.

7.2 Expert Evaluation

This section reports about an initial expert evaluation that was administered based on the implemented prototypes and showcases. A shortened version of this section has just been published in Maderer and Gütl (in press). Some parts, especially data lists and summarised recommendations, can match almost literally with this document.

The prototypes are not ready at this time to be tested in a real student context. Thus, the main purpose for this early evaluation was to confirm the concept from the perspective of practitioners, mostly teachers and university lecturers in the field of physics. Beyond that, the evaluation should also provide information for the prioritisation of further research directives.

7.2.1 Method

The expert evaluation was based on a survey consisting of several fixed-response questions where some especially invited for further comments in written form. Due to the early development state of the prototype and the reduced focus on usability, the prototype was not tested by the experts themselves. Instead, the three showcases were presented through a live demonstration, including further background information about the usage options of immersive 3-D virtual worlds. Further interesting comments or attitudes have been noted as keywords besides the structured questionnaire, as the demonstration featured also aspects of an informal interview. The procedure was separated into three phases:

1. Completion of a pre-questionnaire to obtain background information about the experts. This includes experience with e-learning, opinion on assessment and feedback, as well as questions related to 3-D (learning) environments;
2. Explanation of the project idea followed by the demonstration of the prototype setting and showcases;
3. Evaluation of the assessment and feedback aspects of the prototype through a post-questionnaire, including information about preferences on the usage of such a system and ideas for further application and development.

7.2.2 Results

The group of experts ($N = 9$) that participated in the evaluation of the prototype consists of the following persons:

- Experienced school teacher and expert for subject didactics in physics; university lecturer in terms of teacher training (male; age group 50-59);
- Five teacher trainees (students) in the domain of physics (all 20-29, except one 30-39; only one female);
- University lecturer/researcher in experimental physics (male; 30-39);
- Teacher trainee in school physics *and* university lecturer/researcher in chemistry (female; 20-29);
- Computer science expert who has research experience in the field of 3-D learning environments (male; 40-49).

All participants were German native speakers. Due to that, not all comments were written in English. The remaining section uses close translations as well as minor linguistically improved statements. The original material is available on the attached data disc.

Pre-Questionnaire

Table 7.1: Topic related entrance questions

Question or statement	**Mean (SD)**
Please estimate your experience with e-learning solutions. *almost unfamiliar (1) – inexperienced – moderate – experienced – very experienced (5)*	3.11 (0.99)
Do you think traditional electronic assessment formats, such as multiple choice questions are effective to evaluate the learning outcomes of students in science and related disciplines? *strongly disagree (1) – disagree – neutral – agree strongly agree (5)*	2.89 (0.87)
Do you think it is important to conduct low stake assessment (formative assessment) during learning activities in general? *irrelevant (1) – rather unimportant – moderate – important – very important (5)*	4.11 (0.74)
How important would you consider timely (immediate) feedback for learners? *irrelevant (1) – rather unimportant – moderate – important – very important (5)*	4.56 (0.50)
How important would you consider adapted feedback for individual learners and groups based on their individual knowledge/skill/competency levels and learning progress? *irrelevant (1) – rather unimportant – moderate – important – very important (5)*	4.44 (0.50)

The first set of questions (see Table 7.1), which was mainly designed to eliminate a potential bias against e-learning and technology, did somewhat match up with the author's perception. The familiarity with e-learning is generally rather average. Despite the computer science expert, only two experts would consider themselves experienced with e-learning solutions in general. This being said, rather a confirmation of the literature review is that only multiple-choice questions have been used in the context of e-assessment – in the role of a teacher or student – by almost all participants (see Table 7.2). Other formats, such as gap texts or free text answers, have only been selected by half of the experts. Only the computer science expert has selected an integrated achievement system in simulations. No other assessment formats were named. When it comes to the effectiveness of multiple-choice questions in science education the field of answers is rather average and broad. Comments included that it was generally a controversial topic and it would also depend on what should be assessed. It was rather seen effective for the assessment of knowledge but would also depend on the quality of the questions and especially the distracting wrong answers. The importance of feedback and adapted feedback for individual learners was with no doubt determined as important or very important by each expert. Only the importance of assessment during learning activities was rated slightly less important in average. This suggests that the idea of assessment was rather interpreted as a formal concept of tests rather than feedback being an integral concept of formative assessment.

Table 7.2: Experienced or administered e-assessment formats (predefined, multiple choice)

Assessment Format	**Count**
Multiple (Single) Choice Questions	8
Gap Texts	5
Label association on pictures	4
Free text answers	5
Calculation task with numeric answer field	4
Simulations with integrated achievements	1
Other forms	0

The second set of questions (see Table 7.3) was dedicated to the usage of 3-D virtual environments, such as computer games and simulations. There are no participants who play 3-D computer games on a regular basis. Also the usage of simulations is located in the average sector; only one student and the experienced teacher confirm a regular usage in the context of learning activities. The idea of 3-D virtual worlds to be used as learning environments is practically unknown to most of the experts. However, the greater part believes that 3-D learning environments can motivate and improve learning.

Post-Questionnaire

After a brief introduction to the idea of 3-D virtual worlds – client-server based systems with focus on collaborative aspects – the prototype was demonstrated to the experts. It was emphasised that this is a prototype that demonstrates only a small set of showcases

Table 7.3: Experience with 3-D virtual learning environments

Question or statement	**Mean (SD)**
Do you personally play 3-D computer games? *never tried out (1) – hardly ever – sometimes – regularly – very often (5)*	2.22 (0.63)
Have you used computer simulations for learning activities (as a teacher or student)? *never tried out (1) – hardly ever – sometimes – regularly – very often (5)*	2.78 (0.79)
How much do you know about 3-D virtual worlds to be used as (authentic) learning environments? *completely unaware (1) – heard of – tried out – already used them (professionally)(4)*	1.78 (1.03)
Do you believe 3-D learning environments can motivate and improve learning? *strongly disagree (1) – disagree – neutral – agree – strongly agree (5)*	3.78 (0.63)

and that the evaluation should be focused on the feedback and assessment aspects of the scenario. In addition to the showcases described in the previous sections, the assessment system was manually switched to a different assessment model to exemplarily demonstrate an increased frequency of feedback under the assumption of lower learner competencies.

All questions regarding the evaluation of the prototype are summarised in Table 7.4. The experts were first asked to give an overall impression of the demonstrated assessment and feedback aspects. Most of the experts found the concept good or very good and one participant was particularly impressed by the overall demonstration. The approach was furthermore denoted as *"well done"*; one participant called it a *"practical experiment"*, *"easy to handle"* and that it is a *"good idea with the helping PDF"*. It was further stated that the *"different colours for positive and negative feedback are fine for 'visual types' "*. Positive comments also mentioned the immediate and individual feedback; and it was further considered an interesting approach. One participant further commented *"challenging tasks as motivation for students"* as a positive aspect.

Nevertheless, two participants – which both already had a more neutral opinion on the value of 3-D learning environments in general – were not particularly convinced by the approach and it was not easy for them to focus on the actual question. As a matter of fact the demonstration itself was already criticised as he or she stated: *"3-D graphics is in my view not necessarily required (with the pendulum)"*. One of these participants also stated the opposite of what the approach intends to achieve: the *"feedback is very generic"* and it would be *"complicated to arrange feedback individually"*. Besides, the feedback was experienced too small and it should be centred. Another comment suggested to *"change the green [feedback] colour to something a little bit*

Table 7.4: Prototype evaluation in terms of assessment and feedback aspects

Question or statement	**Mean (SD)**
What is your overall impression of the demonstrated assessment and feedback aspects? *insufficient (1) – barely sufficient – satisfactory – good – very good (5)*	4.11 (0.74)
I think the example is authentic. *strongly disagree (1) – disagree – neutral – agree – strongly agree (5)*	4.33 (0.47)
I think the textual feedback provided at the bottom of the window was helpful. *strongly disagree (1) – disagree – neutral – agree – strongly agree (5)*	4.44 (0.68)
I think the feedback provided would improve the outcomes/results of students. *strongly disagree (1) – disagree – neutral – agree – strongly agree (5)*	4.33 (0.47)
I think the feedback provided would improve the understanding of students. *strongly disagree (1) – disagree – neutral – agree – strongly agree (5)*	4.00 (0.67)
I think the kind of player actions evaluated – measurement activity and calculation – can be used in accordance with competency-based learning models – i.e. the approach is valid to reflect on skills and competency levels of the learners. *strongly disagree (1) – disagree – neutral – agree – strongly agree (5)*	3.94 (0.68)
I think the different intensity of feedback messages is appropriate to catch up with the different competency levels of students. *strongly disagree (1) – disagree – neutral – agree – strongly agree (5)*	4.22 (0.63)

brighter" as well as to *"force students to read the instructions (e.g. by implementing a control task)"*.

Besides, the following concrete ideas and recommendations for further improvements have been given:

- Sound should be added, and maybe also *"laboratory music"*;
- An important recommendation was that the computer should read the feedback aloud;
- *"more comments and hints regarding the expected actions would be fine; I guess that students who are not that talented could be disappointed since they might have problems with starting their own exploration of the virtual world"*;
- Movements should be combined with numerical representations;
- A pocket calculator should be added in-world;

- To facilitate a game-based approach, for instance, the explanation of a formula could be released as a reward for achievements in the practical exercise;
- The considerations of external influences, in the context of a pendulum simulation for instance an eddy current brake.

The exemplary scenario was commonly perceived as an authentic environment. But one participant expressed that the surrounding of the experiment as well as the 'assignment box' would not be included in this rating. Nevertheless, while most of the experts agreed on the usefulness of the feedback messages, as well as the improvement of the learning outcomes, the improvement of understanding was rated slightly lower.

Especially with the new wave of competency-based learning in schools it was asked if such behaviour evaluation could be used in accordance with those teaching approaches. This question was generally answered more differentiated. On the one hand, the subject didactics expert as well as some students (strongly) agreed, as the showcase fits the context of (model-based) experimentation, measurement and data evaluation – which is in fact present in the Austrian competency model for science education in schools (see "Kompetenzmodell Naturwissenschaften 8. Schulstufe", 2011). Particularly one participant noted examples, for instance, in terms of frequency: *"understand the concept of a period"* and *"apply skills to measure accurately"*. On the other hand, the computer science (and research project) expert was rather neutral about it and stated that it should be more clearly expressed what competencies are required.

Table 7.5: Desired information for teachers about students (predefined, multiple choice)

Type of Information	**Count**
Overview of the students' problems in different rubrics	9
Detailed report about each student	2
Summary of entire learning group / class	7
Activity Reports	3
No information	0
Other	1

Regarding the outlook for further development and application the experts have rather expressed interest in compact information (see Table 7.5). One expert expressed additional interest for data about students' collaborative interactivity, highlighting also the communication competencies. Nevertheless, an informal comment included that more information would always be good, but time would be the critical factor in any cases. The same being also true for the design of own assessment rules with a graphical editor (see Table 7.6), which was generally agreed on – independent of the personal programming skills. However, the two experts that were not remarkably convinced by 3-D learning environments – and the assessment approach – answered neutral again. Most of the experts also agreed that a challenging aspect in such simulation environments would also contribute to more motivation and a better learning experience. The incorporation of assessment information into the grading schema was again rather neutral and differentiated among the participants; one remark expressed legal concerns.

Table 7.6: Further considerations for future improvements

Question or statement	**Mean (SD)**
I could imagine designing assessment rules – e.g. for feedback provision – based on a graphical editor on my own. *strongly disagree (1) – disagree – neutral – agree – strongly agree (5)*	4.00 (0.67)
I think a challenging aspect – e.g. progress information between student groups – would improve the motivation and learning experience. *strongly disagree (1) – disagree – neutral – agree – strongly agree (5)*	4.44 (0.68)
I think the assessment information obtained from such virtual activities could be incorporated into the grading process. *strongly disagree (1) – disagree – neutral – agree – strongly agree (5)*	3.78 (0.92)

Table 7.7: Usage of immersive 3-D virtual worlds for teaching (predefined, multiple choice)

Type of Information	**Count**
Preparation for real practical lessons	4
As support for courses/activities which do not feature practical (laboratory) lessons	9
If embedded in a greater context, as replacement for real practical activities	3
Not at all (please provide comment)	0
Other usage	6

The usage of immersive 3-D virtual worlds (see Figure 7.7) focused on supportive aspects (most often selected). But also several other ideas have been given, including homework exercises, comparison with real experiments, *"support after real practical lessons"*, simulation of non-available experiments, as well as *"trail-and-error"*. Finally, additional value in the usage of such 3-D learning environments and assessment approaches was seen in the explanation of concepts, training, comparison between model and reality, as well as fostering of communication skills and collaboration.

8 Lessons Learned

The aim of this penultimate Chapter is to discuss lessons learned from the theoretical, technical and user perspective.

8.1 Theoretical Aspects

The general issues regarding science education are not new to the author due to his personal background as a teacher trainee in school physics. These topics as well as constructivism, the problems of misconceptions and inquiry-based learning approaches and authentic experiences have been discussed in corresponding courses. Based on that, also a general interest in using computer technologies and simulations, particularly with the additional background in computer science, is quite understandable.

The field of immersive education and the usage of immersive 3-D virtual worlds has been emphasised as promising technology for learning. Psychological concepts – such as spatial presence and flow – explain the perception of being immersed into digital environments. Especially the term virtual world has been clarified as a new approach to real-time online interaction through avatars with focus on social activities; although the principles behind have existed for some time. The literature review has highlighted several pedagogical and practical benefits of such environments, especially contextual and collaborative learning. But also computer games and dedicated simulations such as physics experiments belong to this area of research. Nevertheless, the real benefits of educational games seem less conclusive at this time; whereas simulations have stronger empirical support to improve learning outcomes.

The importance of assessment and feedback has further been stressed from multiple perspectives. Formative assessment and timely feedback are important for the success of learning but also psychological concepts such as flow are strongly related to challenge, assessment and feedback – aspects also immanent in computer games and simulations. Besides, proper e-assessment is supposed a necessary requirement to keep up with the increased usage of technology for learning activities and to enable authentic and novel assessment approaches.

8.2 Technical Perspective

The complexity of virtual world platforms was practically confirmed through the technical research and development phase of the prototypes. First experiences with the integrated proof of concept support the feasibility of the generalised behavioural assessment approach and demonstrated the general applicability in the context of a time-critical, dynamic simulation of a physics experiment. The stack of systems was capable of providing very timely feedback, also in situations where all three components – virtual world client and server, as well as assessment system – where separated through public internet connections. However, a specific problem that has been identified – especially during the first online tests – regards the dynamic state of the pendulum whenever a player takes measurements with the stopwatch. What has not been anticipated in the first place is the message delay between the user interaction on the OWL client and the creation of the semantic event on the server-side. This delay leads to a significant mismatch between the players' actual perception of the dynamic state of contextual objects and the perceptual situation that is seemingly accurate for the assessment system. There is further no guarantee when a Darkstar task is really exactly executed on the server-side. A first solution approach to this problem could be to manage a brief history of past simulation data on the server and compile events always based on the original triggering time of an event or player action.

Being not completely new information, as it was already known to the start of this thesis, the behavioural assessment approach of the SOFIA evaluation engine requires significant improvement. On the one hand, a separation between behaviour patterns and actual assessment rules would be important. At this time, the multiple instances of match processes can lead to unexpected results. And sometimes an assessment rule may require to be based on several independent behaviour sequences. However, a special behaviour pattern could be the trigger for the assessment rule, which could than further revise other behaviour patterns that were detected before. The second major drawback refers to the missing support of objects. A redesign of the evaluation mechanism should therefore be based on contextualisation of objects which are required for a specific set of behaviour patterns. For instance, one might define that a pendulum and a stopwatch are required objects in the current context of the player in order to apply behaviour patterns that evaluate the skills on measurement. This makes no sense with neither of these objects being present. Further, it is possible that there a multiple objects of the same kind available in the context. Therefore also an appropriate autonomous selection mechanism would be required, such as considering only the most recent used pendulum of the player, which could then be identified by a fixed alias name within the context of the assessment rule. Nevertheless, such concerns seem important for any kind of assessment system, even if it is based on advanced artificial intelligence approaches.

8.3 User Perspective

This thesis was primarily concerned with the technical challenges involved in the realisation of a flexible assessment system and its prototypical application in the context of 3-D virtual worlds and STEM education. Thus, it was only possible to conduct a concept evaluation with a small number of experts and showcases.

The first part of the evaluation – focusing on the experts' background and opinions – has without any doubt confirmed the importance of formative assessment and feedback from the perspective of acting as well as soon-to-be teachers. However, it also confirms that the experience and knowledge regarding e-learning, more precisely e-assessment and particularly the usage of immersive education technologies is rather low.

The demonstration of the exemplary showcases was altogether well received. Several comments about the assessment and feedback aspects were rather positive. But also the negative feedback confirms the need for an increasingly flexible and learner-adaptive assessment framework. Also concrete suggestions for improvement such as sound – even speech synthesisation for feedback messages – provide interesting ideas for future improvements. Beyond that, also data about the potential practical usage of such environments as well as desired information about student progress was collected. This can help to define priorities for future developments.

9 Conclusion

9.1 Summary

STEM education is an important aspect of economy and society, but is facing severe challenges. Beside social-cultural issues; the lack of interest, motivation and engagement is strongly connected to the way content is presented in relation to perspectives of modern society. Furthermore, authentic teaching and assessment approaches with focus on competencies are considered important to counteract these problems. Immersive environments – such as 3-D virtual worlds – are supposed to provide new approaches to learning by focusing on (social) constructivist learning theories and providing an authentic context in collaborative situations. In comparison to real world learning settings, virtual environments can mimic situations that would be hardly possible under real circumstances. In addition, embedding contents and simulations into game-based environments can further improve motivation and engagement.

Due to the importance of formative assessment and feedback – and based on the current status of e-assessment in general – a literature research on the latest progress of assessment approaches integrated into immersive environments has been conducted. There are still several issues: In contrast to serious games, no established assessment approach has been reported for 3-D virtual worlds that would focus on complex player behaviour. Despite that, approaches partially lack authentic tasks, are strongly bound to specific platforms and would require both – environment developers and teachers – to be an expert on each other's field. Finally, also adaptivity for individual players should be considered.

Based on these issues as well as first approaches in game-based learning, a flexible assessment approach has been proposed, targeting different kind of immersive environments and application domains. With focus on player behaviour, this semantic-enabled approach is supposed to externalise the assessment process and provide immediate feedback and guidance from different assessment systems. The primary aspect is an assessment module, developed for each platform and responsible for event tracking and feedback delivery. Therefore, an exemplary assessment module has been developed for Open Wonderland (OWL) to provide a proof of concept in the domain of physics education. This included also the development of an assessment-compliant simulation of a simple pendulum, as well as the integration with the earlier developed prototype of a context-dependent rule-based assessment system. The prototype implementation was particularly concerned with an authentic, dynamic and smooth simulation of the pendulum in accordance with time-critical assessment measurements. This was a particular challenge due to the client-server based infrastructure of OWL but could be satisfactory resolved from the standpoint of a first prototypical approach. The assessment showcases of the prototypes have been presented to a couple of experts, mostly related to physics teaching. The approach received mostly positive feedback, as well as several suggestions for improvement.

9.2 Outlook

This thesis could only provide the first steps for a flexible assessment system in immersive environments. Due to the expert responses it seems reasonable to pursue this approach in the future. However, a lot of work needs to be done. On the one hand, it is required to provide a more sophisticated and richer scenario to actually approach students and conduct usability experiments. Beyond that, also the learning success must be evaluated and teachers should interact with the design of assessment plans and access student information via LMS. On the other hand, technical improvements are necessary on several dimensions. For instance, the assessment module requires further improvements and could possibly be separated into more fundamental libraries, also integrated with a simulation framework that simplifies the development of time-critical and assessment-compliant OWL cells (3-D objects). This includes also the integration of more sophisticated feedback mechanisms, such as non-player characters (NPC) coupled with speech synthesis.

When it comes to the assessment system, the SOFIA evaluation engine requires significant improvements, which may include a complete redesign that is compatible with all information provided by an immersive environment. Nevertheless, also other software agents which rely on sophisticated artificial intelligences algorithms should definitely be tested with the proposed interfaces. In addition, it is important to actually connect these systems with LMS and learner profiles to implement the automated adaption to learner preferences and learning progress of individual learners.

Finally, it was found that semantic descriptions of 3-D environments is a desired development in the future. Nevertheless, the scope of this work did not allow further research regarding this topic. Thus, a major aspect in future research would involve an extensive literature research regarding standards for semantic descriptions of 3-D objects, as well as if standards only exist for static description of objects or if and how far this involves also timely changes and interaction with objects. Based on this, the semantic knowledge repository as discussed in the conceptual architecture of the flexible assessment approach may be realised as an integrated part.

List of Abbreviations

AEMT	Advanced Educational Media Technologies
ALICE	Adaptive Learning via Intuitive/Interactive Collaborative and Emotional systems
AM	Assessment Module
API	Application Programming Interface
HUD	Heads-up Display
IBL	Inquiry-based learning
ICT	Information and Communication Technology
JAR	Java Archive
jME	jMonkeyEngine
LMS	Learning Management System
MTGame	Multi-threaded Game Engine
NPC	Non-player Character
OWL	Open Wonderland
PDS	Project Darkstar
SAS	Shared Application Server
SOAP	Simple Object Access Protocol
SOFIA	Service-Oriented Flexible and Interoperable e-Assessment
STEM	Science, Technology, Engineering and Mathematics
XML	Extensible Markup Language

Bibliography

Abbasi, A. & Baroudi, U. (2012). Immersive environment: an emerging future of telecommunications. *MultiMedia, IEEE*, *19*(1), 80–86. doi:10.1109/MMUL.2012.7

Akilli, G. K. (2007). Games and simulations: a new approach in education. In D. Gibson, C. Aldrich, & M. Prensky (Eds.), *Games and simulations in online learning: research and development frameworks* (Chap. 1, pp. 1–20). Information Science Publishing.

ALICE. (nodate). *Adaptive learning via intuitive/interactive collaborative and emotional systems*. Retrieved May 2013, from http://www.aliceproject.eu/

Arroyo, D. M., Rodríguez, P. S., Calle, D. P., Kloos, C. D., Espiga, M. I., & Hernández-Leo, D. (2010). Assessment in 3d virtual worlds: qti in wonderland. Congreso Iberoamericano de Informátia Educativa, IE2010, Santiago de Chile, Chile. Retrieved from http://www.researchgate.net/publication/228855500_Assessment_in_3D_virtual_worlds_QTI_in_Wonderland/file/79e4150cb694142af8.pdf

Bartle, R. A. (2003). *Designing virtual worlds*. New Riders. Retrieved March 17, 2013, from http://my.safaribooksonline.com/book/web-development/usability/0131018167

Becker, K. (2007). Pedagogy in commercial video games. In D. Gibson, C. Aldrich, & M. Prensky (Eds.), *Games and simulations in online learning: research and development frameworks* (Chap. 2, pp. 21–47). Information Science Publishing.

Bell, M. (2008). Toward a definition of virtual worlds. *Journal of Virtual Worlds Research*, *1*(1). Retrieved from http://journals.tdl.org/jvwr/index.php/jvwr/article/view/283/237

Blunt, R. (2009). Do serious games work? results from three studies. *eLearn*, *2009*(12), 1. doi:10.1145/1661377.1661378

Bybee, R. W. (2010). What is stem education? *Science*, *329*(5995), 996–996. doi:10.1126/science.1194998

Carlile, O. & Jordan, A. (2005). It works in practice but will it work in theory? the theoretical underpinnings of pedagogy. *Emerging Issues in the Practice of University Learning and Teaching. Dublin: AISHE*, 11–26. Retrieved from http://www.aishe.org/readings/2005-1/carlile-jordan-IT_WORKS_IN_PRACTICE_BUT_WILL_IT_WORK_IN_THEORY.pdf

Chang, W.-C., Wang, T.-H., Lin, F. H., & Yang, H.-C. (2009). Game-based learning with ubiquitous technologies. *Internet Computing, IEEE*, *13*(4), 26–33. doi:10.1109/MIC.2009.81

Charman, D. & Elmes, A. (1998). *Computer based assessment (volume 1): a guide to good practice*. SEED Publications, University of Plymouth. Retrieved from http://www.gees.ac.uk/resources/hosted/seed/cba1.pdf

Colburn, A. (2000). Constructivism: science education's 'grand unifying theory.'. *Clearing House*, *74*(1), 9. Available from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=3537008&site=ehost-live.

Cook, J. & Jenkins, V. (2010). Getting started with e-assessment. Retrieved from http://opus.bath.ac.uk/17712/1/Getting_started_with_e-assessment_14Jan2010.pdf

Crisp, G. (2007). *The e-assessment handbook*. Continuum.

Crisp, G. (2012). Meaningful assessment within virtual worlds–what should it look like? Retrieved from http://www.inter-disciplinary.net/wp-content/uploads/2012/02/crispepaper.pdf

Crisp, G., Hillier, M., & Joarder, S. (2010). Assessing students in second life–some options. *Curriculum, technology & transformation for an unknown future. Proceedings ASCILITE Sydney*, 256–261. Retrieved from http://ascilite.org.au/conferences/sydney10/procs/Crisp-concise.pdf

Dalgarno, B. & Lee, M. J. W. (2010). What are the learning affordances of 3-d virtual environments? *British Journal of Educational Technology*, *41*(1), 10–32. doi:10.1111/j.1467-8535.2009.01038.x

De Freitas, S. & Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education*, *46*(3), 249–264. doi:http://dx.doi.org/10.1016/j.compedu.2005.11.007

Dickey, M. D. (2003). Teaching in 3d: pedagogical affordances and constraints of 3d virtual worlds for synchronous distance learning. *Distance education*, *24*(1), 105–121. Retrieved from http://mchel.com/Papers/Dickey_TeachingIN3D.pdf

Djaouti, D., Alvarez, J., Jessel, J.-P., & Rampnoux, O. (2011). Origins of serious games. In M. Ma, A. Oikonomou, & L. C. Jain (Eds.), *Serious games and edutainment applications* (pp. 25–43). Springer London. doi:10.1007/978-1-4471-2161-9_3

Dunwell, I., De Freitas, S., & Jarvis, S. (2011). Four-dimensional consideration of feedback in serious games. In S. De Freitas & P. Maharg (Eds.), *Digital games and learning* (pp. 42–62). Retrieved from http://sgi.cueltd.net/publications/papers/chapterv9[1].pdf

Dunwell, I., Petridis, P., Hendrix, M., Arnab, S., Al-Smadi, M., & Gütl, C. (2012). Guiding intuitive learning in serious games: an achievement-based approach to externalized feedback and assessment. In *Complex, intelligent and software intensive systems (cisis), 2012 sixth international conference on* (pp. 911–916). doi:10.1109/CISIS.2012.205

Egenfeldt-Nielsen, S. (2006). Overview of research on the educational use of video games. *Digital kompetanse*, *1*(3), 184–213. Retrieved from http://egenfeldt.eu/papers/game-overview.pdf

European Commission. (2008). The european qualifications framework for lifelong learning (eqf). *Luxembourg: Office for Official Publications of the European Communities*. Retrieved from http://ec.europa.eu/education/pub/pdf/general/eqf/broch_en.pdf

Gredler, M. E. (2004). Games and simulations and their relationships to learning. *Handbook of research on educational communications and technology*, *2*, 571–581. Retrieved from http://www.coulthard.com/library/Files/gredler_2004_gamesandsimsandrelationtolearning.pdf

Gütl, C. (2008). Moving towards a fully automatic knowledge assessment tool. *International Journal of Emerging Technologies in Learning (iJET)*, *3*(1). Retrieved from http://www.editlib.org/p/45126

Gütl, C. (2011). The support of virtual 3d worlds for enhancing collaboration in learning settings. In *Techniques for fostering collaboration in online learning communities: theoretical and practical perspectives* (pp. 278–299). doi:10.4018/978-1-61692-898-8.ch016

Howland, G. (1999). Game design: immersiveness. Retrieved August 8, 2013, from http://archive.gamedev.net/archive/reference/articles/article262.html

Ibáñez, M. B., Crespo, R. M., & Kloos, C. D. (2010). Assessment of knowledge and competencies in 3d virtual worlds: a proposal. In N. Reynolds & M. Turcsányi-

Szabó (Eds.), *Key competencies in the knowledge society* (pp. 165–176). Springer. doi:10.1007/978-3-642-15378-5_16

Immersive Education Initiative. (nodate). *About immersive education*. Retrieved May 2, 2013, from http://members.immersiveeducation.org/about

Kaplan, J. & Yankelovich, N. (2011). Open wonderland: an extensible virtual world architecture. *Internet Computing, IEEE*, *15*(5), 38–45. doi:10.1109/MIC.2011.76

Kickmeier-Rust, M. D. & Albert, D. (2010). Micro-adaptivity: protecting immersion in didactically adaptive digital educational games. *Journal of Computer Assisted Learning*, *26*(2), 95–105. doi:10.1111/j.1365-2729.2009.00332.x

Kickmeier-Rust, M. D., Steiner, C. M., & Albert, D. (2009). Non-invasive assessment and adaptive interventions in learning games. In *Intelligent networking and collaborative systems, 2009. incos'09. international conference on* (pp. 301–305). IEEE. doi:10.1109/INCOS.2009.30

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86. doi:10.1207/s15326985ep4102_1

Kompetenzmodell Naturwissenschaften 8. Schulstufe. (2011, October). Retrieved May 19, 2013, from https://www.bifie.at/system/files/dl/bist_nawi_kompetenzmodell-8_2011-10-21.pdf

Kowald, D. (2012). *Combining computer-supported, collaborative learning with e-assessment. enhancing a wiki system with flexible assessment methods* (Master's thesis, University of Technology Graz.).

Machet, T., Lowe, D., & Gütl, C. (2012). On the potential for using immersive virtual environments to support laboratory experiment contextualisation. *European Journal of Engineering Education*, *37*(6), 527–540. doi:10.1080/03043797.2012.721743

Maderer, J. & Gütl, C. (in press). Flexible automated assessment in 3d learning environments: technical improvements and expert feedback. In *Proceedings of the 3rd european immersive education summit*. London.

Maderer, J., Gütl, C., & AL-Smadi, M. (2013). Formative assessment in immersive environments: a semantic approach to automated evaluation of user behavior in open wonderland. In *Proc. of ied 2013 boston summit*. Available from http:

//jied.org/proceedings/iED_2013_PROCEEDINGS.zip. Journal of Immersive Education.

Martínez-Ortiz, I., Moreno-Ger, P., Sierra, J. L., & Fernández-Manjón, B. (2006). Production and deployment of educational videogames as assessable learning objects. In W. Nejdl & K. Tochtermann (Eds.), *Innovative approaches for learning and knowledge sharing* (Vol. 4227, pp. 316–330). Springer. doi:10.1007/11876663_25

Miller, C. H. (2008). *Digital storytelling: a creator's guide to interactive entertainment* (2nd Edition). Focal Press. Retrieved April 14, 2013, from http://proquestcombo.safaribooksonline.com/book/-/9780240809595

Mills, D. L. (1991). Internet time synchronization: the network time protocol. *Communications, IEEE Transactions on*, *39*(10), 1482–1493. doi:10.1109/26.103043

Mono. (nodate). *Mono. crossplatform, open source .net development framework.* Retrieved June 1, 2013, from http://mono-project.com

Moreno Ger, P., Sancho Thomas, P., Martínez Ortiz, I., Sierra, J. L., & Fernández Manjón, B. (2007). Adaptive units of learning and educational videogames. *Journal of Interactive Media in Education*, *2007*. Retrieved from http://www-jime.open.ac.uk/article/2007-5/320

Moreno-Ger, P., Blesius, C., Currier, P., Sierra, J., & Fernández-Manjón, B. (2008). Online learning and clinical procedures: rapid development and effective deployment of game-like interactive simulations. In Z. Pan, A. Cheok, W. Müller, & A. Rhalibi (Eds.), *Transactions on edutainment i* (Vol. 5080, pp. 288–304). Lecture Notes in Computer Science. Springer Berlin Heidelberg. doi:10.1007/978-3-540-69744-2_22

Moreno-Ger, P., Burgos, D., Sierra, J. L., & Manjón, B. F. (2007). A game-based adaptive unit of learning with ims learning design and< e-adventure>. In *Creating new learning experiences on a global scale* (pp. 247–261). Springer. doi:10.1007/978-3-540-75195-3_18

Nakamura, J. & Csikszentmihalyi, M. (2002). The concept of flow. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 89–105). Oxford University Press.

Nareyek, A. (2001). Review: intelligent agents for computer games. In T. Marsland & I. Frank (Eds.), *Computers and games* (pp. 414–422). Springer. doi:10.1007/3-540-45579-5_28

Network Time Protocol. (2013, May 6), In *Wikipedia, the free encyclopedia*. Retrieved May 10, 2013, from http://en.wikipedia.org/wiki/Network_Time_Protocol

Nicol, D. J. & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in higher education*, *31*(2), 199–218. doi:10.1080/03075070600572090

Open Wonderland Documentation Wiki. (nodate). Retrieved June 9, 2013, from http://code.google.com/p/openwonderland/wiki/OpenWonderland

Open Wonderland Foundation. (2012). Open wonderland website. Retrieved March 15, 2013, from http://openwonderland.org/

Open Wonderland: JavaDoc. (nodate). Retrieved March 15, 2013, from http://javadoc.openwonderland.org/

OpenSimulator Website. (2013, April). Retrieved April 14, 2013, from http://opensimulator.org/

Osborne, J. & Dillon, J. (2008). Science education in europe: critical reflections. London: The Nuffield Foundation. Retrieved from http://www.nuffieldfoundation.org/sites/default/files/Sci_Ed_in_Europe_Report_Final.pdf

OVOS. (2012). Ludwig [computer game]. Retrieved from http://www.playludwig.com

Pendulum (mathematics). (2013, April 11), In *Wikipedia, the free encyclopedia*. Retrieved April 23, 2013, from http://en.wikipedia.org/w/index.php?title=Pendulum_(mathematics)&oldid=549870074

Pfeiffer, A. (2012). *Endbericht "spielend lernen"[final report "playful learning"]*. Donau-universität Krems. Retrieved from http://www.playludwig.com/resources/files/2012/8/13/941/sparkling-science-endbericht.pdf

Pirker, J., Berger, S., Gütl, C., Belcher, J., & Bailey, P. (2012). Understanding physical concepts using an immersive virtual learning environment. In *Proceedings of the 2nd european immersive education summit* (pp. 183–191). Paris, France: Immersive Education. Retrieved from http://europe.immersiveeducation.org/sites/default/files/documents/2nd_European_Immersive_Education_proceedings_2012.pdf

Prensky, M. (2001). Simulations: are they games? Retrieved June 18, 2013, from http://www.marcprensky.com/writing/Prensky%20-%20Simulations-Are%20They%20Games.pdf

Pritchard, A. & Woollard, J. (2010). *Psychology for the classroom: the social context*. Routledge. Retrieved 20 5, 2013, from http://www.myilibrary.com?ID=258966

RedDwarf Contributors. (2010). Reddwarf application tutorial. Retrieved May 1, 2013, from http://sourceforge.net/apps/trac/reddwarf/attachment/wiki/Documentation/RedDwarf%20ServerAppTutorial.odt

RedDwarf Website. (2013, March 12). Retrieved May 16, 2013, from http://sourceforge.net/apps/trac/reddwarf/wiki/WikiStart?version=44

Ridgway, J., McCusker, S., & Pead, D. (2004). *Literature review of e-assessment*. DU. Report number 10. Bristol: Futurelab. Retrieved May 20, 2013, from http://dro.dur.ac.uk/1929/

Rogers, L. T. (2008). Information technology for understanding science (it for us). also published in GIREP 2008 Conference Proceedings. Retrieved from http://lsg.ucy.ac.cy/girep2008/papers/INFORMATION%20TECHNOLOGY%20FOR%20UNDERSTANDING.pdf

Rutten, N., van Joolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education*, *58*(1), 136–153. doi:10.1016/j.compedu.2011.07.017

Sambell, K., McDowell, L., & Montgomery, C. (2012). *Assessment for learning in higher education*. Routledge. Retrieved May 14, 2013, from http://www.myilibrary.com?ID=390234

Sandvik, A. W. (2012). *Numerical solutions of classical equations of motion*. Retrieved June 16, 2013, from http://physics.bu.edu/py502/lectures3/cmotion.pdf

Sauvé, L., Renaud, L., Kaufman, D., & Marquis, J.-S. (2007). Distinguishing between games and simulations: a systematic review. *Educational Technology & Society*, *10*(3), 247–256. Retrieved from http://www.ifets.info/journals/10_3/17.pdf

Savery, J. R. (2006). Overview of problem-based learning: definitions and distinctions. *Interdisciplinary Journal of Problem-based Learning*, *1*(1), 3. Retrieved from http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1002&context=ijpbl

Schmeil, A., Eppler, M. J., & de Freitas, S. (2012). A framework for the design of avatar-based collaboration. In R. Hinrichs & C. Wankel (Eds.), *Engaging the avatar. new frontiers in immersive education* (pp. 15–48). Information Age Publishing.

Schroeder, R. (2008). Defining virtual worlds and virtual environments. *Journal of Virtual Worlds Research*, *1*(1). Retrieved from http://journals.tdl.org/jvwr/index.php/jvwr/article/view/294

Sherman, W. R. & Craig, A. B. (2002). *Understanding virtual reality*. Morgan Kaufmann. Retrieved March 17, 2013, from http://my.safaribooksonline.com/9781558603530

Shermis, M. D. & DiVesta, F. J. (2011). *Classroom assessment in action*. Rowman & Littlefield Publishers. Retrieved 18 5, 2013, from http://www.myilibrary.com/?ID=304635

Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning. In *Serious games: mechanisms and effects* (pp. 295–321).

Sjøberg, S. & Schreiner, C. (2010). *The rose project: an overview and key findings*. University of Oslo. Retrieved from http://roseproject.no/network/countries/norway/eng/nor-Sjoberg-Schreiner-overview-2010.pdf

AL-Smadi, M. & Gütl, C. (2008). Present and future of e-assessment-towards a flexible e-assessment system. In *Proceedings of icl2008*. Villach, Austria. Retrieved from http://info.iicm.edu/about/Homepages/cguetl/publications/2008/Al-Smadie%20et%20al.%202008%20-%20%20CAF%20-%20ICL.pdf

AL-Smadi, M., Gütl, C., Dunwell, I., & Caballe, S. (2012). *D5.2.2: enriched learning experience v2*. Project Deliverable, ALICE (Adaptive Learning via Intuitive/Interactive Collaborative and Emotional systems).

AL-Smadi, M., Gütl, C., & Helic, D. (2009). Towards a standardized e-assessment system: motivations, challenges and first findings. *International Journal of Emerging Technologies in Learning (iJET)*, *4*(6). doi:10.3991/ijet.v4s2.919

AL-Smadi, M., Höfler, M., & Gütl, C. (2011). Enhancing wikis with visualization tools to support groups production function and to maintain task and social awareness. *Procs. of ICBL*.

AL-Smadi, M., Wesiak, G., & Gütl, C. (2012). Assessment in serious games: an enhanced approach for integrated assessment forms and feedback to support guided learning. In *Interactive collaborative learning (icl), 2012 15th international conference on* (pp. 1–6).

Suskie, L. (2010). *Assessing student learning: a common sense guide*. Jossey-Bass. Retrieved 18 5, 2013, from http://www.myilibrary.com/?ID=277439

Tipler, P. A. & Mosca, G. (2004). *Physik. für wissenschaftler und ingenieure.* (Zweite deutsche Auflage) (D. Pelte, Ed.). 2nd German edition. Elsevier, Spektrum Akademischer Verlag.

Tsai, F.-H., Yu, K.-C., & Hsiao, H.-S. (2012). Exploring the factors influencing learning effectiveness in digital game-based learning. *Educational Technology & Society*, *15*(3), 240–250. Retrieved from http://www.ifets.info/others/download_pdf.php?j_id=56&a_id=1270

Tutenel, T., Bidarra, R., Smelik, R. M., & Kraker, K. J. D. (2008, December). The role of semantics in games and simulations. *Computers in Entertainment*, *6*(4), 57:1–57:35. doi:10.1145/1461999.1462009

Twilleager, D. (nodate). *Mtgame programming guide.* Retrieved February 2013, from https://sites.google.com/site/openwonderland/archive/MTGameProgGuide.pdf?attredirects=0&d=1

van Staalduinen, J.-P. & de Freitas, S. (2011). A game-based learning framework: linking game design and learning outcomes. In M. S. Khine (Ed.), *Learning to play: exploring the future of education with video games.* New York: Peter Lang Publishers. Retrieved from http://sgi.cueltd.co.uk/papers/Chapter_Staalduinen_Freitas_-_Final.pdf

Vince, J. (2004). *Introduction to virtual reality.* Springer.

Weibel, D. & Wissmath, B. (2011). Immersion in computer games: the role of spatial presence and flow. *International Journal of Computer Games Technology*, *2011*. doi:10.1155/2011/282345

Wesiak, G., Al-Smadi, M., & Gütl, C. (2012). Towards an integrated assessment model for complex learning ressources: findings from an expert validation. In *Interactive collaborative learning (icl), 2012 15th international conference on* (pp. 1–7). doi:10.1109/ICL.2012.6402093

White, B. A. (2007). *Second life: a guide to your virtual world.* Que Publishing.

Winterton, J., Delamare-Le Deist, F., & Stringfellow, E. (2006). *Typology of knowledge, skills and competences: clarification of the concept and prototype.* Office for Official Publications of the European Communities Luxembourg. Retrieved from http://www.cedefop.europa.eu/en/Files/3048_EN.PDF

Wirth, W. et al., Hartmann, T. et al., Bocking, S., et al., Vorderer, P., et al., Klimmt, C., et al., Schramm, H., et al., … Et al., et al., et al., Gouveia, F. R., et al.

(2007). A process model of the formation of spatial presence experiences. *Media Psychology*, *9*(3), 493–525. doi:10.1080/15213260701283079

Wonderland Modules: Subversion Repository. (nodate). Retrieved August 17, 2012, from http://openwonderland-modules.googlecode.com/svn/trunk/. (svn checkout -r2010)

Wonderland: Subversion Repository. (nodate). Retrieved July 30, 2012, from http://openwonderland.googlecode.com/svn/trunk/. (svn checkout -r4865)

Woollard, J. (2010). *Psychology in the classroom: behaviourism*. Routledge. Retrieved May 14, 2013, from http://lib.myilibrary.com?ID=273328

Zierer, K. & Seel, N. (2012). General didactics and instructional design: eyes like twins a transatlantic dialogue about similarities and differences, about the past and the future of two sciences of learning and teaching. *SpringerPlus*, *1*(1), 15. doi:10.1186/2193-1801-1-15

Appendix: Data Medium

The attached data medium (DVD) contains the following resources:

1. *PDF version of this document*
2. *Used references* (see Bibliography; as far as available in digital form)
3. *Expert evaluation*:
    - Interview/feedback questionnaire
    - Results/raw data
4. *Developed software*:
    - Assessment module (binary; `assessment.jar`)
    - Pendulum module (binary; `pendulum.jar`)
    - Latest version of the SOFIA evaluation engine and web service (.NET binaries; including assessment models for the presented test cases)
    - Source codes
5. *Reference software/runtime*:
    - Used source codes of the OWL projects (Wonderland and Wonderland Modules repositories)
    - Java Development Kit (JDK) 1.7.0 (binary; Linux, AMD 64-bit)
    - Mono runtime and XSP webserver (source code; precompiled for Linux, AMD 64-bit)
    - Open Wonderland server (binary; `Wonderland.jar`)
6. *Tutorials and instructions to set up a test environment*
7. *Video demonstration (showcase selection)*