FLEXIBLE AND INTEGRATED E-ASSESSMENT IN COMPLEX LEARNING RESOURCES

A SERVICE-ORIENTED APPROACH FOR COMPUTER-BASED AND COMPUTER-ASSISTED KNOWLEDGE ASSESSMENT

Dissertation in fulfillment of the requirements for the academic degree
Doctor of Technical Sciences (Dr. techn.) in Computer Science
Submitted by
Mohammad AL-Smadi

Institute for Information Systems and Computer Media (IICM)
Doctoral School of Computer Science
Graz University of Technology
A-8010 Graz, Austria
May 2012

First reader: Univ.Doz. Dipl.-Ing. Dr. techn. Christian Gütl
Second reader: Prof. Vanessa Chang

FLEXIBLES UND INTEGRIERTES E-ASSESSMENT IN KOMPLEXEN LERNRESSOURCEN

EIN SERVICE-ORIENTIERTER ANSATZ FÜR COMPUTERBASIERTES UND COMPUTERGESTÜTZTES KNOWLEDGE ASSESSMENT

Dissertation zur Verleihung des akademischen Grades
Doktor der Technischen Wissenschaften (Dr. techn.) in Informatik
vorgelegt von
Mohammad AL-Smadi

Institut für Informationssysteme und Computer Medien (IICM)
Fakultät für Informatik
Technische Universität Graz
8010 Graz, Österreich
Mai 2012

Erstbegutachter: Univ.Doz. Dipl.-Ing. Dr. techn. Christian Gütl
Begutachter: Prof. Vanessa Chang

Abstract

In the so-called 'knowledge age' information and communication technology (ICT) has become the main infrastructure for the post-industrial society. Advances in ICT have caused our life to face changes - political, economical, social, and ethical - as well as rapid developments in educational technology. Likewise, learning has changed from being repetitive to more constructive and based on understanding and active learning through learners' empowerment and skills improvement. This shift in learning paradigm requires aligned assessment practices when it comes to meet the goals of education. Therefore, assessment is no more considered to discriminate between students, rather than it is used to enhance students learning and encourage further learning. Moreover, students play major roles assessment where alternative forms of assessment have been adapted.

Online learning has been provided over the last decades, through which not only content is delivered using technology, but also aspects that include saving instructional time and effort, students' empowerment and engagement, providing feedback, supporting new forms of learning and assessment have been achieved. However, online learning has been influenced by the revolution of emerging technology. Thus, several learning environments have been developed - including commercial and open source - to provide online learning. However, the lack of standards compliance, flexible software architecture, and pedagogy consideration in educational tools development has caused to have many educational - thus assessment - tools that are standalone and used for limited application domains. Therefore, in order to provide flexible and interoperable e-assessment - including content, tools, and services - that is aligned with learning and instruction and meets the educational goals, it is required to pay more attention to assessment frameworks and reference models, standards and specifications, flexible software architectures, and application contexts.

This doctoral dissertation aims to provide an enhanced approach for integrated and flexible e-assessment. In particular, to provide a solution approach in which service-oriented flexible and interoperable e-assessment (SOFIA) can be used to provide integrated forms of assessment with complex learning resources (CLR). Main contributions include an integrated model for e-assessment (IMA), a framework to design standard-conform e-assessment services and tools, and a service-oriented software approach to develop integrated and flexible standard-conform e-assessment forms integrated to CLRs. Thereby, education goals can be met through an alignment between instruction, learning, and assessment - which is achieved by IMA - during the design of the assessment forms. This alignment requires a clear guidance and understanding for e-assessment process and services which is achieved by the framework. The framework is used to develop flexible and standard-conform e-assessment on top of a service-oriented architecture that facilitates content and services accessibility and interoperability. The provided solution approach aims to support students learning and to provide them timely and quality feedback, as well as to support teachers to achieve didactic goals through the use of CLRs. Moreover, SOFIA was tested in several studies covering different application contexts - such as self-directed learning, collaborative writing and serious games. The first findings show that, despite the difficulty in developing integrated, flexible, and interoperable assessment with respect to standards conformation, SOFIA has proven to achieve the required flexibility - in terms of pedagogy and technology. Moreover, flexible and interoperable e-assessment forms integrated to CLRs support students leaning and progress, engage them and motivate deep learning. Nevertheless, the approach supports different teaching strategies and didactic goals achievement.

Kurzfassung

Informations- und Kommunikationstechnologie (ICT) entwickelte sich im sogenannten "Wissenszeitalter" zur Hauptinfrastruktur der post-industriellen Gesellschaft. Politische, wirtschaftliche, soziale und ethische Änderungen, aber auch eine rasche Entwicklung von pädagogischen Technologien sind die Folge. Auch der Lernprozess hat sich von einem sich wiederholenden zu einem konstruktiven Prozess weiterentwickelt und basiert vielmehr auf Verständnis, aktivem Lernen und Fähigkeitsentwicklung. Dieser Paradigmenwechsel benötigt ausgerichtete Assessment-Anwendungen, um den pädagogischen Zielen entgegenzukommen. Folglich soll die Beurteilung kein Unterscheidungskriterium zwischen Lernenden mehr sein, sondern zielt darauf ab deren Lernerfolge zu verbessern und sie zu weiteren Fortschritten und Erfolgen zu ermutigen. Lernende nehmen vermehrt die Hauptrolle in dieser neuen Beurteilungskultur ein, in der sich viele neue und alternative Formen entwickelt haben.

Innerhalb der letzten Jahrzehnte wurde Lernen online angeboten, wodurch nicht nur Inhalte übermittelt, sondern auch Lehrzeit und Aufwand gespart wurden. Weiter werden Lernende befähigt und motiviert Feedback zu geben. Auch werden dadurch neue Formen des Lernens und der Beurteilung, wie z.B. das aktive und soziale Lernen, oder auch Performance und Portfolio Assessment unterstützt. Online Lernen wurde durch neue Technologien stark beeinflusst. Kommerzielle, aber auch open-source Lernumgebungen wurden entwickelt, um technologiebasierte Bildung anzubieten. Durch das Fehlen von Standards, flexiblen Softwarearchitekturen und pädagogischen Ansätze wurden viele Bildungs- und Assessment-Tools in sich abgeschlossen und nur für einen limitierten Anwendungsbereich entwickelt. Um flexibles und interoperables e-Assessment, bestehend aus Inhalten, Tools und auch Services, welche den Fokus auf die Lernziele setzt, anzubieten, ist es notwendig das Augenmerk auf Assessment Frameworks und Referenzmodelle, Standards und Spezifikation, flexible Softwarearchitekturen und Anwendungsbereiche zu richten.

Diese Dissertation zielt darauf ab einen erweiterten Ansatz für integriertes und flexibles e-Assessment vorzustellen. Im Besonderen soll ein Lösungsansatz entwickelt werden, welcher Service-orientiertes, flexibles und interoperables e-Assessment (SOFIA) verwendet, um eine integrierte Form der Beurteilung mit komplexen Lernressourcen (CLR) zu bilden. Hierbei umfassen die wichtigsten Beiträge ein integriertes Modell für e-Assessment (IMA), ein Framework für das Design von standardkonformen E-Assessment Services und Tools, und einen Service-orientierten Softwareansatz, um integrierte und flexible standardkonforme e-Assessment Formen mit eingegliederten CLRs, zu entwickeln. Während des Designs der Assessment Formen können Bildungsziele durch einen Abgleich von Unterricht, Lernen und Beurteilung erreicht werden. Dies wird mittels IMA erreicht. Der Abgleich erfordert klare Richtlinien und Verständnis für E-Assessment Verfahren und Services, was mit Hilfe des Frameworks erreicht wird. Dieses wird für die Entwicklung flexibler und standardkonformer E-Assessment Methoden auf einer serviceorientierten Architektur verwendet, was die Zugänglichkeit und Kompatibilität der Inhalte und Services vereinfachen soll. Der zur Verfügung gestellte Ansatz fokussiert auf die Unterstützung der Lernprozesse von Lernenden und bietet zeitliches und qualitatives Feedback. Außerdem werden Lehrer dabei unterstützt didaktische Ziele mit Hilfe von CLRs zu erreichen. SOFIA wurde in zahlreichen Studien mit verschiedenen Anwendungsbeispielen getestet. Dazu gehören selbstgesteuertes Lernen, kollaboratives Schreiben und auch Spiele-basiertes Lernen. Die ersten Ergebnisse zeigen, dass trotz der Schwierigkeiten bei der Entwicklung integrierter, flexibler, interoperabler und standardkonforme Assessmentsysteme, diese in Kombination mit CLRs den Lernerfolg und Fortschritt unterstützen.

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.

…………………………… date

……………………………………………….. (signature)

Acknowledgment

I am indebted to my colleagues, friends, and family members for their inspiration and support during the research and writing of my doctoral dissertation. This thesis would not been possible without YOU, Thank you!

First of all, I would like to express my gratitude to my supervisor Christian Gütl as well as to the second reader of this work, Prof. Vanessa Chang for support and constructive feedback. Christian, along the years of this research I am really thankful for working with you, your vast knowledge and expertise, understanding, support, and patience have made my ambition and dreams come true. In addition, I am thankful to the trust you gave to me to manage projects, supervise master students, and lead the development team in the research group for Advanced Educational Media Technologies (AEMT).

Along the way, everyone at the IICM has provided me with incredibly valuable help, support and constructive feedback, Thank you all. My special thanks go to the members and ex-members of our research group AEMT. Further, it must be emphasized that the results presented in this thesis were only possible to achieve with the help and collaboration of my research colleagues and students. I am also grateful to ALICE project members for insightful collaboration and research which enriched the outcome of this thesis.

I am deeply and forever indebted to my parents for their love, support and encouragement throughout my entire life. I am also very grateful to my brothers and Sisters for endless love, support and encouragement. I cannot thank you enough for all the support and love you have given me. I never would have made it here without you. Thank you for everything!

Finally, I am also grateful to my uncle and his family in Austria for continuous support and encouragement. My Uncle, above all words fail to express my gratitude to you. However, I hope my success expresses my gratitude to you and I owe my success to you.

Table of Contents

Chapter

1

1. Introduction

1.1 Motivation
1.2 Goals and Research Questions
1.3 Methodology and Structure

This chapter provides an overview on the main aspects of the research presented in this dissertation. Firstly, the scopes for the research work are introduced, whereby the targets of investigation and application are outlined. Secondly, the main research goals with a summary of the achieved results are presented. Finally, this chapter concludes by describing the deployed research methodology and the structure of this dissertation.

1.1. Motivation

Members of our society are affected by rapid changes in every part of our modern life. Terms such as 'post-industrial society', 'information society' and 'knowledge society' have been used to identify and understand the extent of these changes. Knowledge has become a primary resource for production instead of capital and labor. As a result the knowledge society creates, shares, and uses knowledge to improve and to have a well-being of its people. Another term of "global society" with a shared knowledge is one of the aims of globalization and using new Information and Communication Technologies (ICT). Therefore, "global learning" is needed as a primary mean of delivering this shared knowledge to the society people. The shift in education systems - including learning and assessment - to build on emerging technologies and ICT in particular has fostered formal and informal learning, thus demands more flexible learning and assessment. (Gütl & Chang, 2008)

In order to define the research scope of this doctoral dissertation and to specify the main components of target research, a conceptual view of the educational system is depicted in Figure 1.1. As depicted in the figure the *education* and *assessment* - as an integrated component - represent the instructional and learning aspects aligned with assessment forms to meet education goals, moreover, the *technology* part through which e-education - content services - can be supported and provided to *stakeholders* in different educational *contexts*. However, in order to meet education goals, *pedagogy* - including instruction and learning objectives, theories of learning, teaching strategies - should be considered. The colored components in this Figure namely education, assessment, context, and technology represents the general scope of this research. However, the other components are important to explore the domain of e-education and to provide a complete depiction for the educational process. Assessment as depicted in the figure lies in the heart of the educational process. Assessment is no more considered as separated component rather than it is integrated within the learning process (Biggs, 1999). Moreover, assessment is designed to consider pedagogical approaches supporting theories of learning – e.g. active learning, blended learning, experiential leaning, exploratory learning, and collaborative learning – which are provided through learning environments – i.e. learner-centered, knowledge-centered, assessment-centered, and community-centered - that considers learning context and available technology.

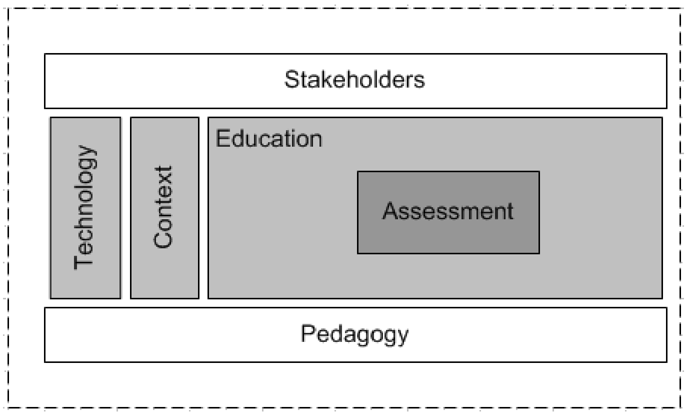


FIGURE 1.1. Conceptual view for e-Education

Having a closer look to Figure 1.1, despite the abstraction of representing the components, combining them provides influential aspects to the general scopes of this research as follows.

Technology-based Education: the influence of ICT on educational systems

The emergence use of ICT in education lacks for some extent the consideration of learning theories and pedagogy (Watson, 2001). Thus, a variety of e-learning tools have been developed for different educational contexts which makes it complicated to decide which tool is most appropriate for a learning practice (Ravenscroft, 2001). Therefore, designers of educational models – including learning and assessment –should consider theories of learning - cognitive, affective, and social learning - during the design of their models, moreover they should provide a framework or guidelines of how to use their models in designing and developing educational tools (AL-Smadi, Hoefler, & Guetl, 2011a).

Information age and Stakeholders: the need to adapt and enhance educational systems to suite the Net generation

In the age of the so-called "information age", learners grow up with technology dominating most of their life activities. They use technology anywhere, anytime, and they are faced with the challenge of needing to be engaged and motivated in their learning (Prensky, 2001). The emergence of Web 2.0 and the influence of ICT have fostered e-learning – e-learning 2.0 - to be more interactive, challenging, and situated. As a result, learners feel empowered when they are engaged in collaborative learning activities and self-directed learning. The learners are also provided with e-learning systems that would maintain their social identity and situated learning experience. As a result, modern learning settings of learner-centred practices have become more dominant.

Technology and Assessment:
the need for flexible e-assessment, models, forms, tools, and systems in terms of flexible pedagogy, flexible technology, variety of learning contexts, and target stakeholders

The shift in education paradigm towards more technology (*cf.* Bartley, 2006) requires an alignment in provided assessment forms with instruction and learning (*cf.* Biggs, 1999) thus to meet education goals (*cf.* Birenbaum, 2003). Consequently, a new culture of assessment of integrating measurement with instruction (Shepard, 2000) to address requirements of assessing skills such as cognitive (e.g. problem solving, critical thinking), meta-cognitive (e.g. self-reflection and self-evaluation), social (e.g. leading discussions and working within groups), and affective aspects (e.g. motivation and self-efficiency) have arisen (Dochy & McDowell, 1997). In this new culture of assessment, students play major roles in the assessment where new forms of assessment have been adapted to suit the learning styles of the modern learners. Such forms include interviews, performance assessment, exhibitions, portfolio assessment, process and product assessment, directed assessment, authentic assessment, performance assessment, alternative assessment, collaborative assessment and self- and peer-assessment. Given the different learning styles of students, educators are faced with the challenge of having to develop assessments which are required to appraise the students' learning process (AL-Smadi, Guetl, & Chang, 2011).

To this end, assessment forms provided in e-learning activities have to be aligned with the type of learning - such as reflective-learning, experiential-learning, and socio-cognitive learning (Dochy & McDowell, 1997; Elliott, 2008), moreover should reflect learners' preferences and consider their knowledge and skill state. Nevertheless, assessment should be integrated with the learning process and designed based on the learning outcomes and the didactic objectives.

e-Assessment is influenced by several aspects which caused to have a great variety in e-assessment tools and systems. Such aspects are, (a) assessment domain - cognitive, affective, and psychomotor (*cf.* Bloom, 1956), (b) assessment referencing – norm-related, criterion-related, or ipsative (*cf.* McAlpine, 2002), (c) assessment strategy – individual assessment, group assessment, self-assessment, peer-assessment, instructor-based assessment, and system-based assessment (*cf.* Dochy & McDowell, 1997), (d) assessment type – diagnostic, formative, and summative assessment (*cf.* AL-Smadi & Gütl, 2008; Crisp, 2007; Bransford, Brown, & Cocking, 2004), (e) assessment practice – behavioral assessment, performance assessment, portfolio assessment, and rubric-based assessment (*cf.* Buzzetto-More & Alade, 2006), (f) assessment adaptation – micro-adaptive assessment, or macro-adaptive assessment (*cf.* Kickmeier-Rust, Steiner, & Albert, 2009), (g) assessment method - quantitative, or qualitative (*cf.* AL-Smadi & Gütl, 2008; Crisp, 2007; Culwin, 1998; Bloom, 1956), (h) assessment's feedback – feedback type, format, frequency, and content (*cf.* Nicol, Milligan, 2006; Wiggins, 2001; Black & Wiliam, 1998; Charman, & Elms, 1998).

Nevertheless, e-assessment is influenced by standards and specifications used to represent assessment content, services, and its practice. Recently, conforming assessment to standards has gained more interest (*cf.* AL-Smadi, Guetl, & Helic, 2009b). However, several assessment tools and services have been developed to provide specific forms of assessment in application contexts which are not considered in e-assessment standards and services. For instance, alternative assessment forms - i.e. behavioral, performance, portfolio assessment, or rubric-based assessment - are not covered in the widely used e-assessment specification - i.e. Instructional Management System Question and Test Interoperability (IMS QTI, 2008). Therefore, technical specifications for tools interoperability have been used as an alternative approach to extend the e-assessment platform with third-party tools providing assessment forms for those application contexts. However, having specifications for flexible and interoperable e-assessment tools and systems is still in its early stages and requires more research and testing.

Further to this discussion, educators are faced with the challenge to design and develop flexible - applicable in different learning settings - and standard-conform e-assessment forms that are integrated to complex learning resources which have been designed to meet specific didactic and learning goals. This doctoral dissertation research aims to address this challenge and related problems. The next sub-chapters explore the research questions and goals and the research methodology.

1.2. Goals and Research Questions

The research conducted for this doctoral dissertation investigates issues of flexible and integrated educational assessment, and designing and developing flexible and integrated e-assessment in complex learning resources. Thereby, in accordance to the aspects discussed so far in this chapter, special focus is set on the following main research goals:

- **To what extent e-assessment practices fulfill multi-purpose e-Education.** A rich and comprehensive literature survey has been conducted using terminological and fictional e-assessment aspects, through which the theoretical and technological background of educational assessment provided online has been investigated. Based on that, main research motivations, challenges, and problems have been identified.

- **How to achieve high level of e-assessment flexibility in terms of: technology - secure, reusable, and accessible assessment content, tools and services - and pedagogy - assessment that is aligned with instruction and learning.** Findings from the comprehensive survey have been used to propose solution approach in which flexibility - in terms of pedagogy and technology - has been considered. Thus, the proposed solution has been used in several application scopes as part of distinct research projects.

- **To what extent flexible and integrated e-assessment with complex learning resources supports students' learning.** The developed solution approach within the context of this study has been used to conduct several studies in different learning settings - i.e. self-directed learning, collaborative learning, and game-based learning - and aspects such as tools usability, students' motivation, their emotional states, and their knowledge acquisition level have been investigated.

In addition to these general research goals other specific ones have been investigated and reported in the dissertation chapters.

1.3. Methodology and Structure

In order to meet the research goals discussed earlier, an applied, evaluative and quantitative research analysis has been conducted to investigate aspects related to educational assessment assisted by the use of technology. More precisely, the research represents an evaluative study for a significant contribution in the context of flexible use of technology - mainly computers - to support integrated and interoperable computer-assisted assessment in complex learning resources - such as collaborative writing assessment, automatic assessment for self-directed learning, and flexible assessment in serious games.

For the sake of this research a three phases' methodology has been used. As depicted in Figure 1.2 the study methodology starts with a survey phase - conceptual framework and technological investigations - to explore the literature in terms of e-education, e-assessment, and educational standards, specifications and guidelines. In the second phase, the solution approach - framework and proof of concept - have been designed and developed. The last phase represents the evaluation phase, in which complex learning resources namely enhanced approach for peer-assessment, and automated and integrated assessment in self-directed and collaborative learning have been used in real learning settings. For the purpose of this study, a literature survey was undertaken during the first phase (see Figure 1.2) through searching related work and technological investigations in international online bibliographic databases.

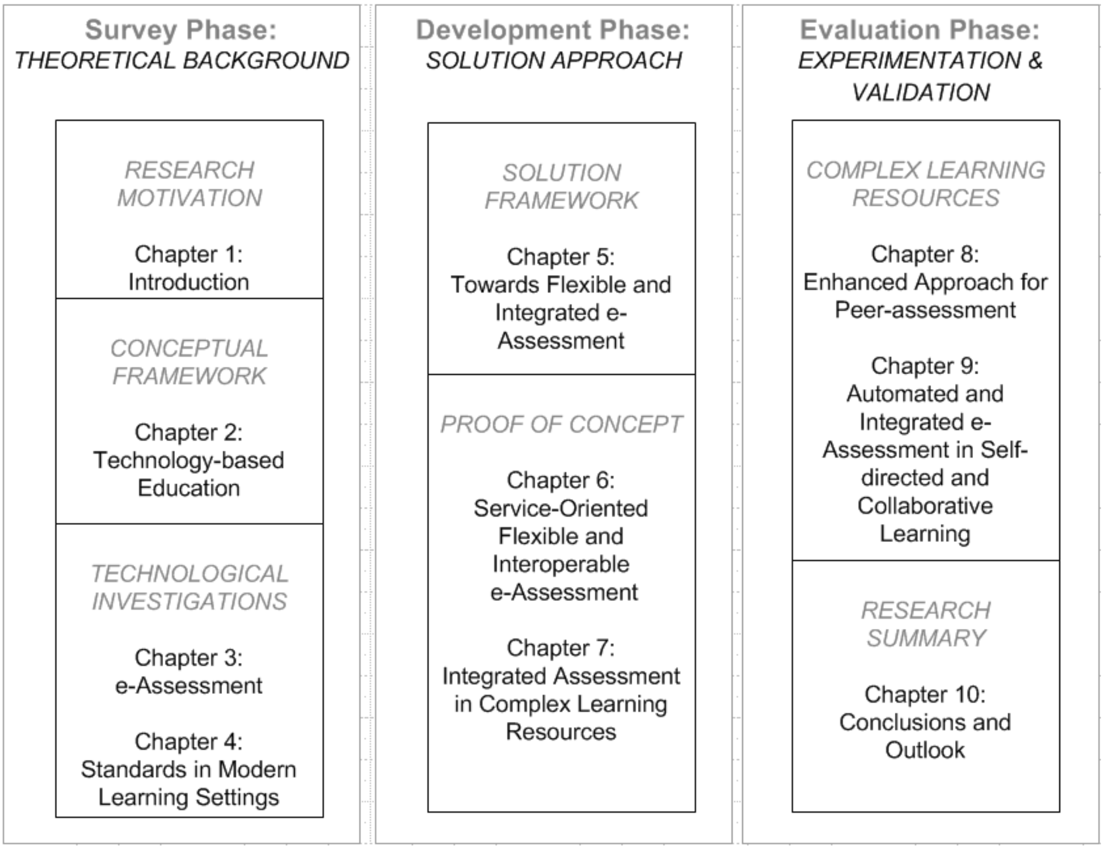


FIGURE 1.2. Study Phases and Structure

Findings from the survey phase - the context of e-education (see Chapter 2), e-assessment (see Chapter 3), and standards and specifications (see Chapter 4) - were used to define the solution approach as depicted in the second part of this dissertation - namely Chapters 5, 6, and 7. In the solution framework an integrated model for e-assessment (IMA) (see Chapter 5) has been proposed. IMA considers aspects from psychology and education as well as affective and emotional domain to design integrated standard-conform assessment forms which can be developed and embedded to complex learning resources - e.g. self-directed learning courses, and game-based learning activities. Nevertheless, the assessment process as a crucial component of education has been used to design a service-oriented framework for assessment (SOFA) (see Section 5.3) in order to support instructional designers as well as learning tools developers with a common understanding of the assessment cycle represented by assessment services - e.g. author, deliver, mark, grade, moderate, etc. – and common services - e.g. user management, group-management, authentication, authorization, etc. Moreover, to discuss aspects of standard-conform e-assessment system and how its components - assessment content and services - can be represented using standards and specifications.

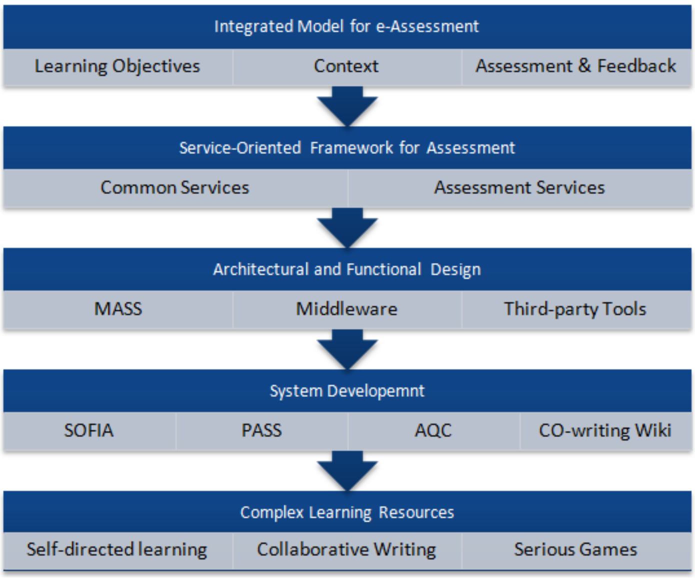


FIGURE 1.3. Development Phase (Chapters 5, 6, & 7)

Findings from the technological investigation as well as the identified software requirements have been used to design the software architecture (see Chapter 6). A service-oriented approach has been followed to design the solution architecture and to develop the proof-of-concepts. As flexibility - in terms of pedagogy and technology – is a main software requirement SOFA and the service-oriented architecture have been used within the context of IMA to develop a service-oriented flexible and interoperable e-assessment system (SOFIA) (see Section 6.4). SOFIA has been extended with third-party tools which develops contextualized assessment forms designed based on IMA to provide integrated and flexible assessment in complex learning resources - i.e. self-directed learning, collaborative writing, and serious games - as depicted in Figure 1.3. The complex learning resources integrated with flexible e-assessment (see Chapter 7) were used to conduct research studies in real learning settings as discussed in the third part of this dissertation (see Chapters 8 and 9). Moreover, Chapter 10 provides a summary of this doctoral research, reflects on research questions through research results, and discusses lessons learned and further work.

Findings form third phase - i.e. experimentation and validation - show that the solution approach has proven to achieve the required flexibility - in terms of pedagogy and technology - and SOFIA and its CLRs were used to conduct studies in different learning settings - i.e. self-directed learning and collaborative learning - and aspects such as tools usability, students' motivation, their emotional states, and their knowledge acquisition level were investigated. First findings show that SOFIA with the CLRs supported students for better and deep learning, and empowered students with learner-centred assessment forms such as self, peer-assessment, and rubric assessment. More detailed information about the studies is provided in the third part of this doctoral dissertation (Chapters 8 and 9).

Part

A

Theoretical Background

Chapter

2

2. Technology-based Education

e-Education - as a synonym for technology-based learning and teaching - is an emerging field of research in which technologies are used to support and provide learning and teaching. This chapter aims to give an overview about education - including instruction and learning - with an emphasis on assessment. Moreover discusses the influence of technology on education. To this end, this chapter presents a broad overview of theoretical foundations of learning in general, constructive, cognitive and social theories of learning in particular. Learning cannot be covered solely apart of assessment and technology. Learning (and accordingly assessing) does never occur separately in a cognitive, social, emotional or motivational way. Nevertheless, previous research often focused on cognitive aspects of learning and often lacked motivational aspects. Therefore, Motivation as an essential factor for education is discussed briefly in this chapter.

Moreover, this chapter sheds the light on assessment in education and focuses on the alignment between instruction, learning, and assessment (Biggs, 1996). Nevertheless, the necessity to consider learning styles and learners' preferences in learning design, content provision, and assessment design is illustrated. The chapter closes with highlighting the shift in education paradigm to e-education and the influence of information and communication technology (ICT) on education in general and assessment in particular. This chapter is based on (AL-Smadi et al., 2011; AL-Smadi, Guetl, & Chang, 2011).

2.1. Education

Educational goals and learning environments changed in the last decades. Students now are no longer seen as passive recipients of knowledge but are rather actively involved in creating their own learning environments. Furthermore, due to increasing requirements in the working environment, lifelong learning is essential for individuals to be competitive in the working place. This factor also increases the individual responsibility to acquire new knowledge and skills. However, learning should not only been seen as a behavior due to which information and facts are memorized. Rather, students must not only have a deep foundation of factual knowledge but they have also to understand facts in the context of a conceptual framework; they have to organize this knowledge to facilitate its retrieval and application (Bransford, Brown, & Cocking, 2004). But how can learning be described in general? As described later on in this chapter, there are several theories on how learning occurs. For instance, whereas behaviorist theories suggest that a learner begins as a clean state and that learning can be defined as a change in behavior, cognitive theories stated that learners are rational beings and that the actions of the learners are the consequence of their thinking. This is in also in accordance with constructivist theories that suggest that learning is a rather active and constructive process due to which new information is linked to already existing knowledge. Before describing this traditional learning theories more detailed, however, we will first introduce four models or frameworks, respectively, to give an idea of about how learning can be defined and described.

In general, Kolb (1984; see also Kolb & Kolb, 2005) posited six assumptions about learning:

1. Learning can be described as a process, not in terms of outcomes. The focus should be on engaging students in a process that includes feedback to enhance their learning.
2. All learning is relearning. Learning is facilitated when students' beliefs and ideas are included so that they can be tested and integrated with new ideas.
3. Conflicts, differences, and disagreement enhance the learning process because opposing modes of reflection, action, thinking and feeling are necessary.
4. Learning is a holistic process of adaption to the world; it involves not only cognition but also thinking, feeling, perceiving and behavior.
5. Learning results from synergetic transactions between an individual and his or her environment.
6. Learning is the process of creating knowledge.

From these statements it can be clearly seen that learning does not take place in one dimension but that it is rather multidimensional. Therefore, it has also to be considered that learning depends on the environment in which it occurs. For instance, Bransford *et al.* (2004) presented four perspectives of learning environments, which need to be seen as a system for interconnected components: learner-centered environments, knowledge-centered environments, assessment-centered environments, and community-centered environments.

In particular, learner-centered environments pay attention to the knowledge, skills, attitudes, and beliefs of the learners. The importance of building on the conceptual and cultural knowledge of the students is emphasized. Knowledge-centered environments support learning that leads to understanding and subsequent transfer. They focus on information and activities that help students developing an understanding of disciplines and include an emphasis of sense-making. Assessment-centered environments should provide opportunities for feedback and revision as feedback is an important factor in context of learning. Community-centered environments include not only the degree to which students, teachers, and administers feel connected to a larger community (e.g. homes, states, the nation etc.) but also aspects of classroom and school as a community. In the latter case, learning is, for instance, enhanced by social norms due to which students have the opportunity to make mistakes in order to learn. With respect to connections to larger communities, the family as one of the most important learning environments has to be considered (e.g., children learn from their family members in various ways).

Another approach to describe learning is the Eight Learning Events Model (8LEM; LeClerc & Poumay, 2005). It emphasizes that learning occurs in eight basic events. The purpose of the model is to analyze and enhance existing training strategies or teaching sequences, respectively, on the one hand, and to provide a framework for the planning of new training strategies or teaching sequences. Accordingly, the eight events in which learning occurs are:

1. *Imitation/Modeling:* learning from observation, impregnation, and imitation.

2. *Reception/Transmission:* learning from intentional communication.
3. *Exercising/Guidance:* learning by practicing-
4. *Exploration/Documenting:* learning by exploration.
5. *Experimentation/Reactivity:* learning by manipulating the environment.
6. *Creation/Confortation:* learning by creating something new, by producing concrete works.
7. *Self-reflection/Co-reflection:* learning by judgments, analysis, and operations.
8. *Debate/Animation:* learning during social interactions.

A further framework that classifies statements of educational goals and objectives to be achieved during learning is the Taxonomy of Bloom (1956). Bloom's taxonomy supports the definition and planning of learning objectives and their assessment. In his original taxonomy, Bloom presented taxonomies for three domains: the cognitive domain which includes skills and knowledge, the affective domain which includes emotional aspects and attitudes, and a psychomotor domain which includes manual and physical skills. As the cognitive domain is the most relevant for our purposes here, we will describe it more detailed. The cognitive domain (as revised by Krathwohl, 2002) consists of six levels:

1. *Remember*: refers to behavior emphasizing recognition and recalling.
2. *Understand*: refers to behavior emphasizing interpretation and classification.
3. *Apply*: refers to behavior emphasizing executing and implementing.
4. *Analyze*: refers to behavior emphasizing differentiating, organizing, attributing.
5. *Evaluate*: refers to behavior emphasizing checking or critiquing.
6. *Create*: refers to behavior emphasizing generating, planning, producing.

TABLE 2.1. Bloom's Taxonomy, cognitive domain (revised by Krathwohl, 2002).

Level	**Group**	**Verbs**
Remember	Recognizing, Recalling	Recognize, recall, name, describe, list etc.
Understand	Interpreting, Exemplifying, Classifying, Summarizing, Inferring,	Interpret, exemplify, classify, summarize, infer, explain, identify, generalize etc.
Apply	Executing, Implementing	Select, schedule, apply, use, utilize, practice, demonstrate etc.
Analyze	Differentiating, Organizing , Attributing	Analyze, elicit, examine, experiment, test etc.
Evaluate	Checking, Critiquing	Assess, evaluate, estimate, score, check, critique etc.
Create	Generating, Planning, Producing	Create, collect, plan, formulate, compose, check, generate, produce etc.

The levels are arranged in a hierarchical order with increasing difficulty (see also Table 1.1). For instance, the simplest behavior during learning is "remember" whereas the most complex behavior is "creating". Action verbs are assigned to each of the six levels. These action verbs describe abilities of the respective level more detailed. As mentioned before, the taxonomy supports the definition of learning outcomes (e.g. a student should be able to *apply* a formula rather to simply *remember* it) but can also be used when knowledge assessment is planned.

2.2. Learning Theories

This sub-chapter focuses on theoretical foundations of learning in cognitive and social theories of learning as well as motivational and emotional aspects. Learning cannot be covered solely through one dimension, learning - accordingly assessment - does not occur separately in a cognitive, social, emotional or motivational way. However, previous research often focused on cognitive aspects of learning apart from emotional or motivational aspects.

In order to provide quality assessment - as an essential part of learning - it is first important to understand how learning occurs. Many theories have been discussed in literature of how people learn. However the following sections describe a selected set of learning theories and help to understand how people learn.

Behaviorist Theories of Learning

Behaviorist theories of learning dominated the perspective of how people learn in the 1960s and 1970s. The behaviorist ideas were adopted by the research of Pavlov, Thorndike, Watson, and Skinner. According to Skinner (1974, p.2) "*learning is a change in observable behaviour caused by external stimuli in environment*". Learning environment plays a major role in learning determination. Hence, learning occurs based on interaction with the learning environment accordance to external stimuli. Rewards - such as praise - and punishments are seen as useful tools for supporting learning in terms of habits formation (James, 2006). Nevertheless, intelligence and mind are not necessarily to learn. Rather than, mastering skills and information memorization based on learning domain are dominant factors to build habits and to learn faster. "*the mind as a black box, in the sense that a response to a stimulus can be observed quantitatively, totally ignoring the effect of thought processes occurring in mind*" (Atkins, 1993).

Cognitive Theories of Learning

Apart from behaviorist explanation of learning, cognitive theories of learning requires active engagement of the Lerner and highlights the role of mid and brain in learning determination (James, 2006). Learning is explained - with respect to cognitive theories - as "*an internal process that involves memory, thinking, reflection, abstraction, motivation, and meta-cognition*" (Ally, 2004). In contrary with behaviorists, constructivists believe that learning is an active process of knowledge construction with major influence of 'understanding' and 'prior knowledge'. Furthermore, knowledge cannot be received from teachers or external environment, rather than knowledge is constructed by receiving information through different senses and mapped into memory based on cognitive process. Moreover, individuals' differences among learners are important and hence different learning styles (Kolb, 1984) determine how learners perceive and interact with learning material.

The variety of research in this domain has come up with several theories that build on the cognitive understanding of learning. Examples of these theories are explained as follows:

Cognitive Load Theory

This theory assumes that humans have two types of memory related to the cognitive system, a large and permanent store which is related to long-term memory (LTM) and time-limited as well as capacity - limited store related to working memory (WM). Cognitive load theory assumes that available knowledge structures in LTM are essential to the cognitive process and play major role in preventing WM overload - in which learner's intended cognitive processing exceeds the learner's available cognitive capacity (Mayer & Moreno, 2003). Moreover, learning is successful when its activities consider the learners WM limitation and provides learning material sequentially. (Kalyuga, 2009)

Cognitive Theory of Multimedia Learning

Mayer and Moreno (2003) discuss the cognitive theory in multimedia learning activities with respect to the following assumptions: limited-capacity of working memory based on dual-channel assumption of verbal and non-verbal material cognition, and learning happens when humans build connections between the verbal and non-verbal related materials. Thus, learning according to this theory includes three phases: *selection*, *organization*, and *integration*. Accordingly, learner first selects relevant information - of verbal and pictorial - from both channels, organizes it in each limited store into coherent representation, and integrates both representations - i.e. verbal model and pictorial model - with prior knowledge from long-term memory.

Experiential Learning Theory (ELT)

This learning theory has been proposed by Kolb (1984) and defines learning as "*the process whereby knowledge is created through the transformation of experience. Knowledge results from the combination of grasping and transforming experience*" (Kolb & Kolb, 2005; Kolb 1984, p.41). ELT assumes that the process of constructing knowledge is an ongoing process in which: 'concrete experience' activates 'observation and reflections'. The 'observation and reflections' are formatted into 'abstract concepts and generalization'. The 'implications' of these 'concepts and generalization' are actively tested in 'new situations' and lead to create 'new experience'.

Constructive, Socio-cultural, and Situated Theories of Learning

In contrary with behaviorist theories, constructivist paradigm considers learning as an active process by which knowledge and understanding are constructed based on pre-existing knowledge (Bransford et al, 2004 p.10; Vygotsky, 1978). Socio-cultural and situated theories of learning are based on this constructivist paradigm. Thus, examples can be found in situated learning, experiential learning, and self-directed learning.

Learning with respect to socio-cultural theories occurs based on an interaction between learners and the social environment. Vygotsky (1978) developed an approach based on socio-cultural interaction for cognitive development. According to Vygotsky's approach learning is a collaborative activity in a socialized environment where individuals develop their thinking together. Hence, group work is not optional or extra to the learning process (James, 2006). Examples of theories based on this understanding are discussed in following sub-sections.

Social Learning Theory (Bandura)

Learning according to Bandura's theory is derived from the behaviorist theories of learning complemented with cognitivist understanding of learning. "*Learning would be exceedingly laborious, not to mention hazardous, if people had to rely solely on the effects of their own actions to inform them what to do. Fortunately, most human behaviour is learned observationally through modelling: from observing others one forms an idea of how new behaviours are performed, and on later occasions this coded information serves as a* *guide for action.*" (Bandura, 1977; cited after Monfardini et al, 2008). The theory describes behavior in terms of interaction between behavioral observation, cognitive modeling, and environmental interaction. Moreover, the theory discusses the following categories as factors influences observation modeling:

- *Attention:* this category deals with aspects that affect the amount of attention paid by learner. For instance, distinctiveness, affective valence, prevalence, complexity, functional value, and individual's characteristics - e.g. sensory capacities, arousal level, perceptual set, and past reinforcement- affect attention.

- *Retention:* is related to the ability to remember what you have paid attention to. Aspects such as, symbolic coding, mental images, cognitive organization, symbolic rehearsal, motor rehearsal are considered in this category.

- *Reproduction:* relates to reproducing what you have remembered based on attention you paid to. This includes physical capabilities, and self-observation of reproduction.

- *Motivation:* considers motives such as past - i.e. traditional behaviorism, promised - i.e. imagined incentives- and vicarious -seeing and recalling the reinforced model.

Social Development Theory (Vygotsky)

Vygotsky (1978) believes that individual's social interaction plays a major role in cognitive development. Moreover, he argues that social learning occurs before development by which individuals collaborate with others or works under guidance of teacher or older adult - he named them "More Knowledgeable Other (MKO)" - to share experience and solve problems. Furthermore, Vygotsky argues that learning occurs in a zone - the notion of the "Zone of Proximal Development (ZPD)" - which represents the distance between student's ability to solve problems under adult guidance and/or peers collaboration and the ability to solve the same problem independently.

Situated Learning Theory (Lave)

Situated Learning - which refers to learning by solving problems based on authentic contextualized activity of learning often solved in collaborative group work - has its roots of research in the work of (Collins, Brown, and Neuman, 1989; Lave, 1990, 1993; Lave and Wenger, 1991; Rogoff, 1990, 1995) as cited in (Billett, 1996). According to Lave and Wenger (1991) learning is situated, unintentionally occurs, embedded within an authentic activity, for a specific context, and influenced by culture. Learners receive authentic and contextualized learning activity and work collaboratively in groups - Lave refers to as "Community of Practice"- affected by socio-cultural factors. According to Billett (1996) situated learning forms as a bridge between socio-cultural and cognitive theories, considering the complementary relation between the cognitive and socio-cultural paradigms.

2.3. Emotion and Learning

Despite the neglecting of emotions and their influence on cognition and learning, emotions are essential for learning and memory. Hascher (2010) stated that there is *"rarely any learning process without emotion"* (p.13). However, there is no common definition for the term "emotion" (Cabanac, 2002). For instance, Kleinginna and Kleinginna (1981; cited after Cabanac, 2002) listed more than 90 definitions of emotion. In particular, there is also no general distinction between "emotion", "affect" and "feeling" which are often used interchangeably (e.g., Städler, 2003). Hascher (2010) described three widely accepted characteristics of emotions: (1) an emotion is an affective reaction; it can be determined and described relatively precisely and can be attributed to a cause; (2) the experience of an emotion is related to for situations that are important for a person, and (3) when an emotion is experienced, it becomes the center of the individual's awareness.

However, affective aspects in the students' interactions, has gained more interest in recent years. This is due to a clear evidence of correlation between affect and learning. The relationship between emotion and learning is difficult to explore. O' Regan (2003) has reviewed two separate perspectives:

- Emotion is relevant to learning in that it provides a base or substrate out of which healthy cognitive functioning can occur. Although there are learning theories that discuss cognitive and affective domains, they identify them separate.

- Emotion is being associated with cognition in some kind of parallel way. H.Gardner's theory of multiple intelligences (2006; including intrapersonal and interpersonal intelligences) and Goleman's theory of emotional intelligence (1995) both construct emotion as analogous to the more traditional cognitive 'intelligence'.

With respect to the relation between emotion and affect, research provides an evidence for the claim that emotion, together with cognition and motivation are the key components of learning (D'Mello et al., 2005). Thus, it is important to help students know how and when their "emotional intelligence" works to help or hinder their success. According to Salovey and Mayer (1990), emotional intelligence refers to: *"The ability to monitor one's own and others' emotions, to discriminate among them, and to use the information to guide one's thinking and actions"*. Thus, emotional intelligence is important to foster students as self-regulators as they should be capable to: know their emotions -i.e. self-awareness, manage their emotions, and motivate themselves – i.e. self-motivation.

To this end, which emotions should be considered when it comes to discuss learning? How students manage their emotions and motivate themselves in case of difficult learning activities? Kort & Reilly (2002) describe the relation of learning and emotions through what they called Four Quadrant model, moreover they state "*A typical learning experience involves a range of emotions, cycling the student around the four quadrant cognitive-emotive space as they learn. It is important to recognize that a range of emotions occurs naturally in a real learning process, and it is not simply the case that the positive emotions are the good ones*" (see Figure 2.1).

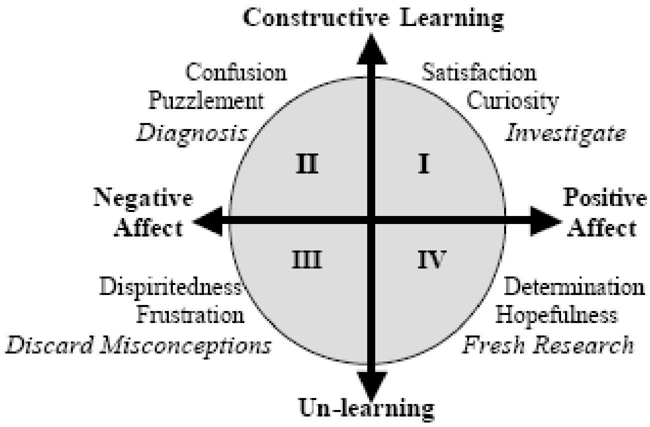


FIGURE 2.1. Four Quadrant model relating phases of learning to emotions (Kort & Reilly, 2002)

Positive emotions do not always lead to good results in learning whereas negative emotions lead to worse learning results (see Hascher, 2010). Moreover, Pekrun et al. (2006) noticed that positive affect is positively related to mastery goals, and that negative affect (e.g. test anxiety) is positively related to performance-avoidance goals. According to Hascher (2010) the following eight factors should be considered to analyze emotion's quality: *Valence* (pleasant - unpleasant), *Arousal level* (activating - deactivating), *Intensity* (intense - low), *Duration* (short - long), *Frequency of occurrence* (seldom - frequent), *Time dimension* (retrospective, actual, prospective), *Point of reference* (self-related; related to others; referring to an activity), and *Context* (during learning, achievement, etc.).

TABLE 2.2. Taxonomy of achievement emotions proposed by Pekrun et al. (2006)

	Valence	
Object Focus	**Positive**	**Negative**
Activity	Enjoyment	Boredom
Outcome		Anger
Prospective	Hope	Anxiety, Hopelessness
Retrospective	Pride	Shame

According to Pekrun et al. (2006), two further important determinants of emotions (with respect to achievement) are the *perceived controllability* and the *subjective value* of the activities and outcomes. High controllability and subjective value lead to positive emotions whereas low controllability and low subjective value lead to negative emotions. Moreover, two important dimensions for emotions with respect to achievement are *object focus* and *valence*. In their taxonomy of achievement emotions, Pekrun et al. summarized their assumptions (see Table 2.2). Regarding object focus, activity-related emotions such as enjoyment or boredom can be distinguished from outcome-related emotions. These outcome-related emotions can be either prospective (e.g. hope) or retrospective (e.g. shame). Regarding valence, positive emotions regarding achievement can be distinguished from negative emotions.

2.4. Motivation and Learning

Motivation has been considered as essential factor for effective learning. Motivation for learning has been described as the 'engine' that drives teaching and learning (Stiggins, 2001, p. 36; cited after Harlen, 2006, p. 61). Moreover, according to Bransford *et al.* (2000) motivation affects the time and effort learners plan or consume to learn or to solve problems. Nevertheless, motivation is considered as an important outcome of education (Harlen, 2006, p. 61), thus teaching and learning activities should be carefully designed to promote motivation. For instance assessment is considered as one of these factors that affect motivation, however Stiggins (2001, p.36, cited after Harlen, 2006, p. 62) argues that assessment practices that are provided within a course can enhance or destroy students' desires to learn more quickly and more deep than any other tools.

Research distinguishes between 'intrinsic' and 'extrinsic' motivation when it is applied for learning. Intrinsic motivation is usually used when the learning process itself satisfies learners, where extrinsic motivation is often derived by potential gains such as money, rewards, or praise (Harlen, 2006, p. 62). The research of (Kellaghan, Madaus, & Raczek, 1996) shows that there is an evidence of the influence of intrinsic motivation on learns engagement that leads to 'deep' learning through higher level thinking skills and the conceptual understanding. Moreover, Crooks (1988) highlights the problems associated with extrinsic motivation as it leads to 'shallow' instead of 'deep' learning.

Harlen (2006) discusses the following aspects as they influence the time and effort required for learning: interest, goal orientation, locus of control, self-esteem, self-efficacy, and self-regulation. Interest is more related to result of interaction between the learner and the learning environment, interest is influenced by the goals the learners set to participate in a learning task and accordingly how much effort and time they require to achieve those goals, thus learners prefer to control locus internally as their success or failure is determined by their abilities and effort rather than external locus based on teacher or luck, this leads to how those students value themselves and how they can show the confidence that they are capable to learn and to overcome problems, thus students can be aware about their self-efficacy when they are confident about their abilities to succeed or fail particular learning task.

2.5. Assessment

Assessment with respect to education and learning has been discussed over decades now as a major part of the learning process. Assessment is considered to be a central practice of education and learning (Gouli, Gogoulow, & Grigoriadou, 2008; Pellegrino, Chudowsky, & Glaser, 2001). Rather than learning and assessment practices are considered solely, assessment practices should be designed to align with learning theories and learning goals (see for e.g. Bloom's Taxonomy (Krathwohl, 2002; Bloom, 1956).

According to (Dochy and McDowell, 1997) there is a change from the so-called 'testing culture' in which instruction and testing are considered to be two separate activities towards an 'assessment culture' in which instruction and assessment are integrated in one process. In this new culture of assessment, more focus on assessing competencies related to: *cognitive* (e.g. problem solving, critical thinking), *meta-cognitive* (e.g. self-reflection and self-evaluation), *social* aspects (e.g. leading discussions and working within groups), and *affective aspects* (e.g. internal motivation and self-efficiency) have been considered (Dochy & McDowell, 1997). Moreover, students play major roles in the assessment where new forms of assessment have been adapted to suit the learning styles of the modern learners. Such forms include interviews, performance assessment, exhibitions, portfolio assessment, process and product assessment, directed assessment, authentic assessment, performance assessment, alternative assessment, collaborative assessment and self- and peer-assessment (Buzzetto-More & Alade, 2006; Martell & Calderon, 2005).

Therefore, assessment rather than being considered as an extension is an integral part of teaching and learning by which information about learner performance is collected and evaluated (Rovai, 2000). Moreover, Sluijsmans, Prins, and Martens (2006, p.46) suggested that: *"because the goals as well as the methods of instruction are oriented towards integrated and complex curricular objectives, it is necessary for assessment practices to reflect this complexity and to use various kinds of assessments in which learners have to interpret, analyse, and evaluate problems and explain their arguments"*.

Furthermore, Wiggins (1990) suggested that assessment should be provided in a way that is rather based on the real world. With respect to the methods of assessment, there are several possibilities such as traditional (paper-pencil) tests, instructor observations, writing samples, discussions, analysis of student work or portfolios in which a variety of samples covering the student work are included (e.g., Boston, 2002; Buzzetto-More & Alade, 2006). Moreover, with respect to e-assessment, a variety of methods and tools can be found (see chapter 3). For instance the use of assessment to diagnose the learning style and thus provide personalized learning activity is discussed in next section.

Learning Styles

According to (Palmer & May, 2004), learning Styles has a major influence on technology-based education in terms of: *Products* (different designs of educational products based on learning styles), *Services* (impact on services configuration, e.g. course configuration, delivery methodology, and assessment type), *Environments* (impact on the role LMS plays in learning delivery), and *Practices* (impact on the possible interaction between learners in the learning process based on different user roles. i.e. instructional designers, technical and management staff, teachers, and students). Assessment plays a major role when it comes to identify learning styles. For instance, in 1986 Honey and Mumford developed their learning style questionnaire as a diagnostic tool to support the learning process based on individual learners preferred learning style (Palmer & May, 2004; Honey & Mumford, 1992), and to support the alignment between learning, teaching, and assessment (Crisp, 2007p.25). According to (Honey & Mumford, 1992) learning style model learners are divided into: *activists* (prefer practical learning activity e.g. games and simulations), *pragmatists* (prefer situated learning activity- related to the real world simulations and group discussions), *reflectors* (prefer to read, compare to find alternatives e.g. case studies, research-based tasks), and *theorists* (prefer to read, analyze to find implications e.g. structured content, lectures and demonstrations).

The research of Honey highly depends on Kolb's Experiential Learning Theory (ELT) (Kolb & Kolb, 2005; Kolb, 1984). In which Kolbs argues that learning cycle happens in the following modes: concrete experience, reflective observation, abstract conceptualization, and active experimentation. Kolb and Kolb (2005) developed the Learning Style Inventory (LSI) based on the experiential learning theory. When the learning style of a person is specified with the LSI, a score for each of the ELT modes is generated. Based on these scores, four different learning styles can be distinguished: *Divergers* (prefer both the concrete experience and the reflective observations), *Assimilators* (prefer a combination of reflective observation and abstract conceptualization), *Convergers* (prefer abstract conceptualization and active experimentation), and *Accommodators* (prefer combinations of concrete experience and active experimentation).

Palmer & May (2004) adapted the learning style model from (Honey & Mumford, 1992) to consider how learning objects (LO) can be presented to learners based on a learning style diagnosis step. According to this adapted model the learning style diagnosis step should take place before the design of the LO and defines the 'dominant learning style' of the learner for this LO. A cognitive pre-assessment follows in the second step by which designer or system measures the required knowledge for the identified learning style as well as for the different phases of learning content based on learners' learning style and knowledge state. Moreover, the pre-assessment can raise issues on the 'dominant learning style' and suggests a second possible learning style according to which the learning content phases can be moderated.

Learning Analytics

Learning Analytics is an emerging field in which analytics tools are used to: (1) track learners interactions with the learning environment, learning content, their peers and teachers, (2) analyze those interactions, (3) predict possible interactions based on extracted knowledge from collected data, (4) visualize the extracted knowledge to students and teachers, (5) refine the way learning have been designed to more support students achieving learning goals, and (6) share the students learning experience with other application domains and contexts.

Learning analytics is related to other fields such as: business intelligence, web analytics, academic analytics, and action analytics. Academic analytics has been first discussed by Goldstein and Katz (2005) to describe the application of business intelligence tools and principles in academia. More precisely to discuss how institutions use academic analytic tools. Norris et al. (2008) argues that improving performance in educational institutes requires the deployment of educational analytics (tools, solutions, and services) in order *"to produce actionable intelligence, service-oriented architectures, mash-ups of information/content and services, proven models of course/curriculum reinvention, and changes in faculty practice that improve performance and reduce costs."* (p. 44). Thus, this deployment includes: to capture meaningless data based on students interactions, report it as information, to enable prediction of future action and to support making decision based on intelligent actions, so to refine and improve the learning process in an ongoing process (Campbell, De Blois, & Oblinger, 2007). Based on that Elias (2011) utilizes the 'Collective Application Model' (Dron and Anderson, 2009) to define the learning analytics process as: data gathering which includes data selection and capture, data processing which includes aggregation processes and information reporting which can be used to make predictions, and application of information in which the use, refinement, and sharing knowledge to support decision-making and system improvement.

To this end, learning analytics aims to use analytics tools in education to improve learning and teaching. Moreover, to develop systems capable to adapt learning content and learning process in a personalized approach by capturing, reporting, processing and acting on collected data in an ongoing process improve learning and support students (Elias, 2011). The use ICT in Higher Education (HE) facilitates the tracking of students' interactions as well as the visualization of the extracted knowledge from these interactions. However, the challenge is to identify learning measures that can be used to analyze the captured data, moreover to align the selected data to be captured with desired pedagogical action to be implemented (Dawson, Heathcote, & Poole, 2010).

2.6. e-Education

e-Education - as a synonym for technology-based learning and teaching - is an emerging field of research in which aspects related to using technologies in learning and teaching are discussed. Terms commonly used to define online education include computer-based training, e-learning, technology enhanced learning, Internet learning, network learning, distributed learning, virtual learning, web-based learning, computer-assisted learning, and distance learning (Ally, 2004; Anderson, 2004; Cross, 2004; Dror, 2004). However, in most of these terms learning and teaching are supported by technology (usually computers) which is used by learners at a distance from tutor or instructors to access learning material, interact with tutors or instructors as well as other learners, and receives feedback and support (Ally, 2004).

As cited in (Cross, 2004) the term 'online learning' goes back to late 1997 when Elliott Masie - a learning expert - said, "*Online learning is the use of network technology to design, deliver, select, administer, and extend learning.*". In 1998 Jay Cross wrote "*eLearning is learning on Internet Time, the convergence of learning and networks. eLearning is a vision of what corporate training can become. eLearning is to traditional training as eBusiness is to business as usual.*".

e-Education has evolved revolutionary through the emergence influence of information and communication technology (ICT) and web 2.0 to have the so called "e-learning 2.0" with a variety of e-learning - including assessment - tools, platforms, and services. However, using ICT in education has been led by the emergence of technology instead of learning theories and pedagogy (Watson, 2001). According to (Ravenscroft, 2001), with respect to e-learning tools no straightforward answer for which tool is most suitable to a learning practice, and thus learning theories should be more considered in the design of those tools. Moreover, Ravenscroft proposes dialogue models as 'design as theory' by which learning theory, technology, and context are considered in the design of educational interactions in a way that utilizes design as a theory to be developed, validated, evaluated, and adapted rather than just delivered. Thus, learners' interactions can be predicted, evaluated and validated rather than just try the tools in real context and see consequences.

Nichols (2003) has proposed fundamental principles for e-learning, towards a theory of e-learning in a discussion and the post-discussion version is as follows:

1 E-learning is means of implementing education that can be applied within varying education models (for example, face to face or distance education) and educational philosophies (for example behaviorism and constructivism).
2 E-learning enables unique forms of education that fits within the existing paradigms of face to face and distance education.
3 Whenever possible the choice of e-learning tools should reflect rather than determine the pedagogy of a course, *how* technology is used is more important than *which* technology is used.
4 E-learning advances primarily through the successful implementation of pedagogical innovation.
5 E-learning can be used in two major ways: the presentation of education content and the facilitation of education processes.
6 E-learning tools are best made to operate within a carefully selected and optimally integrated course design model.
7 E-learning tools and techniques should be used only after consideration has been given to online vs. offline trade-offs.

8 Effective e-learning practice considers the ways in which end-users will engage with the learning opportunities provided to them.
9 The essential aim of education, that is, enabling the learner to achieve planned learning outcomes, does not change when e-learning is applied;
10 Only pedagogical and access advantages will provide a lasting rationale for implementing e-learning approaches.

Based on these principles, e-learning is considered as a *mean* for education instead of a *mode* of education. Moreover, it facilitates the education process by supporting different paradigms of education - e.g. face to face, distance education, and blended learning. The selection of e-learning tools should be derived by the pedagogy to support both content presentation as well as the educational process in a context influenced by different learning theories to support a variety of educational models. Nevertheless, end-users - as 'digital natives' - should be considered in the design of the whole educational process not only on the selection of the e-learning tools. Thus, they can achieve their learning goals despite the mean of learning - i.e. e-learning tools.

In the age of the so-called "information age", learners grow up with technology dominating most of their life activities. They use technology anywhere, anytime, and they are faced with the challenge of needing to be engaged and motivated in their learning (Prensky, 2001). The emergence of Web 2.0 and the influence of information and communication technology (ICT) have fostered e-learning 2.0 to be more interactive, challenging, and situated. As a result, learners feel empowered when they are engaged in collaborative learning activities and self-directed learning. The learners are also provided with e-learning systems that would maintain their social identity and situated learning experience. Given the different learning styles of students, educators are faced with the challenge of having to develop assessments which are required to appraise the student's learning process. Assessment forms provided in e-learning activities have to be aligned with theories of learning so that they can foster different types of learning such as reflective-learning, experiential-learning, and socio-cognitive learning (AL-Smadi, Guetl, & Chang, 2011; Dochy & McDowell, 1997; Elliott, 2008).

Dagger et al. (2007) discuss the evolution of e-learning platforms and they distinguish between three main generations as depicted in Figure 2.2. The first generation is stand-alone monolithic systems for specific learning activities with support content-only inter-operation. The second generation represents web-based modular systems with more interest on users and their associated profiles. The second generation lacks flexible services and tools interoperability, where the need for exchange of semantic representation of e-learning domain has been addressed. The third generation holds great promise when it comes to having interoperable learning services and tools within more personalized and adaptive e-learning platforms. This generation highly depends on service-oriented architectures (SOA) where its services support federated exchange (information and control), various levels of interoperability (intra-domain and inter-domain), and service composition (orchestration and choreography).

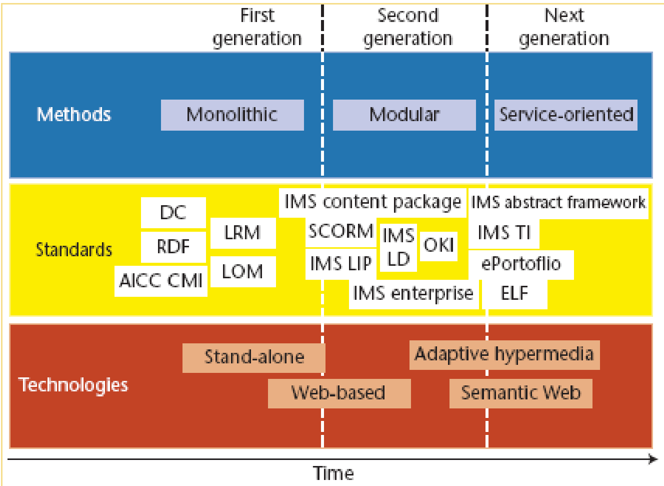


FIGURE 2.2. Evolution of learning management systems (Dagger et al., 2007)

To this end, e-learning tools and environments should be developed with highly attention to adopt pedagogical approaches – that supports theories of learning aligned to suitable assessment forms (AL-Smadi, Guetl, & Chang, 2011; Bransford, Brown, & Cocking, 2004; Dochy & McDowell, 1997). Moreover, instructional designers should give attention to align emergence technology and aspects from pedagogy and learning when it comes to design and develop learning environments (Watson, 2001; Ravenscroft, 2001) capable to support flexible learning - that considers learning styles, learner preferences, and learner knowledge and skill state. Therfore, flexible software architectures - such as service-oriented architecture - should be more considered to modularize educational services, use them within complex learning resources through guidelines from related pedagogical approaches and theories of learning. Educational standards, practices, and guidelines should be considered to achieve technical and pedagogical flexible learning tools to be used in different learning settings through interoperable and accessible approaches (AL-Smadi, Guetl, & Helic, 2009).

2.7. Summary

Over the last decades, our modern life has been influenced by the shift to more global and knowledge-centered society with a rapid development in technology. Educational systems - including teaching and learning - have been struggling to cope with this shift and challenges. Therefore, new and modern teaching and learning styles, settings, and practices have emerged to meet those challenges. These modern settings require people to improve their skills as well as their expertise to cope with the rapid changes in their societies (Gütl, & Chang, 2008).

The learning process has changed from being repetitive to a new form of learning based on understanding, independency, learners' empowerment and skills improvement (Bransford, Brown, & Cocking, 2004). The learning theories have also changed from being associative and behavioral to be more cognitive and constructive, with an emphasis on social, collaborative, affective and emotional theories of learning. In this shift of learning understanding, measurement have evolved from being scientific measurement - i.e. separated from the instruction activity - to have a new culture of assessment - i.e. measurement and instruction are integrated (Shepard, 2000). Moreover, learners are required to master new competencies and skills such as: cognitive competencies (e.g. problem solving, critical thinking, formulating questions and inventing and creating new things), meta-cognitive competencies (e.g. self-reflection and self-evaluation), social competencies (e.g. leading discussions and working in groups) and affective dispositions (e.g. internal motivation and self-efficiency) (Dochy & McDowell, 1997).

New age of information has appeared where information and communication technology (ICT) plays a major role in education and learning society. Over the last decades education has evolved to be administered and provided using technology. However, this shift of education paradigm to e-education has been dominated by the technology with shy attention to pedagogy and theories of learning (Watson, 2001). As a result, educators are faced with the challenge to select among a variety of resulting educational technology and tools to meet their educational goals (Ravenscroft, 2001). Therefore, experts in education and educational professional organizations recommended designing frameworks to foster education through: (a) standards and guidelines to administer, develop, and provide quality e-education, (b) an emphasis on pedagogy and theories of learning as well as types of learning such as, collaborative learning, self-directed learning, intuitive learning, and social learning, (c) adaptive and personalized educational tools and services through which learners are provided learning environments that maintain their social identity, their learning progress, and life-long learning skills. Moreover assessment forms provided in e-education should be aligned with instruction and learning (Biggs, 1999) thus to meet the education goals. The next chapter sheds the light of educational assessment and discussed assessment rationales and purposes through insights of assessment strategies. Nevertheless, discusses several practices of e-assessment provided for e-education and e-learning.

Chapter

3

3. e-Assessment

Over the last decades, the learning process has changed from being repetitive to a new form of learning based on understanding, independency, learners' empowerment and skills improvement. Assessment is no more considered to discriminate between students, rather than it is used to provide them quality feedback, engage them, and enhance students learning and encourage them for further progress and success. Therefore, several assessment systems have been developed to foster this change. This chapter aims to discuss educational assessment in general, assessment rationales, and in particular focuses on e-assessment as an integrated part of e-learning.

To this end, this chapter discusses the use of technology in supporting educational assessment. However, motivations and rationales for using e-assessment are discussed. Moreover, insights from history of e-assessment are presented as well as look ahead on the future of e-assessment. Nevertheless, e-assessment models are illustrated in order to show how assessment is linked to learning objectives and learning tasks. Emerging trends of e-assessment - such as e-assessment in collaborative learning and in serious games and simulations - are highlighted. Instructional rubrics, their types - i.e. holistic and analytic - and use rationales are discussed as well. Feedback as an essential component of e-assessment is discussed in this chapter.

This chapter is based on (AL-Smadi et al., 2011; AL-Smadi, Guetl, & Chang, 2011; AL-Smadi & Gütl, 2008).

3.1. Assessment

Assessment with respect to education and learning has been discussed over decades now as a major part of the learning process. Assessment is considered to be a central practice of education and learning (Gouli, Gogoulow, & Grigoriadou, 2008; Pellegrino, Chudowsky, & Glaser, 2001). Orrell (2005, p.17) defines assessment as "*Assessment practice is a deeply complex phenomenon that defines educational goals and outcomes and shapes student learning. Assessment processes make profound demands on students and teachers alike in terms of time, resources and emotions*". Assessment evaluates students learning based on educational goals, checks the achieved learning outcomes, and concerns different stakeholders –e.g. teachers and students. Moreover, aspects such as, (a) time - required by students to master a learning skill or competence, or time consumed by teachers to assess, or follow student's progress, and provide feedback, (b) used resources, and (c) emotions are important to the assessment process. Moreover, assessment lies at the hurt of the learning process, "*assessment, rather than being something added, is an integral, ongoing aspect of teaching and learning. It is the process of gathering, describing, or quantifying information about learner performance*" (Rovai, 2000 p. 142). Thus, assessment is considered as a continuous process, "*assessment is an ongoing process that involves planning, discussion, consensus building,* *reflection, measuring, analyzing, and improving based on the data and artifacts gathered about a learning objective.*" (Buzzetto-More & Alade, 2006; Martell & Calderon, 2005).

3.1.1. Purpose of Assessment

A major purpose of assessment in education is to improve learning. Bone (1999, p. 3) states that: *"The main purpose of assessment is to discover if students have achieved the learning outcomes of the course studied. The term assessment is derived from the Latin phrase ad sedere: to sit down beside. Primarily then assessment should provide guidance and feedback to the learner".* However, assessment is not only used to evaluate the level of learning outcomes achievement, but also to provide timely feedback, and support learning process. Kellough and Kellough (1999; quoted after Buzzetto-More & Alade, 2006) proposed seven purposes of assessment:

1. Improve students' learning
2. Identify students' strength and weakness
3. Review, assess, and improve the effectiveness of different teaching strategies
4. Review, assess, and improve the effectiveness of curricular programs
5. Improve teaching effectiveness
6. Provide useful administrative data that will expedite decision making
7. To communicate with stakeholders.

According to (Gibbs,1999 p. 47) assessment has six main functions: "*(1) Capturing student time and attention, (2) Generating appropriate student learning activity, (3) Providing timely feedback which students pay attention to, (4) Helping students to internalise the discipline's standards and notions of equality, (5) Generating marks or grades which distinguish between students or enable pass/fail decisions to be made, and (6) Providing evidence for other outside the course to enable them to judge the appropriateness of standards on the course*".

3.1.2. Types of Assessment

Assessment can be categorized according to its purpose into: diagnostic assessment, formative assessment, and summative assessment (AL-Smadi and Guetl, 2008; Crisp, 2007 p. 39; Bransford et al., 2004 p. 140). *Diagnostic assessment* is performed before the learning activity and used to identify the current state of learners' knowledge and skill. Thus, learning activity can be adapted accordingly. *Formative assessment* is part of the learning process; this assessment is used to give feedback to both students and teachers in order to guide their efforts toward achieving the goals of the learning process, whereas *summative assessment* is performed at the end of specific learning activity; and used to judge the students' knowledge and skill and also to discriminate between them.

Another classification is related to assessment delivery type, according to this classification assessment can be either traditional – i.e. paper-based, paper-pencil-test, or oral test based on face-to-face delivery - or it can be computer-based delivered. However, assessment is also classified according to the nature of response to the test items into, Convergent - *fixed response* - assessment and Divergent - *free response* - assessment. Fixed response assessment - also referred to as *objective* assessment - forces students to have a fixed response by selecting an answer from a pre-prepared list of solution alternatives. In the free response assessment – *subjective* assessment - unanticipated answers formulate the user's response. In subjective assessment skills like programming, writing essays, and meta-skills are assessed, where in objective assessment factual knowledge is more likely to be assessed. (AL-Smadi and Guetl, 2008; Crisp, 2007 p. 26; Culwin, 1998).

Nevertheless, assessment can be classified based on:

- **Assessment domain:** cognitive, affective, and psychomotor (*cf.* Bloom, 1956).
- **Assessment type:** diagnostic, formative, and summative assessment (*cf.* AL-Smadi & Guetl, 2008; Crisp, 2007; Bransford, Brown, & Cocking, 2004).
- **Assessment strategy:** traditional assessment, individual assessment, group assessment, slef-assessment, peer-assessment, instructor-based assessment, and system-based assessment (*cf.* Dochy & McDowell, 1997).
- **Assessment referencing:** norm-related, criterion-related, or ipsative (*cf.* McAlpine, 2002).
- **Assessment practice:** behavioral assessment, performance assessment, portfolio assessment, and rubric-based assessment (*cf.* Buzzetto-More & Alade, 2006).
- **Assessment adaptation:** micro-adaptive assessment, or macro-adaptive assessment (*cf.* Kickmeier-Rust, Steiner, & Albert, 2009).
- **Assessment method:** quantitative or qualitative (*cf.* AL-Smadi & Guetl, 2008; Crisp, 2007; Culwin, 1998; Bloom, 1956).
- **Assessment feedback:** feedback type, format, frequency, and content (*cf.* Nicol, Milligan, 2006; Wiggins, 2001; Black & Wiliam, 1998; Charman, & Elms, 1998).

Learning objectives, didactic objectives, and teaching strategies play a crucial role of the provided assessment and its application domain. For instance, Bloom (1956) distinguishes among the domains of educational assessment - i.e. cognitive, affective, and psychomotor - and provides taxonomy for the level and type of the assessment forms in the cognitive domain (see Chapter 2). Moreover, assessment practices differ according to the target pedagogy and the theory of learning. For instance, in some specific types of learning - e.g. collaborative learning - implemented in a class learning activity of group work, traditional forms of assessment such as paper-pencil-test cannot evaluate high levels of skills such as student ability for group-wok. Thus, assessment forms should be carefully designed with consideration of learning objectives, theories of learning, and pedagogy. The next sections explore the domain of assessment based on the aforementioned aspects.

3.1.3. Strategies of Assessment

Assessment varies with respect to its strategy into conventional assessment or alternative assessment. The conventional assessment often refers to the assessment type in which students select an answer whereas in the alternative forms students create an answer (Biggs, 1996). The next sub-sections discusses in details examples of the two categories and theirs role is learning and students support.

Traditional Assessment

Assessment has different forms according to control and participation, traditional assessment (instructor-based): in this form of assessment instructor is controlling the whole assessment process, defines assessment scheme, evaluates the performance of students and learning outcomes, and provides feedback and guidance to learners (Gouli, Gogoulow, & Grigoriadou, 2008). Emerging forms of assessment consider more participation of students and empower them with assessment methods to reflect on themselves, evaluate their peers, and to provide feedback (Gouli, Gogoulow, & Grigoriadou, 2008; Dochy and McDowell, 1997). Examples of such alternative forms of assessment are, self-assessment, peer-assessment, and collaborative (co-assessment) assessment.

Instructor assessment costs teachers extra time and workload; they are required to control and handle the assessment process themselves. They have to define assessment criteria according to which they will evaluate the student's progress and the learning outcome and accordingly provide feedback to their students. With the increase of students in classes a practical problem has been appeared. Thus, the need for assessment forms that engage the learners into the assessment process is important. (AL-Smadi & Guetl, 2008; Charman & Elms, 1998).

With respect to learning theories and the shift in learning understanding, instructor assessment is neither valid nor applicable in the emerging types of learning - i.e. self-regulated learning, collaborative learning, and experiential learning. Moreover, Dochy and McDowell (1997) state that teachers should not be considered as knowledge holders which they have to transfer to their students, rather than they should behave as mentors and support their students to learn based on understanding and doing. However, in alternative assessment students are given the chance to construct an answer instead of selecting one as well as they are empowered thus engaged through defining the assessment schema such as in collaborative assessment. Examples of alternative assessment are discussed in next sub-sections.

Self-assessment

Klenowski (1995, p. 146), defines self-assessment as *"the evaluation or judgment of the worth of one's performance and the identification of one's strengths and weaknesses"*. Self-assessment refers to learning activities in which learners define their own assessment criteria, and evaluate their learning performance accordingly (Gouli, Gogoulow, & Grigoriadou, 2008; Dochy, Segers, & Sluijsmans, 1999). Self-assessment requires students to reflect on their own work and learning progress with respect to assessment criteria they have defined themselves for that. Thus, they evaluate their self-awareness about their learning and improve their skills of self-reflection and meta-cognitive skills in general (Dochy and McDowell, 1997). However, Roberts (2006) noticed that students are not skilled to effectively assess themselves and they need practice and guidance to enhance their meta-cognitive skills.

Peer-assessment

According to (AL-Smadi, Guetl, & Kappe, 2009) peer assessment is not new and it can be rooted back to a long time ago, when George Jardine the professor in the University of Glasgow from 1774 – 1826 prepared a pedagogical plan that included some peer-assessment methods and advantages (Topping, 2003). Peer-assessment has been defined as *"an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of the learning of peers of similar status"* (Topping, 1998, p. 250). From this definition, peer-assessment is not considered a method for measurement but it is a form of assessment that can be utilized within a framework side by side with other methods (Brown, Bull, & Pendlebury, 1997).

Peer-assessment refers to the assessment activities in which learners evaluate their peer's work and provide them feedback. According to Kim (2004) peer-assessment can be used as an alternative assessment method for traditional assessment - i.e. linear - as well as a method for learning - i.e. iterative. Moreover, Kim argues that linear assessment has three main stages: planning, assessing, and ends with receiving peer feedback or grade, whereas in the iterative one - i.e. as a method for learning - two other stages are required in an iterative approach: reviewing/reflection, and revising, as follows:

In the **planning stage**, assessors define their assessment goals and criteria with respect to the course and the assessment activity objectives in general (Boud, 1995; Topping et al, 2000). Accordingly, assessors set strategies and procedures to achieve their defined goals. Defining assessment criteria is crucial in this phase, by which assessors can explicit tacit knowledge to other and improve the quality of formative feedback (Kim, 2008; Rust, Price, & O'Donovan, 2003). Thus, peer-assessment can be either alternative assessment method when only teacher is involved in defining the assessment criteria, or can be a method for learning when students and teachers collaborate in defining the assessment criteria (Orsmond, Merry, & Reiling, 2002, 2000; Dochy & McDowell, 1997).

In the **assessing stage**, the defined assessment criteria in planning stage are used by assessors to evaluate their peer's work. By practicing peer-assessment, assessors learn from other mistakes and progress within their contribution (Roberts, 2006 p. 7, Race, 1998, 2001). Thus, they use their pre-knowledge to evaluate their peer's performance and products which not only lead to develop new knowledge and understanding of the learning domain, but also improves their meta-cognitive skills -i.e. self-awareness, and self-reflection (Orsmond, Merry, & Reiling, 2002; Topping et al, 2000; Dochy & McDowell, 1997). Moreover, this makes the whole learning process authentic and provides a positive effect on the process of learning in general and student's motivation and engagement in particular (Roberts, 2006 p. 6; McConnell, 2000 p.127). In this stage the assessment process differs based on the assessment strategy. In case of conducting peer-assessment as an alternative assessment method, assessors are required to use assessments' criteria to mark their peers' products which will be later on used to grade their learning outcome. While in the other case of using peer-assessment as a method for learning, assessors are required to provide formative feedback - covers both strengths and weaknesses of their peers' products (Roberts, 2006 p. 7; Race, 2001 pp. 94 - 95) - which their peers will use to scaffold their learning performance and product.

In the **receiving feedback stage**, the type of feedback and its purpose differs based on the assessment strategy. In case of receiving a final grade students will be informed about their level of progress and the achievement of learning outcomes. While in formative feedback, students (assessee) are provided comments and hints by which they can scaffold their learning process and product. Both cases have been discussed in research where aspects such as, accountability of students as assessors, the quality of provided feedback, and the validity and reliability of the peer-assessment have been highlighted (Kim, 2008; James, 2006; Orsmond, Merry, & Reiling, 2002, 2000; Topping et al, 2000; Falchikov, 1995).

The **reviewing/reflection stage** is only used in peer-assessment as a method for learning. In this stage the connection between peer-assessment and self-assessment is represented by asking assessors to self-assess their evaluation through self-reflection (Dochy & McDowell, 1997). Moreover, assessees use what they got as feedback to improve their learning process and product. Thus, using peer-assessment as a method of learning is an iterative approach which highly depends on the feedback provision and use. Provided feedback helps students to adjust their learning progress towards the learning objectives and goals (Kim, 2008).

In the **revising stage**, students (assessees) utilize the formative feedback to improve their product, which can be peer-assessed again in an iterative way to improve it towards the objectives of the final product.

Contrary to quantitative assessment students selects answers from pre-defined ones - e.g. fixed response question or objective testing in general, formative peer-assessment is concerned with providing qualitative feedback to learners with more emphasis on the strengths and weaknesses of their work (Kim, 2008; Topping, Smith, Swanson, & Elliot, 2000). Thus, peer-assessment is considered as useful tool for learning through which students can learn by assess other's work (i.e. assessor role), and/or by receiving formative feedback (i.e. assessee role) (Orsmond, Merry, & Reiling, 2002; Topping et al, 2000; Dochy and McDowell, 1997). Assessment can be used to support students learning by means of formative assessment - assessment *for* learning and assessment *as* learning. Knight (1996) cited after (Dochy & McDowell, 1997) has identified the following advantages of formative assessment:

- Feedback provision that can be used to improve the learner's performance.
- It indicates quality and improves it, since it is about improving students learning.
- It can be formed as co-, self-, and peer-assessment where tutors are not urgently needed.
- It is not high-stakes assessment.
- It leads to a continuous communication between the assessor and the assessed, where understanding of the criteria and the learning activity is required.

Collaborative assessment

Collaborative assessment - as a synonym for co-assessment or cooperative assessment - refers to "*the participation of the students with the staff in the assessment process, is a way of providing an opportunity for students to assess themselves while allowing the staff to maintain the necessary control over the final assessments*" (Hall; 1995 cited after Dochy, Sergers, & Sluijsmans, 1999).

In collaborative assessment students and teachers work together in defining assessment scheme, and evaluate the performance of students (Gouli, Gogoulow, & Grigoriadou, 2008). Collaborative assessment empowers students with taking a role of a teacher to plan and define their assessment criteria and participate in the evaluation of themselves –i.e. self-assessment – and their peers –i.e. peer-assessment. Therefore, collaborative assessment is often related to self, peer-assessment forms (Dochy, Sergers, & Sluijsmans, 1999) as they not only participate in defining assessment criteria but they also use them to reflect on themselves and to evaluate their peers.

Rubrics-based Assessment

Assessment rubrics – often referred as instructional rubrics or marking rubrics- are a list of criteria each represented by levels of quality and their description in terms of mark scale or quality scale of poor to excellent (Andrade, 2000). The list of criteria is often defined based on the learning outcomes and what is going to be evaluated. The levels of quality or mastery levels provide description for the exemplary criterion with a possibility to provide informative feedback as well as a mark for each level.

Assessment rubrics are either *analytic* rubrics or *holistic* rubrics, in the analytic rubric different marking scales can be assigned to the rubric criteria where in the holistic one a single scale is used with all criteria (Hazari, 2004). However, assessment rubrics are designed based on the learning outcome and the application domain. The analytic rubrics help to figure the students strengths and weaknesses however its time consuming as they focus on variety of marking criteria. Andrade (2000) identified some rationales to use instructional rubrics:

- Instructional rubrics are easy to use and to explain.
- Instructional rubrics make teachers' expectations very clear.
- Instructional rubrics provide students with more informative feedback about their strengths and areas in need of improvement than traditional forms of assessment do.
- Instructional rubrics support learning.
- Instructional rubrics support the development of skills.
- Instructional rubrics support the development of understanding.
- Instructional rubrics support good thinking.

Literature research shows that the use of rubrics in assessment enhances assessment transparency, reliability, and validity (Andrade, 2005; Tiereny & Marielle, 2004). Moreover, rubrics aim to support students with feedback based on clear levels of quality and marking criteria. Nevertheless, the use of rubrics can support students focus their effort and tasks to promote meta-cognitive skills such as, self-awareness and self-reflection as well as higher level of thinking (Andrade & Du, 2005; Tiereny & Marielle, 2004). However, creating an instructional rubric is a challenging task itself and teachers may find it time-consuming. Therefore they adapt already available rubrics on the internet or from other learning activities. Thus, the new rubric or the amended one may be neither clear nor consistent in terms of its performance criteria descriptors. Therefore, Tiereny & Marielle (2004) proposes some guidelines to provide clear, usable, and consistent criteria descriptors:

- Criteria are explicitly stated: the rubric's criteria should be aligned with what is being taught and assessed. They should be carefully and clearly designed to reflect the learning activity dimensions and the learning outcomes in terms of what should be assessed.
- Criteria attributes are explicitly stated: a precise language should be used to describe the quality levels of each criterion. Research findings have stressed the importance of the clarity of criteria descriptors and their impact on the clarity and reliability of the rubric (Mertler, 2001; Popham, 1997), moreover the clear difference between the criterion levels through description (Moskal, 2003).
- Consistent descriptors of criteria progression from one level to the other: the descriptors for each criterion level should evaluate the same performance criteria

and hence the criteria progression through levels maintains consistency. Thus, clear verbal qualifiers – e.g. few, some, most, and all - should be used to explain and describe the progress thorough the criterion performance scale. Three scale measurement often used to descript progression levels: *amount* (e.g. few, some, most, and all), *frequency* (e.g. seldom, sometimes, usually and always), and *intensity* (e.g. slightly, moderately, mainly, and extremely) (Tiereny & Marielle, 2004; Rohrmann, 2002).

3.1.4. Referencing of Assessment

Referencing of assessment refers to the basis of evaluation. According to McAlpine (2002), there are three types of referencing in assessment: *norm-related* referencing, *criterion* referencing, and *ipsative* referencing. In norm-related referencing, the person's performance is compared with the performance of his/her peers. Despite its frequent use, this form or referencing does not consider the actual abilities of candidates. In criterion referencing, individual's learning is evaluated based on pre-defined criteria. According to Crisp (2007, p. 24) criterion referencing is used when it comes to have effective assessment of the educational process. Ipsative referencing is more related to individual progress improvement where learners can improve their learning progress by comparing their own progress –in same or different area - over time.

3.2. e-Assessment

Using technology to assist assessment has been an interesting research topic for decades. However, developments have mainly transferred traditional assessment approaches into computer environments. In order to automatically grade students' assignments, types of assessment approaches have been further limited (Elliot, 2008). Our life has been influenced by a revolution in the field of information and technology. Consequently, the rapid increase of using technology in learning settings expedites also the need for new technology-based assessment. Culture has changed significantly in the recent years, consequently, pedagogy has become affected and educationalists have also started redesigning educational systems (Prensky, 2001).

Learning is no more divided; there is no separation between schools' education and workplace experience. Acquiring knowledge is a continuous learning process. According to (Jegede, 2005), Learning is a continuous process over lifetime, it is a lifelong process. Therefore a new paradigm for assessment in lifelong learning is becoming important. Changing education from memorizing facts to higher levels of comprehension and synthesis requires building and assessing critical-thinking skills. According to (Haken, 2006), measuring knowledge is important but is not enough. The academic programs should work on building and assessing students' critical-thinking skills.

In general, assessment has different strategies according to its purposes. For instance, diagnostic assessment investigates the knowledge and skill state at the beginning of a learning activity and mainly used to provide personalized learning - based on learning styles and learner preferences (see Section 2.6). Formative assessment is part of the learning process and used to give feedback to both students and teachers in order to guide their efforts toward achieving the goals of the learning process; whereas summative assessment is performed at the end of specific learning activity and used to judge the students' progress and also to discriminate between them (Bransford, Brown, Cocking, 2004). However, assessment is dynamic continuous process in which students are provided quality feedback which they can use to reflect on themselves and scaffold their learning progress as well as feedback can be used to enhance the assessment practice through an iterative approach (*cf.* Kim 2004; Buzzetto-More & Alade, 2006) as discussed earlier. According to (Bannett, 2002), technology is an essential component of modern learning system, thus is also essential for the assessment process. According to (Buzzetto-More & Alade, 2006) "*The use of information technologies and e-learning to augment the assessment process may include: pre and post testing, diagnostic analysis, student tracking, rubric use, the support and delivery of authentic assessment through project based learning, artifact collection, and data aggregation and analysis*"

e-Assessment has been defined by the Joint Information Systems Committee (JISC) - in their report 'Effective Practice with e-Assessment' (JISC, 2007p. 6) - as "*e-Assessment is the end-to-end electronic assessment processes where ICT is used for the presentation of assessment activity, and the recording of responses. This includes the end-to-end assessment process from the perspective of learners, tutors, learning establishments, awarding bodies and regulators, and the general public*". e-Assessment is also known as Online assessment, Computer Based Assessment (CBA), or Computer Assisted Assessment (CAA) which are often used interchangeably. CBA can be understood as the interaction between the student and computer during the assessment process. In such assessment, the test delivery and feedback provision is done by the computer. Where CAA is more general, it covers the whole process of assessment involving test marking, analysis and reporting (Charman & Elms, 1998). The assessment lifecycle includes the following tasks: planning, discussion, consensus building, reflection, measuring, analyzing, and improving based on the data and artifacts gathered about a learning objective (Martell & Calderon, 2005).

e-Assessment systems can be classified according to the users' response on the test items into: *fixed response systems* and *free response systems*. According to (Culwin, 1998) fixed response systems which also referred to as objective forces the user to have a fixed response by selecting an answer from a pre-prepared list of solution alternatives; whereas in the free response systems, unanticipated answers formulate the user's response. In such type of systems skills like programming, essays writing and meta-skills are assessed rather than factual knowledge assessment which represents the main domain of the first type.

e-Assessment is not only applicable for individuals, but it is also used for groups. Assessment of groups, also referred to collaborative assessment, is used to assess the participation of individuals in group work and their behavior of how they collaborate with each other to solve problems (Reimann & Zumbach, 2003).

3.2.1. *e*-Assessment Motivations and Rationales

Motivations and rationales of using e-assessment instead of paper-based assessment in higher education are discussed in this section. According to (Charman & Elms, 1998), the practical and pedagogic rationales are the main motivators for adopting e-assessment in higher education.

Practical Rationales

Increasing number of students supervised by the same staff resources causes an increase in the staff workload. Accordingly, time spent by the teachers to assess students is also increasing. Therefore, a step toward the e-solutions becomes a real need. Although many e-learning environments have been developed in universities to overcome the workload problem, most system have not adequately solved the assessment tasks. Therefore, reducing time and efforts spent on students' assessment is a strong rationale to use the e-assessment technology. (Charman & Elms, 1998)

Another practical rational is the education paradigm shift to use ICT. The emerging use of technology in learning has influenced assessment practices as well. As discussed earlier assessment lies in the heart of the learning process. According to Bennett (2002; cited after Buzzetto-More & Alade, 2006), "*technology is central to learning and, as a result, is going to prove to be central to the assessment process*". Therefore, the shift to foster learning through technology would practically demand similar shift in providing assessment (Bartley, 2006). Moreover, in alternative forms of assessment such as behavioral assessment it is required to track students' behavior and analyze it to provide feedback and guidance thus computers are required to facilitate and support such forms of assessment.

Pedagogic Rationales

There is an increasing pressure on teachers and higher education in further to provide assessments that are fair, reliable, efficient and effective. Brown and Race (1996) illustrate that each assessment must take proper attention of a set of quality dimensions-i.e. fair, equitable, formative, well timed, redeemable, efficient, valid, incremental, and demanding. Many of the values are achieved by CBA systems in nature (Fair: offer fair opportunity for success; Equitable: be indiscriminating between students; Formative: give many opportunities to learn through feedback; Well timed: provide learning stimulus and be fair; Redeemable: allow a series of opportunities; and Efficient: be manageable within the constraints of resources); some others (Valid: accurately assess the delivered material, Reliable: promote consistency between assessment tasks, Incremental: increase reliability and consistency over time; and Demanding: challenge students and ensure high standards) depend on the experience of the assessment designer or the system designer (Charman & Elms, 1998).

According to (Dietal, Herman, & Knuth, 1991), appropriate assessment information provides an accurate measure of student performance to enable students, teachers, administrators and other key stake holders to make effective decisions. Therefore, any CBA/CAA system should satisfy the quality dimensions outlined above. Moreover, Formative assessment is an important component of any assessment activity. In such type of assessment, the current state of students is diagnosed. This ongoing process involves a continuous feedback to both teachers and students which allows them to enhance their teaching and learning activities to satisfy the learning goals (Bransford, Brown, Cocking, 2004). However, providing feedback to each student is an overwhelming task and requires extra workload. Therefore, in order to gain the pedagogical values of formative assessment and to avoid practical problems the use of technology to assist assessment is mandatory.

3.2.2. e-Assessment History and Evolution

Computers have been used for decades to assist assessment. The first attempts to use e-assessment can be found in 1930s when machine-scoring was first applied for the Strong Vocational Interest Blank. The use of computers to assist testing was firstly implemented in 1960s when in 1965 computer-assisted test interpretation system was implemented for interpreting the Minnesota Multiphasic Personality Inventory (Moreland, 1992 cited after Carlson & Harvey, 2004).

The history of e-assessment can be rooted to the use of computers to automatically assess the students' programming assignments (Douce, Livingstone, & Orwell, 2005). One of the early attempts of using computers to automate the process of assessing students' programming assignments was the "Automatic Grader" (Hollingsworth, 1960). Rather than using this program as a compiler for the programming assignments, it also helped the student to better learn programming, and also facilitates the teacher to supervise a larger number of students in the same course. Forsythe & Wirth (1965) presented another system for automatically assessing programming exercises written in Algol. The system was used by the students of a numerical analysis course at the University of Stanford to assess their programming exercises. The system was responsible for supplying data, monitoring running time and keeping a "grade book" for recording problems.

Assessment plays a main role for enhancing the performance of learners as well as the quality of instructional materials. According to Reiser (2001), in 1960's, formative evaluation was applied to the drafts of instructional materials before they were in their final form. The shift towards using computers in instruction for instance computer assisted instruction (CAI) has started in the same time. In the 1960s and early 1970s computers was used to assist instruction and assessment as an integrated process. Examples from that era are, the Programmed Logic for Automatic Teaching Operations (PLATO) project was started at the University of Illinois (Woolley, 1994). Time-Shared, Interactive, Computer-Controlled, Information Television (TICCIT), started in 1967 is another example of a large-scale project for using computers in education (Hayes, 1999).

Assessment also plays a major role in the domain of adaptive systems and in particular educational adaptive systems in which diagnostic assessment is used to diagnose learners' individual needs (such as personality factors, knowledge state and background, and leaning style) and thus adapt provided instructional activities provided to them (Park & Lee, 2003). For instance in the domain of computer-managed instruction (CMI) examples of such systems are, Program for Learning in Accordance with Needs (PLAN) and the Computer-Assisted Instructional Study Management System (CAISMS), details can be found in (Alessi & Trollip, 1991). In the domain of Intelligent Tutoring Systems (ITS) examples such as, PACT in the domain of Algebra, SHERLOCK in the domain of electronics can be found in (Graesser, VanLehn, Rose, Jordan, & Harter, 2001). In the domain of Adaptive Educational Hypermedia (AEH) a shift towards providing adaptive learning navigation in addition to adaptive learning content (known as adaptive presentation) started in the early 1990s. Examples such as, ISIS-Tutor, ITEM/PG, and other Web-based examples are, AHM, MetaLink, RATH, AHA! and ART-WEB. Further details and other examples can be found in (Brusilovsky, 2000; Brusilovsky & Peylo, 2003; Park & Lee, 2003; Sadat & Ghorbani, 2004).

With the emergence of technology and the use of web 2.0, adaptive educational hypermedia evolved to have more enhanced tools of the so-called "Enhanced Adaptive and intelligent Web-based Educational Systems" (Brusilovsky & Peylo, 2003). Examples for recent development are, KBS-Hyperbook (Henze & Nejdl, 2001), ActiveMath (Melis et al., 2001) in adaptive educational hypermedia (adaptive navigation support), whereas in intelligent tutoring German Tutor (Heift & Nicholson, 2001) is an example of intelligent diagnosis systems but not adaptive (see Brusilovsky & Peylo, 2003) , in the support of teachers and students in the teaching and learning process MLTutor (Smith & Blandford, 2003), and HyperClassroom (Oda, Satoh, & Watanabe, 1998; Merceron & Yacef, 2003) for intelligent class monitoring and education support.

Assessment as a main part of the instructional design and media was affected by the revolution of micro-computers in the 1980's. According to Reiser (2001), the 1980's marked an increasing interest of using computers in instruction, and computers were used in automating some instructional design tasks. Examples of assessment systems from 1980's in scientific disciplines are mathematics (Rottmann & Hudson, 1983) and chemistry (Myers, 1986).

The 1990's was affected by the important impact of the World Wide Web (WWW). With the emergence of the Web, assessment systems started to be web-based systems. There are several open-source and commercial assessment systems and tools. Some of them are integrated to learning management systems, where others are stand-alone assessment systems. Assessment tools vary from open-source to commercial and from low-level tools with minimal technical requirements to high-level systems designed for enterprise applications and institutions. Systems such as QUIZIT (Tinco at el, 1997), ASSYST (Jackson & Usher, 1997) and PILOT (Bridgeman at el, 2000) are also examples of web-based systems with ability of online testing and grading. Recent examples such as, Lei (2006) presented a web-based assessment system that applies Bloom's taxonomy to evaluate the outcomes of students and the instructional practices in the educators in real time. In a step towards a fully automatic knowledge assessment, Guetl (2007) introduced the e-Examiner as a tool to support the assessment process by automatically generating test items, marking students' answers and providing feedback.

As discussed in (Douce, Livingstone, & Orwell, 2005) e-assessment tools have three generations along history: (1) the so-called 'mechanically-oriented' testing, examples of this generation are: 'Automatic Grader' (Hollingsworth, 1960) and the system for automatically assessing programming exercises written in Algol (Forsythe & Wirth, 1965). (2) the so-called 'tool-oriented' systems , examples of this generation are: TRY system (Reek, 1989) which allows students to test their programming assignment using a testing program. The ASSYST system developed by (Jackson & Usher, 1997). In addition to testing programing assignments ASSYST opened the doors to automatic grading through providing tools to assign weights to the assignment based on pre-defined testing aspects. Another example from the mid-eighties is the Ceilidh system, developed at Nottingham University (Higgins et al. 2003; cited after Douce, Livingstone, & Orwell, 2005). Ceilidh represents a CAA management system through the phases of development, deliver, and reporting of the assessment process. (3) the so-called 'web-oriented' systems which have been influenced by the revolution of WWW, examples of this generation are: the BOSS system developed at the University of Warwick in the UK. BOSS developed after ASSYST using similar specifications (Joy & Luck 1998; Luck & Joy 1999), RoboProf deployed at Dublin City University (Daly 1999; Daly &Waldron, 2004) to assess Java-based assignments in programming courses. RoboProf is enhanced with a web-based user interfaces to provide Java-based code and to test them and returns a report. Moreover the students receive automatic programming assessment as multiple-choice questions generated randomly to avoid cheating.

According to (Sanz-Santamaria, Gutierrez Serrano, & Vadillo Zorita, 2007) reviewing the domain of e-assessment tools is challenging especially when it comes to explore application domains and purposes. However, in their review they considered the following aspects: computer adaptive test support, widely used in academia, and still used and evolving. Moreover the study identified 10 leading assessment tools and evaluated them with respect to standards conformation, adaptive aspects and support, and pedagogy and learning theories considerations (see Figure 3.1).

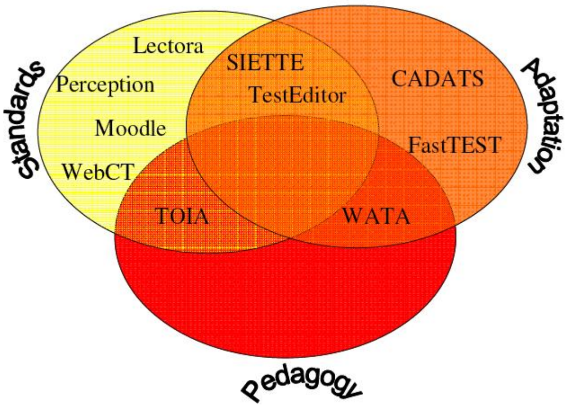


FIGURE 3.1. Evaluation of some assessment systems according to standards-conformation, adaptation support, and pedagogy aspects (Sanz-Santamaria, Gutierrez Serrano, & Vadillo Zorita, 2007)

Crisp (2007p. 69) in his book "e-Assessment Handbook" identifies over 50 e-assessment tools and systems. Examples of these systems include: (a) the Assessment Management System[1] (AMS) is an example of e-assessment system that can be used for different delivery modes such as online, off-line, LAN based, and CDs. AMS can be integrated with a LMS or can be used a standalone assessment system. It also has some abilities to author questions for particular disciplines such as mathematics and chemistry. For instance, an editor for built-in equations requires the Mathematics Markup Language (MathML) or the Chemical Markup Language (CML) so that instructors can create algebraic questions as well as chemical formulas and exercises. (b) Blackboard™[2] and Questionmark Perception™[3] are examples of enterprise-level of LMSs that includes an e-assessment engine. Such enterprise systems can conduct assessment for different purposes such as, diagnostic assessment, formative assessment, and summative assessment. They also designed to cover different assessment forms such as, self, peer-assessment and collaborative learning assessment. Moreover, the students responses are recorded in a grade-book and analyzed, as well as features such as item analysis and custom grading scales are available. (c) For free non-commercial examples, the Hot Potatoes[4] suite enables you to create interactive multiple-choice, short-answer, crossword, matching/ordering and gap-fill exercises. Hot Potatoes was developed by the research and development team at the University of Victoria Humanities Computing and Media Centre in Canada. Hot Potatoes suite includes six applications of: JMatch- creates matching exercises which can include pull down menus or drag-and-drop, JMix- You can split up sentences into parts, words or even words into letters, JCross- creates a crossword where you click on the numbers to enter the words, JQuiz- gives you the ability to make quizzes using multiple-choice, multiple-select, short answer or hybrid questions, JCloze- makes fill-in the blank (cloze) exercises, and Masher is a sixth tool that enables you to easily combine exercises. A license is required to use the Masher application. (d) AiM[5], CABLE[6], Maple T. A[7]., and

---

1 http://www.excelindia.com

2 http://blackboard.com/

3 http://www.questionmark.com

4 http://hotpot.uvic.ca/

5 http://sourceforge.net/projects/aimmath/

WaLLiS[8] are examples of web-based assessment software for delivering mathematical assessments. These tools provide a mathematical editor where student and instructors can write mathematical formulas and expressions in related format. Maple is capable to compare two algebraic expressions and determine if they are mathematically equivalent. Therefore, other tools such as AiM and WaLLiS use Maple to determine mathematical equivalence in order to deliver mathematical assessments.

Additionally, portfolios can also be used to assess learning outcomes as well as to diagnose limitations in curriculum that should be enhanced (Buzzetto-More & Alade, 2006). Moreover, they argue that "*Portfolios are an effective form of alternative assessment that encourages students and educators to examine skills that may not be otherwise accessed using traditional means such as higher order thinking, communications, and collaborative abilities*". According to (Chun, 2002), portfolios represent the highest point of students' learning, what they collect, assemble and reflect on samples are represented in their portfolios. As cited in (Buzzetto-More & Alade, 2006), Cooper (1999) identified six considerations of the portfolio building process: "*identification of skill areas, design of measurable outcomes, identification of learning strategies, identification of performance indicators, collection of evidence, and assessment*". Examples for e-Pertfolois include: TaskStream, LiveTech, TK20, Foliotek, FolioLive, ePortfolio, TrueOutcomes, and SpringBoard (Buzzetto-More & Alade, 2006)

The evolvement in software technology influences the technology used in developing e-assessment tools. Learning tools in general and e-assessment in particular evolved from being monolithic and used as stand-alone applications to be more flexible based on services with clear interfaces. Futuristic e-assessment tools will adopt the service-oriented architecture paradigm to modularize contextualized assessment service capable to flexibly support different pedagogical approaches and learning types - e.g. self-directed learning, collaborative learning, and game-based learning (Al-Smadi & Gütl, 2010; Daggerel.al. , 2007; Millard et al, 2005). However, this will be faced with the challenge of accessibility of these tools and services to be used in institutional learning environments (Douce, Livingstone, & Orwell, 2005). Therefore, specifications and standards should be widely used to guide the development, and represent e-assessment content, tools, and services (AL-Smadi, Guetl, & Helic, 2009) (see chapter 4). This will guaranty acceptable levels of durability, scalability, affordability, interoperability, reusability, manageability, and accessibility of e-assessment tools.

Assessment systems varies according to their purpose or the context they used in. the next section gives insights from literature for the use of computers to assist assessment in different contexts such as: peer-assessment, assessment of essays and short free-text answers, assessment in CSCL in particular the use of wikis to support collaborative writing, and Game-based learning.

Computer-based Peer-Assessment

The use of computers to support peer-assessment activities has gained more interest as peer-assessment considers students as an important part of the assessment process, in which students are not only treated as 'assessee' but also as 'assessor' where students and tutors work

[6] http://sourceforge.net/projects/cable/ ; http://www.cable.bham.ac.uk

[7] http://www.maplesoft.com/products/mapleta/

[8] http://www.maths.ed.ac.uk/wallis

together collaboratively on the assessment model (Orsmond, 2004). This manner of working together in peer-assessment mode may decrease staff load and time spent on the assessment process. This collaborative nature may also develop certain skills such as communication skills, self-evaluation skills, observation skills and self-criticism for the students (Dochy & McDowell, 1997).

One of the first systems with the peer-assessment methods was a tool for collaborative learning and nursing education based on multi-user database, which was called Many Using and Creating Hypermedia (MUCH). In the same period a Macintosh application was developed which included a peer-review process for an assignment to be reviewed by two peers (Rada, Acquah, Baker, & Ramsey, 1993; Gehringer, 2000). In the late 1990's, Network Peer Assessment System (NetPeas) was implemented and Artificial Intelligence (AI) was used to develop the tool of Peer ISM that combined human reviewing with artificial ones (Bull, Brna, Critchley, Davie, & Holzherr, 1999; Gehringer, 2000; Tsai, Lin & Yuan, 2002).

Computer-assisted peer-assessment systems were also affected by the revolution of the Web as several web-based systems were developed. An example of the first reported web-based system was a web-based tool for collaborative hypertext authoring and assessment via e-mail (Downing & Brown, 1997). Other systems such as a web-based system for group contributions on engineering design projects (Eschenbach & Mesmer, 1998), the Calibrated Peer Review (CPR) was introduced in 1999 (Carlson & Berry, 2005), the Peer Grader (PG) as a web-based peer evaluation system (Gehringer, 2000), the Self and Peer Assessment Resource Kit (SPARK) which is an open-source system designed to facilitate the self and peer assessment of groups (Freeman & McKenzie, 2002). The computerized Assessment by Peers (CAP) is another example (Davies, 2003). Further examples, OASIS has automated handling for multiple-choice answers and peer assessment for free-text answers, the Online Peer Assessment System (OPAS) is another example of a system that has the abilities to upload and review assignments as well as group management and discussions (Trahasch, 2004). An improvement for this system was the introduction of a Web-based Self and Peer Assessment (Web-SPA). This system included methods of scoring and the recording of the workflow of the assessment process (Sung, Chang, Chiou, & Hou, 2005).

Recent examples of peer-assessment developments are the enhanced open-source implementation of WebPA system which was originally developed in 1998 (WebPA, 2009), as well as the Comprehensive Assessment of Team Member Effectiveness (CATME) system which assesses the effectiveness of team members' contributions (Ohland, Pomeranz, & Feinstein, 2006).

Automated Essay and Free-text Answers Grading

This section focuses on the rationales for using computers to in the domains of automated essay grading (AEG) and the grading of free-text answers. However some tools and systems have been used for both of them, the boundaries between them are still distinguishable. Some of the systems discussed in this section are used to evaluate one of these domains while others are used for both of them.

According to Grondlund (1985), students should be capable to express themselves in writing. Therefore, writing essays is an important activity in higher education. Essays are subjective in their nature which leads to a variance of their grads provided by humans. This variance is considered by students to be unfair. Computers are free of judgments myths, false believe and value biases (Streeter et al, 2003). Furthermore, the assessment process factors such as, reproducibility, consistency, tractability, item specification, granularity, objectivity, reliability and efficiency can be improved by using computers as grading tools (Williamson et al, 1999). According to Valenti et al. (2003), AEG can be used to face this problem; it is at least consistent how the grader marks the essay. As well as, an essay marking is a time consuming activity, therefore it is recommended to use computer based techniques to handle this activity (Mason & Grove-Stephensen, 2002). The rapid increase in the number students supervised by the same staff is one of the practical rationales of using CAA/CBA. Furthermore, e-learning provides new possibilities and new modern learning settings which forms another motivation for e-assessment in general. Therefore, it is recommended by researchers that using CAA/CBA for free text answers in higher education will reduces the time, costs and efforts devoted by teachers to mark essays.

Rather than the fairness and the efficiency, computers can be used to improve the learning process. The useful interaction between the student and the assessment tool encourages students for further progress and learn. Once it is possible for the student to engage in an interesting and valuable learning and assessment activity, and get fast response marking and valuable feedback, it will develop their writing skills as well as help them achieving the learning goals. Feedback provision is one of the major motivators for CAA/CBA for essays and free text answers. Computerized marking avoids the problem of grading variance done by humans, make the process faster and provides immediate and valuable feedback to the students (Conlon, 1986).

Benefits such as developing effective instructional materials, plagiarism detection and challenging students are also some rationales to adopt AEG and e-assessment in general for higher education. According to Hearst (2000) AEG could be used to improve reading, writing and other communication capabilities of the students. AEG system can detect plagiarism in students' answers more easily than human experts (Palmer, Williams, & Dreher, 2002). One of the critics of e-assessment is the possibility of fooling the machine. According to (Dessus, Lemaire, & Vernier, 2000) students that are capable to fool the machine are the ones with good knowledge and skills of the domain so, they deserve the score.

Essays and free-text answers assessment has been a field of interest during the last 50 years. Several applications based on different techniques have been developed during this period. Page (1994), has distinguished between assessing content and style of the answer. Content refers to the answer's body, what does the answer say, where style is related to the syntax and mechanics of writing. As cited in (Valenti, Neri, & Cucchiarelli, 2003) in order to grade a free-text answer, both content and style are important (Christie, 2003). However, there are different techniques that only assess one of them during their grading process. Free-text answers scorers can be classified according to the technique they utilize. Some of these systems are based on natural language processing methods (NLP), where some of them are based on statistical methods. A combination between both methods can be found in other systems. As cited in (Pérez-Marín, Pascual-Nieto, and Rodríguez, 2009) another classification can be found in Chung and O'Neill (1997), where such systems are classified into systems that depend on documents classification, systems of this category are multilingual and do not perform any linguistics processes. The other category is the systems that assess the text meaning where a semantic, morphological and/or syntactic analysis is performed.

Several research publications have elaborated the history of using computers in AEG and to assess short free-text answers (Wresch, 1993; Whittingdon and Hunt, 1999; Hearst, 2000; Darus and Stapa, 2001; Williams, 2001; Valenti, Neri and Cucchiarelli, 2003; Gütl, 2008;

Pérez-Marín, Pascual-Nieto, and Rodríguez, 2009; Karanikolas, 2010). The use of computers to assess free text answers goes back to the 1960s where a pioneer system Project Essay Grader (PEG) was developed (Page, 1994). PEG was based on the deployment of the computers statistical capabilities in the process of textual features detection. Page identified some variables related to the text features such as, "word length, essay length in words, number of commas and number of uncommon words". Page also believed that some of these features could not be directly extracted by computers but they could be approximated and he referred to them by "proxes", and termed the ones evaluated by human raters as "trins" (Page, 1994). According to Wresch (1993), most of the teachers did not know that there was software for automatic assessment of students' essays at that time. In the 1970s, Slotnick and Finn had some improvements in the AEG arena. Slotnick used Page's approach with little changes in identifying "trins" and "proxes", while Finn evaluated the correlation between the low frequency words and the writing quality (Wresch, 1993).

In the 1980s, there has been more interest in providing feedback to the students about their essays. Two main tools had been developed for this purpose, The Writer's Workbench tool (WWB) which was developed by AT&T was used to evaluate students writing abilities in terms of "spelling, diction and readability" (Kukich, 2000). The other one was the Writer's Helper (WH) developed by Conduit for writing evaluation with reference to "word frequency, sentence variety, and transition word and paragraph development".

The 1990s was influenced by the ideas of the 1980s (Wresch, 1993). Two efforts were made to advance the free text answers evaluation research. The first one was the Hal Hellwig's tool to grade business writing by using the idea of Semantic Differential Scale (SDS). Set of 1,000 commonly used words have been used to construct the scale for evaluating the writing quality. The second effort which is based on the Hellwig's one was the Alaska Assessment Project. The system was based on textual features detection and variable lists building. An expansion to the variables' lists used by Page's system with two additional readability indexes. "Fogg readability" and "Flesch readability" indexes had been used to in the process of reading level determination. According to Wresch (1993), the project had better results than Page's PEG, with a higher correlation between the system score and the human rater's one.

Newbold (1990) stressed on the importance of using computers for AEG, but with new techniques rather than the ones used for style grading. The deployment of other techniques such as Natural language processing (NLP) and Information Retrieval (IR) has motivated the researchers to develop new ideas. According to Pérez-Marín, Pascual-Nieto, and Rodríguez, (2009) in 1997, Page's system has become commercially available. Three new systems were introduced in the same year. The Intelligent Essay Assessor (IEA), which was developed at Colorado University in USA to assess the content of the students' essays via a Latent Semantic Analysis (LSA) (Foltz, Laham, & Landauer, 1999). E-rater, which is an enhanced version of the Educational Testing Service I (ETS I) combines between NLP and statistical techniques to measure the organization and the sentence structure rather than essay content (Burstein et al., 1998). The Vantage Learning Technologies, which is an American company developed a new system to assess both the style and the content. This system is based on Artificial Intelligence (AI) approach and called IntelliMetric (Vantage Learning Tech., 2000). A year later ETS developed a new system for content grading and they called it C-rater (Burstein, Leacock, & Swartz, 2001). Since 1999, E-rater has been used in the GMAT exam. Two Years later, ETS invested over a million dollars in the Criterion project to produce the Criterion 1.0 web interface, which is based on E-rater. In 2002, Criterion 1.2 has been integrated with Critique and Criterion 2.0 was presented soon later. Over 200 institutions have purchased the system to have approximately 50.000 users that time.

In 1998, Larkey (1998) presented a new system that depends on text categorization techniques, text complexity features and linear regression methods to automatically grade essays. A year later, the Schema Extract Analyze and Report (SEAR) was presented by Christie (1999). SEAR uses pattern matching techniques to automatically grade the essays content. In 2000, Apex Assessor was developed by Dessus, Lemaire and Vernier (2000). The system is similar to IEA where both of them are based on LSA. In the same year Ming, Mikhailov and Kuan (2000) created IEMS based on the Indextron technique (Mikhailov, 1998). A year later the Automated Text Marker (ATM) was developed at the University of Portsmouth (UK) (Callear et al., 2001). The system looks for concepts in the text and their dependencies with two independent scores, one for the content and the other for the style.

In 2002, several systems were presented. Automark is based on deploying NLP techniques to perform an intelligent based search of answers with reference to a predefined scheme of answers. The scheme is a set of answers that were marked by computers (Mitchell, Russel, Broomhead, & Aldridge, 2002). Lütticke (2006) discussed an approach that uses semantic network to map candidate answers, assess the answers against a model of correct ones, identifies wrong and incomplete answers, and provides feedback in natural language. According to Lütticke (2006) the approach has been successfully used to assess free-text answers since 2002. Rudner and Liang (2002), developed another system called Bayesian Essay Test Scoring sYstem (BETSY), based on statistical analyses. In the same year, the Paperless School free text Marking Engine (PS-ME) was developed by Mason and Grove-Stephenson (2002). PS-ME uses Bloom's taxonomy (Bloom, 1956) and NLP to assess the answers.

In 2003, Auto-marking which is based on NLP and pattern matching methods was presented (Sukkarieh, Pulmand, & Raikes, 2003). In the same year CarmelTC was presented by Rośe, Roque, and VanLehn (2003) to grade students' writing based on machine learning classification methods. CarmelTC is uses a rule-learning text classification method, and it combines results from syntactic functional analyses of text with "bag of words" classification approach. Moreover, the research of Mitchell, Aldridge, and Broomhead (2003) where they used the Intelligent Assessment Technologies (IAT) a commercial assessment engine to conduct a "progress test" in the Medical School at the University of Dundee in 2003. IAT employs NLP techniques to assess candidate answers against pre-defined computerized mark scheme template of answers. In 2004, Williams and Dreher (2004) developed a system at Curtin University of Technology. They called it MarkIT which is underpinned by NLP and pattern matching techniques.

E-Examiner (Gütl, 2008) is an example for automatically grading short free-text answers; the system was developed at Graz University of Technology in the year 2007. E-Examiner is web-based and uses a hybrid approach built on a natural language pre-processing chain and based on ROUGE (Recall-Oriented Understudy for Gisting Evaluation) (Lin, 2004) characteristics. ROUGE defines a set of statistical measures to automatically determine the quality of a summary by comparing it with reference summaries.

According to Hearst (2000) using computers to assess free-text answers can support the educational community with effective instructional material for improving reading, writing, and communications abilities. As e-assessment main rationales are to save time and costs, to reduce staff workload, and to provide valuable feedback, the scoring process of free-text answers should be automated and integrated to the learning process (Valenti, Neri, & Cucchiarelli, 2003).

Automated Test Item Creation

Automated test item creation from textual learning material has raised the interest of the community for quite a while; however research results and products in the past were quite limited and basic. This section gives insights from literature of recent work on different approaches.

One of the first approaches to automatically create test items is the automatic question generation system (AUTOQUEST) (Wolfe, 1976). AUTOQUEST uses natural language processing (NLP) techniques in particular pattern-matching approach to syntactically generate questions from textual material. The learning material is delivered to students per paragraph and each paragraph is used to generate questions to the student. According to the student answers the next paragraph is delivered or the same one is repeated with different questions. The system uses a dictionary of articles which covers 1700 verbs terms to identify the verb from the sentence and generate syntactically questions using pattern-matching approach.

Another approach is the REAP system (Brown, Frishkoff, & Eskenazi, 2005) for automatic generation of questions to assess English language vocabulary. The system generates 6 types of questions namely: definition, synonym, antonym, hypernym, hyponym, and cloze questions. REAP uses the English lexical resource WorldNet synsets in questions generation as a synset may have relations with other synsets based on synonym, antonym, hypernym, hyponym, or syntactic/semantic relations. Part of speech (POS) annotation is used to annotate the textual material and then used with specific approaches with WorldNet to get synsets of words which are used to generate vocabulary questions.

In same context of language learning, ArikIturri is a system to automatically create questions for Basque language assessment (Aldabe, de Lacalle, Maritxalar, Martinez, & Uria, 2006). ArikIturri uses corpora represented in XML mark-up language for the Basque language and NLP techniques to generate fill-in-the-blank, word formation, multiple choice, and error correction question types. Questions are represented in XML markup language and can be imported to assessment systems in order to be provided as tests to students.

Stanescu, Spahiu, Ion, and Spahiu (2008) introduce a semi-automatic approach for test items creation based on textual material. The approach is basically requires teacher to provide tags (e.g. <what is>, or <Define>) to the learning material. Each tag is associated to class of questions with a similar type. The system uses three steps to generate questions: (1) tags definition done by the teacher in an interactive way (e.g. <what is>), (2) defining question templates for a specific tag (e.g. "What is a/an #"), (3) parsing the text to generate questions, the sign '#' is replaced with the word or phrase tagged by the teacher with the tag '<what is>'.

The authors in (Wang, Hao, & Liu, 2008) present an automatic questions generator based on medical textual material. The approach uses questions templates to generate questions based on medical terms such as 'Disease', and 'Symptom'. The approach uses a medical parser named MMTx to parse the medical text and extract the medical terms with respect to the Unified Medical Language System (UMLS). The extracted concepts are then used to provide semantic interpretation of the medical sentence. Related templates and the semantic representations are used to generate questions automatically. According to the authors the automatically created questions are just factual and with less quality than the manually created ones. Moreover, its time consuming to parse the medical articles to semantically interpretate the medical concepts and use them to compare with available question templates to generate questions based on most suitable template.

The authors in (Papasalouros Kotis, & Kanaris, 2008; Papasalouros Kotis, & Kanaris, 2011) describe an ontology-based approach and prototype to automatically create multiple choice test items. The domain ontology is represented in the OWL format which is a standard Web ontology language based on description logic knowledge representation formalism. The concrete structure of the ontologies applied for question creations is compiled of concepts or classes which can have different relationships or properties, also known as roles. For creating multiple choice questions (MCQ) distractors are automatically created based on two strategies: class-based strategies take advantages of so called individuals in hierarchic structures which are members of classes (is-a relationships); correct distractors are created by actual 'is-a' relationships and wrong one by individuals not member of a certain class. Property-based strategies take advantages of properties and roles which describe relationships between individuals in a given ontology; in general a property has a so-called valid domain which specifies the member of individual which a certain property can be applied, a range which describes the valid values. Both information of domain and range can be used to create wrong or correct distractors. The prototype implementation focuses on one type of question which is find the correct sentences.

The authors in (Sanz-Lobera, González Roig, & González Requena, 2011) have proposed a parametric approach to create variants of exercises. In this parametric approach, mathematic formulas and models are the base for distractors of multiple choice test items. The application domain of the created test items are engineering and physics topics. The applied methodology consists of: (a) *question parameterization* defines the variable values and the ranges of variation; (b) *parametric resolution* executes the solution of all parameters defined in (a); (c) *alternative generation* selects different variants which may include multiple correct and/or wrong answers; (d) *questionnaires creation and maintenance* creates and manages the actual test items by combining text and computed values of variables; (e) *results spreading and evaluation* concerns the actual assessment activities. In this approach, multiple choice questions are created by means of templates.

The AEGIS system (Mine, Suganuma, & Shoudai, 2000) creates automatically test items from annotated documents. This system can create multiple choice exercises, fill-the-gap questions and error-correcting questions based on tagged learning content. The teachers can add tags in the learning content to indicate the chunk of content to be a potential test item. Teachers also can define one or more hidden regions which will be used to create a fill-the-gap exercise or can add candidate list to create multiple choice or error-correction answers. The AEGIS system can import such tagged learning content, extracts potential content, creates automatically test items, administers online tests, and provides results and feedback to the students.

The authors in (Goto et al., 2010) describe an approach and prototype to automatically create multiple choice questions from English textual material for native or foreign language assessment. The learning/assessment environment is designed to receive texts from students and creates based on that test items. The approach applies machine learning techniques (preference learning) to extract potential sentences, estimates blank parts based on the discriminative model (conditional random field), and creates distractors based on statistical patterns of existing questions. The data and process flow within the system is: textual input is tagged by a part of speech tagger, followed by the three above mentioned process steps, and finally a number of candidate distractors are selected and test items are created.

The authors in (Cubric & Tosic, 2010) extend an existing approach using ontologies to automatically create test items (e.g. (Papasalouros Kotis, & Kanaris, 2008) discussed above) by the two following interesting aspects: a meta ontology to model and creating different question types, and a semantic interpretation on question types and respective levels based on the Bloom's taxonomy. Finally, the approach applies question template for the test item creation process. The described approach make use of concepts and their "is-a" relationships only, a proof of concept is available as *Protégé* plugin.

Heilman (2011) introduces an approach which focuses on the automatic creation of factual questions based on an unseen input text. The goal is to create questions for assessing a reader's or student's knowledge of information in the text. The approach is composed of three stages: (1) *natural language processing transformations* are applied to transform a sentence or a set of sentences into a simpler declarative statement. (2) *The question transducer* component turns the simplified declarative sentences into a set of questions by executing a series of well-defined syntactic transformations. (3) *The question ranker* module scores the created candidate questions according to features of the source sentences, question type and transformation rules applied in the creation process. The output is a list of open-ended factual questions.

The Automatic Question Creator (AQC) (Gütl, Lankmayr, Weinhofer, & Höfler, 2011) is a tool for semi-automated (interactive) and fully automated creation of various types of test items from learning content. AQC supports various learning scenarios, and can be used as a tool (stand-alone ore integrated in a LMS) or as a service. The tool can support test item creation (interactive mode) in self-directed learning (fully automated) and supports the creation of several assessment types (multiple choice, true- false, fill-in-the-blank, etc.) (see (Al-Smadi & Guetl, 2011)). The architecture of AQC has three main modules: The p*reprocessing* module deals with format conversion, text cleaning methods and transformation into an internal XML schema which contains all necessary data for further processing. The *concept extraction* module performs structural, statistical and semantic analysis, runs term weighting and finally extracts the most suitable phrases. The q*uestion creation* module determines the most appropriate sentence for each phrase and adds the previous and the following sentences to provide sufficient context information. Moreover the module identifies distractors and antonyms, creates question items and reference answers, and finally transforms those items in QTI standard.

CAA in Computer-Supported Collaborative Learning

Computer-supported collaborative learning (CSCL) is an emerging learning science concerned with studying how people learn together with the support of computers (Stahl et. al., 2006). The emergence of Web 2.0 has supported CSCL with a variety of collaboration tools and software. Examples of such tools include discussion forums, blogs, wikis, social networks, VOIP, and virtual words (Elliott, 2008; Crisp, 2007). According to Elliott (2008), CSCL is pedagogically rooted to the social constructivist theories (Vygotsky, 1978) and the experiential learning theories (*cf.* Kolb, 1984; see Section 2.2). Learning is taking place through students' interactions with others, with text and content, as well as with teachers. Such interactions support students to build and construct knowledge in a collaborative way (Murphy, 1994; Elliott, 2008). CSCL provides an environment for social negotiation and discussion where students are encouraged to reflect on other responses in a way to facilitate collaborative construction (Jonassen, 1994; Huang, 2002). Moreover, it improves student's interpersonal skills and social skills by providing tools and software that might overcome the barriers of students' interactions and facilitate the reflection and knowledge construction (Huang, 2002).

Despite the pedagogical advantages of CSCL, integrating CSCL activities within a course influences the assessment forms and procedures (Knight, 1995; Macdonald 2003). Learning activities that encompass assessment tasks attract student's attention (Macdonald, 2003; Reimann & Kay, 2010). Assessment of CSCL activities lacks the recognition of individual's effort within the group. Traditional work group lacks the fairness quality in general (Elliott, 2008). Computers are capable of recording all the individuals' interactions within the group work which facilitates the assessment of individual's contribution by gathering their interactions and analyzing them in comparison to assessment criteria. Assessment of CSCL is a challenging task where relationships between group partners have to be considered, their performance within the group has to be evaluated, as well as affective aspects such as motivation and self-confidence are measured (Macdonald, 2003). Macdonald (2003) has distinguished between two means of assessing CSCL activities: *product* where the output of the group work has to be evaluated against the CSCL learning objectives, *process* where individual's performance and contributions are measured against the CSCL learning activities, or *both.* Moreover, Macdonald (2003) has suggested the following guidelines for the assessment of CSCL:

- CSCL activities have to be linked to assessment procedures. Using a series of linked assessments can support the skills development during the CSCL activities.

- If the CSCL activities have designed to improve the students IT and interaction skills then their practice has to be assessed within the CSCL activity assessment.

- The development of skills during the CSCL activity and their impacts on the course content, have to be covered by the course objectives and weighted appropriately in the assessment.

- The CSCL activity product has not to be assessed rather than, it has to be subject of peer-review where individuals can improve their peer-review skills as well as their management and negotiation skills.

Learning in general is concerned with *cognition* and *affect* aspects where cognition is concerned with skills and processes such as thinking and problem solving, whereas affect is concerned with emotional areas such as motivation, attitudes, and feelings (see Chapter 2). Affect is very important to the CSCL activities, aspects such as motivation and emotional state may influence the level of knowledge acquisition (Jones & Issrof, 2005). The constant and fast processing and feedback provision of qualitative and quantitative interactions within the CSCL activity as well as their systematic analysis may positively affect the motivation, emotional state, and problem-solving abilities of learners and thus may enhance their knowledge acquisition (Zumbach, Hillers, & Reimann, 2003; Daradoumis, Martínez Monés, & Xhafa 2006; Caballé, Daradoumis, & Xhafa, 2008).

Providing feedback using visualization aspects - textual and graphical - have been recommended as a possible solution in order to support CSCL in both the collaborative learning process itself and group learning scaffolding (Janssen *et al.*, 2007; Zumbach & Reimann, 2003; Reimann & Kay, 2010). Designing a suitable visualization highly depends on the following: *what* information it will visualize: CSCL related information can be either task-related (e.g., How many problems have been solved by the group?) or social-related (e.g., How many messages have been sent by each group member, or how much each group member have contributed to the CSCL product?) or both. Moreover, selecting information related to the aforementioned production function, member-support, and group well-being functions (McGrath, 1991; Zumbach & Reimann, 2003); *why* is it important to visualize those selected information; and finally *how* those information will be visualized: regarding this question possible visualization can be textual representations (e.g. tables or hints) or graphical representations (e.g. graphs and charts) or a combination of both. However, visualizations have to be carefully selected and designed so that group members and easily perceive and interpret them correctly (Keller & Tergan, 2005). Furthermore, visualization aspects in CSCL can be used to scaffold task/social group activities in such a way to foster them to provide evidence for the assessment process (Reimann & Kay, 2010).

The literature shows that learning activities linked to assessment more attracts students and increase their motivation (Macdonald, 2003; Reimann & Kay, 2010). According to Reimann & Kay (2010) assessment has not been in the focus of research on computer-supported interaction analysis. Moreover, they argued that "*Unfortunately, what students do in the course of their collaboration with peers does not relate to how they are assessed, and the outcomes of assessment rarely affect what they will do next*" (Reimann & Kay, 2010, p. 184). However, the use of computers in collaborative learning activities supports with logging and tracking individuals' interactions within the group work, the extraction of valid assessment evidences out of those log files is a challenging task. Therefore, alternative forms of assessment -e.g. automated assessment, peer-assessment, rubric-based assessment - are required to evaluate individual/group progress.

According to (Reimann & Kay, 2010), assessing group work automatically is challenging; however it can be done when group artifact has formal semantics. For instance expert solutions can be used calculate the similarity between the concept's map extracted from the group artifact (e.g. wiki page) with a reference one extracted from a reference text. Moreover, the authors argue that "*assessing group performance requires normative reference models of what constitutes "good teamwork", what processes characterize a good software team*". For instance the relationship between the 'Student Model' and the 'Task Model' in the Evidence-centered Assessment Design (Mislevy, Steinberg, & Almond, 1999), where this relationship is maintained by an evidence model that determines which of the students interactions to register and how to use the registered interactions to update the student model (see Section 3.2.3). In order to make this feasible a detailed understanding and representation of the task model should be available. However, Reimann, Frerejean, & Thompson (2009) propose an approach by which the student model can be updated based on a graphical model of team practices. The research discusses how transition diagrams can be used to formalize a graph of team decision making process automatically identified from the observations (even logs) and can be used as basis for formative and summative assessment.

Nevertheless, the literature of CSCL assessment shows that peer-assessment has been usually used to assess the collaborative learning processes (*cf.* Crisp, 2007). Group partners can evaluate their interactions and contribution using peer-assessment activity. Process assessment of the CSCL activities can be performed as peer-assessment in a formative way where individual's performance is assessed by their peers and valuable feedback can be provided out of this assessment. Examples of peer-assessment tools that can be used for group-work assessment are Web-SPA (Sung et al., 2005), and Self and Peer Assessment Resource Kit (SPARK). SPARK is open-source assessment software designed to facilitate group work assessment (Freeman et al., 2002) for more examples you can refer to section about using computers in peer-assessment discussed earlier.

In order to more discuss how e-assessment is supporting CSCL, two main domains namely, assessment in online discussions and assessment in wiki-based CSCL have been explored in the next sub-sections.

Assessment in Online Discussion

Discussion forums (also referred to an online discussion) are forms of computer mediated communication (CMC) where students and teachers can interact in asynchronous manner. The assessment of online discussion posts can be either formative or summative; the assessment criteria should reflect the task goals and consider both qualitative and quantitative participations (Caballé, et. al., 2007; Crisp, 2007). A possible assessment criterion can be based on content understanding, participation rate, and participation quality. Feedback is important especially in the early stages of the discussion which may support students to get a better understanding of the task and the content. Teachers may take passive or active role in the discussion, and this will be based on the task nature, the task objectives, and the level of scaffolding as well as the learning outcomes form group participation. However, the teacher or e-moderator, has to provide clear guidelines regarding the number, size, type of posts, the type of content and language, whether literature referencing is required or not, the deadlines for initial and final entries (Salmon, 2000; Crisp, 2007: 194).

In order to have a quality online discussion, the online discussion activities have to be linked to assessment (Swan et al., 2006). However, the question of how to assess online discussion activities remains challenging. Possible solutions may include analyzing the content of the online discussion or by using assessment rubrics. Examples from literature for content analysis are the use of grounded theory and theoretical codes (Glaser, 2005), content analysis and ethnography (Stemler, 2001), and content analysis by categorizing the discussion post (Garrison et al., 2001; Hara et al., 2002). Such approach of content categorization supports the teacher to categorize each discussion post and assess knowledge construction, critical thinking, and how students use others posts to build their own contributions (Crisp, 2007). Examples for using rubrics to assess online discussions can be found in, the work of Baron and Keller (2003), Pelz (2004), (Ho, 2004).

Assessment in Wiki-based CSCL

Wikis are websites that can be authored in a mass collaboration of users, where they are capable to add, edit, delete, and rollback to previous versions of the wiki-page. Moreover, wikis provide features such as, e-mail and Really Simple Syndication (RSS) notifications of page edits as well as pre-, post-comments of page content (Judd et al., 2010). Wikis are also easy to use, have pedagogical advantages, and they provide students with the so called "structured bulletin board" (Leuf & Cunningham, 2000). Students can reflect and receive feedback and wikis also support different learning styles as they form as "inherently democratic medium" (Leuf & Cunningham, 2001) Wikis also provide students with 24/7 interaction medium, facilitate knowledge acquisition, and prepare them to be more than readers and writers but also editors and reviewers (Cubric, 2007; Judd et al., 2010).

Several researches have been conducted to investigate the validity of using wiki systems in CSCL. Despite that wiki constitutes from semiotic contributions, wiki plays an interesting double role of medium and product of the collaborating (Reimann & Kay, 2010). Wikis prevent users from editing the same page simultaneously which may be a disadvantage in some scenarios like using wikis for co-writing. However, this may be avoided in distance learning as the probability of simultaneous editing for the same page is less than in-campus learning. Wikis are designed to log all the users' edits and comments, with the ability of page editing notifications (e-mail, RSS). Such ability of automatically logging users' contributions and activities can be used to analyze and interpret the nature, scope, context of user contributions (Swan et al., 2006; Trentin, 2009; Judd et al., 2010).

Despite the technical and pedagogical advantages of wikis, additional work is required to promote collaboration and participation among students (Judd et al., 2010). According to Ebner el al (2008) and Cole (2009) if a piece of work and tool (e.g. wiki page) are assigned to students without the work being assessed, the students will not voluntarily edit or create new pages. This goes in line with literature where CSCL activities have to be linked with assessment activities in order to promote students contributions and participation (Macdonald, 2003). The effective use of wikis in CSCL lacks incentives such as assessment and support of group work (AL-Smadi, Hoefler, & Guetl, 2011b; Judd et al., 2010). The remaining parts of this section give insights from literature on e-assessment integrated with wiki-based CSCL activities.

In the work of Trentin (2009), the author tested an approach for co-writing using wiki. As part of that approach the students used online discussion forum for co-planning and structuring the content for the co-writing phase. Moreover, they used online discussion forum for peer-review where they were required to peer-review their peers contributions and writings. Wiki had been used for the co-writing activities. The student's collaborative activities had been evaluated according to the product of co-writing, the process implemented by groups, and the learning of the subject content. Within the process evaluation, the objective (number of messages and amount of produced material) and subjective (teachers and peers evaluation) data extracted from the wiki logs and discussion forum posts analysis were used to evaluate the co-writing process. 3D graphic projections had been used to visualize both the interaction among participants and among the links between the hypertext pages. Moreover, network analysis techniques had been used to represent the reticular relationships among those interactions.

Reimann et al. (2010) proposes an assessment approach for team practices in CSCL - in particular wiki-based collaboration - based on formal process model represented as a transition diagram. Such process model can be formalized automatically based on tracing student's behaviour (log file). Moreover, it can be used to provide feedback - in terms of visualized knowledge, and summative valuation - by comparing created graphs with an optimum one.

Another example of providing assessment activities within wiki-based learning is the work of (Kumar, Gress, Hadwin, & Winne, 2010), in this research the authors discuss an ontological approach to perform assessment in the process assessment within CSCL activities. In this approach the learners interactions are tracked into XML-based log file and mapped onto ontology they named CILT. CILT covers four main domains namely: *content* (refers to the learning content related to the application domain), *interaction* (defines interactions the learner can take within an application), *learner* (identifies the learner knowledge state, skills, and preferences), and *time* (time information imported from DAML-Time ontology). The whole approach has been designed to support students in self-regulated learning trend, to maintain flexibility among different application domains, and to provide sharable and Interoperable framework.

The work of (Khandaker & Soh, 2010) in which the authors tracked and learners' interactions with a wiki designed for education they named "ClassroomWiki". According to the authors, ClassroomWiki assessment approach: (1) tracks students' interactions and textual contributions, (2) evaluation of concept-based contribution, (3) evaluates peer-ratings towards group progress. Moreover, their first findings shows that teachers were capable to better evaluate the individual's contribution, and supported with tools to provide timely feedback and support to the students who are not contributing to the group work- i.e. scaffold learner path.

According to (AL-Smadi, Hoefler, & Guetl, 2011b) integrating assessment forms such as self, and peer-assessment within tools used in co-writing can support students to maintain task-awareness, enhance their contribution towards the group production function, and increase their motivation and their engagement accordingly. The authors discuss their findings based on a study they have conducted using tool they have developed to collaborative writing and peer-review. Therefore, wikis should be enhanced and enriched with new forms of assessment such as self, peer-assessment so that the processes of co-writing can be peer-reviewed. Moreover, some enhanced visualization tools should be implemented to provide both students and teachers valuable feedback about the collaborative learning using wiki. The visualization tools help answering questions such as, how much has each student contributed to the assignment product? How collaboration is taking place? To what extent the students are collaborating within the group? Who did what and when?

CAA in Serious Games

Games content is very interactive and therefore can be utilized for supporting assessment. When players interact with the game they eventually take possible actions pre-defined in the game model of actions. This can be utilized to define assessment model of monitoring the player activities, logging all actions within the game session which can be used to grade the player activities within the game. The game engine has to be carefully designed so that instructors can define assessment rules for specific state transitions during the game session (Burgos, Moreno-Ger, Sierra, Fernández-Manjón, Specht, & Koper, 2008). Assessment and feedback should be seamlessly integrated into games without compromising game-play. '*Micro-adaptivity'* can be considered as an assessment technique where it is applied as non-invasively interpreting the player's (learner) behavior during the game. The level of difficulty may also be applied based on the interventions of adapted scenes (e.g. increased difficulty) (Kickmeier-Rust, Steiner, & Albert, 2009).

Serious games represent a challenging as well a rich domain for assessment practices. However, the efficacy of any assessment approach is highly related to the *target demographic*, *usage context*, *choice of technology*, and *underlying pedagogy* (de Freitas and Oliver, 2005). Hence an attempt to evaluate any assessment model typically results with lack applicability when transferred to other groups of learners, context, and educational situations. Examples of assessment trends in educational games can be found in the work of (Shute, Ventura, Bauer, & Zapata-Rivera, 2009; Kickmeier-Rust, Steiner, & Albert, 2009) to ably non-invasive assessment practices thus not to break the player flow within the game.

In the work of (Shute, Ventura, Bauer, & Zapata-Rivera, 2009) the authors discuss the evaluation of players progress within immersive games via what they called *stealth assessment* -

embedded formative assessment within the immersive game - based on extending the evidence-centered design assessment model (ECD) (see Section 3.2.3) (Mislevy, steinberg, & Almond, 2003) with an *action model* instead of *task model* which can be used by Bayesian networks to track player actions within the game and provide an evidence of progress. The aim of this research is to use what they called stealth assessment approach within immersive games to track players actions and with respect to the ECD model to provide formative and dynamic feedback thus to support students learning. However, what they discussed represents an summative approach by which they evaluates the progress of the player in terms of interactions and used it in comparison with the *evidence model* - part of ECD - to provide an evidence of learning and skills achievement.

In the work of (Kickmeier-Rust, Steiner, & Albert, 2009; Kickmeier-Rust, & Albert, 2010) the authors propose what they called *micro-adaptivity* approach for assessment in educational games. The approach has been developed in the context of the Enhanced Learning Experience and Knowledge TRAnsfer (ELEKTRA[9]) project. The ELEKTRA framework uses the Competence-based Knowledge Space Theory (CbKST) to model the competencies required by the student to achieve a learning goal. The basic idea of CbKST is to associate problems in a domain with skills in order to provide a model of competencies for a specific domain which can be used to update the knowledge and skill state of the learner in a learning domain. The ELEKTRA game and its successor 80Days[10] game tracks the player interactions and uses them to update the competence state represented in the CbKST for the learning domains provided in the games. However, according to (Kickmeier-Rust, & Albert, 2010) the approach demands extra load on authoring aspects to define all required information for the models as well as computational load as the game updates the CbKST based on each player action.

In the work of (Moreno-Ger, Burgos, Sierra, & Fernández-Manjón, 2008) the authors proposes the adventure game engine called <e-Adventure>. In <e-Adventure> assessment rules are defined within the XML model representing the game model. This model contains information about the game storyline, scenes, characters, their associated resources, as well as assessment and adaptive rules. When the player is interacting with the game an associated assessment rule is triggered based on internal flags associated with specific states of the game. Moreover, the authors claims that the <e-Adventure> is flexible to be used with educational modeling languages – i.e. IMS Learning Design (IMS LD)- to design the pedagogical impact of using assessment rules in the game engine to evaluate the progress of the players, and hence to provide personalized and adaptive digital educational games.

According to Burgos et al. (2008) in order to have an adaptive and personalized serious game within the domain of game-based learning, the game engine has to be integrated with a LMS. LMSs use the Log of player interactions within the game session to provide more personalized and adaptable content. The player flow within the game will form like a learning path where a third-party tool is needed to interact with the game engine, retrieves the player state, and communicate with LMS so the learner model can be updated as well as adaptive and personal content can be provided during the game next phases. When students play, they interact with the game by making decisions and taking right/wrong actions and paths. The

[9] http://www.elektraproject.org

[10] http://www.eightydays.eu

game platform should have the possibility to define checkpoints (assessment rules) so that to assess players interactions and decisions. Moreover, it should provide valuable feedback.

Literature reviews of feedback in digital educational games (Shute, 2008, Mory, 2004) highlight the importance of formative models for feedback provision. Dunwell, de Freitas, and Jarvis (2010) discusses the feedback aspects in digital educational games based on Rogers' (1951) classification into *evaluative* (players get a score), *interpretive* (players get a score and the wrong action), *supportive* (players get a score and guidance information), *probing* (players get a score and analysis of why the player did the wrong action), and *understanding* (players get a score and analysis of why the player did the wrong action as well as guidance for supportive steps or learning material) forms. Moreover they propose a four-dimensional approach for feedback provision in serious games. According to their approach the following aspects should be considered:

- **Type:** feedback type differs based on Roger's classification -discussed above- with respect to students, teachers, or technology thus required aspects to classify feedback – such as measure variables, their relationships model, learner model, knowledge model, and domain model- should be considered.

- **Content:** content can be classified with respect to the learning outcomes into *essential* or *desirable*.

- **Format:** the media used to represent feedback (e.g. text, image, voice, etc.).

- **Frequency:** the rate of feedback provision to students differs with respect to instructors, technology, pedagogy, and learner preferences control. Hence, feedback can be *immediate*, *delayed*, or *dynamic* based on the domain and learner action type.

3.2.3. e-Assessment Models

The questions that educators often ask about assessment are whether all assessment forms have the same assessment model and what are the common features between the assessment forms, moreover what are the aspects that should be considered when it comes to design an assessment practice.

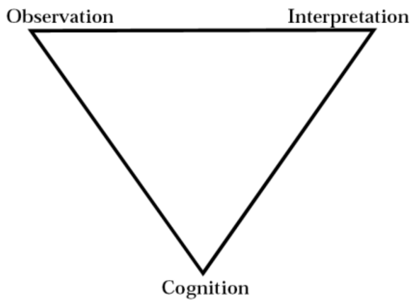


FIGURE 3.2. The assessment triangle (Pellegrino et al., 2001, p. 44).

Pellegrino, Chudowsky, and Glaser (2001) provided what they called 'assessment triangle' that discusses three key elements of assessment in general. As depicted in Figure 3.2, the first element is cognition which is a model for learning and assessment in the domain that represents how students build knowledge and develop competence. The second element is observation which represents a set of beliefs about the kinds of observations that are constructed based on situations and tasks provided to the students so they can interact with and build their knowledge and skills. Observations provide an evidence of students' competencies. The third element is interpretation which is the process of reasoning an evidence of competence achievement based on the observations.

Evidence-centered assessment design (ECD) (Almond, Steinberg, & Mislevy, 2002; Mislevy, Almond & Lukas, 2004) is a framework that explains the structures of assessment arguments, their elements and process, as well as the interrelationships among them. ECD consists of five layers as summarized in Figure 3.3.

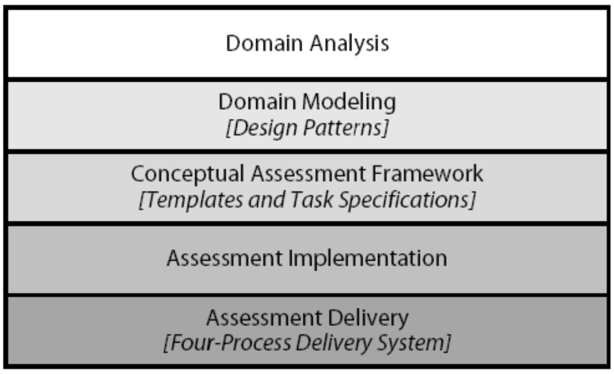


FIGURE 3.3. Graphical representation of ECD layers (Mislevy & Riconscente, 2005).

The *conceptual assessment framework* (CAF) discusses the assessment arguments sketched in *design patterns* in terms of the kinds of elements and processes required to implement an assessment that embodies those arguments. As depicted in Figure 3.4, CAF modules represent the blueprint of the operational elements of an assessment as well as their interrelationships. CAF discusses the substantial, statistical and operational aspects of assessment elements. Moreover, it covers technical details such as, specifications, operational requirements, statistical models, details of rubrics. CAF forms as an intermediate step between the output of the *domain analysis* and *domain modelling* steps - which is a framework specifying the knowledge and skills to be assessed, conditions for assessment and evaluations, as well as type of evidences to assess the provided tasks - and the *assessment implementation* step which describes the requirements for process during the assessment delivery system. CAF consists of a set of modules which provides specifications to answer critical questions such as:

- *What Are We Measuring:* ***The Student Model***
- *How Do We Measure It:* ***The Evidence Model***
- *Where Do We Measure It:* ***The Task Model***
- *How Much Do We Need to Measure:* ***The Assembly Model***
- *How Does It Look:* ***The Presentation Model***

Moreover, these models describe the requirements for the objects in the assessment delivery system. The *Delivery System Model* describes the collection of student, evidence, task, assembly, and presentation models necessary for the assessment and how they will work together.

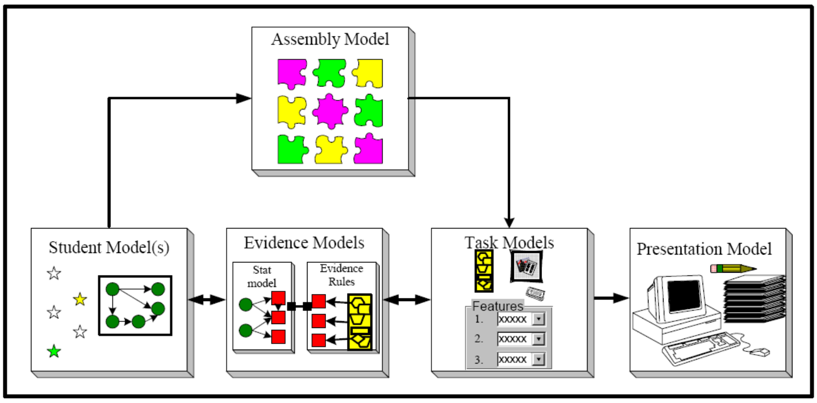


FIGURE 3.4. The Conceptual Assessment Framework (CAF) for the delivery system model (Mislevy et al, 2004).

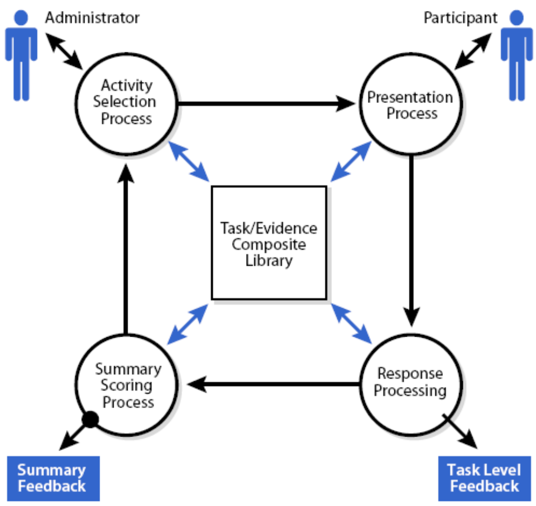


FIGURE 3.5. Four-process assessment architecture (Almond et al., 2002).

The four-process architecture (Almond et al., 2002; Crisp, 2007) discusses common features between different forms of assessment. These processes include *activity selection*, *presentation*, *response processing*, and *summary scoring*, as presented in Figure 3.5. The creation of the assessment task starts by the *activity selection process* where the administrator (instructor) selects and sequence tasks form the *task/evidence composite library* (a database of possible tasks, their description, materials, rules, and evidence parameters). Then information is sent to the *presentation process*, which delivers the assessment task to the participant (student). Relevant materials can be retrieved form the *task/evidence composite library* for instance, assessment paper (traditional assessment) or images, audio/ video files (e-assessment). The *presentation process* records the students responds as a *work product* which can be assessment paper script, or computer file and then delivers this *work product* to the *response processing* section for evaluation. The evaluation process may consist of simple scoring process or more complex series of evaluation for the students' responses. The evaluations are then passed to the *summary scoring process* which updates the *scoring record.* The scoring record contains all the judgements about students' knowledge, skills level, and abilities based on pre-defined evidences provided for all tasks. According to Almond et al. (2002), separating the response processing step from both summary scoring and presentation is vital to an evidence-based focus in assessment design and supports reuse of the task in multiple contexts. Two types of feedback can be delivered based on this architecture: *task-level feedback,* which represents the immediate feedback based on student responses independently of evidence from other tasks, and *summary feedback,* which reports the accumulated *observations* from the *scoring record* based on tasks evidences to the participant (student).

According to (Brinke et al., 2007), Almond's four process conceptual assessment framework (CAF) has a limitation as it was designed for computer-based assessment and more directed to the execution phase of assessment. Moreover, CAF views assessment as a process of two main roles participating in, an *administrator* to setup and maintain the assessment, and a *participant* (student) who's competence, skills, and knowledge are going to be assessed. As cited in (Brinke et al. 2007) any educational model for assessment has to be validated to the following requirements adapted from Koper (2001) for any complete conceptual model:

- *Pedagogical flexibility:* The assessment model can describe assessments that are based on different theories and models.

- *Formalization:* The assessment model describes assessments and its processes in such a formal way that it is machine-readable and automatic processing is possible. The formalization gives the possibility to extend the model if new developments in assessment arise.

- *Personalization*[11]*:* The assessment model describes personalization aspects within its contents and activities can be adapted based on the preferences, prior knowledge, educational needs and situational circumstances of users. Moreover, control on content and activities should be given to students, staff members, and developers as required.

- *Re-usability:* The assessment model supports identification, isolation, de-contextualization and exchange of useful objects (e.g. items, assessment units, competencies, assessment plans) and their re-use in other contexts.

- *Interoperability and sustainability:* The assessment model distinguishes the description standards from the interpretation techniques, thus making the model resistant to technical changes and conversion problems.

- *Completeness:* The assessment model covers the whole assessment process, including all the typed objects, the relations between the objects and the workflow.

- *Explicitly typed objects:* The assessment model expresses the semantic meaning of different objects within the context of an assessment.

- *Reproducibility:* The assessment model describes assessments in such a way that replicated execution is possible.

- *Medium neutrality:* The educational model for assessment, where possible, supports the use of different media, in different (publication) formats, such as computerized assessments on the web or paper and pencil tests.

---

[11] *Personalization* is discussed in the original Koper (2001) reference but not cited in Brinke et al. (2007).

- *Compatibility:* The assessment model matches available standards and specifications.

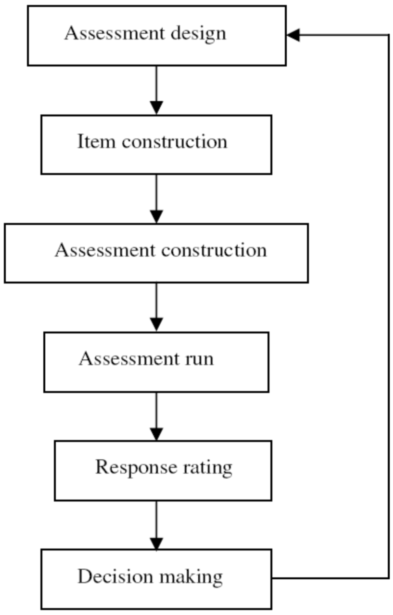


FIGURE 3.6. Main stages of the assessment process (Brinke et al., 2007).

Brinke et al. (2007) have constructed an educational model for assessment in which they covered new types of assessment. The model is designed to have different sub-models each represent a different stage in the assessment process as summarized in Figure 3.6. The model can be used to enrich the IMS Question & Test Interoperability specifications (IMS QTI, 2008) with more features especially for the 'assessment' and 'section' parts of the specification (see Section 4.2.1). Moreover it can be used to fill in the gaps between IMS QTI specifications and other related specifications such as IMS Learning Design (IMS LD, 2008) by providing directions of using both specifications to address teaching, learning, and assessment. However, the model has some limitations as it does not discuss statistical and psychometric information which are more covered in the four process model of Almond et al. (2002).

Another useful framework is the Framework Reference Model for Assessment (FREMA[12]). FREMA was the principal deliverable of the FREMA research project, which ran from April 2005 until October 2006. The project was funded by JISC (Joint Information Systems Committee) as part of its e-learning framework (E-Framework) program (Millard et al., 2005). FREMA explains and visualizes possible activities and entities related to the e-learning assessment domain. The framework uses concept maps to visualize assessment components and their interrelationships in a way to explain possible assessment services, standards, organizations, and use cases. FREMA possible resources and activities have been defined in consultation with the e-assessment community in UK (Millard et al., 2005). A useful view of 'Noun Map' and 'Verb Map' represents the related assessment resources and activities. The 'Noun Map' explains the possible assessment resources as well as stakeholders and their roles in the assessment cycle. The 'Verb Map' represents the possible processes of assessment and what people can do in the context of e-assessment.The FREMA website provides interactive Flash® components to demonstrate the assessment domain as depicted in Figure 3.7.

[12] http://www.frema.ecs.soton.ac.uk

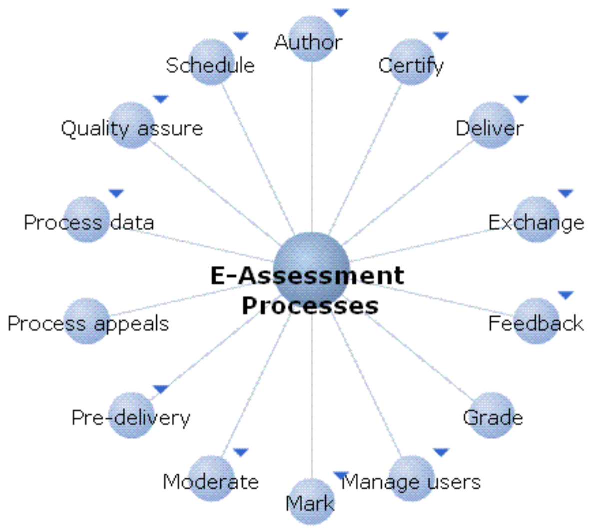


FIGURE 3.7. First level of FREMA e-assessment processes (Millard et al., 2005).

3.3. Feedback

Feedback has been considered to be a mirror for learning (Klassen, 2001). Both learners and educators can use feedback to be aware what they have done and what they have not during the course. Thus, feedback forms as a valuable tool by which learners become aware of the gaps in their knowledge, skill, or performance within a course (Boston, 2002; Garris, Ahlers, & Driskell, 2002). Ramaprasad (1983) defines feedback as "*information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way*" (p.4). Moreover, Black & Wiliam (1998) identified four elements making up a feedback system: data on the actual level of some measurable attribute, data on the reference level of that attribute, a mechanism for comparing the two levels, and generating information about the gap between the two levels, and a mechanism by which the information can be used to alter the gap.

Nicol and Macfarlane-Dick (2006) clarify two types of feedback, 'internal feedback' by which students monitor their performance in a learning task and evaluate it with respect to the task desired goal, and 'external feedback' which is provided by their teachers or their peers to scaffold students' progress towards the final goal. Moreover, they have identified based on literature seven principles for a good feedback practice, Good feedback practice:

1. helps clarify what good performance is (goals, criteria, and expected standards);
2. facilitates the development of self-assessment (reflection) in learning;
3. delivers high quality information to students about their learning;
4. encourages teacher and peer dialogue around learning;
5. encourages positive motivational beliefs and self-esteem;
6. provides opportunities to close the gap between current and desired performance;
7. provides information to teachers that can be used to help shape the teaching.

With respect to principle 3, feedback delivered to students should be carefully designed so that to support them to regulate their learning towards the desired learning goals. Wiggins (2001; cited after Nicol & Milligan, 2006) argues that quality feedback is more descriptive rather than evaluative. Feedback should provide students information about the gap they have in their knowledge and skill based on their current performance state and the desired goals, moreover it should clarify the criteria have been used to evaluate their performance. Nevertheless, Wiggins distinguishes between *feedback* – information about student interactions with learning task, learning material, learning environment, and behavioral impact on other people such as peers and teachers, *guidance* – information provided to students about what can be done to achieve their learning goals, and *evaluation* – is more about judgmental information weather students achieved the learning goals or not. Therefore, praise or blame information (such as 'good work') do not support students to maintain successful progress; instead descriptive information based on clear assessment criteria is more supportive (Nicol & Milligan, 2006; Wiggins, 2001).

The quality of the feedback is a key feature in any procedure for assessment (Black & Wiliam, 1998). Feedback is often linked to formative assessment as a mean for learning (Al-Smadi & Gütl, 2008). One of the major purposes of formative assessment is to provide valuable feedback to both students and teachers in order to support learning and teaching (Gibbs, 1999; Dochy & McDowell, 1997; Knight, 1996).

Technology has a major influence on feedback (Charman & Elms, 1998). The use of computers facilitates the process of tracking user behavior and performing assessments as well as analyzing the results (Al-Smadi & Gütl, 2008; Yeh & Lo, 2009). In (Yeh & Lo, 2009) the authors developed a computer-based tool to provide error correction and corrective feedback based on online annotations in the context of second language writing. The tool has been evaluated in real learning settings and results have shown that students who used the tool identified more errors that the ones they did not use it. Moreover, the use of computer facilitates the process of error correction and corrective feedback provision.

Feedback can be distinguished in terms of its format into *verbal* and *visual*. Visual feedback can be used to scaffold task and social group activities, thus foster them to improve their learning process (Janssen *et al.*, 2007; Reimann & Kay, 2010). Enhanced visualization tools should be implemented to provide both students and teachers valuable feedback about learning progress.

3.4. Summary

Over the last 100 years, the learning process has changed from being repetitive to a new form of learning based on understanding, independency, learners' empowerment and skills improvement. As a main part of this learning process assessment is no more considered to discriminate between students, rather than it is used to enhance students learning and encourage them for further progress and success. In the so-called 'new culture of assessment', rather than teachers are considered to be knowledge carriers which they have to transfer to students, they guide and encourage their students to get the knowledge and skills they have to learn. Students play major roles in this new culture of assessment where new and alternative forms of assessment have been adapted. Such forms of assessment include: observations, interviews, performance assessment, writing samples, exhibitions, portfolio assessment and project and product assessment. Several labels have been used for assessment such as, directed assessment, authentic assessment, performance assessment, alternative assessment, collaborative assessment and self- and peer-assessment, where different methods, goals, forms and administrations are used. (AL-Smadi, Guetl, & Chang, 2011; Bransford, Brown, & Cocking, 2004; Dochy & McDowell, 1997)

Assessment in general has three main types: (a) diagnostic assessment: which takes place at the beginning of the learning activity and usually related to diagnose learners individual needs such as, learning style, learning preferences, knowledge and skill state and background, and personality factors. Diagnostic assessment is common in the domain of educational adaptive hypermedia. (b) Formative assessment: this takes place during the learning activity and used to maintain students' learning progress and to provide quality feedback. Formative assessment is common in low stake assessment and in alternative forms of assessment such as, behavioral and performance assessment. (c) Summative assessment: this takes place at the end of the learning activity and used to judge students learning and to provide an evidence of student's success or fail. Summative assessment is often high stake assessment and common in traditional forms such as, courses comprehensive exams.

The emergence of Web 2.0 and the influence of information and communication technology (ICT) have fostered e-learning 2.0 to be more interactive, challenging, and situated. Nowadays learners use technology anywhere, anytime, and they are faced with the challenge of needing to be engaged and motivated in their learning (Prensky, 2001). Learners feel empowered when they are engaged in learning activities. Given the different learning styles and teaching strategies, educators are faced with the challenge of having to develop assessments which are required to appraise the student's learning process. Assessment forms provided in e-learning activities have to be aligned with theories of learning so that they can foster different types of learning such as reflective-learning, experiential-learning, and socio-cognitive learning (AL-Smadi, Guetl, & Chang, 2011; Dochy & McDowell, 1997; Elliott, 2008).

Designing quality assessment requires considering several aspects which include: (a) assessment domain: cognitive, affective, and psychomotor (*cf.* Bloom, 1956), (b) assessment type: diagnostic, formative, and summative assessment (*cf.* AL-Smadi & Guetl, 2008; Crisp, 2007; Bransford, Brown, & Cocking, 2004), (c) assessment strategy: traditional assessment, individual assessment, group assessment, self-assessment, peer-assessment, instructor-based assessment, and system-based assessment (*cf.* Dochy & McDowell, 1997), (d) assessment referencing: norm-related, criterion-related, or ipsative (*cf.* McAlpine, 2002), (e) assessment practice: behavioral assessment, performance assessment, portfolio assessment, and rubric-based assessment (*cf.* Buzzetto-More & Alade, 2006), (f) assessment adaptation: micro-adaptive assessment, or macro-adaptive assessment (*cf.* Kickmeier-Rust, Steiner, & Albert, 2009), (g) assessment method: quantitative or qualitative (*cf.* AL-Smadi & Guetl, 2008; Crisp, 2007; Culwin, 1998; Bloom, 1956), (h) assessment feedback: feedback type, format, frequency, and content (*cf.* Nicol, Milligan, 2006; Wiggins, 2001; Black & Wiliam, 1998; Charman, & Elms, 1998).

In order to provide quality assessment a set of assessment models have been designed. Assessment models are either general and discuss key elements for assessment in general (e.g. Chudowsky, & Glaser, 2001) or specialized and emphasize on specific aspects of the assessment process (e.g., Almond, Steinberg, & Mislevy, 2002). However, the discussed assessment models lack to some extent aspects such as: (a) pedagogical flexibility and the alignment with theories of learning, (b) the suitable assessment form for the learning activity or task, (c) available technology - in terms of systems, tools, and services, (d) standards, specifications, and guidelines of how to design, and develop assessment for the target learning practice, (e) feedback as a crucial component for quality assessment practice, (f) guidelines or framework of how to use these models to support developing learning tools with integrated assessment.

Standards and specifications as well as guidelines for designing assessment are important to have quality assessment. The next chapter discusses standards and specifications in general and e-assessment content and service specifications in particular. Nevertheless, specifications of how to represent a learning activity and learning design are illustrated.

Chapter

4

4. Standards in Modern Learning Settings

The emergence use of information and communication technologies (ICT) and web 2.0 has fostered the domain of learning platforms with a variety of learning tools. As a result many open-source or even commercial learning management systems (LMS) were developed. The variety of the platforms and approaches used in these LMSs makes it difficult to exchange information between them. Therefore, some of them have become obsolete and dedicated for specific institutions (Bizonova & Ranc, 2008). Moreover, the e-learning content has not been carefully designed. According to (Anane, Crowther, Beadle, & Theodoropoulos, 2004), e-learning content should not be electronic replications of the classroom materials, rather than it should be value-added. Careful design of the learning content with integrated presentations, exercises, and valuable evaluation and feedback, as well as flexible content navigation and usable user interfaces may motivate learners for further success. Learning content reusability and interoperability, learner's information accessibility and share ability, are main maters of quality for any LMS. e-Assessment as main part of any e-learning system also faces the same challenges and problems. Having the variety in e-assessment tools and systems (see Section 3.2.2), minority of these tools consider learning tools interoperability which makes these tools only available as stand-alone and cannot be used to extend the LMS services - based on their application domains - without further redesign and redevelopment.

This chapter aims to investigate educational standards and specifications in general and focuses on e-assessment specifications and standards in particular. To this end, this chapter explores the domain of educational standards and specifications in particular the ones for representing e-assessment content, tools, and services. Moreover, it identifies the limitations and problems in these specifications and standards when it comes to provide flexible and interoperable e-assessment.

This Chapter is based on (AL-Smadi & Guetl, 2011a; AL-Smadi & Guetl, 2011b; AL-Smadi et al., 2009b; AL-Smadi et al., 2011)

4.1. Standards and Specifications

Learning content reusability and interoperability, learner's information accessibility and share ability, are main maters of quality for any LMS. Therefore, LMS should be designed and implemented to be standard-conform. e-Assessment as an important part of any e-learning system also faces the same challenge and problem. Different standards and specifications have been developed to design and develop e-learning content and components. According to (Shepherd, 2006, p.75), "conformance or compliance testing distinguishes specifications form standards". The process of standards releasing starts with the users who contact technologies with their requirements. The technologies reply with a specification proposal where many users can use it for systems building. In order to ensure that all users interpret the specifications the same way, a conformance statement is written and all system have to stick to it. A certification process is followed in which, criteria to ensure systems compliance to the specifications has to be set, a third party has to test the systems against this criteria. This process will improve the specifications, conformance statements, and the test criteria over time to ensure compatibility. To this point, standards are not yet released. The tested and matured specifications are forwarded then to a standard committee such as IEEE Learning Technology Standardization Committee (IEEE LTSC) (IEEE LTSC, 2008), a step before the last approval from an official standards organization as ISO and ANSI to be official standards.

Specialists argue that conforming to standards during the design and development of e-learning tools in general and e-assessment in particular may foster them with the following abilities (AL-Smadi, Guetl, & Helic 2009b; Shepherd, 2006): *Durability:* no need for further redesign or redevelopment even with new versions of the system, *Scalability:* can it grow from small to large?, *Affordability:* is it affordable?, *Interoperability:* are information and services sharable with other systems?, *Reusability:* can it be used within multiple contexts?, *Manageability:* is it manageable?, and *Accessibility*: are the contents accessible and deliverable from anywhere and anytime?.

Although people in the domain of e-assessment recognize the values and importance of specifications and standards, there are a set of problems and challenges that may face them. These challenges can be defined into two main categories (Shepherd, 2006) as follows:

- *Idealists vs. Pragmatists:* this point discusses the debate between two different schools where the former one is looking for the perfect model while the second concerns about the time of having outputs. For example, Academics and long-term thinkers who belong to the former one concern about the quality of standards while, business people and salespeople may belong to the second one and concern about how sooner the standards are ready to be used.

- *Patents and Intellectual Property (IP):* releasing standards with patents works against the aim of standards of having interoperable systems, reduces the enhancements and makes an overload of paying royalties for these patents owners.

Specifications and standards can be classified according to their level of approval into the following (Devedžic, 2006): (a) Official standards: a set of definitions, requirements, formats and design guidelines for e-learning systems or their components that a recognized standards organization has documented or approved, e.g. IEEE LTSC (Learning Technology Standardization Committee), ISO/IEC JTCI (Joint Technical Committee). (b) De facto standards: the same as the official one, but accepted only by the community and industry (e.g. IMS QTI (Question and test Interoperability) (IMS QTI, 2008)). (c) Specifications: the same issues as the official standards, but less evolved; usually developed and promoted by organizations or consortia of partners from academia, industry and educational institutions, e.g. IMS Global Learning Consortium, PAPI Learner (Public and Private Information)(IEEE PAPI, 2003). (d) Reference models: an adapted and reduced version of a combination of standards and specifications focusing on architectural aspects of an e-learning system, definitions of parts of the system and their interactions - e.g. LTSA (Learning Technology Systems Architecture) (IEEE LTSA, 2008), SCORM (Sharable Courseware Object Reference Model; SCORM, 2008).

According to (Shepherd, 2006, p. 80) educational standards and specifications can be grouped into the following categories: (a) Authentication: specifications or standards on how systems can provide single-sign-on access to individuals and tools. (b) Content Packaging: specifications or standards for packaging e-learning or e-assessment content in order to provide sharable content as well as to facilitate content transmission between tools and systems. (c) Data Definitions: specification and standards that provide some kind of schema that represent logical data structures of content items such as courses, assessment items, or learner information. (d) Data Transport: specifications or standards that explains and describes how data can be transferred among systems. (e) Launch and Track: specifications or standards that explains how content (courses, assessments, etc.) can be launched and tracked by LMSs. (f) Metadata: specification or standards that describes data-about-data which mainly used by LMS for content tagging so to facilitate content search and retrieval. (g) Philosophical: specifications or standards that represent a framework for describing the overall learning process, materials, services and tools.

After this brief introduction to standards and specification, the next sub-chapter sheds the light on educational standards and specifications in particular. National and professional organizations and consortia working on standards and specifications as well as guidelines for educational purposes are mentioned with an emphasis on e-assessment standards and guidelines.

4.2. Educational Standards

National organizations for educational standards have been established in several countries. The goal of such organizations is to provide guidelines by which institutions can guarantee to have an acceptable and sustainable higher education. Assessment standards can be used to validate the quality of assessment practices, and items. Crisp (2007, p. 151) gives examples for those organizations: The Quality Assurance Agency (QAA)[13] for Higher Education in UK. QAA has published a suite of interrelated documents which forms an overall Code of practice for the assurance of academic quality and standards in higher education (Code of practice). One of these documents is a code of practice for the assessment of students in UK higher education institutions. The Australian Universities Quality Agency (AUQA)[14], an independent, not-for-profit national agency that promotes, audits, and reports on quality assurance in Australian higher education. AUQA has published a database for good practices to quality assure Australian higher education. The Association of Universities and Colleges of Canada (AUCC)[15] has published principles for institutional quality assurance in Canadian higher education. The Higher Education Quality Committee (HEQC)[16] is responsible to promote the quality of higher education in South Africa.

However, the aforementioned national organizations are not specifically for e-assessment, rather than they are more related to promote the quality of higher education in general. Examples of more e-assessment specific are:

---

[13] http://www.qaa.ac.uk

[14] http://www.auqa.edu.au

[15] http://www.aucc.ca

[16] http://www.che.ac.za/about/heqc

- The British Standards Institute (BSI)[17] which published in April 2002 a standard named "BS 7988: a code practice for the use of information technology in the delivery of assessment". BS 7988 represents a standard code of practice for the use of information technology for the delivery of computer-based assessment. The standard covers the minimum requirements for the institutions to deliver sustainable assessments using computers. The standard has been approved by the national bodies of ISO (the International Organization for Standardization) and IEC (the International Electro-technical Commission), and has been adopted by the Joint Technical Committee ISO/IEC JTC 1, Information technology.

- The Scottish Qualification Authority (SQA)[18] has published guidelines for adopting computer-based assessment in further education. SQA has collaborated with other organizations which regulate qualifications in the UK such as the Qualifications and Curriculum Development Agency (QCDA)[19], the Department for Children, Education, Lifelong Learning and Skills (DCELLS), and the Northern Ireland Council for the Curriculum Examinations and Assessments (CCEA) to develop a report of "E-assessment: Guide to Effective Practice". The guidelines can be used for colleges and training providers but it may also applicable to schools and higher education. Moreover, it covers two key aspects of e-assessment: the management and delivery of e-testing and the use of e-portfolios for assessment.

In e-assessment and e-learning domains, standards, specifications, and reference models can be classified according to their applications into the following (Devedžic, Jovanovic, & Gaševic, 2007):

- *Metadata Standards:* a set of standards used to describe Learning objects' (LO) attributes, Such as the authors, title and languages. This description can be published with the LOs to facilitate their search and retrieval - such as, IEEE Learning Object Metadata (LOM) (IEEE LOM, 2008), IMS Meta-data (IMS LRM, 2008).

- *Packaging Standards:* describes the assembly of LOs and other complex learning units (e.g. online courses) from various texts, media files and other resources. Such assembly can be stored in a Learning Object Repository (LOR) and imported in a Learning Management Systems (LMS) such as IMS Content Packaging and IMS Learning Design (IMS CP, 2008).

- *Learner Information Standards:* formulates the description of the learner information and used to exchange that information between several systems, rather than their use in users modeling and personalization such as, IMS LIP (Learner Information Package) (IMS LIP, 2008) and PAPI Learner (Public and Private Information).

- *Question and Test Standards:* Special types of standards which are used in the assessment systems to represent questions and tests. IMS QTI (Question and test Interoperability) (IMS QTI, 2008) is an example of such standards.

---

[17] www.bsi-global.com

[18] http://www.sqa.org.uk

[19] http://www.qcda.gov.uk

- *Communication Standards:* specify the users' access to the LMS content, assessments, collaborative tasks and services communication, such as IEEE LTSA (Learning Technology Systems Architecture).

- *Quality Standards:* specify the pedagogical, technical, design and accessibility perspectives for the LOs' quality, such as 'BS 7988: a code practice for the use of information technology in the delivery of assessment'. Moreover the Scottish Qualification Authority (SQA) has published guidelines for adopting computer-based assessment in further education.

- *Semantic Standards:* specify how we can organize content and refer to it in the semantic web, such as Resource Description Framework (RDF)[20], W3C Semantic Web Activity[21], Web Ontology Language (OWL)[22].

Several organizations and consortia are working on building standards and specifications for the domains of e-learning and e-assessment. Examples of these organizations are: Dublin Core (DC) (DC, 2008), The Instructional Management System Global Learning Consortium (IMS GLC) (IMS GLC, 2008), The Aviation Industry CBT (Computer Based Training) Committee (AICC) (AICC, 2008), The Alliance of Remote Instructional Authoring and Distribution Networks for Europe (ARIADNE) (ARIADNE, 2008), Advanced Distributed Learning (ADL) (ADL, 2008), and IEEE Learning Technology Standardization Committee (IEEE LTSC) (IEEE LTSC, 2008). Despite the variety in educational standards, few examples among them are completely dedicated to represent e-assessment. The next section focuses on e-assessment content standards and specifications as well as guidelines for quality assessment.

4.2.1. e-Assessment Content Standards and Guidelines

In order to focus on the scope of this doctoral dissertation, an emphasis on e-assessment standards and specifications has been considered. This section explores the guidelines to have quality e-assessment. Moreover, it discusses some available specifications for e-assessment and concludes with limitations and problems collected from literature.

e-Assessment Guidelines

Professional bodies from several countries also publish documents for assessment standards and specifically guidelines deliver assessment online (Crisp, 2007). The International Test Commission (ITC) has published guidelines on computer-based and internet delivered testing (ITC, 2006). The following aspects have been discussed in the document:

- *Give due regard to technological issues in Computer-based (CBT) and Internet Testing:* (a) Give consideration to hardware and software requirements, (b) Take account of the robustness of the CBT/Internet test, (c) Consider human factors issues in the presentation of material via computer or the Internet, (d) Consider reasonable adjustments to the technical features of the test for candidates with disabilities, and (e) Provide help, information, and practice items within the CBT/Internet test.

---

[20] http://www.w3.org/RDF

[21] http://www.w3.org/2001/sw/

[22] http://www.w3.org/2001/sw/wiki/OWL

- *Attend to quality issues in CBT and Internet testing:* (a) Ensure knowledge, competence and appropriate use of CBT/Internet testing, (b) Consider the psychometric qualities of the CBT/Internet test, (c) Where the CBT/Internet test has been developed from a paper and pencil version, ensure that there is evidence of equivalence, (d) Score and analyse CBT/Internet testing results accurately, (e) Interpret results appropriately and provide appropriate feedback, and (f) Consider equality of access for all groups.

- *Provide appropriate levels of control over CBT and Internet testing:* (a) Detail the level of control over the test conditions, (b) Detail the appropriate control over the supervision of the testing, (c) Give due consideration to controlling prior practice and item exposure, and (d) Give consideration to control over test-takers authenticity and cheating.

- *Make appropriate provision for security and safeguarding privacy in CBT and Internet testing:* (a) Take account of the security of test materials, (b) Consider the security of test-takers data transferred over the Internet, and (c) Maintain the confidentiality of test-taker results

The British Psychological Society has also published guidelines for e-assessment (BPS, 2002). The guidelines cover four main components of CBA systems, assessment generation, assessment delivery, assessment scoring and interpretation, and storage retrieval and transmission. The guidelines stressed on the following principles:

- *Principle 1:* That, as with all psychological assessments, users should be made aware of what constitutes best practice in CBA so that they can make informed evaluations and choices between CBA systems offered to them.

- *Principle 2:* That CBAs should be supported by clear documentation of the rationale behind the assessment and the chosen mode of delivery, appropriateness and exclusions for use, and research evidence supporting validity and fairness.

- *Principle 3:* Requirements for administration of the CBA should be clearly documented and should include the knowledge, understanding and skills required for competent administration.

- *Principle 4:* The knowledge, understanding and skills required for interpretation of CBA information and for the provision of such information to a third party should also be clearly stated.

The Association of Test Publishers (ATP) in the USA has sponsored the development of guidelines for computer based testing (ATP, 2002; Crisp, 2007). The guidelines have been published to support CBT with principles, procedures, and best practices to administer and develop these tests. Moreover, to extend the Standards for Educational and Psychological Testing published in 1999 which were prepared by the Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME; cited after Olsen , 2000).

The Joint Information Systems Committee (JISC) supports higher education and research in the UK by providing leadership in the use of ICT (Information and Communications Technology) in support of learning, teaching, research and administration. JISC e-Learning program has published and sponsored several publications of e-learning and e-assessment in higher education. Roadmap for e-assessment (Whitelock & Brasher, 2006), Effective Practice with e-Assessment (JISC, 2007), Effective Practice with e-Portfolios (JISC, 2008), Effective Practice in a Digital Age (JISC, 2009), and Effective Assessment in a Digital Age (JISC, 2010) are examples of those publications and documents. The Effective Practice with e-Assessment report covers the e-assessment technologies, policies, and practices in higher education in UK. Moreover, it covers technical aspects of e-assessment by explaining how tools and standards have been deployed in real case studies from colleges and universities in UK as effective practices of e-assessment. The Effective Assessment in a Digital Age report provides guidelines for technology-enhanced assessment and feedback. It also discusses the Re-Engineering Assessment Practices (REAP)[23] principles of good assessment and feedback, developed as a result of the REAP project funded by the Scottish Funding Council during 2005–2007. The REAP project explored how technology might improve learning outcomes in different disciplines, and provided 12 principles of formative assessment and feedback (JISC, 2010):

- *Help to clarify what good performance is (goals, criteria, and standards):* To what extent do learners on your course have opportunities to engage actively with goals, criteria and standards before, during and after an assessment task?

- *Encourage 'time and effort' on challenging learning tasks:* To what extent do your assessment tasks encourage regular study in and out of class and deep rather than surface learning?

- *Deliver high-quality feedback information that helps learners to self-correct:* What kind of teacher feedback do you provide, and in what ways does it help learners to self-assess and self-correct?

- *Provide opportunities to act on feedback (to close any gap between current and desired performance):* To what extent is feedback attended to and acted upon by learners on your course and, if so, in what ways?

- *Ensure that summative assessment has a positive impact on learning:* To what extent are your summative and formative assessments aligned and supportive of the development of valued qualities, skills and understanding?

- *Encourage interaction and dialogue around learning (peer–peer and teacher–learner):* What opportunities are there for feedback dialogue (peer–peer and/or tutor–learner) around assessment tasks on your course?

- *Facilitate the development of self-assessment and reflection in learning:* To what extent are there formal opportunities for reflection, self-assessment or peer assessment in your course?

- *Give choice in the topic, method, criteria, weighting or timing of assessments:* To what extent do learners have choices in the topics, methods, criteria, weighting and/or timing of learning and assessment tasks on your course?

[23] http://www.reap.ac.uk

- *Involve learners in decision making about assessment policy and practice:* To what extent are learners on your course kept informed or engaged in consultations regarding assessment policy decisions?

- *Support the development of learning groups and learning communities:* To what extent do your assessment and feedback processes help to encourage social bonding and the development of learning communities?

- *Encourage positive motivational beliefs and self-esteem:* To what extent do your assessment and feedback processes enhance your learners' motivation to learn and be successful?

- *Provide information to teachers that can be used to help shape their teaching:* To what extent do your assessment and feedback processes inform and shape your teaching?

Examples of good practices for assessment activities can also found in institutions and organizations. The assessment audit tool from the Higher Education Academy (HEA) Bioscience network is an example of such practices (Fraser, Crook, & Park, 2008). The tool has been developed to support instructors and course designers in the review of assessment practices. The tool has been designed to be developmental, where teachers consider the course content and design with respect to assessment issues in order to further improve the course to achieve the assessment issues. Another example is Managing Assessment: Student and Staff Perspectives, is a practical tool developed by the Managing Effective Student Assessment (MESA)[24] benchmarking club. This document provides stakeholders with practical tools and case studies on assessment issues in higher education. Many other agencies are working on guidelines for assessment, for instance: Assessment Reform Group[25], Assessment Standards Knowledge Exchange (ASKe)[26], Centre for Excellence in Teaching and Learning in Assessment for Learning[27], Institute of Education University of London[28], JISC TechDis[29] (for guidance on inclusivity), the Higher Education Academy[30], and the Higher Education Academy Subject Centres[31].

e-Assessment Content Standards and Specifications

Despite the variance in e-learning content specifications and standards, e-assessment content has a limited number of specifications. The IMS QTI represents a data model for describing question (assessmentItem), test (assessmentTest), and their corresponding results reports. Unified Modelling language (UML) has been used to abstractly describe the data model which

---

[24] http://www.heacademy.ac.uk/resources/detail/ourwork/assessment/MESATool_Resource_Form

[25] http://www.assessment-reform-group.org

[26] http://www.brookes.ac.uk/aske

[27] http://www.northumbria.ac.uk/sd/central/ar/academy/cetl_afl

[28] http://www.ioe.ac.uk

[29] http://www.techdis.ac.uk

[30] http://www.heacademy.ac.uk/ourwork/teachingandlearning/assessment

[31] http://www.heacademy.ac.uk/ourwork/networks/subjectcentres

facilitates the binding with programming tools via the industry standard eXtensible Markup Language (XML) which provides a platform independent interchange and interoperability between different assessment tools and LMSs.

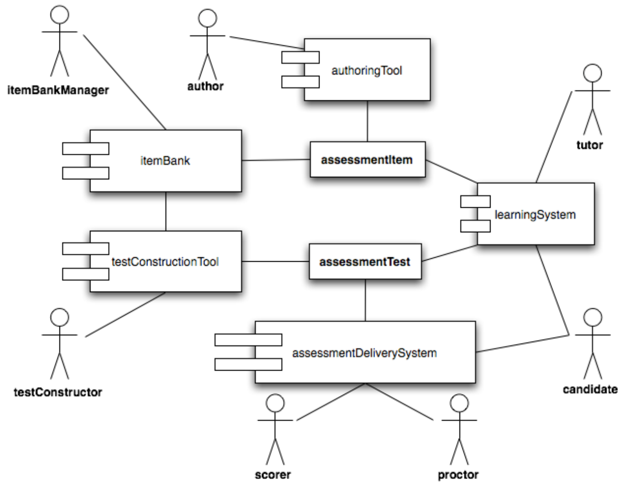


FIGURE 4.1. Assessment (item, test) role in assessment system (IMS QTI, 2008).

IMS QTI is designed to provide a well-formed assessment content where questions can be created, stored, and exchanged independently from the authoring tool. Moreover to support the deployment of item banks that can be used among several assessment authoring tools and LMSs. Similar to questions the specification is designed to provide a well formed representation of tests so that they can be created by selecting questions form item banks, stored, and exchanged between different assessment delivery tools and LMSs. Moreover, QTI specification supports systems with the ability to report test results. Figure 4.1 summarizes the role of 'assessmentItem' and 'assessmentTest' within authoring tools and learning management systems.

The IMS QTI information model consists of two main data structures:

- ASI (Assessment, Section, and Item) data structure for assessment content representation: (a) *Assessment:* represents the test unit, (b) *Section:* is a group representation of sub-sections and assessment items that may share common learning objectives, and (c) *Item*: is the fundamental structure that holds information about the question and how to score it. Scoring is handled within the model by transforming the candidate (student) responses into outcomes using pre-defined response processing rules.

- Results Reporting represent the results from the candidate interactions: (a) *Context:* holds information session variables such as participant username, ID, and institution, and (b) *Assessment Results:* used to report the results of candidate's interaction on both levels test (testResult) and item (itemResult).

QTI defines a test item as "*a set of interactions (possibly empty) collected together with any supporting material and an optional set of rules for converting the candidate's response(s) into assessment outcomes*". QTI items are classified according to their points of interaction into: *simple items, and composite items.* Simple item only have one point of interaction (e.g. single-choice, multiple-choice, cloze, match, hotspot, graphic-order), composite item is the item that contains more than of point of interaction where multiple instances of the same type of interactions or different types of interactions can be provided. Interactions in QTI are classified into:

- **blockInteraction:** extends the Interaction class for the following interactions
    - Simple Interactions: (e.g. choiceInteraction, orderInteraction, associateInteraction, matchInteraction, gapMatchInteraction)
    - Text-based Interactions: (e.g. extendedTextInteraction, hottextInteraction)
    - Graphical Interactions: (e.g. graphicInteraction)
    - Miscellaneous Interactions: (e.g. sliderInteraction, mediaInteraction, drawingInteraction, uploadInteraction)
- **inlineInteraction:** extends the Interaction class for the following interactions
    - Text-based Interactions: (e.g. inlineChoiceInteraction, textEntryInteraction)
    - Alternative Ways to End an Attempt (e.g. endAttemptInteraction)
- **positionObjectInteraction:** The position object interaction consists of a single image which must be positioned on another graphic image (the stage) by the candidate.
- **customInteraction:** The custom interaction provides an opportunity to extend the specification with new interactions.

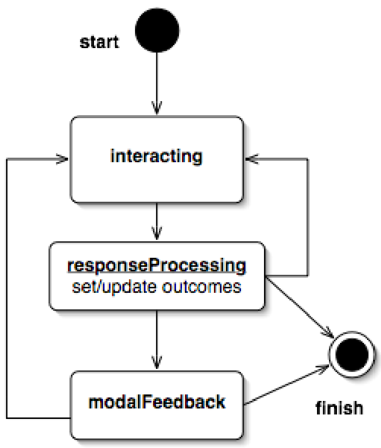


FIGURE 4.2. Response Processing based on candidate interaction(s) (IMS QTI, 2008).

Based on the aforementioned QTI item definition the item consists also with an optional set of rules for converting the candidate's responses based on the discussed interactions into assessment outcomes. The process of converting the candidate responses into outcomes is called *response processing*. Response processing is used for some items automatic scoring and may provide immediate or timely feedback based on the candidate response as depicted in Figure 4.2. Response processing is handled by applying a set of responseRules to evaluate expressions of item variables using responseConditions (i.e. responseIf, responseElseIf, and responseElse). For the sake of simplicity, QTI has standard response processors called *response processing templates*:

- **Match Correct:** uses the 'match' operator (QTI expression) to match the value of a response variable RESPONSE with its correct value. It sets the outcome variable SCORE to either 0 or 1 depending on the outcome of the test.

- **Map Response:** uses the 'mapResponse' operator (QTI expression) to map the value of a response variable RESPONSE onto a value for the outcome SCORE.

- **Map Response Point:** uses the 'mapResponsePoint' operator (QTI expression) to map the value of a response variable RESPONSE onto a value for the outcome SCORE.

QTI provides a set of different question types as summarized in Table 4.1. Question of types extended text, drawing, and upload do not have pre-defined response processing templates as they require complex scoring and grading techniques (e.g. automated essay grading for extended text question type).

TABLE 4.1. QTI question types and their corresponding interaction type and response processing template.

Question Type	Description	Interaction Type	Response Processing
true/false	selecting a response from the choices 'True' and 'False'	choiceInteraction	Match Correct
single response	selecting a single response from the choices	choiceInteraction	Match Correct
multiple response	selecting multiple responses from the choices	choiceInteraction	Map Response
order	reordering the choices that are displayed initially	orderInteraction	Match Correct
associate	pairing up the choices that are displayed initially	associateInteraction	Map Response
match	pairing up choices from a source set into a target set	matchInteraction	Map Response
gap match	filling gaps from an associated set of choices	gapMatchInteraction	Map Response
inline choice	filling gaps from a shared stock of choices	inlineChoiceInteraction	Match Correct
text entry	filling gaps by constructing a simple	textEntryInteraction	Map Response

	piece of text		
extended text	entering an extended amount of text	extendedTextInteraction	
hot text	selecting choices embedded within a surrounding context	hottextInteraction	Match Correct
hot spot	selecting areas (hotspots) in the graphic image	hotspotInteraction	Match Correct
select point	selecting points in the graphic image	selectPointInteraction	Map Response Point
graphic order	reordering the choices that are presented as hotspots on a graphic image	graphicOrderInteraction	Match Correct
graphic associate	pairing up the choices that are presented as hotspots on a graphic image	graphicAssociateInter-action	Map Response
graphic gap match	a graphical interaction of filling gaps from an set of choices	graphicGapMatchInter-action	Map Response
position object	positioning a given object on the image	positionObjectInteraction	Map Response Point
slider	selecting a numerical value between a lower and upper bound	sliderInteraction	Map Response
drawing	using a common set of drawing tools to modify a given graphical image	drawingInteraction	
upload	uploading a pre-prepared file representing the response	uploadInteraction	

Feedback as an important process of any assessment has been considered during the design of QTI specification. QTI handles two types of feedback material, *modal* and *integrated*. Modal feedback is provided to the candidate after response processing has finished and before any subsequent attempt or review of the item. Integrated feedback is embedded into the itemBody and is only shown during subsequent attempts or review. Figure 4.3 explains the two types of feedback using a single-response question.

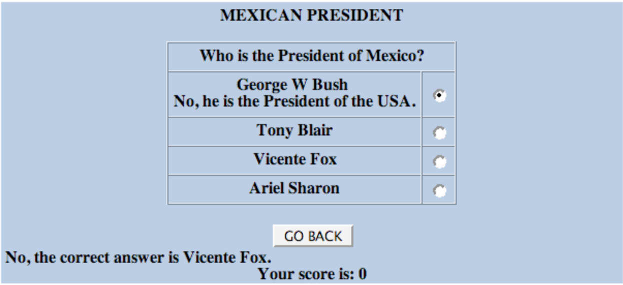


FIGURE 4.3. Both types of feedback with QTI-based item (IMS QTI, 2008).

QTI concentrates on item creation and storing with no APIs for item authoring and delivery. QTI is more applicable for XML based authoring where only QTI professionals can author items. Therefore, the design of items user interfaces is left to the software designer/developer. Some common practices use EXtensible Stylesheet Language Transformations (XSLT) to map the QTI XML item to XHTML so that it can be usefully visualized to candidates. For items that require complex interaction such as 'hot spot' and 'graphical order' items (require dragging and dropping objects), browser applets such as Adobe Flash, Microsoft Silverlight, Curl, and Java applets are attached to the XTML files in order to visualize and handle those interactions. Despite the speed browsing these technologies have over JavaScript, they have some problems such as, plat-form dependency and the need for browser plug-ins, security problems and client-side scripting, browser lack of support for instance mobile browser on running Apple iOS or Android do not run Java Applets. The emergence web 2.0 holds a great promise of visualizing and delivering QTI items with more platform independent manner. For instance, QTI XML items can be mapped into HTML5 based files and visualized independently of the platform (AL-Smadi & Guetl, 2011a).

Although QTI is the leading e-assessment content metadata, it has some limitations and challenges. For instance, the so-called impedance mismatch between the features offered by the standard and the ones needed in a particular application domain (Helic, 2006). IMS QTI has some difficulties in some application domains (e.g. foreign languages teaching). One of these difficulties is that the IMS QTI is designed to formulate general types of questions and does not take into consideration some specific questions (e.g. Crossword puzzle) and test types for a particular domain (Milligan, 2003). According to (Smythe & Roberts, 2000) the QTI specification is not related to didactical issues and tries to be didactically neutral as possible. Moreover, it has proved to have high complexity during assessment authoring and delivering, it does not cover pedagogical aspects such as learning objectives, as well as it has no text and item analysis, in the other hand it has a model for results reporting (Chang, Hsu, Smith, & Wang, 2004).

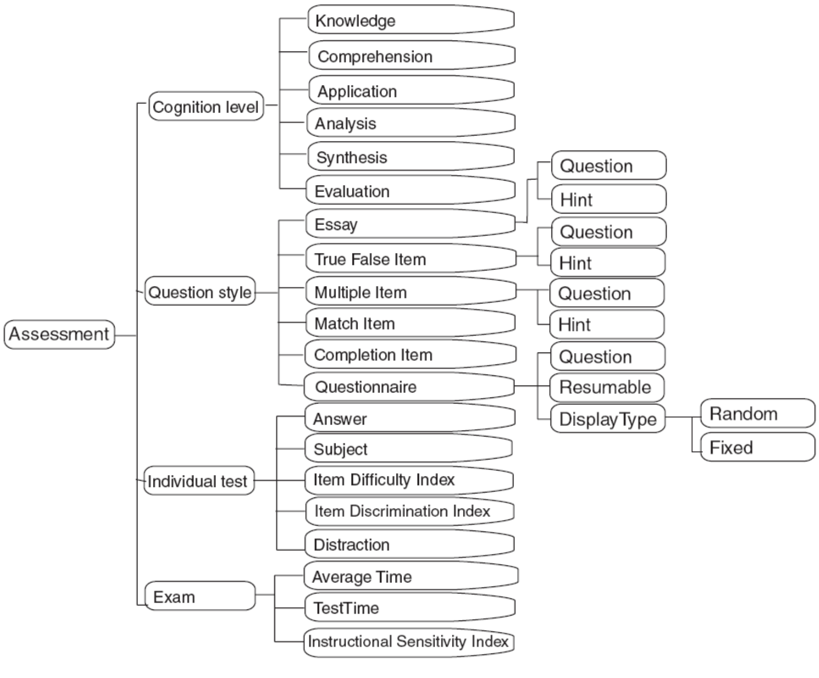


FIGURE 4.4. MINE SCORM e-assessment metadata model (Chang et al., 2004)

Chang et al. (2004) propose a SCORM 1.3 metadata extension for e-assessment content. The metadata model is called MINE SCORM and has been designed to cover cognition level, discrimination, instructional sensitivity and difficulty, different question types, as well as feedback provision and tem/test analysis. The authors used Bloom's taxonomy (Bloom, 1956) of the cognitive domain to classify questions and assessments, and an assessment analysis model that provides useful statistical data about the items to teachers, students, and the system. Figure 4.4 depicts the proposed MINE SCORM as an assessment metadata model.

4.2.2. Interoperability Standards

Interoperability has been always a challenge for e-learning software designers and developers. LMSs have been designed as centralized environments where educational activities are organized and provided to students. Nevertheless, the variance of e-learning application domains has increased the limitation of LMSs to cover those different application domains. For instance in higher education, universities and institutions provide a variety of disciplines where students are required to learn and interact with contents, and perform experiments and collaborate with other students. Therefore, more activity-specific or application-domain specific tools have been developed. As a result, a variety of educational and learning tools are available as standalone tools apart from the centralized LMSs. This has caused people in the domain to think how to reuse and share content among those tools, how to integrate those tools within the centralized LMSs in a way to extend the LMS services by third-party tools and services. As a result interoperability has been decided to be a major requirement for any e-learning content, tool, service, or LMSs.

Several definitions have been provided to the term interoperability. The Oxford English Dictionary[32] defines the word "interoperable" as "*(of computer systems or software) able to exchange and make use of information*". The IEEE defines interoperability as "*the ability of two or more systems or components to exchange information and to use the information that has been exchanged*". Taking into consideration the integration point of view, Merriman (2008) defines interoperability as "*the measure of ease of integration between two systems or software components to achieve a functional goal. A highly interoperable integration is one that can be easily achieved by the individual who requires the result*". Merriman discusses the aforementioned two definitions and argued that both of them don not take integration into consideration. Moreover stresses on the level of achievement of integration goal as a main measure for interoperability. Based on that interoperability is not only the ability of sharing information, rather than it goes deeper to cover the ability of sharing functions and services in flexible way of integration. Bull and McKenna (2004, p. 112) defines interoperability as "*interoperability describes the capacity for different systems to share information and services such that two or more networks can communicate with each other to exchange data in a common file format*". Similar to Bull and McKenna definition, Crisp (2007, p. 158) defines interoperability as "*interoperability is the ability of a system, content or activity to be exchanged or used in a variety of situations with the confidence that it will function in a predictable manner. Interoperability allows efficient use of resources and avoids the necessity to design a system, content or activity de novo for every context*".

Based on these definitions interoperability can occur on two main levels (AL-Smadi & Guetl, 2011b) of information (content, user data) level and on tools level (tools interoperability) as follows:

- *Information interoperability:* has been a major research area for years. Several specifications and standards have been published. For content examples are IEEE

[32] http://oxforddictionaries.com

LOM, IMS Meta-data, SCORM, and IMS QTI. For user data examples are IEEE PAPI, and IMS LIP. Some other supportive standards are IMS CP for content packaging and IMS LD for the learning process design and workflow.

- *Tools interoperability:* is an emerging research where limited examples of specifications are available. Among these specifications we can mention the Open Knowledge Initiatives (OKI)[33] and its Open Service Interface Definition (OSID)[34], and CopperCore Service Integration (CCSI) (Vogten et al. 2006). A more recent and promising research is the IMS Tools Interoperability specifications by which tools and LMSs are provided guidelines of how they can be designed to flexibly be integrated with each other. This decoupling of content and tools as well as building systems using SOAs supports the comprehensive idea of interoperability.

As e-assessment content meta-data have been discussed before the following sections sheds the light on available specifications for learning tools interoperability – including e-assessment tools.

OKI Open Service Interface Definitions (OSID)

The Open Knowledge Initiative (O.K.I.) develops specifications that describe how the components of an educational software environment communicate with each other and with other enterprise systems. O.K.I. specifications enable sustainable interoperability and integration by defining standards for Service Oriented Architecture (SOA). The O.K.I. project was initially launched in 2001 through a generous grant from the Andrew W. Mellon Foundation, and led by MIT and Stanford.

O.K.I has been designed as a layered architecture which fosters the modular development and maintenance of the educational applications independently of each other. As depicted in Figure 4.5 the core of the architecture is the "institutional infrastructure layer" which includes - but not limited to - file systems, databases, and authentication servers. The infrastructure layer provides services to the "educational applications layer". These services are classified into "common services" and "educational services" according to their use. The O.K.I. services are accessed via Application Programming Interfaces (APIs). The APIs must be implemented on both sides of interaction the educational applications as well as the institutional infrastructure. This separation allows the educational applications services to be used as institutional infrastructure ones if needed. Based on that, tools and services can flexibly interoperate with each other as well as with the institution systems.

O.K.I. has a main concern to foster interoperability and tools integration with standards that fits with SOA. Moreover it aims to support the flexible integration and development of educational tools. Therefore, O.K.I. has produced a set of Open Service Interface Definitions (OSIDs) influenced by its layered architecture. The OSIDs are an abstraction layer between the software developer and the enterprise infrastructure systems and tools. The OSIDs are designed based on the common and educational services APIs. For each OSID a set of methods have been defined in an abstraction layer (interface) so that developers are free to develop these methods using different programming languages and frameworks. Using OSIDs holds great promise of software flexibility in terms of, (a) flexible integration: simple

---

[33] http://www.okiproject.org

[34] http://www.okiproject.org)

integration with existing infrastructure, and (b) tools and services interoperability: tools and services can be easily shared among different campuses and universities.

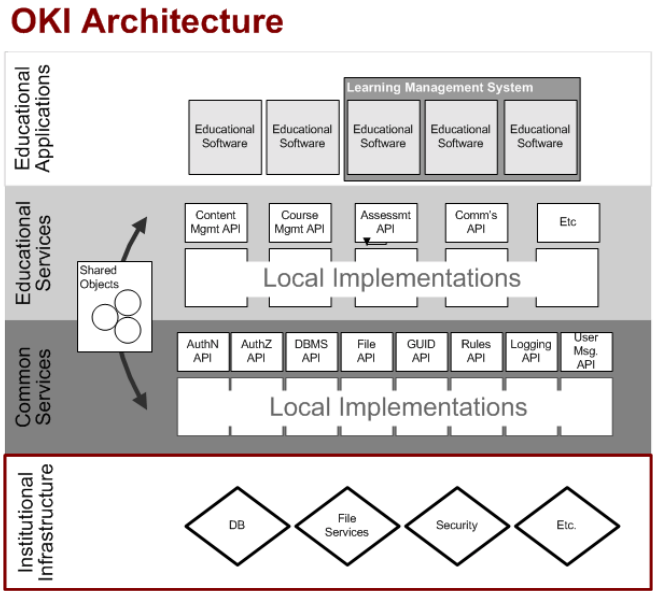


FIGURE 4.5. OKI Architecture.

The first version of OSID was introduced in the year 2003, a year after in 2004 O.K.I. introduced the second version OSIDv2. Two years later a new version OSIDv3 has been introduced in order to tackle some problems and complaints. Troubles with Types, issues with Iterators, to the general challenges raised through applying the OSID model to multiple programming languages have been discussed. The current version of OSIDv3 is provided in different programming languages such as, Java and PHP, C and C#.

An Example of a project that utilizes the O.K.I. OSIDs is the Campus[35] project. The Campus project, promoted by the Secretariat for Telecommunications and the Information Society (STSI) of the Regional Government of Catalonia, based on the agreement signed by the majority of Catalan universities in order to have an open source virtual campus and has been adopted by the Open University of Catalonia (UOC) as a model for its virtual campus. The project aims to develop a technological infrastructure with open source tools to provide online training and to extend the tools and services provided by platforms such as Moodle (PHP) and Sakai (Java) with third-party tools of different pedagogical approaches. The main idea is summarized in Figure 4.6 where a service-oriented architecture has been used to develop a service bus named OKI Bus and OKI-based interfaces to Moodle and Sakai platforms in what they called Campus Middleware.

[35] http://www.campusproject.org

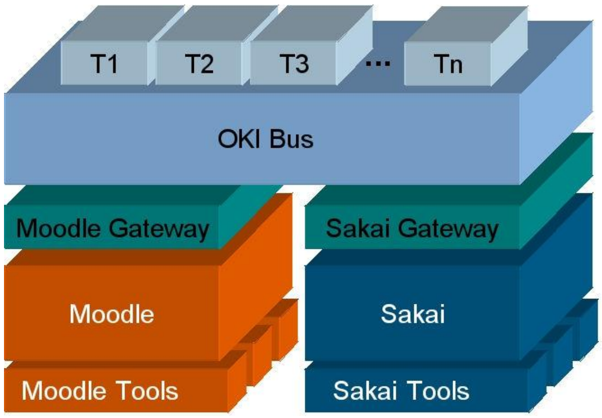


FIGURE 4.6. Campus architecture of layers and components

CopperCore Service Integration (CCSI)

CCSI is a generic framework for e-learning services integration. CCSI has been designed for the IMS Learning Design (LD) as part of JISC e-learning framework (ELF) toolkit project called SLeD2 (Service Based Learning Design System). The project extended an earlier work which provided a LD runtime service (called CopperCore) and a corresponding web based client application called SLeD. (Vogten et. al., 2006).

CopperCore processes units of learning (UOLs) which are IMS content packages for a specific LD. As CopperCore does not provide any user interface and its methods are only available through an Application Programming Interface (API) it must be used as a service integrated into a larger framework or Learning Management System (LMS). Therefore CopperCore utilized the approach of adapter design pattern (Gamma, Helm, Johnson, & Vlissides, 1995; Vogten et. al., 2006) to develop adapters in order to adapt class's interfaces according to clients' needs. In CCSI, each adapter is a software component encapsulating a single service implementation. A dispatcher is the central component, responsible for the orchestration between these services. To make this orchestration possible, all adapters share a common API providing the dispatcher a standard interface to all integrated services. The idea is more explained in Figure 4.7, which depicts how APIS (Assessment Provision through Interoperable Segments) IMS QTI-based service (Barr, 2006) can be integrated with IMS LD (CopperCore).

It is worth mentioning that CCSI framework has been used in the TENCompetence project (2006- 2009)[36]. TENCompetence is a European Commission funded project through the IST (Information Society Technologies) Program. Its goal is developing and using infrastructure to support individuals, groups and organizations in lifelong competence development.

[36] http://tencompetence-project.bolton.ac.uk

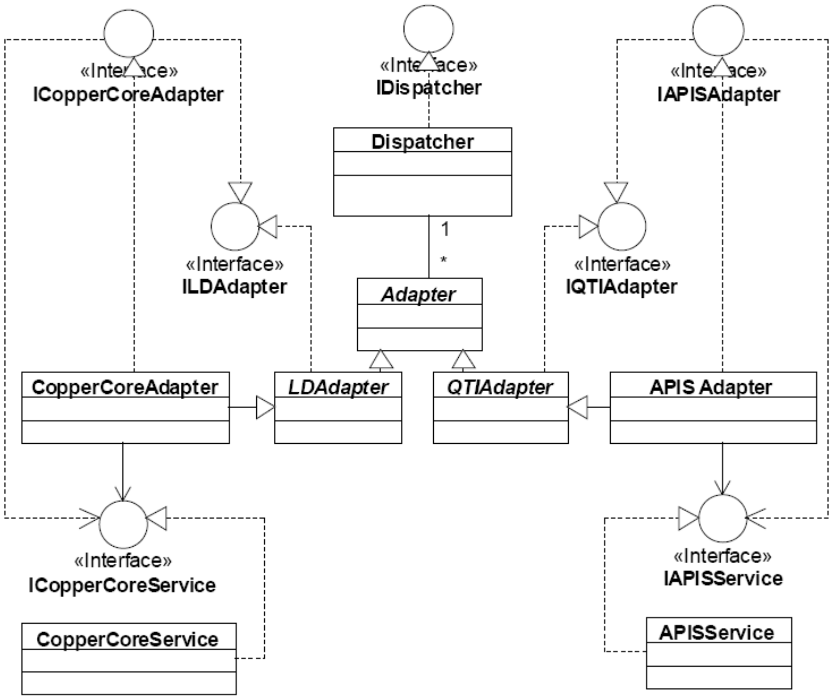


FIGURE 4.7. Integrating APIS with IMS LD using CopperCore Service Integration architecture (Vogten et al., 2006).

IMS Learning Tools Interoperability Specification

Tools interoperability concerns the ability of aggregating third-party tools to cooperate with a LMS platform (tool consumer). Third-party tools (tool provider) can be used to extend the services provided by the core system with services for specialized application domains. One of the possible specifications for tools interoperability is the work provided by the IMS GLC. IMS GLC has provided architecture for tools interoperability as web services. IMS Learning Tools Interoperability Guidelines v1.1 has described this architecture as well as its main components (IMS LTI, 2012). As summarized in Figure 4.8, the suggested architecture has introduced two main concepts:

- *Proxy Tool Runtime:* from its name this tool will be used by the tool consumer to communicate with the tool provider. A standard mechanism for packaging this tool to be deployed to an LMS has been defined by the architecture. The proxy tool is meant to be environment-independent where it does not require specialized code. The proxy tool is entirely a descriptor-based package that describes the deployment, configuration, and runtime context.

- *Web Services:* a set of services that have to be implemented to the hosting environment (i.e. Tool Consumer). These services facilitate the deployment, configuration, and launching of the proxy tool through its main services of, d*eployment Service, configuration Service, launch Service,* and o*utcome Service.*

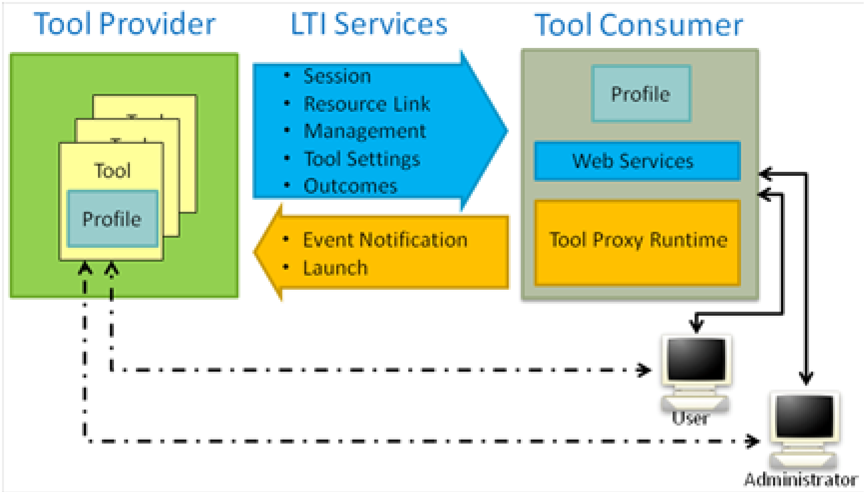


FIGURE 4.8. Tools interoperability architecture (IMS LTI, 2012)

The communication between the TIR/Proxy Tools is handled by a core protocol defined for this purpose. The core protocol is based on the IMS General Web Services (GWS) specifications v1.0 (IMS GWS, 2009) which recommends XML with Web Services Description language (WSDL) for defining services and Simple Object Access Protocol (SOAP) for the base transport protocol. Since the communication between the third-party tools and the core system is performed as web services some kind of web services management is required. Matters such as services provision, services registration, services invocation, and security and accessibility should be taken into consideration. Special standards and specifications can be used to represent these web services descriptions. For instance the IMS General Web Services (GWS) specifications can be used where the WSDL/XSD created files are designed to comply with Web Service Interoperability (WS-I) Consortium Base Profile v1.1 (WS-I, 2009).

The IMS LTI specification requires the tool provider to expose its services as webs services and to provide a descriptor-based package that describes the deployment, configuration, and runtime context. However, the approach is not innovative rather than it adopts the service-oriented paradigm by which the tool provider interfaces with the tool consumer through well-defined web services. The web services can be described in many was such as, (a) using Web Services Description language (WDSL) whereas the developed web services as well as the specification of their messages (SOAP -based) are described, (b) using the Business Process Execution Language for Web Services (BPEL4WS)[37] is an emerging *de facto* specification to describe the business process workflow execution. BPEL – as it is called for short – provides XML-based meta-data to design a workflow of how web services participating in a business process can coordinate among each other. Similar to BPEL is the Business Process Model and Notation (BPMN)[38].

To this end, e-assessment content standards lacks alignment with learning theories and target pedagogy which appears clearly between the guidelines for quality assessment and the specifications provided for the content. Moreover, specifications provided to represent e-assessment covers only content authoring and lacks other aspects such as, content delivery

---

[37] http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsbpel

[38] http://www.bpmn.org/

(platform and browser independent technology specification or the so called "user interface interoperability"), and assessment results analysis and reporting. Nevertheless, e-assessment content specifications focus on traditional test items and ignore alternative types of assessment such as, performance assessment, behavioural assessment, portfolio-based assessment, and rubric-based assessment.

Considering assessment strategies (see Section 3.1.3) in way to align them with discussed e-assessment standards and specification, it is clear that available e-assessment standards and specifications lacks representations for alternative assessment - such as performance assessment, behavioural assessment, self-assessment, peer-assessment, and rubric-based assessment. However, as discussed in Chapter 3, assessment is no more separated from learning; instead assessment is integrated within leaning activities and plays major roles in learning support and learner engagement. Nevertheless, the lack of such specifications and standards side-by-side with the variety in assessment application domains have caused to have several e-assessment tools which are mainly designed as standalone applications.

In a way to tackle this problem, educators, researchers, and standard organizations propose learning design specifications by which the learning activity - in terms of content and tools - can be represented. The next section discusses learning design specification and standards in more detailed.

4.2.3. Learning Design and Workflow Standards

Teaching and learning *workflow* in e-Education refers to the automation of the provision of learning activities controlled by a set of rules that defines the pre-requisite, sequence, and consequence of each learning activity. Teaching and learning workflow – which is also known as *learning design* – is known to the educational community through two main initiatives: the IMS Global Learning Consortium (GLC) Learning Design (LD)[39] and the Learning Activity Management System (LAMS)[40].

JISC for instance has funded several projects to encourage using IMS LD in e-educational, SLED[41] and RELOAD[42] by which an editor and player for IMS LD have been developed. Despite the emerging use of IMS LD in designing learning activities it is common among e-education society that the process is complex and it requires involvement of specialists.

In contrast LAMS is easy to use system which provides a drag-and-drop user interface (UI) enhanced with components that represents units of learning (UoL) which make it useful for non-specialists to easily design a learning activity. Despite the easy to use UI, LAMS in contrast to IMS LD does not consider learning objectives during the design of the learning activity whereas IMS LD provides the ability to define learning objectives in a separate metadata file and to link on the level of the learning activity or on the level of the UoL.

In enterprise solutions some other initiatives are used to design the actions for the business process. For example, the Business Process Execution Language for Web Services

---

[39] http://www.imsglobal.org/learningdesign/

[40] http://www.lamsinternational.com/

[41] http://www.jisc.ac.uk/whatwedo/programmes/elearningframework/demolivhope.aspx

[42] http://www.reload.ac.uk/

(BPEL4WS)[43] is an emerging *de facto* specification to describe the business process workflow execution. BPEL – as it is called for short – provides XML-based meta-data to design a workflow of how web services participating in a business process can coordinate among each other. Similar to BPEL is the Business Process Model and Notation (BPMN)[44]. Despite the impact achieved in terms of flexibility and services orchestration and choreography, these initiatives lake pedagogical aspects – e.g. learning objectives - when it comes to use them to design workflows for e-education. Moreover, they require the e-education platform to be built on top of SOA.

Using such learning design specifications, learning objects - e.g. IEEE LO, SCORM, IMSQTI, related learning and assessment tools, and pre-requisites and consequences of using those tools are represented in standard-conform way to formulate the target learning activity. However, this is faced with the challenge of some of those tools have been designed to work as standalone and in somehow are seen as a black-box to the instructional designer. Therefore, learning tools interoperability standards are required and learning tools should not only be standard-conform in terms on their content but also their services and learning workflow.

To this end, providing standard-conform integrated assessment forms with other leaning objects (LO) in order to achieve specific learning outcomes or to evaluate pre-defined learning and didactic objectives increases the complexity of those LOs. As discussed in (AL-Smadi, wesiak, & Guetl, 2012) the atomic learning object is defined by (Nitto, Mainetti, Monga, Sbattella, & Tedesco, 2006) as "*the smallest unit of reuse for LOs that may or may not be associated to one or more multimedia contents*", whereas the define a Complex LO (CLO) as "*an LO whose instructional material is an aggregation of Learning Objects. Being an LO, a Complex LO can be treated exactly as any other LO*". Accordingly, AL-Smadi, Wesiak, & Guetl (2012) define complex learning resource (CLR) as "*CLR as a composite didactic resource consists of one or multiple learning objects (either atomic or complex). Accordingly, CLR inherits the features of LO of reusability and interoperability provided by the standards and specifications used to represent LOs*".

4.3. Summary

Learning content reusability and interoperability, learner's information accessibility and share ability, are main maters of quality for any LMS. Therefore, LMS should be designed and implemented to be standard-conform. Examples of educational standards include (Devedžic, Jovanovic, & Gaševic, 2007): (a) *Metadata Standards:* a set of standards used to describe Learning objects' (LO) attributes, such as the authors, title and languages. This description can be published with the LOs to facilitate their search and retrieval - such as, IEEE LOM, 2008, IMS Meta-data (IMS LRM, 2008), (b) *Packaging Standards:* describes the assembly of LOs and other complex learning units (e.g. online courses) from various texts, media files and other resources, such as IMS CP, (c) *Learner Information Standards:* formulates the description of the learner information and used to exchange that information between several systems, rather than their use in users modeling and personalization such as, IMS LIP and PAPI, (d) *Question and Test Standards:* Special types of standards which are used in the assessment systems to represent questions and tests. IMS QTI is an example of such standards, (e) *Communication Standards:* specify the users' access to the LMS content, assessments, collaborative tasks and services communication, such as IEEE LTSA, (f) *Quality Standards:* specify the pedagogical,

[43] http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsbpel

[44] http://www.bpmn.org/

technical, design and accessibility perspectives for the LOs' quality, Such as BS 7988: a code practice for the use of information technology in the delivery of assessment. Moreover the Scottish Qualification Authority (SQA) has published guidelines for adopting computer-based assessment in further education, and (g) *Semantic Standards:* specify how we can organize content and refer to it in the semantic web, such as RDF[45], W3C Semantic Web Activity[46], and OWL[47].

Several organizations and consortia are working on building standards and specifications for the domains of e-learning and e-assessment. Examples of these organizations are: Dublin Core (DC) (DC, 2008), The Instructional Management System Global Learning Consortium (IMS GLC) (IMS GLC, 2008), The Aviation Industry CBT (Computer Based Training) Committee (AICC) (AICC, 2008), The Alliance of Remote Instructional Authoring and Distribution Networks for Europe (ARIADNE) (ARIADNE, 2008), Advanced Distributed Learning (ADL) (ADL, 2008), and IEEE Learning Technology Standardization Committee (IEEE LTSC) (IEEE LTSC, 2008). Despite the variety in educational standards, few examples among them are completely dedicated to represent e-assessment. IMS QTI is the widely used specifications for e-assessment content.

AL-Smadi, Guetl, & Helic (2009b) discuss some limitations and problems in the available specifications and standards. For instance, the so-called impedance mismatch between the features offered by the standard and the ones needed in a particular application domain (Helic, 2006). For example, IMS QTI has some difficulties in some application domains (such as, foreign languages teaching). One of these difficulties is that the IMS QTI is designed to formulate general types of questions and does not take into consideration some specific questions (e.g. Crossword puzzle) and test types for a particular domain (Milligan, 2003). According to (Smythe & Roberts, 2000) the QTI standard are not related to didactical issues and tries to be didactically neutral as possible. Another example is what authors of (Recker & Wiley, 2001) have noted about the IEEE LOM (Learning Object metadata), that IEEE LOM from a perspective of metadata does not provide enough information to support the learning process. Another major challenge is the problem of selecting the most appropriate standard in cases of having different types of standards for the same aspect of the Learning Management System (LMS) (Devedžic et al, 2007). For example IEEE PAPI Learner and IMS Learner Information Package (LIP) both of them are related to the issue of learner modelling.

Interoperability standards and specifications are important aspects to be considered when it comes to provide flexible and interoperable e-assessment. Therefore, not only e-assessment content should be standard-conform but also e-assessment tools and services. Learning tools interoperability fosters e-assessment platforms with assessment third-party tools that acts as service, thus can be easily used to extend the main system tools and services. This flexible extension enables e-assessment platforms to provide assessment services for different application contexts - such as, collaborative learning and game-based learning - in a way to support different learning theories, learning types, and pedagogical approaches, as well as provide alternative forms of assessment as mentioned earlier in Chapter 3.

---

[45] http://www.w3.org/RDF

[46] http://www.w3.org/2001/sw/

[47] http://www.w3.org/2001/sw/wiki/OWL

Part

B

Solution Approach

Chapter

5

5. Towards Flexible and Integrated e-Assessment

Over the last decades e-assessment has emerged with the influence of ICT. The so-called 'e-Assessment 2.0' (Elliott, 2008) is a result of this influence as well as the use of web 2.0 technology in e-learning. Rather than pedagogy and student support, the evolution of using ICT in education assessment has been led by technology (*cf.* Watson, 2001). Consequently, a variety of e-assessment tools have been developed for different contexts and application domains (see Section 3.2.2). Thus, selecting the most appropriate one for a specific learning activity has become a complex task (AL-Smadi, Guetl, & Helic, 2009b; Ravenscroft, 2001). Despite the richness in technological capabilities used in e-assessment, developed assessment tools lack to some extent the alignment with theories of learning and pedagogy. Moreover, minor group of e-assessment tools are standard-conform thus supports content and services reusability, share ability and interoperability (see Chapter 4).

The paradigm shift for online learning and assessment has caused researchers to rethink assessment practices. Traditional assessment practices – often based on objective testing – are neither adequate for testing meta-cognitive skills such as critical thinking, creativity, and self-reflection nor to test authentic learning and to support life-long learning (*cf.* Haken, 2006). Thus, rethinking e-assessment practices towards advocating alternative assessment has emerged. Alternative assessment practices - including self and peer-assessment, portfolio-assessment, behavioral assessment, and performance assessment (*cf.* Buzzetto-More & Alade, 2006) - address the lack of considering theories of learning and pedagogy in online assessment through advocating constructive, authentic, contextualized, and deep learning assessment. Consequently, educators are faced with the challenge of having to develop appropriate, authentic, reliable, and ethical e-assessment that is integrated with the learning process, evaluates learning, engages students, appraises the students' learning process, and promotes further learning (Bartley, 2006). (see Chapters 2, & 3)

To this end, this chapter aims to address the challenges of having appropriate assessment practice through a solution approach for providing flexible and integrated e-assessment forms. To achieve this solution approach, the remaining of this chapter discusses the aspects related to flexible and integrated assessment, moreover proposes an integrated model for e-assessment through which aligned assessment with instruction and learning can be designed and closes with recommendations to use e-assessment standards and specifications in a way to assure flexible integration with complex learning resources.

This chapter is based on (AL-Smadi, Guetl, & Helic, 2009; AL-Smadi, Hoefler, & Guetl, 2011a; AL-Smadi et al., 2011).

5.1. Flexible and Integrated e-Assessment

In Chapter 3 a set of frameworks and models - which either discuss key elements for assessment in general (e.g. Chudowsky, & Glaser, 2001) or emphasize on specific aspects of the assessment process (e.g., Almond, Steinberg, & Mislevy, 2002) - has been presented. Additionally, different forms of assessment (e.g. peer-assessment) and assessment tools (e.g. automatic test item creation or assessment of collaborative learning) were outlined. However, the discussed assessment models lack to some extent aspects such as: (a) pedagogical flexibility and the alignment with theories of learning, (b) the suitable assessment form for the learning activity or task, (c) available technology – in terms of systems, tools, and services, (d) standards, specifications, and guidelines of how to design, and develop assessment for the target learning practice, (e) feedback as a crucial component for quality assessment practice, (f) guidelines or framework of how to use these models to support developing learning tools with integrated assessment.

To this end, this chapter addresses the aforementioned problems through: (a) proposing an enriched learning experience as a form of learning experience that is built through the use of complex learning objects – such as collaboration, simulation, and serious games - integrated with e-assessment forms – the considers cognitive and affective theories of learning - and capable to generate an effective kind of learning – such as reflective learning, experiential learning, or socio-cognitive learning. In order to develop such forms of assessment – i.e. suitable for the enriched learning experience - an integrated model for e-assessment (IMA) is used. This model addresses the aforementioned limitations of the available e-assessment models, and considers approaches of e-assessment in traditional and adaptive learning scenarios. Moreover it enriches complex learning resources - such as simulations, collaborative experiences, virtual experiences, and emotional elements - with integrated forms of assessment capable to evaluate the results of those learning experiences and to support and scaffold students learning process. (b) discussing aspects related to standard-conform e-assessment system and the importance of using standards and guidelines in the design, development, and use of e-assessment. Moreover, proposing a solution framework namely Service-Oriented Framework for e-Assessment (SOFA) in order to have a flexible e-assessment system.

More precisely, this chapter discusses three main research questions, How integrated e-assessment forms can be designed to align with theories of learning, instructional strategies, and learning outcomes? What is a standard-conform e-assessment system? Why e-assessment systems must be adhere to standards?, and how standards can support into having a flexible e-assessment system?.

5.2. Enriched Learning Experience

Assessment forms provided in e-learning activities have to foster learners to use their prior knowledge to build enriched learning experiences as they participate in authentic and intuitive learning activities. Moreover, essential aspects such learning styles, learning types and theories, teaching trends, and assessment purpose based on learning outcomes must be considered when it comes to design integrated e-assessment forms for enriched learning experience. Therefore, designers of assessment models should consider theories of learning – cognitive, affective, and social learning - during the design of their models. Moreover assessment models should be provided with a framework or guidelines of how to use them in designing and developing integrated assessment for CLRs (AL-Smadi, Hoefler, & Guetl, 2011a).

This sub-chapter discusses an integrated model for e-assessment (IMA) by which providing integrated e-assessment forms for CLR can be guided, as well as a framework of how to develop the designed assessment and feedback forms using IMA and utilize them in learning activities.

5.2.1. Integrated Model for e-Assessment (IMA)

Figure 5.1 depicts the abstract level of IMA which mainly identifies possible tools, practices, guidelines for providing enhanced forms of e-assessment in complex- learning objects (CLO) - such as serious-games and simulations, virtualized collaborative learning- and considers different learning theories. The assessment and feedback part of the model considers four main components: the *learning objectives*, the used c*omplex learning resources*, *assessment and feedback*, and *indicators for evaluation and validation*. As assessment is considered as continuous process (see Chapter 3) the components represented as a cycle implementing an iterative approach of design, develop, and update assessment forms and feedback. Moreover, the enriched learning experience is influenced by other aspects such as *educational and psychological* aspects, *technology* and existing *standards and specifications* (see red arrows in Figure 5.1). Furthermore, quality criteria have to be defined to ensure a high quality standard of all activities in this complex learning environment. Therefore, *quality assurance* which assures adequate use of all components of the enriched learning experience when it comes to design an integrated assessment forms for enriched learning experiences through indicators such as, *educational efficiency* and *effectiveness*. Finally, in order to ensure that the learning experience supports *adaptation*, the model also considers three other important models: the *learner model*, the *knowledge model*, and the *didactic model*, respectively.

The iterative process for designing and developing integrated e-assessment forms - including feedback - for CLRs used to support learners to build enriched learning experiences has the following phases: (a) prepare inputs to the enriched learning experience, (b) design CLR integrated with assessment and feedback, and (c) evaluation and validation. These phases are explained in more detail as follows.

Phase 1: Prepare inputs to the enriched learning experience

An enriched learning experience is affected by several components such as educational and psychological, and technical aspects. Nevertheless, motivational and emotional/affective aspects are expected to influence learning experiences. It has to be mentioned that there is some reciprocal relationship between the educational and psychological components and use of technologies - although it is not depicted in the graph (*cf.* Watson, 2001). The following IMA components are considered to collect input for the design and development of assessment and feedback.

Educational aspect

In order to describe and provide enriched learning experiences, learning theories - such as reflective learning, experiential learning, and socio-cognitive learning - and learning models have to be consulted. For instance, Blooms' Taxonomy (*cf.* Bloom, 1956; Krathwohl, Bloom & Bertram, 1973) is a valuable framework in order to define learning goals as well as assessment activities. Due to these theories, not only individual learning styles can be considered but also processes that affect types of learning (e.g. collaborative and social learning).

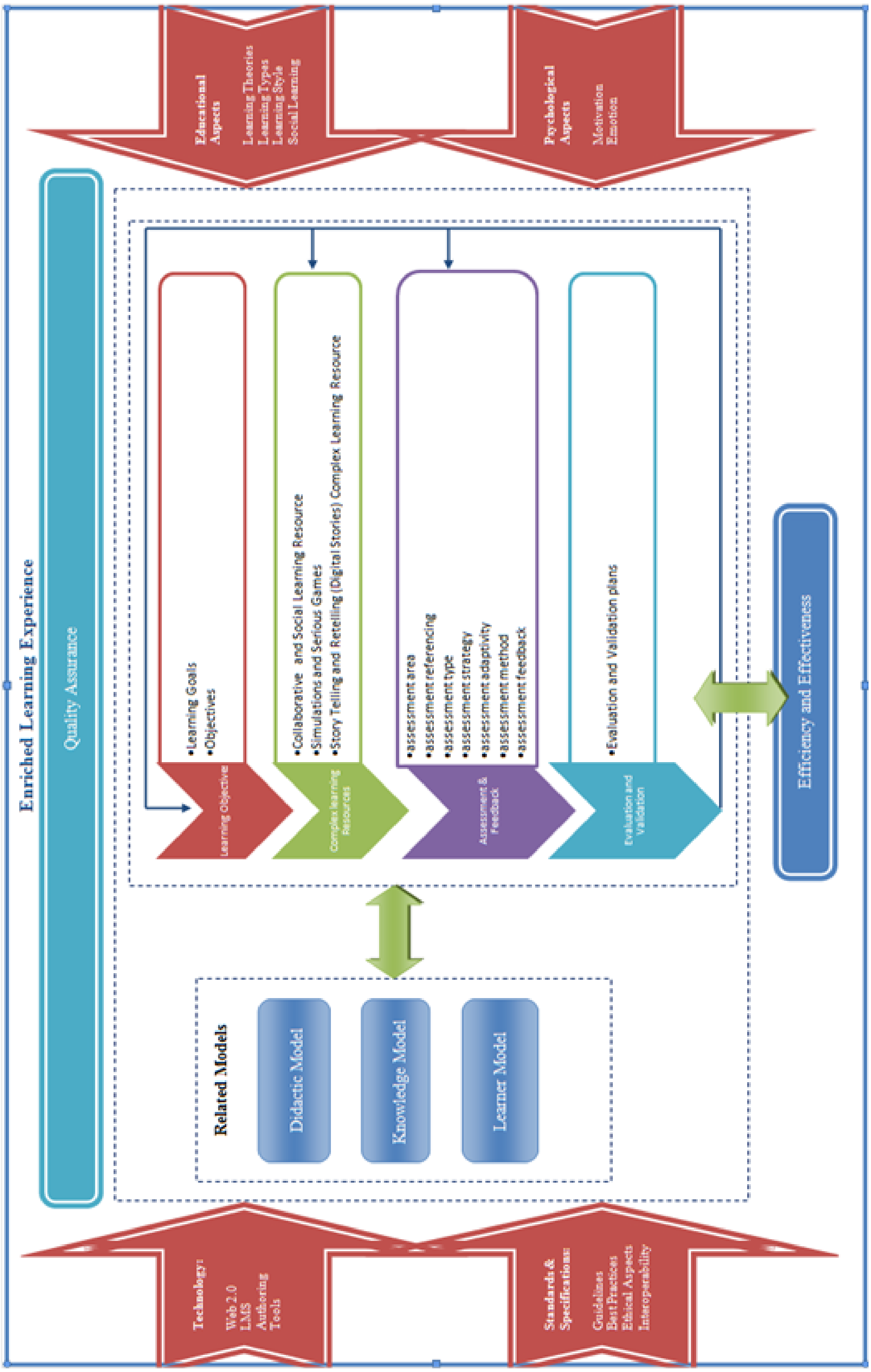


FIGURE 5.1. Abstract view of the integrated model for e-assessment (IMA).

Psychological aspects

Other important issues are motivational and emotional aspects during learning and assessment (see Section 2.4). Due to the measurement of the emotional/affective status of a learner he or she can be supported in a suitable and personalized manner in order to enhance his or her affective/emotional inclination and hence, to stimulate the learners' attention and learning.

Technological aspects

From a technological viewpoint, learning and assessment activities can be supported in many ways. For instance, enhanced technologies cannot only generate CLRs but also able to flexibly adapt the learning path with respect to the individuals needs and learning progress. One aim of IMA is to support in the development of e-assessment tools that support assessment activities in the enriched learning experiences. These tools will not only consider the assessment of individuals and self-regulated learning but also peer-assessment and group assessment. Furthermore, they will provide an adaptive learning path and consider emotional and motivational aspects based on the outcome of the assessment activities.

Standards and Specifications

Educational standards and specifications such as learning content reusability and interoperability, learner's information accessibility and share ability are essential for e-assessment and also for quality assurance. Different standards and best practices have been provided to design and develop e-assessment content and components have been discussed in Chapter 4.

Efficiency and Effectiveness

From an instructors' viewpoint, efficiency and effectiveness of an enriched learning experience are important criteria. For instance, the theory of constructive alignment (Biggs, 1996) describes the compatibility between instruction, learning, and assessment. According to this theory, teaching is more effective when there are alignments in between what teachers want to teach, how they teach, and how they assess students' performance. According to Kellough and Kellough (1999, quoted after Buzzetto-More & Alade, 2006) one aim of assessment is to improve teaching effectiveness. Hence, during the assessment process, effectiveness and efficiency should be considered.

For instance, a factor that might affect efficiency and effectiveness is the question of which tool should be used for which CLR and assessment. Not all tools provided for CLRs and assessment might be useful. Therefore, when selecting an assessment tool, both CLR and didactic objectives have to be considered (*cf.* Ravenscroft, 2001).Moreover, efficiency and effectiveness aspects can be investigated through questions such as, to what extent did the students learn from the provided learning activity? Should there be an individual assessment, or a group assessment, or peer assessment? Should the assessment activity be formative or summative? What exactly should be assessed? The knowledge of the learner or whether he or she can apply the knowledge or even create new applications based on the knowledge they acquired?

Quality assurance

Learners profit from an enriched learning experience most when the standard of the quality is high for activities within the learning experience. Therefore, quality assurance is essential in order to guarantee that the learning experience meet the requirements. The quality can be assured when several aspects are considered. For instance, learning and assessment activities should consider the state-of-the art of best practices and standards in the field. As discussed in Chapter 4, such guidelines should be consulted when assessment is generated but also when it is delivered, scored, and interpreted (*cf.* BPS, 2002). It is also necessary to consider ethical aspects. Such ethical standards are not only addressed to issues like plagiarism or cheating but also the fact that personal information (emotional and/or motivational status, behaviour etc.) is used to adapt the learning and assessment activity – often without the explicit knowledge of the learner. Hence, factors like anonymity, voluntariness, and transparency of the assessment activities are important aspects that have to be covered carefully during the assessment. Furthermore, results from regular evaluation and validation are also valuable indicators in order to measure and improve the quality of a learning experience. A comprehensive framework for e-learning quality, which includes criteria for infrastructure, technical standards, content development, pedagogic practices, and institutional development, as well as a specification of ten pedagogical principles for e-learning, is outlined in (Anderson and McCormick, 2005).

Furthermore, results from regular evaluation and validation are also valuable indicators in order to measure and improve the quality of a learning experience.

Phase 2: Design CLRs integrated with assessment and feedback

After related aspects of educational and psychology, technology and standards, and quality assurance are explored with respect to a specific enriched learning experience, the following aspects are considered to design integrated assessment forms and feedback with CLR.

Learning objectives

The first step designing assessment and feedback considers the *learning objectives*. Learning and assessment activities highly depend on those objectives. Typically, the main learning objective is to achieve the immediate learning goal which is usually defined by the instructor of a course or the stakeholders. Learning goals can be defined - e.g. using Bloom's Taxonomy (Bloom, 1956; Krathwohl, Bloom & Bertram, 1973; see also Chapter 2). Didactic objectives affect the type of learning resources and assessment activities that are chosen in an enriched learning experience. However, there might be some relating didactic objectives during the learning process that are not immediately linked to the learning goals in a narrower sense. Such further objectives could be, gaining social competences (due to collaborative work) or meta-cognitive skills (due to self-regulated learning activities). As those skills might also be very important it is necessary to consider them at the beginning of a learning experience.

Complex learning resources (CLR)

As discussed in Chapter 4, complex learning resource (CLR) can be defined as "*a composite didactic resource consists of one or multiple learning objects (either atomic or complex), accordingly, CLR inherits the features of LO of reusability and interoperability provided by the standards and specifications used to represent LOs*" (AL-Smadi, Wesiak, Guetl, & Holzinger, 2012).

According to constructivist learning theories we actively create knowledge by building explanations of ourselves and our environments (see Section 2.2). To address the needs of an "active learner" who is actively involved in the learning process, an enriched learning experience is made up of complex learning resources (CLR). Those CLR are expected to add moments of collaboration, simulation, and storytelling to support the learners in achieving the learning objectives. Examples of CLRs are, *Collaborative Learning Resources:* collaboration can enhance the learning efficacy because people learn from one another - due to observational learning, imitation, and modelling. Moreover, due to collaboration, learners can be supported in the achievement of specific skills – such as, communication, problem solving, decision making. *Simulation and Serious Games Learning Resources:* Serious games are intuitive learning systems used to train - individual but also groups of - learners while achieving their potential learning goals and through including aspects of education – such as, teaching, training and informing. *Story Telling:* is defined by learning objectives that consist of story scripts which are composited from various situations. The story telling can represent a method of intercultural training mediation in order to foster a cooperation based on training, sharing of knowledge and experiences. The digital story tales are interactive didactic elements, oriented to a student-centred teaching approach able to involve learners emotionally, provide guidance and support reflection.

Interaction with other related models

In order to provide adaptive and personalized learning, IMA is interacting with three other models namely *learner model*, the *knowledge model*, and the *didactic model* (*cf.* Capuano et al., 2009). The *knowledge model* is able to formally represent the information associated to the available didactic resources. In particular it allows the teachers to define and structure disciplinary domains by constructing domain dictionaries - including relevant concepts, and ontologies - organising concepts through different kind of relations. Ontologies are used in synergy with metadata associated to the learning resources in order to allow the dynamic personalisation of learning paths and the automatic evaluation of the students - gaps and competencies evaluation and assessment. The *learner model* is able to capture the knowledge acquired by each learner during learning activities as well as his/her learning preferences - considered as cognitive abilities and perceptive capabilities - with respect to important pedagogical parameters such as: kind of media, didactic approach, interaction level, semantic density. The *didactic model* defines the rules that the system must follow in order to build the best sequence of learning activities to be performed by a specific learner in order to let him/her acquire the selected domain concepts with respect to his/her learner model and according to a given knowledge model.

Assessment and Feedback

Assessment activities represent the core step of this phase. There is a need for forms of assessment which meet the high demands arising from the CLR. Therefore, innovative assessment activities are considered to base on the Bloom's taxonomy (Bloom, 1956) of educational objectives and effective kinds of learning such as reflective learning and experiential learning (Kolb, 1984) as well as socio-cognitive learning (Bandura, 1977).

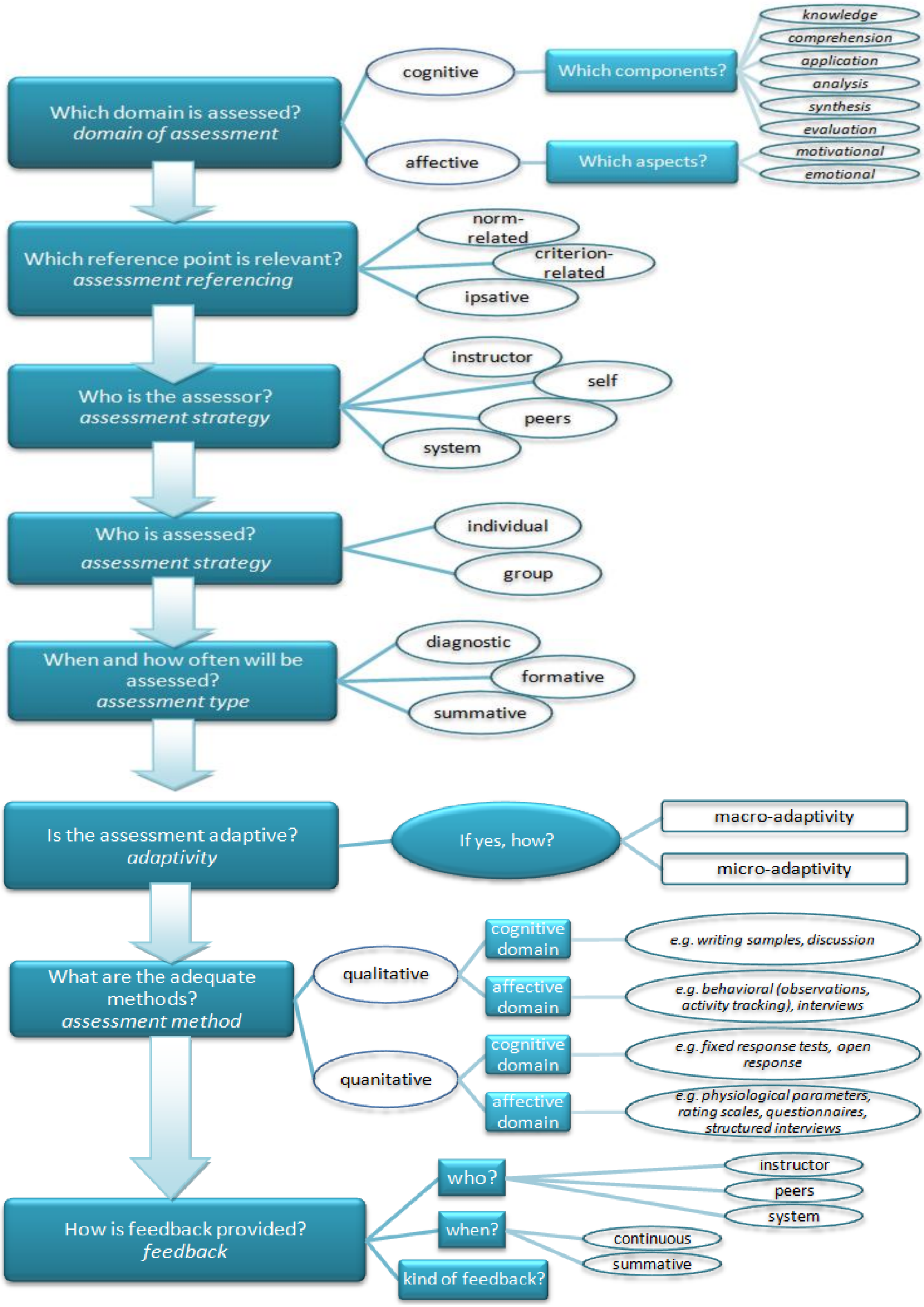


FIGURE 5.2. Methodology to design assessment and feedback.

Besides traditional summative assessment, also diagnostic and formative assessment activities should be considered. Thus, learners are assessed more or less regularly during the learning activity and are supported in reflecting their learning progress. Assessments' forms are designed based on CLR and should involve students in the assessment process. For instance, students can be asked to assess their own work (self-assessment) and/or the work of their peers (peer assessment (*cf.* Prins, Sluijsmans, Kirschner, & Strijbos, 2005; Roberts, 2006). The involvement of students in assessment activities enables students to develop meta-cognitive skills and to find criteria that reflect the quality of their work or the work of their peers. Moreover, forms of assessment should consider emotional and motivational aspects of the learner in order to enhance their learning outcome. Finally, giving feedback is an important issue in context of assessment because learners become aware of their gaps in knowledge, skills or understanding of a topic (*cf.* Boston, 2002), and can accordingly change their learning behaviour.

As depicted in Figure 5.2 in order to design and develop integrated assessment forms and feedback for CLRs, the following aspects should be considered.

Domain of Assessment

Traditionally, *cognitive assessment* activities mainly consider the assessment of knowledge: Learners have to demonstrate that they reached the learning goal by passing a knowledge test at the end of the learning activity. In line with the learning theories that built the background of the enriched learning experience (e.g., Bloom's Taxonomy; see Bloom, 1956), not only knowledge, but also role, skill, and behavioural assessments should be considered. In order to choose an adequate method for the assessment, it is also necessary to specify the level of difficulty, i.e. which competence should be assessed. For this, the six levels according to Bloom's Taxonomy can be used. Additionally, innovative forms of assessment should always cover the learner's affective disposition in order to enhance the learning outcome. Regarding the *affective assessment,* it can be differentiated between the assessment of motivation and the assessment of emotions:

- *Assessment of motivation*: To assess the motivation of the learner, mostly interviews, questionnaires, or self-reports are applied. However, also the methods of observation or activity tracking can be used. The assessment of learners' motivation can give important information about the underlying reasons for their learning progress or missing progress, and is especially important within personalized learning systems, where the choice of learning objects is adapted to the needs of the individual learner. Furthermore, knowledge of learner's motivational state at different points in the learning process can help to improve the learning resources and/or learning environment.

- *Assessment of emotions:* There are different emotional assessment systems, which can be divided into three main areas, namely, psychological (e.g. rating scales, checklists, questionnaires, semantic differentials), physiological (e.g. skin conductance, heart activity, papillary dilation), and behavioural (e.g. facial expression, posture, gestures) assessment. As for the assessment of motivation it mostly relies on self-assessment and should therefore be interpreted with caution. However, also the emotional state of the learner can give important hints on the reasons for a specific learning state as well as on possible improvements from the instructor's side (e.g. if the provided learning content or assessment process causes frustration or irritation on the side of the learners).

Assessment Referencing

This point refers to the reference point that is used to evaluate a learner's status of knowledge. Whenever a student's performance is compared with the performance of peers – i.e. norm-related referencing - or the comparison concerns the individual's actual status with a pre-defined domain – i.e. criterion-referencing – or ipsative referencing by comparing the actual performance with his or her own performance in the past. This latter method of referencing has the advantage, that the individual progress can be monitored (see Section 3.1.4).

When assessing motivation or emotion, the reference can be used to: set an intervention whenever the learner falls below a specified motivational/emotional threshold (criterion-related) or whenever the individual curve shows a downward trend over a longer period of time (ipsative referencing).

Assessment Strategy

The assessment strategy can be classified with respect who could be the assessor into: instructor/tutor, self-, peer-, group-, or system based assessment. Another important issue in the context of assessment is the role of the involved persons. Usually, the learner is assessed by the instructor. However, integrated forms of assessment with CLRs involve students in the process of assessment – such as self-assessment, peer-assessment, and collaborative assessment. Also the performance of a whole group can be assessed. Such assessment activities may also facilitate the work for the instructors, though self- and peer assessment activities need guidance and practice as well. Additionally, in for specific CLR such as serious games, system based assessment can be used. In this case the system or the tool itself detects a pattern of actions which triggers a change of the learning path, a change in the components of a scene, or the whole scene in a non-invasive way (micro-adaptive). Regarding the assessment of motivation and emotion mostly rating scales are used, which are self-assessment strategies. However, affective assessment can also include the measurement of physiological or behavioural parameters, and thus be instructor or system based.

As already mentioned, it is important to focus on the role of the involved persons. Firstly, we described who could be the assessor. So in a second step, we also should have a look on who is assessed. As described above, it can be distinguish between self-, peer and group assessment. In case of self- and peer assessment, the individual assesses him or herself or is assessed by his/her peer(s). In the context of a group assessment, a group's working product or learning process is assessed by students or an instructor. Hence, the individual or the whole group can be assessed.

Assessment Type

Assessment type as discussed in Section 3.1.2 can be classified into: (a) diagnostic assessment concerns students' knowledge and misconceptions as well as affective status at the beginning of the learning process. It is also known as pre-assessment, which can, for example, be used for comparisons with a student's cognitive or affective state at the end of a learning activity, (b) summative assessment takes place at the end of a learning activity and checks whether a learner has reached the learning goal, (c) formative assessments might provide a more valuable outcome for the learning process (see Chapter 3).

In social settings (such as in collaborative learning contexts), the assessment of knowledge can be divided into performance and immediate assessment, where the latter one basically corresponds to formative assessment. In performance assessment, the collaborative process is tracked and used to: evaluate the learner behaviour, progress in terms learning goals achievement, and provide continuous feedback to the learner/group in order to improve their awareness and social experience. On the emotional side, assessments can occur before the collaborative task, in real-time during the task, and retrospectively after the task, which corresponds to the diagnostic, formative, and summative assessment types.

Adaptive Assessment

e-Assessment has the great advantage that it allows personalized testing. Thereby, it can be differentiated into macro-adaptive (concerning the adaptive presentation of learning content and adaptive navigation support) and micro-adaptive (concerning non-invasive interventions effecting the presentation of learning objects). The adaptive assessment step is an important aspect within the whole assessment model and is supported through the interaction with learner, knowledge, and didactic models as explained earlier. Adaptive assessment on the one hand support in providing an adaptive and personalized learning content (such as when it used to diagnose learning style to adapt learning content; see Section 2.5), workflow and learning path support, and on the other hand it also entails an update of the learner model - representing the knowledge state of a learner - and the didactic model - representing the learning preferences.

Assessment Method

There is a wide variety of assessment methods, varying from simple tests, instructor observations, or writing samples to discussions or the analysis of student work. Generally, we can differentiate between (a) quantitative such as points or percentage achieved in a test, ratings, physiological parameters and (b) qualitative assessment such as open ended questions in interviews, behavioural observations methods. For e-assessment, fixed-response (multiple choice; for assessing knowledge) or free response (essay, programming assignments) formats are common. The chosen assessment method depends on: the assessment domain such as multiple choice items for knowledge tests or rating scales for motivational assessments, the assessment type such as instructor, self, or peer as well as individual or group assessment, the assessment type such as diagnostic, formative or summative and last but not least the learning objective as described in Bloom's taxonomy, (Bloom, 1956).

Feedback

Providing feedback is essential practice in the context of assessment by which learners become aware of gaps in their knowledge, skills or understanding of a topic (e.g. Boston, 2002; Garris, Ahlers, & Driskell, 2002) and can hence enhance their learning progress (see Sections 3.3). e-Assessment can be used to automatically provide personalized feedback. However, the quality of the feedback is important in any procedure for assessment (Black & Wiliam, 1998). Hence, feedback should be provided continuously, although not intrusively in a formal or informal way in order to support the learners (Bransford et al., 2004).

Phase 3: Evaluation and Validation

To ensure that learning and assessment activities have a high quality standard, these activities should be evaluated and validated regularly. Evaluation means that a method, procedure, etc. are assessed using predefined quality criteria. However, it is risky to confound results of successful assessment with successful assessment itself. For instance, even if all students have passed a course because they have completed a test successfully this does not mean that the assessment itself was reliable. Perhaps the test was simply too easy. Therefore, evaluation criteria should consider best practices and standards (see Chapter 4) as well as the learning objectives.

Validation means that the measure provides a valid conclusion about the status of a learner. Thus, the underlying assessment activities (and also their underlying technologies) should be validated regularly in order to ensure that they are valid. Results from those evaluation and validation processes form important indicators for the quality of the enriched learning experience in order to adapt/enhance it.

To this end, aspects of how to design integrated forms of assessment and feedback to CLR using IMA through the consideration of educational and psychological aspects, technology and standards aspects, as well as the interaction with other related models – i.e. learner model, knowledge model, and didactic model – have been discussed. However, in order to use this model in modern learning settings a framework is required of how to develop the designed assessment and feedback forms and utilize them in learning activities. The next section proposes a framework to explain how IMA and the framework can be used in developing a CLR integrated with assessment and feedback forms.

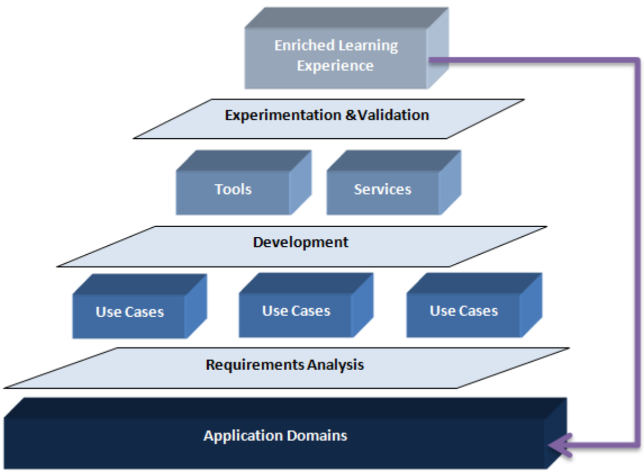


FIGURE 5.3. Bottom-up framework to use IMA.

5.2.2. Framework to Use IMA

The complexity of IMA is driven from its coverage of enriched learning experiences with high complexity and integrity where assessment forms are embedded within complex learning resources such as virtualized collaborative sessions and serious games. Moreover, IMA is affected by the external aspects such as educational settings, technology and standards, and affective aspects. Therefore, a framework has designed and implemented to facilitate using IMA in modern learning settings (AL-Smadi, Hoefler, & Guetl, 2011). The framework uses methodology of how to use IMA in developing educational tools that represents the CLR with integrated forms of e-assessment through a bottom-up layered framework (see Figure 5.3).

The provided framework is built up of layers representing continuous process and implementing an iterative approach to design, develop, and evaluate integrated assessment and feedback forms with CLRs. On the top of the framework is the enriched learning experience as a main outcome of the experience that target users will gain through using the developed assessment forms within the CLRs in modern learning settings (see Part 3 of experimentation and validation). The framework is a continuous process where the output of each layer can be fed back to the one before in order to update and enhance the methodology. As depicted in Figure 5.3 the framework provides a methodological approach where the following steps are followed:

- Define *Application Domain:* in this step required information about the application domain is collected. This step answers questions about IMA main components of learning and didactic objectives, complex learning resources, assessment activities (including feedback), and indicators for these forms evaluation and validation. Moreover it investigates inputs to the IMA represented by psychological and pedagogical aspects, affective and emotional aspects, available technologies and specifications and standards. More precisely this step investigates questions such as, what are the learning objectives? What is the learning style? What kind of learning? What are the available tools and software? What are the available specifications and standards? Are there any recommended guidelines? Moreover, questions about didactic objectives and suitable assessment forms can be answered. Nevertheless, information about target users as well as whether the learning scenario is personalized, adaptive, or not. In this step experimentation and validation planning starts, these plans can be updated later on in next steps based on the progress of the learning scenario design and tools development.

  However, collecting all of these information can be a tedious task and requires experts to participate in this a step. Therefore, it is recommended to model application domains based on the relationships extracted from the aforementioned questions into an ontology. An ontology is simply the collection of classes and relationships representing the domain based on common understanding and agreement. Having application domains ontologically modeled and represented on Semantic Web facilitates the extraction of answers and useful knowledge regarding the modeled application domain.

- Conduct *Requirements Analysis:* the collected information in the previous step is mapped into functional and non-functional requirements. Requirements are used to develop or enhance available tools to be applied in the learning scenario. Moreover, to identify suitable quality assurance policy and to identify technical efficiency and effectiveness parameters related to these requirements. Not only instructional designers and tool developers should participate in this step but also target users such as students and teachers should participate as well.

- Build *Use Cases:* the identified requirements from last step are used to build Personas that represent possible use cases. Use cases explain the interaction within the context of the learning scenario. The interrelationships among possible users, the tools, and the context are more explained. Once the use cases are built the experimentation and validation plans are updated and finalized.

- Develop *Tools and Services:* the information extracted from the first step regarding available software and technologies as well as specification and standards with the use cases and requirements from the other steps are used to develop and improve software and tools. In case of available tools they can be used as they are, or enhanced to achieve the technical and pedagogical requirements by embedding the assessment forms identified in the first step. If no tools are available, then tools and services are developed from scratch following guidelines and standards if possible and using the requirements and use cases from previous steps as well as by embedding or integrating assessment forms or assessment tools.

- Conduct *experiments and Validate results:* the experimentation plans are used to conduct experiments based on the designed learning scenario and the developed tools and services. Results are then analyzed against the predefined hypotheses in the experimentations plan. Evaluation and validation plans are used to evaluate the tools and to validate the learning scenario in general.

- Update *IMA:* the outcome of this framework is an enriched learning experience of integrated assessment forms with CLR to achieve specific learning goals in a learning scenario. The findings from last step are used to update the data model (could be an ontology) representing IMA based on conducted evaluation and validation plans. It is worth mentioning that in each step a feedback can be provided to the steps before to update and continue with the framework. This is depicted by the back arrow - which closes the cycle and represents the continuity of the process.

For the scope of this doctoral dissertation, IMA and its bottom-up framework have been used to support in the design and development of two CLRs, an automated and integrated assessment in self-directed learning (see Section 7.2) and collaborative writing and peer-review (see Section 7.4).

As discussed earlier in this sub-chapter considering available standards and specifications as well as best practices and guidelines for e-assessment is an important aspect that IMA considers. However, considering the findings from literature in the domain of educational standards from Chapter 4, it is still not clear how assessment tools and system should adhere to standards and specifications. Moreover, Chapter 4 discusses some guidelines of to provide quality e-assessment with respect to pedagogy, technology. To this end, the next section answers the following questions: What is a standard-conform e-assessment system? Why e-assessment systems must adhere to standards?, and how standards can support into having a flexible e-assessment system?.

5.3. Standard-conform e-Assessment System

In order to have flexible e-assessment systems, standards conformation is an essential requirement (see Chapter 4). According to (AL-Smadi, Guetl, & Helic 2009; Shepherd, 2006), standards foster e-learning systems in general and e-assessment systems in particular to be: *interoperable* - the ability of different systems to share information and services in a common file format (see Section 4.2.2), *reusable* - refers to the ability of using the learning content by different tools and platforms, *manageable* - how much the system is able to keep track on the learning experience and activities, rather than the ability of tracking how learning objects are created, stored, assembled and delivered to users, *accessible* - the ability of customize, access and deliver learning contents from anywhere and anytime, *durable* - means that the learning content does not need any redesign or redevelopment even with new versions of the system, *scalable* -refers to the ability of the system to grow from small to large, and *affordable* -is the system affordable?

Thorne (2004), identifies a set of standard-conform levels in order to have flexible and interoperable tools and services: (a) *Data and information (content):* e-assessment content has to be represented using common specifications and standards (e.g. IMS QTI, IMS LIP) so that different tools can share and reuse their content in a flexible manner. (b) *Communication (transport and protocols):* tools have to use common platform independent communication protocols (SOAP, HTTP) so that they can easily communicate to share functions, activities, or content. (c) *Software Interfaces:* that forms as a contract between service provider and consumer (e.g. OKI OSID). Moreover, interfaces represent an abstraction level to tools and services which make them easily integrated into LMSs. (d) *Domain Models:* that provides a common conceptual understanding of the problem domain in general and e-learning domain in particular. Domain models help developers to have common understanding with input/output data, data representation, possible services, and their workflow to achieve specific goals. Examples of this are the e-learning Framework (ELF) and Framework REference Model for Assessment (FREMA).

5.3.1. Conceptual e-Assessment System

AL-Smadi, Guetl, & Helic (2009) define standard-conform e-assessment system as "*the system that its components are designed and implemented according to specific standards*". The term components makes the definition general, thus to more clarify it Figure 5.4 depicts the conceptual e-assessment system and its three main parts. The first one is the core e-assessment system which should be flexible to work as a stand-alone system or to be integrated with other system – e.g. LMS. The second part is the interface which is responsible for the external communication between the core system and the other external systems and tools. The interface should adhere to standards which keeps the core system flexible and interoperable – such as content standards (SCORM, IEEE LOM, IMS QTI, etc.), and interoperability standards (OKI OSID, CCSI, IMS LTI), see Chapter 4 for more information. The last part is the system target users which could be end users – i.e. standalone approach- or other systems – i.e. integrated approach.

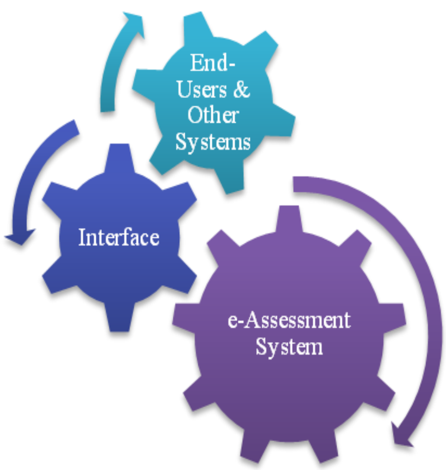


FIGURE 5.4. A Conceptual e-Assessment System (AL-Smadi et al., 2009b).

In order to have a flexible e-assessment system, two levels of standard-conform have been defined (see Figures 5.5, 5.6). The consideration of these two levels fosters the overall e-assessment system to be standard-conform. Consequently, it will be flexible and interoperable. The two levels are explained as follows.

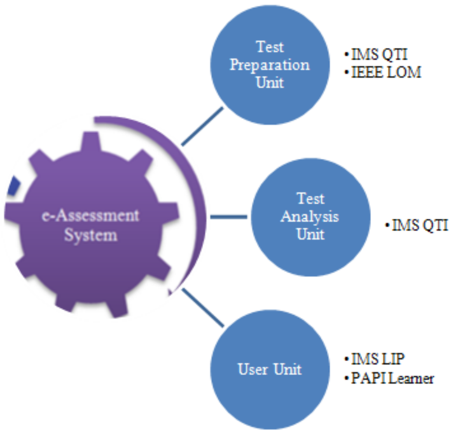


FIGURE 5.5. Internal level of standardization (AL-Smadi et al., 2009b)

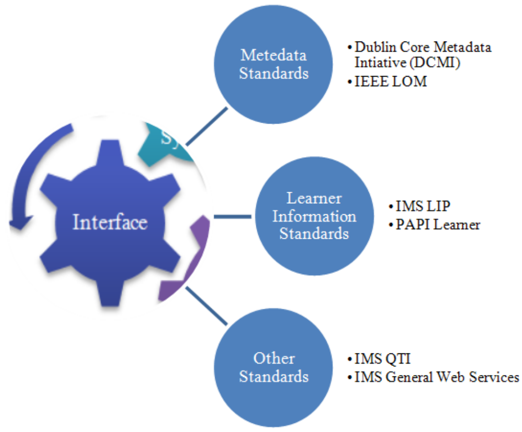


FIGURE 5.6. External level of standardization (AL-Smadi et al., 2009b)

The Internal Level

In this level, the core e-assessment system' components should be conformant to specific types of standards and/or specifications (see Figure 5.5). As discussed in (AL-Smadi, Guetl, & Kummar, 2010), in general e-assessment system has four main modules: Authoring, Scheduling, Delivering, and Reporting. For each of these modules, a set of processes has been defined. The use of these processes as well as their modules during the assessment life cycle depends on the requirements of the application domain as well as the needs of the assessment system. Figure 5.7, shows those main phases as well as their corresponding processes, and they are as follows:

- **Authoring:** as part of this phase the author user starts with selecting the form of the assessment which can be a test, quiz, assignment, or a survey. After that, she/he defines the objectives behind the assessment items and the assessment process in general (*cf.* Bloom's, 1956). These objectives shall guide the assessment process to achieve the learning goals. Based on those objectives the items' types are chosen. For each item, a feedback can be defined. The feedback can be pre-defined or it can be provided based on the later on marking, grading, or analysis of the learners' answers in the reporting phase. Marking criteria can be defined in this module for each of the assessment items. For instance, rubrics can be used to handle the marking process. In case of the author is preparing a test, the ability of importing items or selecting pre-prepared items from items' banks as well as arranging the items in tests is available. Finally, the items are assured for quality so they can be ready for the next phases as well as to be exchanged with other systems. This process is done by quality assurance bodies who consider matters such as learning goals and objectives and standards-conformation.

- **Scheduling:** after the assessment has been prepared and quality assured, it will be ready for the delivery phase. Before delivery, managing the users who will set the assessment as well as the assessment environment is done during this module. Such processes mainly are done by the 'timetabler' user-role. The 'timetabler' selects the place where the assessment will be taken and the timing matters such as the assessment date, how long it will stay and how many times it can be repeated or taken?. In case of the assessment will be delivered in a controlled environment the 'timetabler' chooses the invigilators who will monitor the learners during taking the assessment. In the other case of web-based delivery matters such as security, and plagiarism detection should be part of the assessment system.

- **Delivering:** in this module assessments are delivered in different forms: paper-based, web-based, offline delivery, or via third-party such as LMS and different assessment systems. In case of some assessment items are imported from different systems or exported to different systems, matters of exchange should be considered. Web services can be used to deliver such items in both cases of import or export.

- **Analyzing & Reporting:** once the assessment is taken by the learners, answers are gathered and stored in a related database. The answers are then marked according to the defined marking criteria ranging from semi- to fully-automated marking with reference to items types degree of complexity. For instance, multiple choice questions can be fully-automated marked whereas free answers may require human interventions – i.e. tutor/teacher - for final judgments. After marking the assessment the final marks are assigned to grades and reported to the learners. The learners' results can be further analyzed and reported to decision makers as well as used to moderate and adapt the learning and assessment activities with regards to their goals and objectives.

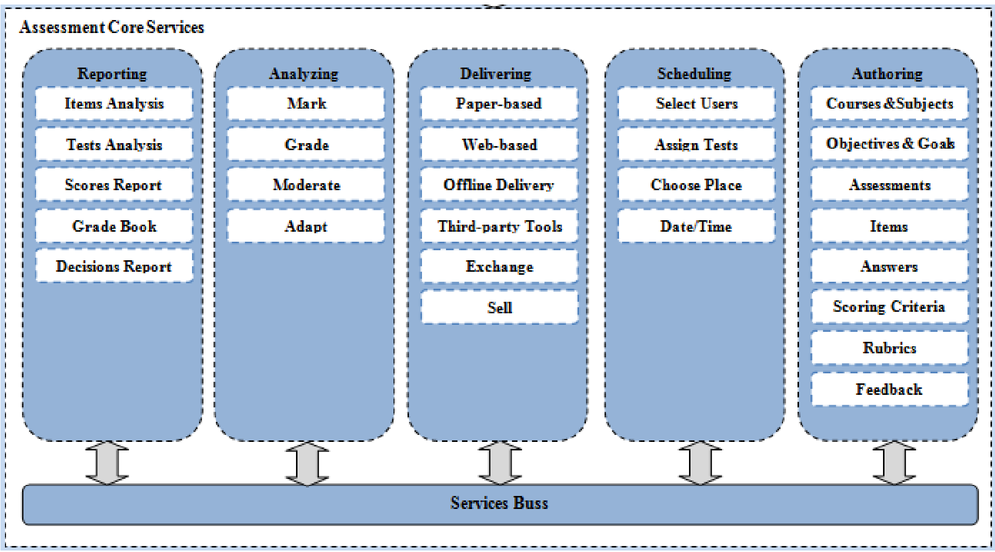


FIGURE 5.7. e-Assessment Modules and Services (Al-Smadi & Guetl, 2010).

Figure 5.5 shows some of e-assessment system levels and the possible standard(s) and/or specifications that can be used for representing them. The Test Preparation Unit is responsible for the purposes of tests Authoring and Delivery. A specification such as IMS QTI is used by this unit during the test authoring. In cases of having learning objects related to the test we may use the IEEE LOM standards. The tests can be analyzed by the use of Test Analysis Unit which is based on the same type of specifications to provide a feedback (timely or immediate) to the system users (individuals or organizations). The system users are managed by the User Unit which is a standard-conform to provide some services as user personalization and modeling. Standards such as PAPI Learner or IMS LIP can be used.

The External Level

The external level is represented by the functionalities of the interface that is underpinned by a set of available standards and/or specifications. The interface is responsible for the external communication between the e-assessment system (the internal level) and the target users. Through this interface information such as questions/exercises and answers, users' information, list of enrolled students, courses information and learning objectives can be shared with other systems and tools. The more standards this interface supports the much more flexible the e-assessment system will be. As shown in Figure 5.6, different examples of possible standards and specifications that the interface should be flexible to support.

Referring to Thorne (2004), levels of standard-conformation mentioned earlier, the conformation to content standards are depicted in the internal level e-assessment standard-conformation, whereas the communication and software interfaces standards are covered in the external level through the e-assessment interface. So what about the domain models and frameworks? Domain models help developers to have common understanding with input/output data, data representation, possible services, and their workflow to achieve specific goals.

A framework represents a rich vocabulary that is used to support people in the domain as well as software developers to overcome the problems encountered through the software development. A framework is also used to create a shared language that will describe the problems and their solutions. In the domain of e-learning and e-assessment, frameworks are not used to build generic LMS/e-assessment systems. Rather than, they encourage 'coherent diversity' where common service definitions are provided and used to achieve the diverse goals of the organizations. Therefore, the organizations unique infrastructures will stay coherent and consistent with respect to each other. A framework also supports organizations to develop service-oriented architectures by identifying the main services that these organizations may need to develop their applications. The main aim of a framework is to have the ability to identify the services as well as to assign one or more open standards and specifications for each service. (Davies & Davis, 2005)

Considering the importance of having frameworks to support standard-conform e-assessment, the next section proposes a service-oriented framework for e-Assessment (SOFA) in order to achieve the latest aspect from Thorne recommendations as well as to foster the standard-conform e-assessment system with a clear understanding of possible services through depicting the two levels of standard-conformation discussed earlier.

5.3.2. Service-Oriented Framework for Assessment (SOFA)

According to (Davies & Davis, 2005) a service-oriented framework consists of a set of services, each of which represents a pattern to solve a particular problem. Services can be represented in different ways, as a web service, as an API, as well as a manual business process. The process of defining the framework services is called "factoring services", where an ongoing process is done by the organizations to identify a service to solve a specific problem and then group the related services - e.g. domain related services, problem related, etc. - in a framework for further enhancement and development. The more fine-grained the services are the more flexible and supportive the framework. There is no current best practice for factoring services, but the fine-grained they are the more useful they will be. Moreover, there is no concern about how the service is implemented as the concern of the service functionality and behavior towards the other services. A service-oriented framework may support e-assessment systems to easily share and exchange test between each other's. Services for tests, items, results, users information, etc. can be easily implemented in the system and they are flexible to be used by other authorized services or systems. For example, students that are registered for a specific test can only attend the e-learning course in other system and vice-versa.

Service-oriented architectures foster the development of modular and flexible systems (Milligan, 2003), where the components of the system are flexible to be added, replaced or removed. As well as, new systems can be composed from a collection of suitable services. Based on what we have discussed earlier and a step towards the flexible e-assessment system design and implementation, a Service-Oriented Framework for Assessment (SOFA) has been designed.

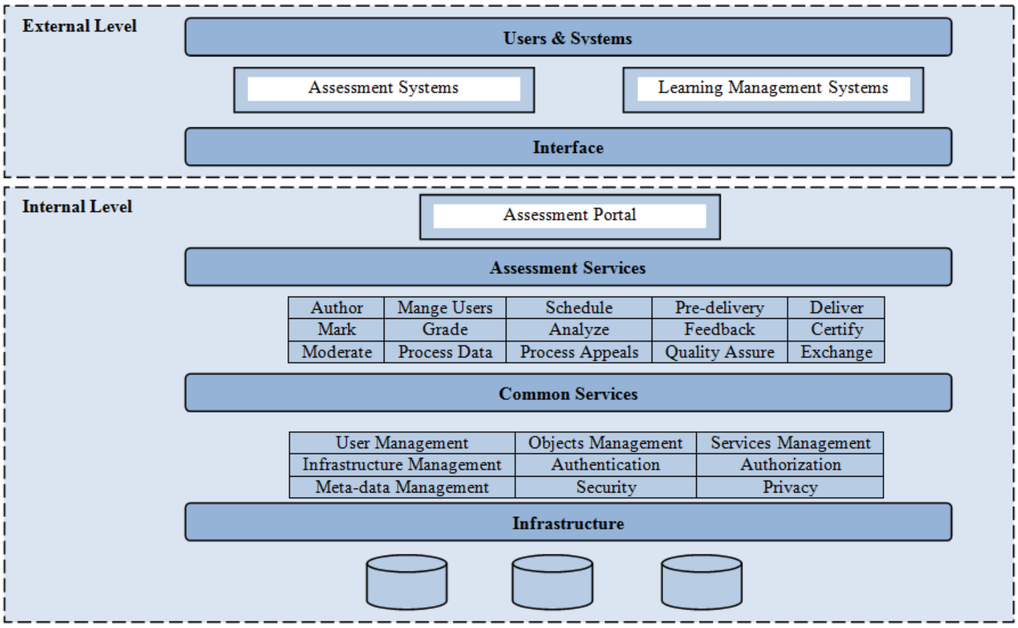


FIGURE 5.8. Service Oriented Framework for Assessment (SOFA) (AL-Smadi et al., 2009b)

Figure 5.8 shows the SOFA layers and services. SOFA has five abstraction layers as follows:

- *Users and Systems* represent the external possible users, tools, and systems that may interact with the e-assessment system. Such as, assessment systems and LMS as well as any other authoring tools.

- *Interface* as discussed earlier, the interface is used for the external communications between the e-assessment system and the other external systems, users, and tools. The interface layer should be underpinned with a set of specifications and standards in order to facilitate the integration and communication between the core e-assessment system and the external user agents.

- *Assessment Services* represent the fundamental services for any e-assessment system. The services in this layer are used to perform the main functionality of the assessment process from authoring the items until exchanging them. Special interfaces are used to make the interaction between these services and the assessment portal and users. For which specifications and standards of the internal level of the e-assessment system are used – e.g. IMS QTI, IMS LD.

- *Common Services* a lower level of services those are not assessment-specific such as authorization and authentication.

- *Infrastructure* represents the internal networks, storage and processing capabilities that the e-assessment system requires.

SOFA has two layers of services in the internal level: assessment services and common services. The assessment services in SOFA have been identified based on FREMA (Framework REference Model for Assessment) processes concept map (Millard et al., 2005). All of the services in this group are assessment services and work together in order to support the assessment process. The group of Common Services is a set of services that may be found in any assessment system or any other system such as e-learning systems. The services are organized in a set of layers based on the IMS GLC Abstract Framework (IAF) (Smythe, 2003) which consists of four main layers, the "Application Layer", the "Application Layer Services", the "Common services" and the "Infrastructure Layer". The assessment services are described as follows:

- *Author* this service is responsible for creating the tests, items, templates and items' pools. This service should be standard-conform in order to have an interoperable components. An example of possible standards and specifications is IMS QTI.

- *Manage users* this service maintains the secure logins and passwords to the e-assessment system. This process is handled by identifying the possible groups of the users as well as the roles for them. Possible specifications and standards can be IMS LIP or PAPI.

- *Schedule* sets up the test where the required hardware and software systems are identified as well as the candidate users of the assessment. This service is usually used by the timetabler.

- *Pre-delivery* handles the necessary operations after authoring and scheduling of a test and before delivering this test.

- *Deliver* presents the items to the candidate. According to FREMA definitions, this service involves the following processes: (1) deciding next item to be presented (2) retrieving the item file and its related resource files (3) displaying the item, which may require specialized software (4) reading in (and perhaps validating) the candidate's response (and perhaps confidence level).

- *Mark* assigns a mark or a score to the candidate response.

- *Grade* assigns the final mark to a grad, for example the mark 90 is assigned to a grade "A".

- *Analyze* The candidate responses as well as their marks and grades are analyzed for providing feedback as well as for further moderation and enhancements.

- *Feedback* displays useful information for the candidates during/after the assessment process.

- *Certify* this service is responsible for the different processes of candidate performance recording ranging from the paper certificates to the electronic certificates in e-portfolios.

- *Moderate* checks the assessment process is satisfactory or not.

- *Process data* produces useful information after the test is finished. This information is usually used by the stakeholders and decision makers.

- *Process appeals* allow the candidates to appeal against their final grades.

- *Quality assure* assures the quality of the assessment processes ranging from the fairness of an item to the satisfactory of the assessment processes to the institution goals and objectives.

- *Exchange* responsible for the exchange and purchase of the tests and items.

5.4. Summary

As discussed in Chapter 2, quality education requires an alignment between instruction, learning, and assessment (Birenbaum, 2003; Biggs, 1999). Therefore, it is required to consider teaching strategies, learning objectives, learning theories and pedagogy when it comes to design assessment. Assessment has evolved to advocate alternative forms of assessment - such as, performance assessment, self and peer-assessment, behavioural assessment, portfolio-based assessment, and rubric-based assessment - through which high level of metacognitive skills are evaluated, on-going feedback is provided, and students are more engaged in the learning process (Birenbaum, 2003).

Figure 5.9 shows the assessment aspects and the shift in the assessment paradigm for the so called "new culture of assessment". In this new culture the assessment is considered as a tool for learning. Moreover, it becomes part of the learning process and represented as integrated assessment forms. Nevertheless, students have more responsibility in the learning process in general and in the assessment activities in particular. They become more engaged in developing assessment criteria, participating in self, peer-assessments, reflecting on their own learning, monitoring their performance, and utilizing feedback to adapt their knowledge, skills, and behaviour (see Chapter 3).

In addition to aspects depicted in the Figure 5.9, Segers, Dochy, & Cascallar (2003) discuss another important aspect based on assessment relation to learning. In this aspect assessment has changed from being assessment *of* learning (summative) - provides an evidence if the student has achieved the learning goals, and certifies the whole learning process - to assessment *for* learning (formative) - integrates assessment into the learning process - where teachers have more control than students and use different tools to integrate assessment in the learning process and provide students with feedback to improve their learning experience, and assessment *as* learning (formative) - defines the learning process as it is experienced by students - in which students are more responsible in relation to their learning and assessment; they control their learning, define assessment criteria, participate in self, peer-assessment, reflect on the learning process, monitor their performance, and decide what to do next during their learning experience. (AL-Smadi & Guetl, 2011a)

In order to address the challenges and problems which outcome from this shift in assessment paradigm, new forms of learning experiences - outlines constructive, cognitive, and social learning - enriched with complex learning resources - designed based on instructional and learning objectives - integrated with new forms of assessment - e.g. performance assessment, self and peer-assessment, and behavioural assessment - should be considered. However, designing and developing such forms of assessment requires a comprehensive understanding of the aspects that may influence assessment, learning and instruction. Despite the richness in assessment models and practices - see Chapter 3 - they lack to some extent the alignment to instruction and learning. Therefore, this chapter proposes an integrated model for e-assessment (IMA) through which limitations of the available e-assessment models are addressed and new forms of assessment are considered.

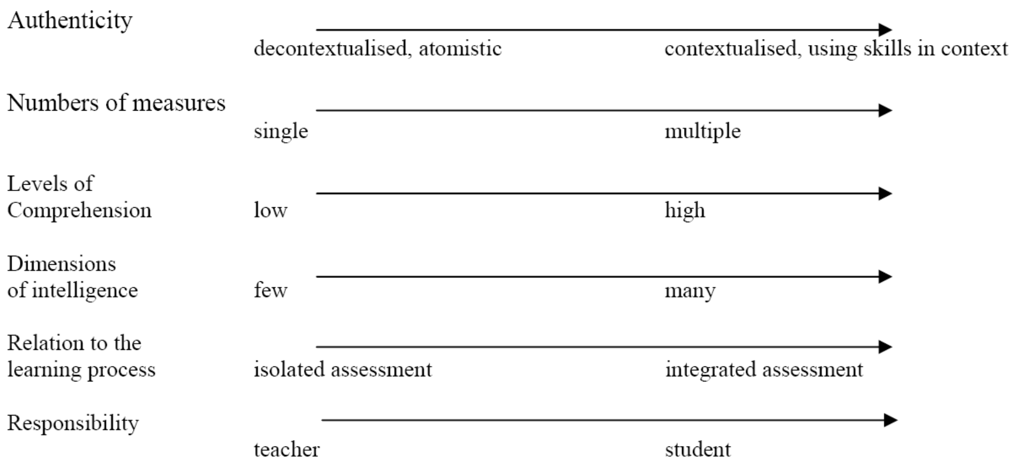


FIGURE 5.9. The shift of assessment paradigm (Segers, Dochy, & Cascallar, 2003, p. 3).

e-Assessment standards - see Chapter 4 - lack the representation of alternative forms of assessment as well as the integration with CLR represented by complex learning objects. Therefore, this chapter discusses the importance of having standard-conform e-assessment system which not only adheres to content standards - e.g. IMS QTI - but also considers learning tools interoperability specifications - e.g. IMS LTI - when it comes to provide integrated assessment to CLRs. In addition to that, providing assessment that is aligned to instruction and learning requires flexible technology which leads to flexible pedagogy. Adopting flexible and accessible software architecture fosters assessment tools to be flexible and to be used in several application domains thus, meeting different pedagogical requirements (AL-Smadi & Guetl, 2011; Dagger et al, 2007). Therefore, this chapter proposes a solution framework namely Service-Oriented Framework for e-Assessment (SOFA) in order to have a flexible e-assessment system.

This chapter forms the first part of the solution framework towards providing flexible and interoperable integrated forms of e-assessment for complex learning resources. The part of how to design integrated assessment with CLR is covered in this chapter whereas the next chapter explains how interoperable software architecture - i.e. SOA - has been used to develop integrated forms of e-assessment - such as self, peer-assessment, automated assessment, rubric based assessment, and performance assessment - for CLRs - such as collaborative and virtualized learning resources and serious games and simulations.

Chapter

6

6. Service-Oriented Flexible and Interoperable e-Assessment

6.1 The Application Scopes
6.2 Architectural and Functional Design
6.3 Development of the System

Within the context of this doctoral dissertation, the notion of flexible and interoperable e-assessment has been clarified in the last chapters as the e-assessment system that is (a) standard-conform (see Section 5.3) and (b) underpinned by a rich and comprehensive model for e-assessment covering essential aspects such as educational and psychological, technology and standards, and a clear guidance of how to design integrated forms of assessment - including feedback - with CLRs (see Section 5.2), thus to be (c) pedagogically flexible to support different application contexts such as collaborative writing, game-based learning and self-directed learning. Moreover, this e-assessment system (d) should be developed based on flexible software architecture that provides a technical flexibility through accessible services interfaces and transparent data transportation and communication.

Based on these general requirements, this chapter concentrates on the latest point of flexible software architecture and describes the design and development of a service-oriented flexible and interoperable e-assessment environment. The Service-Oriented Flexible and Interoperable e-Assessment system (SOFIA) is the result of research and development on distinct research projects at Institute for Information Systems and Computer Media at Graz University of Technology (IICM). SOFIA represents a flexible and standard-conform e-assessment system that can be used as a standalone tool or can be extended by third-party tools and services to provide assessment for different learning contexts. SOFIA has been lately applied in the EC-funded project entitled "Adaptive Learning via Intuitive/Interactive, Collaborative and Emotional System" or ALICE[48] from 2010 to 2012 to provide integrated assessment forms for complex learning resources.

In order to place the developed system in its theoretical and application context, the first part of this chapter provides an overview on the application scopes. This part sheds the light on possible application scenarios for e-assessment through which the diversity of application contexts for e-assessment can be depicted as well as the limitation and problems in standards and practices of e-assessment can be identified. The second part of this chapter discusses the architectural and functional design of the system which has been used to cover the development aspects discussed in the third part of this chapter. The chapter concludes with a summary and remarks on to which extent the developed system fulfills the essential requirements.

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

This chapter is based on (AL-Smadi, & Guetl, 2011a; AL-Smadi, & Guetl, 2011b; AL-Smadi & Guetl, 2010; AL-Smadi, Gütl, & Kannan, 2010; AL-Smadi, Guetl, & Helic, 2009a; AL-Smadi, Guetl, & Helic, 2009b; AL-Smadi, Guetl, & Kappe, 2009).

6.1. The Application Scopes

In the steps have been taken towards providing flexible, interoperable, and integrated e-assessment for CLR, an integrated model for e-assessment have been designed to consider aspects such as, educational and psychological, technology and standards during the design and development of the e-assessment forms. Moreover, the aspects related to standard-conform e-assessment have been defined and possible e-assessment standards have been identified. SOA as a flexible software architecture has been selected to underpin the technological aspects for tools and services development. However, to investigate the applicability of the solution framework different application scopes have been explored to provide application domains for experimentation and validation.

In order to identify application scopes, the next sub-sections discuss application scenarios for e-assessment in general and align these scenarios with available standards and specifications from Chapter 4 in a way to figure problems and challenges.

Application Scenarios for e-Assessment

In order to have a better understanding of why we need a flexible and interoperable e-assessment, and to identify the main requirements for SOFIA, as well as to show the diversity of possible application contexts, moreover to figure the limitations of the available e-assessment systems and standards and guidelines, this section outlines a set of application scenarios for e-assessment in modern learning settings:

- **e-Assessment services for Job recruitment:** WebSys is a software company that requires any job applicant to have a specific certificate related to their system. They are looking for a tool that can be integrated to their system with the ability to prepare tests to evaluate the new applicants. In order to handle this need and to prepare factual knowledge questions based on the selected content, the SOFIA must have a modular design that facilitates the process of integration with the current system. Also SOFIA should adhere to e-assessment standards to facilitate the integration process especially when it comes to provide sharable and accessible test material and services.

- **Semi-automated knowledge assessment and feedback:** Ali is a lecturer in a university who teaches Management Information Systems for students of the second year in the college of Management and Administration. His didactic objectives include the evaluation of understanding level of the factual knowledge through a continuous assessment. To do that, he decided to use SOFIA to deliver tests and to analyze the results through providing continuous feedback during the course. SOFIA should have flexible and user-friendly interfaces to help him to author his tests and deliver them to his students. As well as helping him to (semi-) automatically generate the tests based on the selected contents and to assess the results. Furthermore, SOFIA should be designed to analyze the answers of the students and provide a feedback which makes it useful for Ali to conduct continues assessment during his courses.

- **Semi-automatic test item creation for teaching Algebra:** Sara is a university lecturer and teaches Algebra to undergraduate students. One of her didactic objectives is to use computers to assess and assist students during here courses. She believes that when her students practice Algebra on computers and do more and more on-the-fly provided exercises they can easily pass the course. In this situation, SOFIA should provide her with flexible and easy to use interfaces to author algebraic questions and save them in the item bank. Then, the SOFIA itself can generate a set of exercises to the students and assess their answers based on the answers had been prepared by Sara before, or based on the algebraic engine that the tool should have - i.e. step by step evaluation and feedback. Moreover, SOFIA must provide a feedback to the student about her/his competence level in Algebra and an appropriate feedback about the progress of the student during the course.

- **Online rubrics-based scoring:** Jake is a teacher in a high school and he is interested in applying a set of online rubrics to assess the students' results according to a specific criteria. Regards online rubrics, SOFIA should be flexible to help Jake to design a set of assessment rubrics to (semi-)automatically grade the students' results based on the rubrics criteria and mastery levels.

- **Performance assessment for Online Collaborative Learning.** Anna is a student in Computer Science Department and she has to participate in a collaborative learning activity within a group of peers as part of Software Engineering course. Anna may not have previous experience in online collaboration within a group. She has to participate within a group in a collaborative writing activity where her performance will be assessed. SOFIA provides a third-part tool for collaborative writing integrated with forms of assessment to evaluate Anna's performance and provide her continuous feedback.

- **Online collaborative learning with guidance for distance learning**: Elena is a student in Computer Science and she has to participate in a small virtual group (4-5 members) to carry out a software development project at a distance. She has experience in computer programming, however the project sets high level requirements and needs that demand intensive collaboration during the whole quarter. Elena may not have previous experience in collaborating with other people, especially at a distance. She will certainly need guidance and support by her teacher who should be able to monitor individual and group work throughout the experience. SOFIA provides third-party tools for collaborative learning in distance and supports with integrated assessment and feedback.

- **Collaborative writing with continuous assessment for blended learning**: Eric is an assistant-professor in Computer Science Department. Eric is teaching Software Engineering for undergraduate students. Eric has been teaching the Software Development course for more than five years. Over the years he identified problems regarding a great variety of student's knowledge and motivation; he also has somehow to deal with different types of students, from inexperienced fulltime students to experienced part time students. This year he is intended to offer blended learning activities and improve the course with collaborative writing activities. Moreover, he is interested to continuously evaluate student's performance and knowledge competencies as part of theoretical and practical software development activities. SOFIA provides an extension to use a third-party tool for collaborative writing through providing services interfaces to maintain user groups and to facilitate assignment authoring and assessment.

- **Semi-automatic test item creation for blended learning**: John is Lecturer at the Medical University in Graz; he has to provide several test items of different types for each of the learning content items both for self-assessment for the students and to prepare the final test. SOFIA processes the content, extracts concepts and creates test items, and enables John interactively to change concepts and/or test items.

- **Automated test item creation for self-directed learning**: Caterina interested in the subject history of Art. She retrieves quality content form open online repositories and websites. Because of her learning style, she wants to take a pre-test before reading the content and afterwards to take a post-test to get an indication of the level of knowledge acquisition. SOFIA provides Caterina appropriate test items for each selected document or learning content.

- **Automated test item creation as service**: Julia has completed her PhD on e-education and e-assessment. During her studies, she found out the market potential of an e-assessment service to support creating and administering knowledge assessment based on given learning content. She offers both the semi-automated creation of test items in an interactive approach and automated creation of test items. Her customers can either administer the test as part of the service offered or they can download the created test items in a standard-conform format. SOFIA should be designed to act as a service and supports customers through its assessment modules for tests authoring, deliver, scoring, and reporting.

- **Dynamic assessment and feedback for game-based guided intuitive learning:** Mark is an elementary school teacher and responsible for civil defence training, in a scenario to train pupils how to act during fire evacuation, he uses game-based learning approach of using a serious game in blended learning paradigm. In order to provide dynamic feedback and guidance he wants to define assessment rules with feedback as a consequence which can be provided to pupils as they interact with the game. Moreover, he is interested to track pupil's behaviour within the game and conduct post-analysis within the e-assessment system to compare pupils' progress with themselves and with their peers. SOFIA provides and flexible author for these rules of assessment and helps Mark to define feedback as well as provides and assessment engine to praise pupils' interactions with the game and their post evaluation.

To this end, after showing the variety of e-assessment applications scenarios, a set of application scopes have been indentified to investigate the solution approach proposed for this doctoral dissertation. These application scopes are, self-directed learning, collaborative learning, and game-based learning. The selected application scopes cover different pedagogical approaches and learning theories. Thus, gives more possibility to investigate the applicability of the solution approach. A short overview about each application scope is as follows.

Self-directed Learning

The terms self-regulated learning and self-directed learning are used in research interchangeably and basically have the same meaning. As cited in (Bracey, 2010) Knowles defines self-directed learning as "*self-directed learning is a process in which individuals take the initiative, with or without the help of others, in diagnosing their learning needs, formulating learning goals, identifying human and material resources, choosing and implementing appropriate learning strategies, and evaluating learning outcomes*". Where self-regulated learning has been defined by Pintrich and Zusho (quoted after Nicol, & Macfarlane-Dick, 2006) as "*Self-regulated learning is an active constructive process whereby learners set goals for their learning and monitor, regulate, and control their cognition, motivation, and behavior, guided and constrained by their goals and the contextual features of the environment.*"

By referring to both definitions, self-regulated learners set their learning goals, plan and use effective strategies to achieve their goals, manage resources based on that, monitor their understanding, and assess their progress towards their goals (Zimmerman, 2002). However, self-directed learning is more used when it comes to discuss e-learning or distance learning without the support of course instructor (Bracey, 2010). A primary goal for higher education is to support students and lifelong learners to be independent, self-motivated, as well as self-regulated learners (Bracey, 2010; Zimmerman, Bonner, Kovach, 1996). Thus, self-regulated learning fosters students with the skills required for life-long learning (Dynan, Cate, & Rhee, 2008).

According to (Nicol, & Macfarlane-Dick, 2006) the more learning process become self-regulated, the more students have control over their learning and the less dependent they are on the course instructor. However, self-regulated learning has to link with suitable assessment activities. Assessment is a main part of the learning process. Therefore, learning strategies of self-regulation and self-direction have to be supported with enhanced and suitable assessment forms (AL-Smadi, and Guetl, 2011). Such forms of assessment should support learners to monitor their performance and understanding. Moreover, should provide students valuable feedback so that to adapt their learning strategies towards their learning goals.

Collaborative Learning

'Collaborative Learning' has been defined by Dillenbourg (1999, p.1) as "*a situation in which two or more people learn or attempt to learn something together*". In this broad definition, Dillenbourg refers to three dimensions in collaborative learning: the *scale* of the collaborative situation (group size and time span), the type of *learning* may occur in collaboration, and the depth of *collaboration.* However, Dillenbourg discusses four aspects for the theory of collaborative learning: learning *situation*, *interactions*, learning *mechanisms* or *processes*, and *effects* of collaborative learning. Interactions are influenced by the learning situation conditions and learning situation is categorized – individual or collaborative - based on interactions taken place in that situation. Nevertheless, there is a relation between interaction and learning process, and learning process and effect. The interaction with a situation activates specific cognitive processes which influence affects collaborative aspects such as group shared knowledge.

Cooperative and collaborative are used synonymously in research to describe learning in groups. However, Dillenbourg (1999, p.8) distinguishes between them based on the mechanism the learners use to divide and accomplish work. In cooperation, the work is divided by the partners into sub-tasks on which they work individually and the final product is provided by assembling these individual works. Moreover, in collaboration partners work 'together', with some spontaneous division of tasks more likely happen even they work together.

Game-based Learning

The use of games technology for learning is not new and online games have been available for more than a decade. According to Kriz (2003), interactive-learning environments foster knowledge transfer, skills and abilities improvement in general and social skills in particular. Variety of educational games has become available to increase learners' motivation, support collaborative learning, and may foster students to gain knowledge (Gütl, 2010). Sophisticated and simultaneous online games exist for a small group of players to multiplayer online games (MMOG). World of Warcraft is the most popular MMOG Western title reaching over $1.4 billion in consumer spending in 2008 in North America and Europe and since 2005, the cumulative spending on subscriptions is reaching over $2.2 billion (Screen Digest, 2009). (AL-Smadi, Guetl, & Chang, 2011)

As discussed earlier in Section 3.2.2, providing assessment for game-based learning holds some challenges especially when it comes to provide dynamic evaluation and feedback for player's – i.e. students – progress and interactions within the game (AL-Smadi, Guetl, & Chang, 2011; Shute, Ventura, Bauer, & Zapata-Rivera, 2009; Kickmeier-Rust, Steiner, & Albert, 2009). Therefore, in this chapter an enhanced approach of integrated assessment for game-based learning is discussed.

The application scopes - i.e. self-directed learning, collaborative learning, and game-based learning - cover a variety of learning types and pedagogical approaches such as, collaborative and social learning, experiential learning, problem-based learning, and self-regulation (see Chapter 2 for related learning theories). Therefore, CLRs covering these learning types and pedagogical approaches have been developed and integrated with alternative forms of e-assessment designed based on IMA and SOFA (see Chapter 5).

In order to identify the limitations of e-assessment standards as well as the limitations of available assessment tools to provide assessment services not limited to but cover the aforementioned scenario the next section discusses these aspects and helps to emphasize the need for integrated flexible and interoperable e-assessment.

6.2. Problems and Challenges

This section discusses the problems and challenges for developing integrated, flexible and interoperable e-assessment system based on the previously discussed chapters - i.e. e-education, e-assessment, and standards and specifications for e-assessment - the application scenarios in previous section.

Based on the scenarios discussed in previous section, this section shows some recommendations and limitations on the available standards. In the scenario of the Company, the e-assessment system should be flexible to work as a standalone system or to be integrated with other systems such as the case in this scenario. To make this applicable the e-assessment system must have a modular design so that some modules can be integrated with other systems such as the system in this company. In this situation the problem of standards appears where the e-assessment system must support the standards used in the other system. Therefore, it is recommend that the e-assessment system should support as much as possible of the available standards.

The second scenario is a traditional one where the e-assessment system is applied as a standalone system to deliver and assess the students' tests and provides feedback. The limitation of standards appears again in the third scenario where a mathematical representation of the question (symbolic representation) is needed. For example, when the student is going to solve an equation the e-assessment system should have the capability to support symbolic representation for the solution. Furthermore, a standard such as IMS QTI do not have the ability to represent the solution as a set of symbolic representation for the equations using XML. Therefore, no reference answer is available to automatically assess the students' candidate answers and provide them valuable feedback. One of the other limitations of IMS QTI specification is rubrics representation. The problem appears in the fourth scenario where online rubrics are needed to assess the students answers based on a specific criteria.

The application scenarios depict the target context as it covers mainly - but not limited to - two contexts of (a) industrial and commercial domain, in which e-assessment systems are required to act as services to support customers to manage assessment activities, and (b) educational domain, in which e-assessment systems have to provide assessment forms for different types of learning - such as blended and distance learning - considering pedagogical aspects such as - learning goals and outcomes as well as learning styles - in a way to foster learning trends - such as collaborative learning, game-based learning, and self-directed learning - through extending the e-assessment system services with third-party tools which represents CLR - such as automatic question creator, collaborative writing and peer-review, and game-based learning assessment engine.

Another limitation in e-assessment standards appears in the domain of game-based learning assessment or in the performance assessment in general - when it is required to track behavior and evaluate progress. Thus, the limitation of IMS QTI as it represents more generalized item types and does not consider learning objectives and didactics. Although, IMS QTI supports interactive item types such as 'Hot spot' or 'Graphic-order' these item types cannot be used especially when it comes to maintain non-invasive feedback and guidance within the game (*cf.* Kickmeier-Rust, Steiner, & Albert, 2009). Despite the current practices to provide assessment for game-based learning (*cf.* Moreno-Ger, Burgos, Sierra, & Fernández-Manjón, 2008; Shute, Ventura, Bauer, & Zapata-Rivera, 2009; Kickmeier-Rust, & Albert, 2010), the approaches they used are not widely recommended as specification or even good practices.

Providing an assessment environment that can be extended with domain-based tools and services to provide integrated assessment services designed for specific application contexts has become a need. As depicted in the application scenarios, it is to some extent challenging to provide a single e-assessment environment to assess all application domains. However, extending the e-assessment services through interoperable domain-based - such as collaborative learning assessment, group-assessment, rubric-based assessment, automated assessment - could be a promising solution. The problem is that e-assessment services considers content standards and in somehow lacks learning tools interoperability standards (see Chapter 4 for more information). Recently, research groups started considering the support of learning tools interoperability as an essential requirement for standard-conform e-assessment systems (*cf.* AL-Smadi & Guetl, 2011b; AL-Smadi, Guetl, & Helic, 2009b).

As discussed in Chapter 4, one of the most important problems and challenges of designing a standard-conform system is the so-called impedance mismatch between the features offered by the standard and the ones needed in a particular application domain (Helic, 2006). For example, IMS QTI is a specification that provides a questions/test description for the authoring tools. In addition to that it supports the development of question/test databases that have a common schema which makes them easily sharable and interoperable; it also provides a common definition for interfaces that facilitates the creation and retrieval of tests and results (Davies & Davis, 2005). Even though the IMS QTI has these features it still has some difficulties in the application domain (such as, foreign languages teaching, teaching Mathematics, or Programming assessment). One of these difficulties is that the IMS QTI is designed to formulate general types of questions and does not take into consideration some specific questions and test types for a particular domain (Milligan, 2003). For instance, Crossword puzzles which are used in the domain of foreign language teaching are an example of those not supported question types by the QTI (Helic, 2006). According to (Smythe & Roberts, 2000) the QTI standard are not related to didactical issues and tries to be didactically neutral as possible.

Another example is what authors of (Recker & Wiley, 2001) have noted about the IEEE LOM (Learning Object metadata). They noted that IEEE LOM from a perspective of metadata don't provide enough information to support the learning process. According to (Devedžic et al, 2007) some developers find parts of IEEE LOM too *restrictive* or *imprecise*. And they also argue that the amount of metadata is not enough to facilitate the search and retrieval of the LOs. Another major challenge is the problem of selecting the most appropriate standard in cases of having different types of standards for the same aspect of the Learning Management System (LMS) (Devedžic et al, 2007). For example IEEE PAPI Learner and IMS Learner Information Package (LIP) both of them are related to the issue of learner modeling. Even though they look similar but there are a lot of differences in the way how they model the learner. Therefore, the developer should have a good understand of the current available standards and the main requirements that helps him to choose the most appropriate standard.

Summarizing the aforementioned problems and limitations, the following aspects can be highlighted:

- Limitations in e-assessment standards such as IMS QTI, (a) to provide test items for specific domains such as, mathematics assessment, programming assessment, or (b) to provide representation for alternative assessment such as performance assessment, behavioral assessment, or rubric-based assessment. Moreover, (c) no consideration of assessment referencing (see Section 3.1.4) - norm-related, criterion-based or ipsative- in the response processing and scoring, (d) there is no support for specific types of assessment such as – self, peer, group-assessment – on the level of test item authoring and scoring.

- Lack of pedagogical aspects in e-assessment standards - such as IMS QTI – as they do not consider, learner preferences, learning style, didactic objective, and learning objectives.

- Available e-assessment systems and tools lacks to some extent the alignment with learning theories and learning outcomes, questions such as what to assess? How to assess? Why to assess this way? When to assess? Who will assess? Who will be assessed? What is the learning type? What is the learning style?, are often not considered in the design and the development of assessment tools. Thus, e-assessment lacks the support and consideration of feedback in terms of type, frequency, format, and content.

- e-Assessment tools and systems usually adhere to e-assessment content standards such as IMS QTI and lacks learning tools interoperability standards conformation. Thus, extending e-assessment services with domain-based services through third-party tools requires extra effort in terms of redesign and redevelopment.

- e-Assessment standards and tools lack to some extent adaptive aspects neither on the level of content presentation nor on the level of the services and functions navigation.

To this end, by considering these problems and challenges as well as the application scopes discussed in Section 6.1, requirements in terms of pedagogical and technical flexibility, as well as alignment with learning theories should be considered in defining the requirements for SOFIA. Next sub-chapter discusses this in detail and identifies essential requirements to meet the main goal of providing integrated, flexible and interoperable e-assessment. Moreover, it explores some application contexts - represented by CLR - and provides architectural and functional requirements for SOFIA.

6.3. Architectural and Functional Design

In order to define the requirements for SOFIA this sub-chapter explores the application contexts emerged from the application scopes discussed in Section 6.1. Based on that, the requirements for SOFIA are illustrated and a comprehensive architecture is provided by which the solution for extending MASS assessment modules with the third-party tools - represented by the application contexts developments - services and tools is depicted.

6.3.1. Architectural Overview of the Application Contexts

SOFIA evolved through the development of requirements emerged from distinct research projects covering different application contexts as follows:

a) a flexible e-assessment system that can be used as a standalone or to be used with other systems and tools. As a result a Modular ASsessment System for Modern Learning Settings (MASS; (AL-Smadi, Gütl, & Kannan, 2010)) has been developed.

b) an enhanced approach for Peer-ASSessment (PASS) (AL-Smadi, Guetl, & Helic, 2009a; AL-Smadi, Guetl, & Kappe, 2009), by which new features of candidate answer marking have been used. Using PASS students are capable to select specific parts from the candidate answer and mark them as *correct*, *wrong*, or *irrelevant*. Special colors are used to tag the selected part of the candidate answer in order to help students giving a reasonable final score and to provide visual feedback for the answer owner. A web based flexible tool has been developed and used as third-party tool to extend SOFIA for the domain of peer-assessment.

c) an Automated and Integrated Assessment in Self-directed Learning in which standard-conform tests - i.e. QTI-based - are created automatically using an Automatic Question Creator tool (AQC; (Gütl, Lankmayr, Weinhofer, & Höfler, 2011)) based on textual learning material. The scenario aims to support self-directed learners with automated assessment and feedback to scaffold their self-awareness about their knowledge state (AL-Smadi & Guetl, 2011a; AL-Smadi, Hoefler, & Guetl, 2011).

d) an Enhanced Approach for Collaborative Writing and Peer-review integrated with self, peer-assessment which can be used in internal peer-review based on groups' actions or for groups-assessment underpinned with assessment rubrics to maintain group's production function and task/social awareness as well as to provide feedback (AL-Smadi, Hoefler, & Guetl, 2011b).

e) an Integrated Assessment Approach for Game-based Learning by which flexible assessment scenarios are provided to evaluate players - i.e. students - progress within the game environment based on post evaluation of tracked behavior or dynamic guidance represented by formative 'stealth' assessment and feedback (Dunwell I., AL-Smadi, & Guetl C., 2012).

As a result from the multiple requirements stated so far, SOFIA has been designed and developed through service-oriented approach to be flexibly extended via third-party tools to provide integrated and interoperable assessment tools for different application domains – i.e. e-assessment in general, self-directed learning, collaborative writing and peer-review, and game-based learning. To this end, next sections give an overview about those application contexts and their associated SOFIA third-party tools.

Modular Assessment System for Modern Learning Settings

The variety in e-assessment application domains has caused several academic and commercial assessment tools to be developed. Some of these assessment tools are limited to specific application domains. Moreover, they are not carefully designed to be standard-conform as well as they barely coordinate to other tools and systems in different application domains of assessment. This has caused universities and higher education institutes to have more than one assessment tool. Managing several assessment tools are money and time consuming as well as requires extra resources. In order to tackle such problems and challenges possible solution could be a flexible e-assessment system that can be used as a standalone or to be extended through aggregating tools from different assessment application-domains.

The Modular Assessment system for Modern learning Settings (MASS) has been developed as a generic and flexible e-assessment system. Form the architectural point of view (see Figure 6.1), MASS is designed and developed based on the notion of standard-conform e-assessment system discussed in Chapter 5. Thus, MASS has the main modules for e-assessment namely: *author, schedule, deliver,* and *report,* moreover MASS uses IMS QTI 2.1 to represent assessment content - i.e. items, tests, and feedback. MASS modules have been developed to provide the two levels of services - i.e. assessment services and common services - based on the service-oriented framework for assessment (SOFA; see Section 5.3).

In the context of e-assessment, the term interoperability mainly refers to the ability of sharing learning objects (LO) between different assessment systems. For the sake of LOs interoperability, LOs should be standard-conform. Once they are standard-conform they can be easily shared between different standard-conform e-assessment systems (AL-Smadi, Guetl, Helic, 2009b; see Chapter 5). Furthermore, these LOs can be reused to create new assessment activities without a need to build them from scratch. According to (Costagliola, Ferrucci, & Fuccella, 2006) almost all of content standards require the following aspects regarding LOs communication: (a) *launch:* the requirements for launching an LO in a web-based environment, (b) *application programming interface (API):* the interface of methods to be invoked by an LO to communicate with LMSs, and (c) *data model:* the dataset for the communication process.

These LOs communication aspects have been considered in developing MASS where the IMS QTI data model has been used to develop a 'QTI_Manager_Service' which is used to interpret and validate imported or authored items and tests. Moreover, MASS has assessment player which can be used to play QTI-based items and tests. Nevertheless, MASS is able to import/export IMS QTI 2.1 meta-data representing items and tests. Thus, tests authored by

MASS can be delivered in a standalone approach or can be shared and reused by other assessment tools.

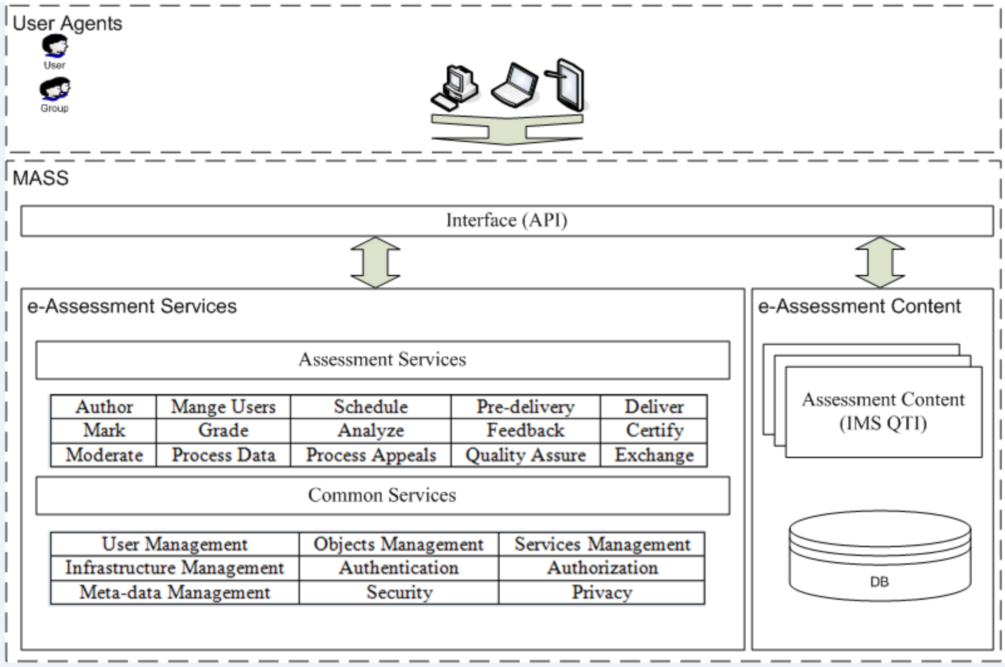


FIGURE 6.1. MASS Architecture based on SOFA.

Enhanced Approach for Peer-assessment

The requirements for this prototype have emerged from an internal research project to develop an enhanced approach for Peer-ASSessment (PASS) (AL-Smadi, Guetl, & Kappe, 2009). PASS provides new features of candidate answer marking by which students are capable to select specific parts from the candidate answer and mark them as *correct*, *wrong*, or *irrelevant*. Moreover, special colors are used to tag the selected part of the candidate answer in order to help students giving a reasonable evaluation and to provide visual feedback for the answer owner. A web based flexible tool has been developed and used as third-party tool to extend MASS for the domain of peer-assessment.

Figure 6.2 demonstrates the overall architecture of PASS. PASS has three main modules: *user management module*, *test management module* and *results analysis and feedback module*. PASS has been developed using .Net 3.5 framework in particular C# for the code-behind logic and ASP.Net for the front-end presentation layer, and MySQL database[49]. The system applies the MVC (Model-View-Control) approach[50]. Moreover, PASS has been developed as part of a flexible e-assessment system project - i.e. SOFIA - based on the SOFA framework services and the standard-conform assessment tools (AL-Smadi, Guetl, & Helic, 2009b; see Chapter 5).

---

[49] http://www.mysql.com/

[50] http://en.wikipedia.org/wiki/Model-view-controller

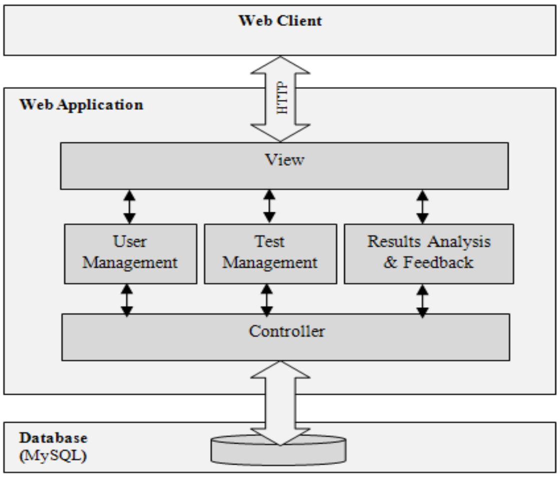


FIGURE 6.2. PASS component based Architecture (AL-Smadi, Guetl, & Kappe, 2009)

The three PASS main modules are:

- **User Management Module**: from its name, this module handles the authority levels of the systems' users. According to the diversity of the systems' users we have identified three main roles: *administrator's role*, *teachers' role*, *tutor's role* and *student's role*. Other roles like parents and decision makers can be easily constructed using this module. This module also handles the login/logon processes based on the users that have been created and the roles that they belong to. Moreover, this module uses the common services form SOFA framework - i.e. user management, group-management, authentication, authorization.

- **Test Management Module:** represents the core module in this application. This module is responsible for tests authoring, assessment activities, items preparation, reference answers, marking and final grading. Teachers have the facility to define an assessment activity based on a specific learning goals, define tests, create test items, assign items from the items pool to specific test(s) with reference to the test goal(s), as well as granting privileges to students and tutors roles or individuals to participate in these tests and activities. This module uses the assessment services form SOFA framework - i.e. author, schedule, deliver, score, and grade.

- **Results Analysis & Feedback provision Module:** this module computes the final grades of the different assessment activities that took place during this experiment. Dedicated results' analysis and mining is conducted in this module to support students, teachers and other related decision makers with a valuable feedback. This module implements the other two parts in terms of feedback provision by using the assessment services of SOFA framework related to score and provide feedback.

This tool has been designed to be used as a standalone whereas in order to extend MASS scoring services a flexible design is required. Therefore, realizing this enhanced approach for peer-assessment requires MASS to have a flexible design to be extended with the third-party tool - i.e. PASS in this scenario. Moreover, PASS should adhere to tools content and interoperability specifications thus to be utilized in complex learning resources such as peer-assessment. Therefore SOFIA environment has been used to support realizing this scenario and flexibly use PASS in the context of peer-assessment and group-assessment.

Automated and Integrated Assessment in Self-directed Learning

The requirements for this scenario emerged from the requirements for the ALICE project. This scenario aims to foster learners with automated assessment and feedback in the context of self-directed learning. In this scenario learners can use some keywords from course curriculum and use them to search and select their learning materials. The keywords represent some concepts that the learners have to understand out of the course. An automatic question creator (AQC; (Gütl, Lankmayr, Weinhofer, & Höfler, 2011)) tool is used to create simple questions (open-ended, single-choice, multiple-choice, and fill-in-the-blank) based on textual content materials. The questions are then used by MASS author module to automatically create tests and deliver them to the learner.

The scenario can be adapted and utilized in different assessment types, for instance the scenario can be used in a formative way to support the student learning, or it can be used to assess the students learning at the end of the learning unit in a summative way. The instructor can be given more control to intervene in the step of creating the test out of the automatically created questions for a specific topic material, by considering learning objectives (*cf.* Bloom, 1956) to select the suitable test items. Hence the scenario can be utilized in a summative approach.

From the technological point of view, the implementation of this scenario requires a flexible extension of MASS by integrating the AQC tool as a third-party tool to support in automatically create test items based on textual learning material. Figure 6.3 depicts the architectural design for this enhanced e-assessment approach. Three main components have been utilized or implemented in order to realize this approach, (i) an e-assessment system capable to interact with LMS and selects learning material – i.e. MASS in this case, (ii) an automatic question creator tool (AQC), and (iii) IMS Question and Test Interoperability (QTI) web service - i.e. QTI_Manager_Service - in order to interpret, validate, and create QTI questions which adhere to the IMS QTI specifications.

For the sake of design flexibility, a web service interface was developed to handle those interactions between the tools and the QTI business logic represented by the class library. In addition to the flexibility influence this decoupling between the QTI class library and the tools implementation facilitates further enhancements and changes especially when a new version of IMS QTI is published. In this case potential changes will take place mainly on the level of the web service. (AL-Smadi & Guetl, 2011)

The AQC tool has been developed by the research group of Advanced Educational Media Technologies (AEMT) at Graz University of Technology. The goal behind developing AQC is to provide a tool which supports the creation of text items or even generates them automatically from the textual learning content. AQC utilizes an automated process to create different types of test items out of textual learning content, more precisely to create single choice, multiple-choice, completion exercises and open ended questions. AQC is capable to process textual learning content stored in various file formats, extracts most important content and related concepts, creates different types of test items and reference answers, as well as exports the those items in IMS QTI format. AQC architecture (see Figure 6.4) has a language dependent data flow and process chain design provides multilingual test item creation, currently English and German are supported, whereas a flexible extension to other languages - e.g. Italian - is possible.

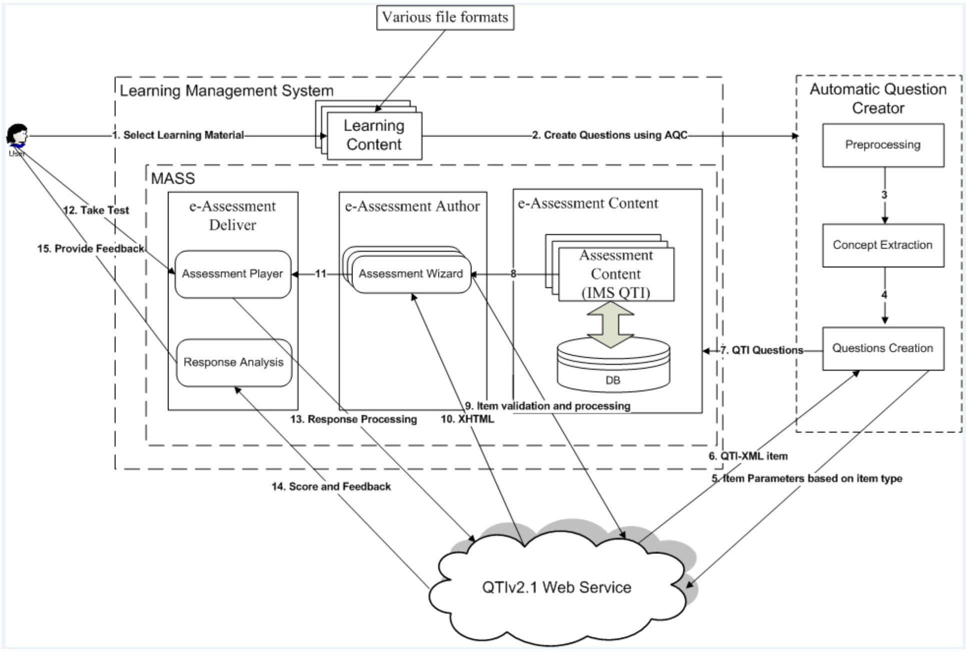


FIGURE 6.3. Architecture for automatically created formative assessments to support self-directed learning (AL-Smadi & Guetl, 2011a).

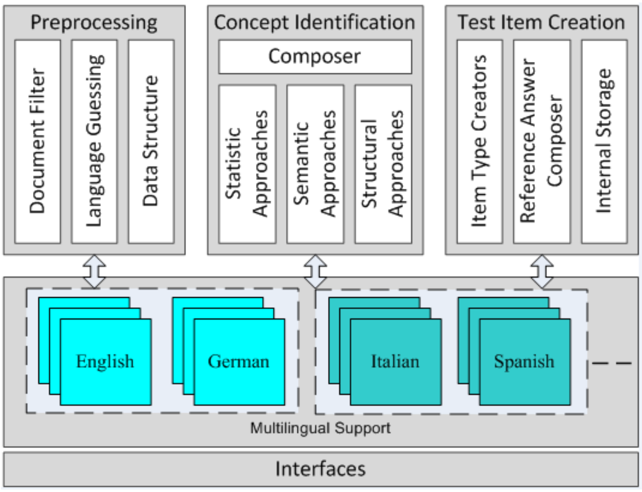


FIGURE 6.4. Enhanced AQC after (Gütl, Lankmayr, Weinhofer, & Höfler, 2011).

From the point view of SOFIA, realizing this enhanced approach for e-assessment requires MASS to have a flexible design to be extended with the third-party tool - i.e. AQC in this scenario. Moreover, AQC should adhere to tools interoperability specifications thus to be utilized in complex learning resources such as self-directed learning courses. Therefore SOFIA environment has been used to support realizing this scenario and flexibly use AQC in the context of self-directed learning.

Enhanced Approach for Collaborative Writing and Peer-review

The requirements for this scenario emerged from the requirements for the ALICE project, in which self, peer-assessment approaches has been suggested to evaluate the learner's performance during a collaborative learning activity. The course instructor should be able to provide collaborative learning activities based on the course learning goals. By finishing those collaborative activities the learner is supposed to have specific competencies and skills.

The scenario aims to provide a new form of assessment for collaborative writing using wiki, where self-, peer-assessment forms are used to evaluate students' contributions. The course instructor uses the author module to create a "collaborative writing assignment" based on wiki. Then he assigns the assignment to groups of students. The students start the collaborative activity by using the discussion forum or the chat tool to divide the work among them. Then they start writing using the wiki based on their tasks. As part of their contribution each student is required to assess his self-contribution using pre-defined comments, and may rate his contribution with a rate out of "5" for instance where "0" is not important and "5" is very important. Moreover, the students should also peer-assess their peers contribution among the group. Similar comments and ratings can be used. Nevertheless, the students should be provided with valuable feedback represents their peer's progress as well as their peers' comments on their contribution. The importance of contribution in both self-assessment and peer-assessment is based on this question "how much do you think that this contribution is important towards the final group product".

In order to support the instructor as well as the students with quality feedback an enhanced visualization tools are provided. The tools retrieve the students' contributions and visualize them in an enhanced way. The tool may support the instructor to mark the student's performance where the following questions could be answered. *How much has each student contributed to the assignment product? How collaboration taking place? To what extent the students are collaborating within the group?*

Co-Writing Wiki - developed for this scenario - is an enhanced wiki for collaborative writing and peer-review. Co-Writing Wiki is enhanced with tools to maintain task-, social-awareness and group well-being (AL-Smadi, Hoefler, Guetl, 2011b). ScrewTurn wiki[51] has been selected to be enhanced with features of the Co-writing wiki. ScrewTurn wiki is open source wiki developed using C# and ASP.Net for the front-end presentation layer. The engine is partitioned in two main blocks. The Core Assembly contains all the business logic, such as data management and caching, content formatting, provider's configuration and loading and system configuration. The Access control is directly performed by the ASP.Net pages, which also take care of the content presentation and user interaction.

As depicted in Figure 6.5, Co-Writing Wiki utilizes the available services from the wiki module to provide enhanced services for collaborative writing. For instance the extensions of *Assignment Manager and Assessment Manger* utilize the group management and document management provided by the wiki system to author and deliver co-writing assignment with peer-review. Moreover, Co-Writing Wiki can be used as a standalone tool or can be integrated other LMS to support co-writing assignment and peer-review.

[51] ScrewTurn Wiki - Free ASP.NET Wiki Software [http://www.screwturn.eu/]

For the sake of integration, the Co-Writing Wiki is fully integrated with SOFIA where a single sign-on (SSO) mechanism is applied as Co-Writing Wiki should be extended with some services in order to realize this scenario. Moreover, the tool interacts with the *learner model* and *knowledge model* services form LMS to maintain personalisation and adaptation of the learning activities.

However, Co-Writing Wiki main requirements and competitive advantages include:

- Enhanced tools to maintain **task and social awareness** and to support group well-being and production function during a collaborative writing assignment.

- **Integrated self, peer, and group-assessment** activities with the use of assessment rubrics designed for scientific writing.

- **Continuous Feedback provision** for learner scaffolding as well as for teachers to follow collaboration progress.

- **Visualization tools** to support both students and teachers to know who did what and when.

- **Motivational Charts** in order to motivate peers to contribute and work in comparison with others in the same group as well as to motivate groups to contribute in comparison with other groups.

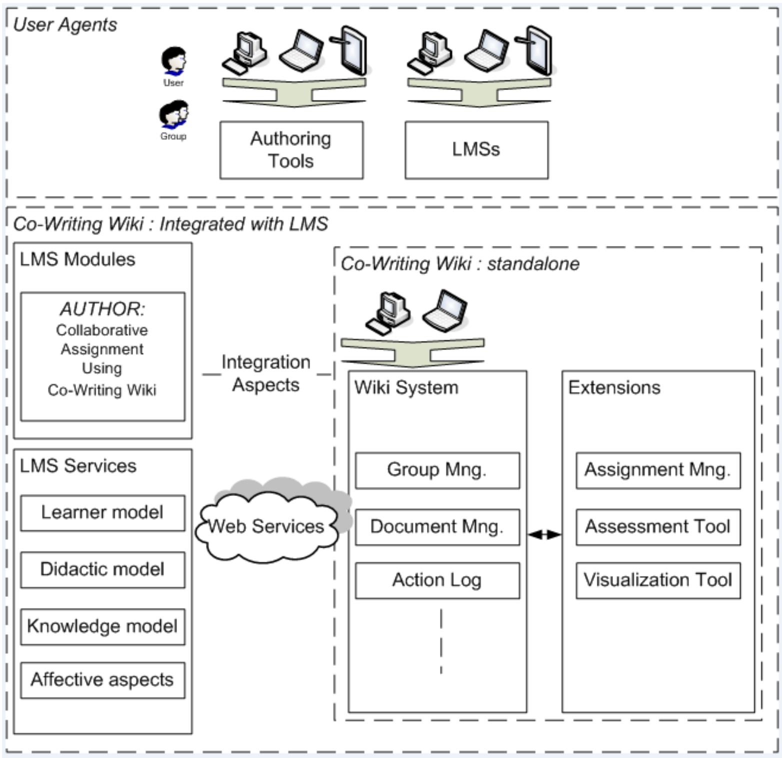


FIGURE 6.5. Co-Writing Wiki Abstract Architecture.

Integrated Assessment Approach for Game-based Learning

The requirements for this scenario have emerged from ALICE project to provide a serious game (SG) for fire evacuation training in the context for civil defence. The game – developed at the Serious Games Institute (SGI) at Coventry University - adopts a freely navigable 3D environment, created within the Unity Engine[52]. The game contains elements of crowd simulation within fire evacuation scenarios, effectively placing the player within the building and monitoring their actions as they evacuate. Hence, provide effective feedback and assessment, it is essential that the game monitors and correctly identifies key actions which may indicate correct and incorrect behaviours. The principal means through which it is proposed is achieved through the implementation of virtual 'checkpoints' within each scenario, recording players' time and state as they pass within a radius of a single point within the virtual space (AL-Smadi, Guetl, Dunwell, & Caballe, 2012).

Providing this scenario for educational purposes requires it to run under a learning environment by which students - i.e. players - preferences and knowledge state are considered, moreover students should get dynamic feedback adapted to specific scenarios during their evacuation thus to scaffold their learning and performance and to achieve micro learning objectives such as students should learn not to use elevator during fire evacuation. Therefore, flexible assessment forms integrated within the game scenarios are required.

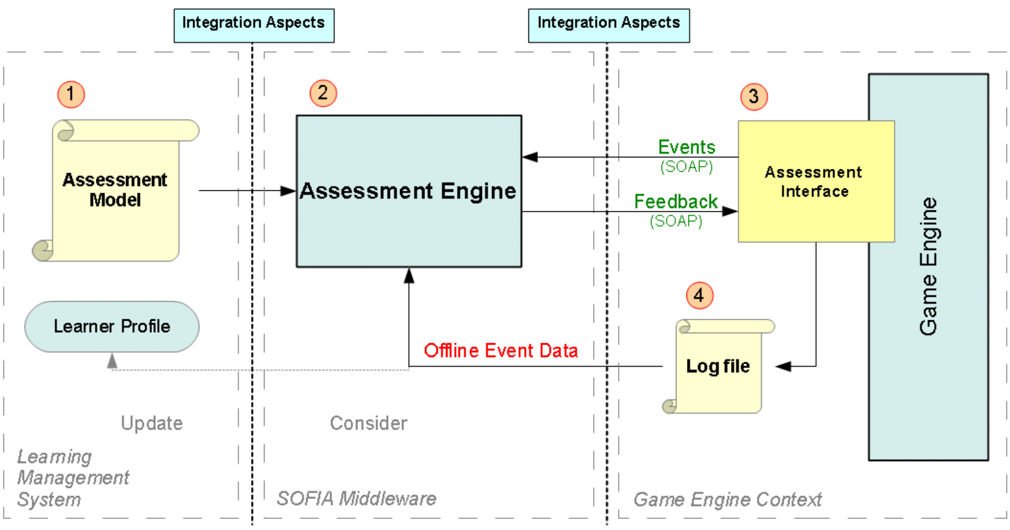


FIGURE 6.6. Integrated Assessment for Game-based Learning.

In order to meet these requirements, an enhanced approach for integrated assessment in 'stealth' mode has been developed through the architecture depicted in Figure 6.6. The architecture has been designed to consider two main scenarios for assessment:

- **Post Evaluation:** in which a 'log file' has been designed to hold all the actions related to the assessment scenario from in specific context - i.e. fire evacuation training - through tracking the players' interactions. Moreover, an 'evaluation engine' is developed to interact with an 'assessment model' to evaluate the players progress -

[52] http://unity3d.com/

represented by log file actions - against a pre-authored assessment rules to assess specific learning objectives - e.g. crawling in Smokey areas during evacuation.

- **Dynamic Assessment and Feedback:** in which an 'assessment plug-in' is attached to the game engine to handles events coming from players interactions and calls the 'evaluation engine' to evaluate those actions - e.g. picking stuff during evacuation - with respect to the pre-defined assessment rules in the 'assessment model', and provides the pre-defined feedback associated to those assessment rules dynamically to the player. The rate of feedback provision to players depends on instructors, technology, pedagogy, and learner preferences control. Hence, feedback can be *immediate*, *delayed*, or *dynamic* based on the domain and learner action type. This has been considered in the design of the feedback block in the 'assessment model' in order to support both scenarios for assessment.

When students play, they interact with the game by making decisions and taking right/wrong actions and paths. The game platform should have the possibility to define checkpoints (assessment rules) so that to assess players interactions and decisions. Moreover, it should provide valuable feedback. As a result, the developed scenario to foster integrated assessment for game-based learning has the following components (see Figure 6.6):

- **Assessment Model:** is an XML based description of behaviour patterns and associated consequences. Behaviour patterns are defined through sequences of possible player actions and conditional matches. While consequences have the primary goal of providing feedback - messages or actions - to the player within the game engine after detecting specific pattern by the assessment engine. Consequences can take a form of action to enable internal measurement operation (e.g. stop watches). The assessment model is authored by the teacher using SOFIA or a target user of LMS. The model is then retrieved by the assessment engine with respect to the learning task. Therefore, the assessment engine should have access to an assessment model service provided by SOFIA middleware.

- **Assessment Engine:** loads the related assessment model once it is invoked. The retrieval of the assessment model is based on the learning task as discussed earlier. Using the model, the assessment engine analyses and match all possible assessment rules when invoked by the game engine by receiving new game flow events. Possible event sources are, log files - for the post evaluation scenario - or direct calls - for the dynamic assessment and feedback scenario - from the game engine.

- **Assessment Interface:** handles the communication between the game engine and assessment engine. The assessment engine is managed through a web service developed as part of SOFIA middleware. For this web service an interface is provided and used to call the assessment engine methods. The service is described using the Web Services Description language (WDSL) and uses the Simple Object Access Protocol (SOAP) for messages communication and transport.

- **Log file:** is created by the game engine which tracks the player interactions and environment changes and logs these in an XML-based log file. The log file is used for post evaluation to provide report based on player behaviour and performance within the game environment.

To this end, the application contexts discussed earlier has been developed as third-party tools to provide specific forms of integrated services for evaluating learners' experiences using CLRs. However, using these scenarios in e-education requires them to be integrated with the learning platform (see Figure 6.7). Having the ability to sequence these third-party tools within a learning activity provided online provides a margin of pedagogical flexibility as educators can define learning and didactic objectives which can be evaluated using the third-party tools, and thus educators can design a learning activity where these tools can be a LO sequenced within the learning path. Therefore, these third-party tools should be flexible - technically and pedagogically, integrated within the learning environment, and interoperable - on both levels of content and services. Summing up the requirements coming out form the application contexts and highlighting the need for a flexible e-assessment environment to meet those needs are depicted in next section.

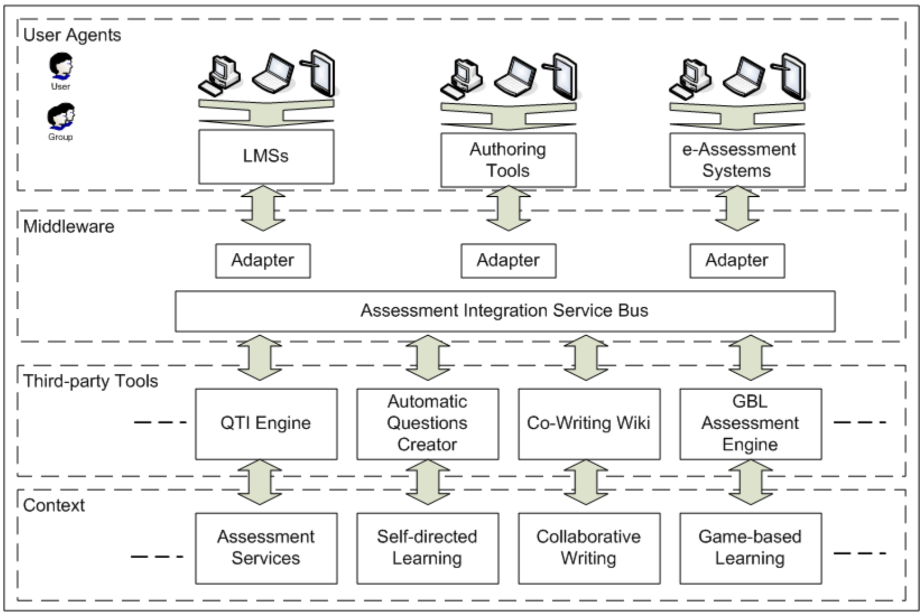


FIGURE 6.7. SOFIA third-party tools and their application scopes.

6.3.2. Definition of SOFIA Requirements

Providing flexible and integrated e-assessment in complex learning resources is the main goal of this study. Thus, providing a flexible and standard-conform e-assessment system - in terms of content and services - that can be extended with third-party tools and services to provide embedded assessment forms within CLR to evaluate enriched learning experiences (see chapter 5) is the general requirement of this research.

In order to fulfill this goal, SOFIA should have a flexible e-assessment system that is capable to interact with different target users represented by e-assessment stakeholders - such as, students, tutors, teachers, administrators, parents, sales, managers - and other systems - such as, LMSs, authoring tools, third-party tools. Therefore, SOFIA should support different user roles represented internally in its services to define access level and service type. In the context of e-assessment, the user roles for CAA systems were considered to be *lecturers*, *students*, and *administrators*. The IMS QTI has nine types of user roles: *assessor*, *scorer*, *candidate*, *invigilator/proctor*, *administrator*, *authority*, *tutor*, *author*, and *psychometrician* (Smyth & Roperts, 2000). Two years later, a work done by (Sclater & Howie, 2002) has identified 21 user roles for what they have called "ultimate" e-assessment engine. The 21 roles are categorized based on: (a) system, user and group administration, (b) questions and tests author, viewer, and validator, (c) questions and test purchaser and sales administrator, (d) test session or instance of a test as timetabler, invigilator, and learner, and (e) responses and results such as marker, scorer, and feedback provider. However, in the context of SOFIA general user roles of student, teacher, and administrator have been used with a permissions to handle other roles such as assessor, scorer, candidate, etc.

To this end and by considering the application scenarios in the application scopes section and challenges and problems, SOFIA should consider the following general requirements (AL-Smadi & Gütl, 2008):

a. *Flexible design* to be used as a stand-alone system or to be integrated with existing systems and tools.

b. *User-friendly interfaces* for both students and educators where a user interaction and online submission of solution and evaluation can be done.

c. Assessment environment for *various learning and assessment settings* which supports guided as well as self-directed learning.

d. *Management and (semi-)automatic support* over the entire assessment lifecycle (exercises creation, storage and compilation for assessments, as well as assessment performance, grading and feedback provision).

e. *Rubrics design and implementation interfaces* to allow the educators to design their own rubrics based on learning objectives to assess learners' performance against a set of criteria.

f. *Support of various educational objectives and subjects* by using various tools sets which for example enables automatic exercise generation or selection, automatic grading and feedback provision.

g. *Results analysis and feedback provision* (immediately or timely) of the current state of user knowledge and meta-cognitive skills for both educators and learners and also for adapting course activities and learning contents based on users' models.

h. *Standard-conform information and services* to be easily sharable, reusable and exchangeable. This will include the tests' questions, answers and students' results, moreover any other required services.

i. *Security and privacy* where required mechanisms to ensure that confidential or private data is used or provided as the user wish, and considering organizational rules and ethical aspects. Moreover, secure user's access based on pre-defined roles and access levels is considered in the design and the development of the system.

6.3.3. Comprehensive Architecture

In the world of technology enhanced learning (TEL) extending the LMS services using third-party tools holds a great promise and challenge in the same time. To what extent third-party tools and LMSs are flexibly designed to be integrated with zero-line codes? What are the levels of integration (integration goals)? What are the main requirements in order to reach such flexibility? These and some other questions have been considered during the research of tools and content interoperability. According to (Thorne, 2004), the following elements have to be considered when it comes to have interoperable tools and services:

- *Data and information (content):* e-assessment content has to be represented using common specifications and standards (e.g. IMS QTI, IMS LIP) so that different tools can share and reuse their content in a flexible manner.

- *Communication (transport and protocols):* tools have to use common platform independent communication protocols (SOAP, HTTP) so that they can easily communicate to share functions, activities, or content.

- *Software Interfaces:* that forms as a contract between service provider and consumer (e.g. OKI OSID). Moreover, interfaces represent an abstraction level to tools and services which make them easily integrated into LMSs. Interfaces decouples between services implementation and access where service providers are free to evolve and improve their services without affecting consumers as well as consumers can switch between different service providers in case those providers share common semantic definitions for their services.

- *Domain Models:* provides a common conceptual understanding of the problem domain in general and e-learning domain in particular. Domain models help developers to have common understanding with input/output data, data representation, possible services, and their workflow to achieve specific goals. Examples of this are the e-learning Framework (ELF) and Framework REference Model for Assessment (FREMA), the Service-Oriented Framework for e-Assessment (SOFA; see Section 5.3) (AL-Smadi, Guetl, & Helic, 2009)

Similar to Thornes' interoperability aspects, AL-Smadi and Guetl (2010) suggested the following requirements for a flexible e-assessment system:

- Clear guidance represented by a well-formed *framework* (see Chapter 5).

- *Standards and specifications* that represent the whole process of assessment as well as the communication between the services and components (see Chapters 4, 5).

- *Cross-domain requirements* analysis in order to define the specific requirements for each application domain (such as, requirements to provide assessment for CLR as discussed in SOFIA application contexts, as discussed earlier in Section 6.2).

- *Web services* that achieve the cross-domain requirements and interact through well-defined interfaces.

Dagger et al. (2007) discussed the flexibility and interoperability challenges for the so-called "next-generation LMSs". Dagger et al., stress on the importance of that LMS should exchange both information's' syntax and semantics which goes in line with IEEE definition of interoperability (see Section 4.2.2) as systems have to be able to share information and to use them as well. Moreover, they argue that semantic exchange is not enough, LMSs have to have control on the shared tools and services so that they can keep their workflows, internal representations, and tracking mechanisms. They also recommend a shared dynamic semantic view about services (such as Semantic Web) instead of APIs as in OKI OSIDs so that services can be easily selected, orchestrated, and consumed based on a common understanding of the learning process.

Based on that, e-assessment systems should be designed to work as standalone applications or to be flexibly integrated with other LMSs. In order to have a "pluggable" and flexible e-assessment system we have to distinguish between two levels of standardization by which we can tackle the challenge of information and tools interoperability (AL-Smadi, Guetl, & Helic, 2009). The authors discuss internal and external levels of standards-conformation. The internal level is usually used where assessment content and user information are designed to adhere to specifications and standards such as IMS QTI and IMS LIP for instance. The internal level is suitable when assessment systems are used as standalone systems and they only share their content and user information. In order to tackle the challenge of tools and services interoperability the authors discussed what they called external level where they recommended a standard-conform interface. This interface represents an abstraction level of interoperability and could conform to any of the learning tools interoperability specifications discussed in Chapter 4.

Taking into consideration those two levels of standards-conformation, what is the suitable architecture that e-assessment should use to be pluggable and flexible? What are the challenges in designing such architecture? Moreover, how this architecture will foster e-assessment with integrating e-assessment third-party tools?

The ultimate goal in having a flexible and pluggable e-assessment system is to design e-assessment Software-as-a-Service (SaaS). According to Dagger et al. (2007) future LMSs will satisfy wide range of needs by providing what they called "interoperable architectures". Accordingly, what could be a possible architecture for the next generation of e-assessment systems? Figure 6.7 depicts a flexible and interoperable architecture for e-assessment system. The architecture reflects the idea of the two levels of standards-conformation based on SOFA (AL-Smadi, Guetl, & Helic, 2009b) discussed in Chapter 5. Those two levels go in line with Daggers' levels of interoperability (*intradomain* and *interdomain*) where this architecture fosters e-assessment systems to be flexibly used as standalone systems or to be integrated with other tools and systems.

As depicted in the figure the architecture addresses four main areas - represented with dashed borders- of *user agents*, *middleware*, *third-party tools*, and *e-assessment system*. The user agents represent the potential users for the flexible and interoperable e-assessment system. Flexible and interoperable e-assessment system could adopt SOA to represent its services (Millard et al, 2005; Al-Smadi & Guetl, 2010). Based on SOFA, two levels of services are available in e-assessment systems: e-assessment services and common services (see Section 5.3). e-Assessment services represent the core assessment services provided by any e-assessment system such as authoring, scheduling, delivering, grading, and reporting, whereas the common services represent the required services that can be available in different systems such as, security services (authentication & authorization) and infrastructure services such as database management and network management. Moreover the architecture shows e-assessment content that should be standard-conform. As discussed before in Chapter 4, little number of specifications is used for e-assessment content, the more widely used is the IMS QTI. Learner's information should also adhere to specifications and standards such as IMS LIP and PAPI Learner so that to be sharable among different tools and systems.

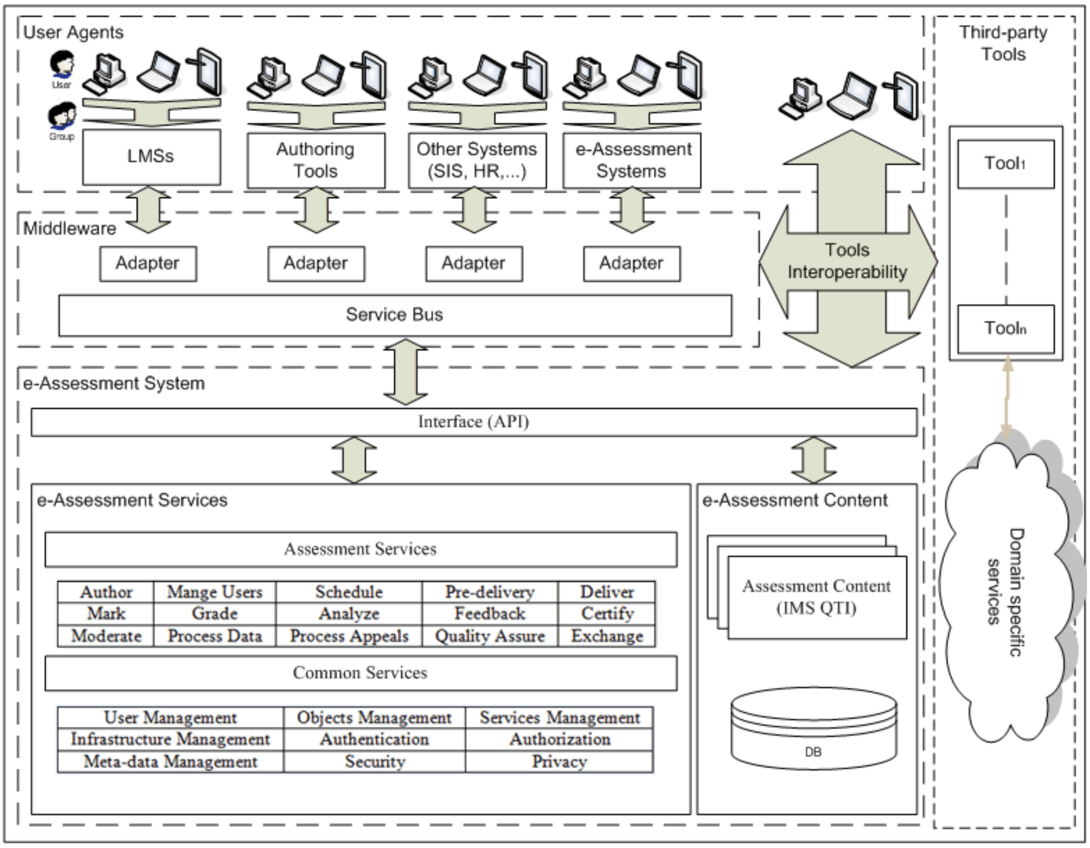


FIGURE 6.8. Architecture for a Service-Oriented Flexible and Interoperable e-Assessment System (AL-Smadi & Guetl, 2011b).

The middleware layer mainly contains a service bus and adapters to integrate the service bus services with other tools and systems. The adapters should adhere to one or more of the previously discussed tools interoperability specifications (i.e. O.K.I OSIDs, CCSI, or IMS LTI). It is worth mentioning that the middleware may contain a *service registry* so that services from the e-assessment system as well as from other domain-specific services can be registered. This will foster the middleware to search, compose or orchestrate suitable services based on the demands of the user agents. The service registry - if required - can be part of the middleware architecture such as using UDDI (Universal Description, Discovery and Integration), or can be provided by an external service such as the JISC's IESR[53] (Information Environment Service Registry) project, which focuses on improving resource discovery mechanisms for electronic resources, thus to make it easier to find materials to support teaching, learning and research.

In order to show the alignment between this comprehensive architecture and the application scopes discussed in Section 6.1 as well as the general requirements from Section 6.3.2, the architecture component of e-assessment system represents the MASS system whereas the third-party tools component are the tools developed for the application contexts discussed earlier namely PASS, AQC, Co-Writing Wiki, and the game-based learning assessment engine. These third-party tools have been used to extend MASS services within SOFIA assessment environment through the support of the proposed SOA-based services middleware, thus to provided flexible and integrated assessment forms to evaluate rich learning experiences using

[53] http://iesr.ac.uk/

CLR such as automated assessment within self-directed courses, collaborative writing and peer-review via self and peer-assessment using assessment rubrics, and integrated assessment for game-based learning. Moreover, the third-party tools have been fully integrated within SOFIA using a single sign-on (SSO) approach through related services in SOFIA middleware. Nevertheless, SOFIA has been used in the context of ALICE project fully integrated using SSO approach with the intelligent web teacher LMS (IWT; (Capuano, Miranda, & Orciuoli, 2009)). The next sub-section explains in more details the SOFIA middleware component and its main services.

Middleware for Flexible and Interoperable e-Assessment

Flexible and interoperable e-assessment systems should be flexibly designed to be used as standalone systems where users can have a secure access, moreover to be integrated with other tools and systems such as LMSs and authoring tools. The whole e-assessment system should behave like a service so that to be used by other related tools and services. This highlights challenges and aspects such as, *accessibility*, *security*, *single sign-on (SSO)* and *software instances management*, moreover, integration aspects such as *process integration*, *control integration, data integration, and presentation integration*. In a step towards tackling such challenges a middleware layer is added to the architecture.

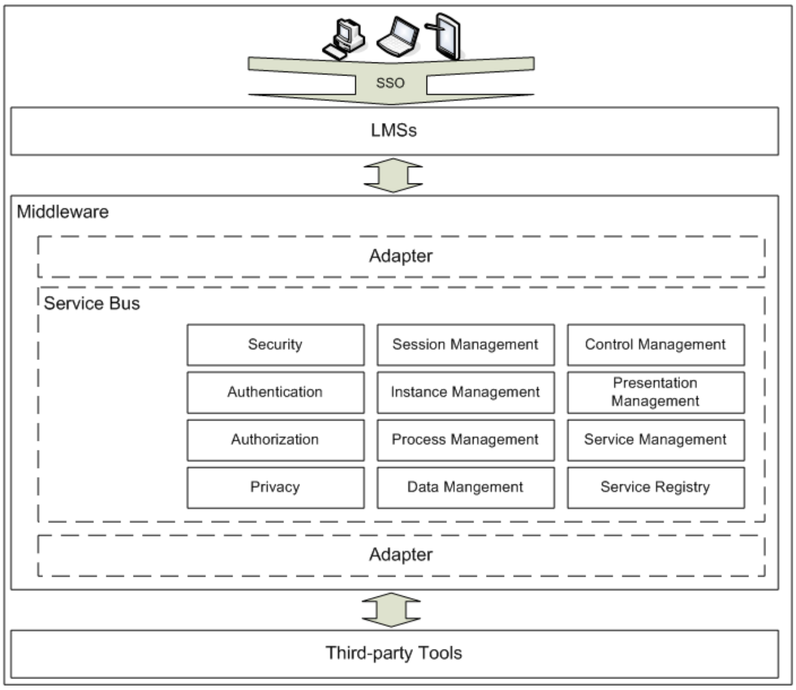


FIGURE 6.9. Architecture for a Middleware to foster Learning Tools Interoperability (AL-Smadi & Guetl, 2011b).

Figure 6.9 depicts the middleware architecture in general and the service bus services in particular. The service bus related services mainly address the challenges and aspects discussed above. Generally, the middleware is suggested to foster third-party tools interoperability with LMSs moreover, to support the integration of MASS e-assessment system with other related tools and services.

As shown in the figure, the service bus related services cover the following aspects: *(1) Single Sign-On (SSO) services:* in order to support SSO, a set of services have been suggested. Those services mainly cover security, authentication, and authorization aspects. Mechanisms of delegated authentication and authorization have to be considered. The same problem has been discussed in (González et al., 2009) where the authors discussed different *Delegation Permits* initiatives such as OAuth (OAuth specification, 2009), Delegation Permits (Hasan et al., 2008), Shibboleth (Shibboleth, 2009), and OpenSSO (OpenSSO, 2009). Moreover, the authors have suggested an adapted technology of OAuth which they called "Reverse OAuth" as a useful model to support SSO among e-learning systems. As cited in (Crisp, 2007, p. 163) Shibboleth has been recommended by the Joint Information Systems Committee[54] (JISC) to be used in higher educational institutions to control authentication and access to educational material and services. Shibboleth and the other initiatives forms as a *middleware* to control access of educational systems' target users to particular resources. Moreover, JISC defines middleware as "*helping institutions to connect people to resources*" as part of their 'Core Middleware: Technology Development Programme ran between April 2004 and March 2007' which aimed to develop central services that is essential to middleware, these are: authentication, authorisation, directory services, and identifiers. *(2) Integration services:* referring to O.K.I. definition of interoperability as measure of ease of integration between to systems to achieve a functional goal, several integration aspects have to be considered. Among these aspects the following have been highlighted:

- **Process Integration:** how much tools are flexibly integrated to support the performance of a process. Moreover, how much they agree on required events for this service as well as how much they enforce constraints. Process integration is required to achieve that tools have to share a common understanding of the application domain, services, learning activities, tasks, standards and specifications, and outcomes. This stresses the requirement for a framework and reference models for e-assessment in particular and for e-learning in general. Moreover, a need to the so-called "semantic exchange" among services in order to explain what services can do, what are their inputs/outputs, how they can be managed and used (Dagger et al., 2007).

- **Control and Presentation Integration:** to what extent tools exchange control on their services, processes, interactions, appearance, and behaviour. How much flexible the tools can adapt their appearance and behaviour according to LMS style. How much flexible to change the tools internal workflow and to utilize it within the LMS workflow.

- **Data Integration:** do tools share the same specifications of their data (content and information)? To what extent they cooperate to maintain data consistency? Do they share the same semantic constraints on the data they manipulate?

- **Management and Supportive services:** this group of services handles aspects related to session management, tools instances management, services registration, services management in general such as services search and orchestration, data and process backup.

For each tool an adapter (interface) has to be provided in the middleware. Those adapters have to be designed to adhere to one or more of the tools interoperability specifications discussed in Chapter 4. This fosters third-party tools in general and e-assessment systems in

[54] http://www.jisc.ac.uk/

particular to expose their services in a flexible way, moreover in order to adapt their classes' interfaces according to clients' needs.

The next sub-chapter demonstrates how the comprehensive architecture for flexible and interoperable e-assessment (see Figure 6.8) has been developed on top of the service-oriented middleware (see Figure 6.9) to foster providing integrated e-assessment forms for CLRs.

6.4. Development of the System

The software development of SOFIA e-assessment environment is the result of fostering flexible and interoperable e-assessment tools through integrated assessment forms for the CLR within the distinct research projects discussed in Section 6.3.2.

The previous sub-chapters depict the main issues concerning the software application contexts as well as the high level requirements which were used to design the overall layered architecture of the system (see Section 6.3). In this sub-chapter, aspects relevant to SOFIA software development are discussed. In particular this sub-chapter focuses on the development aspects of MASS core assessment system - i.e. Standalone SOFIA - and SOFIA middleware. Moreover, this sub-chapter discusses the development of SOFIA middleware through a service-oriented paradigm to extend MASS - i.e. core assessment system - services to provide integrated assessment in CLR based on the target application scopes (see section 6.1). The development of the CLRs represented by SOFIA application contexts is discussed in Chapter 7.

SOFIA standalone assessment system and SOFIA middleware services are developed using .NET 3.5 Framework, in particular C# for the code-behind business logic and ASP.Net for the front-end presentation layer, and MySQL database for the data layer. Microsoft Visual Studio® (the primary development tool for building .NET Framework applications) is used as development suite - similar to Eclipse for Java-based applications. Internet Information Services (IIS) - a host environment for web services and web applications - is used to host SOFIA developed services.

To this end, the remaining of this sub-chapter meets the high level requirements discussed earlier as follows, the next sub-sections discuss the development of SOFIA middleware using the service-oriented approach through which MASS standard-conform resources - i.e. e-assessment content and services - and third-party tools can be accessible from target users. An emphasis on standards conformation is illustrated by discussing aspects related to e-assessment content and tools. Moreover, security and privacy as a main requirement are depicted through highlighting the related developed services. Nevertheless, this sub-chapter closes with showing the standalone SOFIA whereas the proof-of-concept of using SOFIA integrated with other tools and systems are provided in Chapter 7.

6.4.1. Service-Oriented Approach

As discussed in Chapter 3, the emergence of using computers in assisting assessment is not new. Over the last 60 years different e-assessment tools and systems have been developed in a variety of contexts (see Section 3.2.2). In the context of educational assessment, e-assessment tools are designed with a consideration to support different didactical resources - e.g. CLR of serious games and simulations, collaborative learning resources. CLRs are influenced with the evolution of ICT thus e-assessment software should be modular and flexible and built on top of reusable, scalable and accessible architectures (AL-Smadi, Gütl, & Kannan, 2010). Solutions based on service-oriented architecture (SOA) have attracted research community in the last years. The SOA paradigm aims at providing distributed systems with high transparency with respect to platform and the communication heterogeneity.

The SOA paradigm in contrast to other paradigms - e.g. component-based, object-based, etc. -holds great promises when it comes to have transparent communication protocols through well-defined interfaces which provide open connectors and standardized contracts. Thus, solutions which are based on SOA provide services that are loosely coupled and reusable. Moreover, services are accessible to be invoked through platform-independent interfaces. Nevertheless, SOA provides software units with: platform-independent service interface, dynamic invocation of services, and services are self-contained.

As discussed in (AL-Smadi & Guetl, 2010) in the world of SOA there is three main roles of interaction: *service provider, service registry,* and *service requestor*. Service providers are software agents that provide the service. They should **publish** a service description on a services registry. Service clients are software agents that **request** the execution of services. They should be able to **find** service description on the services registry and to **bind** the service. During the bind operation the service requestor invokes a web service at run-time using the binding information in the requested service description to locate this service. This invocation has two main possible scenarios: the first one is direct invocation by the service requestor using the technical information in the service description located on the services registry. The second one is via a service discovery agency where the communication between the service requestor and the service provider goes through the services registry of the discovery agency. (*cf.* Papazoglou, 2008). In the case of SOFIA, the first scenario is used where the invocation of the application contexts web services is done directly through the services technical description.

Table 6.1 summarizes the developed services within the context of SOFIA middleware. SOFIA middleware is developed using a service-oriented approach which provides on the one hand the required usability, flexibility, and interoperability of learning tools and information. On the other hand the middleware provides platform-independent access to SOFIA assessment services and third-party tools. The general services form the table refers to common services that are required in most educational systems, and mainly cover security management, data management, and context management. In the case of SOFIA, different contexts are supported (see Section 6.3.2), for each application context a set of services are developed as part of the context services in the middleware. For instance, in the application contexts of SOFIA, the scenario of automated assessment for self-directed learning utilizes a Java-based standalone third-party tool - i.e. AQC - to automatically create test items based on textual learning material, whereas in the scenario of collaborative writing and peer-review the scenario uses a .Net web-based application - i.e. Co-Writing Wiki - integrated with IWT LMS. In addition to the technical support in terms of flexibility and interoperability, the middleware fosters SOFIA to support different pedagogical approaches - e.g. self-directed learning, game-based learning, collaborative learning, etc. - through the flexible integration of CLR represented by third-party tools and services - e.g. Co-Writing Wiki and AQC.

In order to provide interoperable learning tools, the third-party tools should provide their services as web services as discussed earlier in the middleware section. The middleware 'SOFIA_Context_Manager' service helps to access the third-party tools through each tool adapter provided for the middleware - i.e. Context Services. In the context of SOFIA the developed web services - based on AL-Smadi & Gütl (2010) recommendations for flexible e-assessment system - communicate through standardized way by using the Simple Object

Access Protocol (SOAP) for messages communication and transport. The developed web services as well as the specification of their messages (SOAP -based) are described using the Web Services Description language (WDSL). Despite the necessity of using UDDI server to support SOFIA services discovery and registration and that the middleware architecture supports this through its management services, no UDDI registry is used in SOFIA as the web services interact with each other based on direct service invocation. This is one of the debate arguments of SOAP vs. REST web services choreography (*cf.* Muehlen, Nickerson, & Swenson, 2005). REST (Representational State Transfer) is an HTTP-based architectural and design style for networked solutions that follow the CRUD (Create, Read, Update, and Delete) principle to operate on remote resources (*cf.* Fielding, 2000).

TABLE 6.1. SOFIA-middleware developed services

Category	**SOFIA middleware service**	**Description**
General Services	SOFIA_Service_Manager	Forms as a single point of access for SOFIA target users and handles the despatching of multiple instances requests for services and third-party tools
	SOFIA_Context_Manager	Holds the runtime information of the third-party tools and provides the required information to access, launch, and outcome of those tools
	SOFIA_Security_Manager	Assures secure authenticated and authorized access to SOFIA resources - i.e. content and services.
	SOFIA_Data_Manager	Provides APIs for different data providers - such as MySQL data provider and SQLServer data provider - and maintains data persistence
Context Services	QTI_Manager_Service	A core assessment service that supports SOFIA to author, deliver, interpret, and validate QTI-based items and tests (see Section 6.4.2)
	MASS_Runtime	Holds the runtime information of MASS -SOFIA core assessment system - and provides the required information to access, launch, and outcome of MASS services (see Section 6.4.4)
	PASS_Runtime	Holds the runtime information of PASS and provides the required information to access, launch, and outcome of PASS CLR (see Section 7.1)
	AQC_Runtime	Holds the runtime information of AQC and provides the required information to access, launch, and outcome of AQC CLR (see Section 7.2)
	AQC_QuestionsManager	Manages the type and number of questions to be automatically created by AQC for the selected learnin material. This service uses QTI_Manager_Service to provide IMS QTI compliant test items (see Section 7.2)

	AQC_ConceptsManager	Manages the concepts created automatically by AQC for the selected learning material, as well as the concepts selected manually by the learner (see Section 7.2)
	Rubric_Manager	Manages the authoring and playing of assessment rubrics (see Section 7.3)
	CoWritingWiki_Runtime	Holds the runtime information of Co-writing Wiki and provides the required information to access, launch, and outcome of Co-writing Wiki CLR (see Section 7.4)
	CoWriting_AssignmentAuthor	Manages co-writing assignment authoring in terms groups, topics, peer-review, assessment rubric, and groups assessment configurations (see Section 7.4)
	CoWriting_GroupManager	Manages groups for the authored co-writing assignment (see Section 7.4)
	GBL_Assessment_Runtime	Holds the runtime information of GBL_Assessment and provides the required information to access, launch, and outcome of GBL_Assessment CLR (see Section 7.5)

A major problem with WSDL is that the service interface definition lacks semantics of the service and information on non-functional features. For instance, aspects related to performance and dependability or only covered by the programmer documentation and this is represented by natural language. Using natural language may lead to misunderstanding and thus facing problems in services development and deployment. A possible solution is the research trend of specifying the semantics of the services. A promising solution is to use ontology-based description. Ontologies provide a standardized way to share terminologies and concepts as well as their relationships. Thus, ontologies can be used to explain specific concepts used in the description of the service. The Semantic Markup for Web Services (OWL-S[55]) is an example of ontology-based language to describe web services. The OWL-S - which based on Ontology Web Language (OWL) - consists of three main parts: (a) the service *profile* for registering and discovering services, (b) the *process model* which gives a detailed description of a service's operation, and (c) the *grounding* which provides details on how to interoperate with a service using messages. However, SOFIA services can be modelled using such ontology, but for the current state - as discussed earlier in this section - SOFIA services interact with each other based on direct service invocation.

In addition to lack of services semantics, the third-party tools developed as part of SOFIA practical contexts (see Section 6.3.2) act as a black box to the instructional designer. As discussed in Chapter 4, some specifications have been designed to represent learning design and workflow through which a learning activity can be designed and represented using interactive tool. However, it is required that these third-party tools should expose their services and tools in an accessible way - web services in SOFIA - thus instructional designers can sequence them in learning activities. Through the web services (see Context Services in

[55] http://www.w3.org/Submission/OWL-S

Table 6.1) the third-party tools operations and their required input as well as their output can be utilized in learning activities.

This section discusses how SOA is used to foster SOFIA with a flexible software architecture thus to be used as a standalone or with other tools and systems. Next sections discuss this in more detail and show how the SOFIA development meets other requirements such as standard conformation, and security and privacy. Nevertheless, more focus on SOFIA standalone development and assessment services is provided in Section 6.4.4.

6.4.2. Standards Conformation

In order to meet the requirements of flexibility and interoperability SOFIA, services and tools should adhere to e-assessment standards and specifications as well as should consider best practices and guidelines (AL-Smadi, Guetl, & Helic, 2009; AL-Smadi & Gütl, 2010).

As discussed in Chapter 4, standard-conform e-assessment system should not only adhere to content standards but also to learning tools interoperability. Nevertheless, learning design specifications should be used to represent the learning workflow when it comes to provide integrated assessment forms with CLRs represented by third-party tools. The next sub-sections shed the light on these aspects and show how SOFIA has been developed to consider them.

Interoperable e-Assessment Content and User Interface

Among the discussed e-assessment standards and specifications in chapter 4, IMS QTI has been selected to represent e-assessment content as it is widely used in the context of e-assessment. As discussed earlier in this chapter QTI lacks specifications to represent special application domains such as behavioural assessment - e.g. players flow in game-based learning - as well as special types of items such as puzzles which is widely used in the context of language learning. Despite these limitations QTI provides a rich meta-data to represent items, tests, and results. Nevertheless, QTI provides standardized way to analyse candidates' response on items through the so-called '*response processing templates*' (see Section 4.2.1). Moreover, QTI lacks a platform-independent items player or even guidelines of how to design controls to play QTI items. For instance, in the context of QTI players you can find tools that utilize browser applets such as Java Applets, Silverlight, or Flash controls to design the front-end QTI players. In the context of SOFIA, a QTI player has been developed within MASS author, and deliver modules using ASP.Net controls for simple items - e.g. MCQ, True/False, and FIB, AJAX controls for specific items such as Slider, and HTML5 based player to handle highly interactive items such Hot Spot, Graphic Order, and their combinations, see Figures 6.10, 6.11 for Hot Spot item author and player in SOFIA using HTML5 technology.

Using HTML5[56] technology fosters SOFIA with browser independent authoring and playing of QTI items. According to the World Wide Web Consortium (W3C) introducing HTML5 "*This specification defines the 5th major revision of the core language of the World Wide Web: the Hypertext Markup Language (HTML). In this version, new features are introduced to help Web application authors, new elements are introduced based on research into prevailing authoring practices, and special attention has been given to defining clear conformance criteria for user agents in an effort to improve interoperability.*" Thus, HTML5 aims to facilitate authoring of web applications and to maintain interoperability on the level of user interface. In contrast to XHTML, HTML5 ignores errors in the syntax of the

[56] http://www.w3.org/TR/html5/

document and renders it while in XHTML syntax minor errors in the syntax prevent the document from being rendered. Moreover, in the context of QTI items playing no further applets are needed to render objects and handle interactions. Figure 6.12 depicts SOFIA author for a composite item in terms of its '*response processing template*' as both hot spot and graphic order are combined to provide one question item, see Figure 6.13 for the same item delivery using SOFIA.

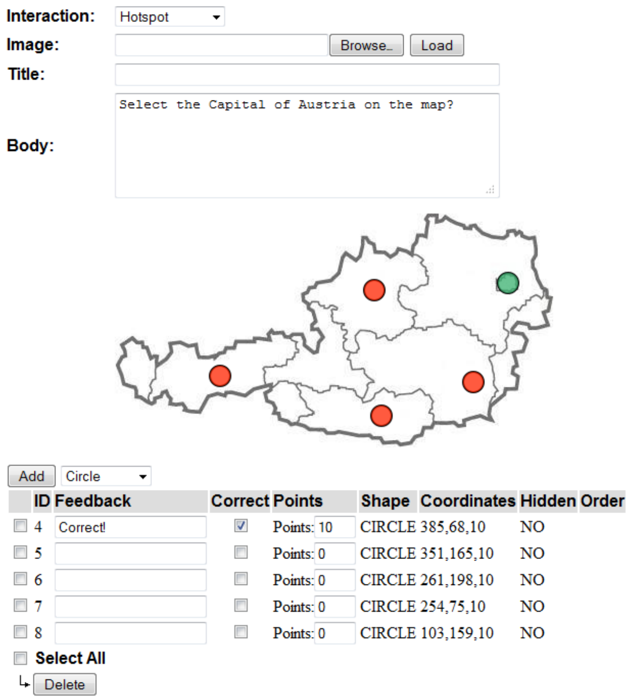


FIGURE 6.10. SOFIA HTML5 based QTI Author - Hot Spot item.

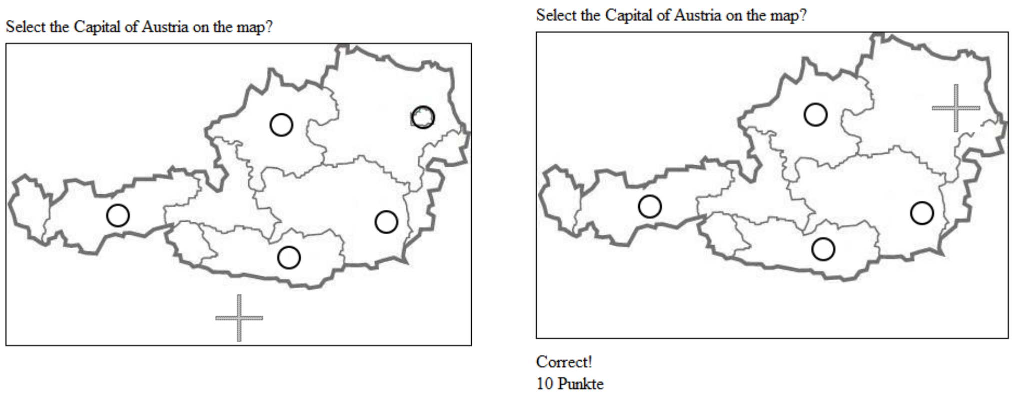


FIGURE 6.11. SOFIA HTML5 based QTI Player - Hot Spot item.

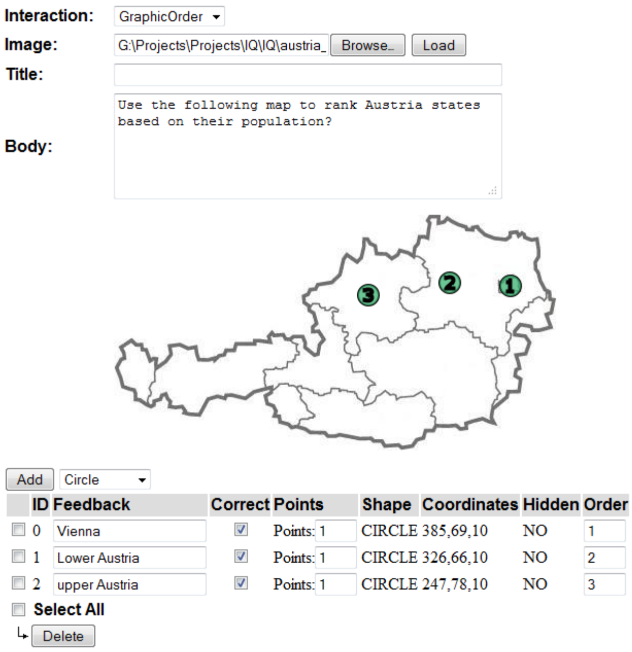


FIGURE 6.12. SOFIA HTML5 based QTI Author - Graphic Order item.

In order to foster SOFIA with QTI based services to author, deliver, interpret, and validate QTI-based items, a web service for QTI management has been developed. The 'QTI_Manager_Service' has been developed - as part of SOFIA middleware - using .Net Framework 3.5 based on the latest version IMS QTI 2.1 specifications. This decoupling between QTI business logic developed as a web service and its application in the context of SOFIA fosters SOFIA to share not only test items syntax but also their semantics which are flexibly accessed via the web service (*cf.* AL-Smadi & Guetl, 2011). Moreover, the 'QTI_Manager_Service' as well as the browser-independent QTI Player (HTML5) discussed earlier have been used to meet some of the requirements for SOFIA practical contexts in particular: MASS as the core e-assessment system, PASS as an enhanced approach for peer assessment, and AQC as an automatic question creator utilized in the self-directed learning scenario.

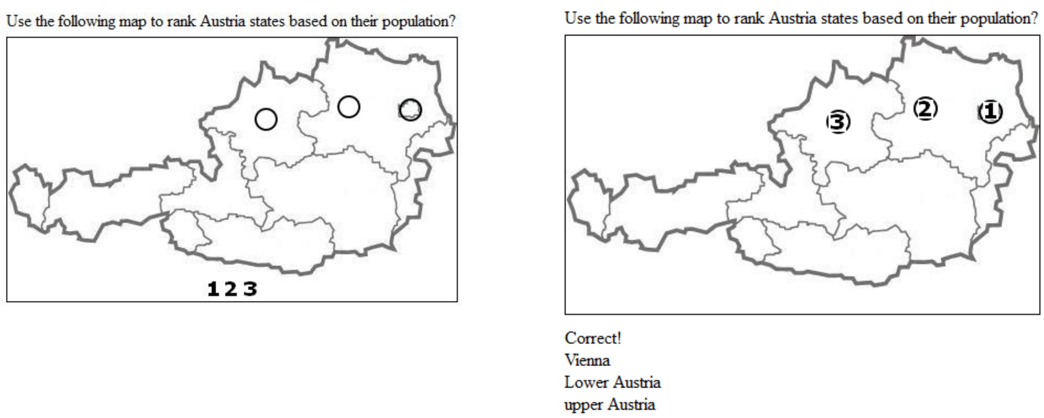


FIGURE 6.13. SOFIA HTML5 based QTI Player - Graphic Order item.

Interoperable Learning Tools

For the application domains in which IMS QTI specification is limited to provide assessment forms - such as, behavioral assessment in game-based learning - a third-party tool that supports such assessment form is required. Thus, e-assessment system services should be extended through the use of the third-party tool. Such extension highlights the challenges of tools accessibility, extensibility, and interoperability. For instance in the context of SOFIA, for the application context of integrated assessment in game-based learning, an assessment engine has been developed to evaluate players' - i.e. learners - behavior and interactions based on flexible assessment model (see section 6.3.1). However, running this assessment engine in the context of SOFIA requires this assessment engine to adhere to learning tools interoperability as well as to have a flexible architecture to support platform-independent access - e.g. SOA (AL-Smadi & Gütl, 2010).

As discussed in Chapter 4, several specifications have been designed to address learning tools interoperability problems. One of these specifications is the guidelines for learning tools interoperability as web services. In this specification IMS GLC provides IMS LTI Guidelines v1.1 (IMS LTI, 2012). IMS LTI recommends a SOA-based web services for each third-party tool and their deployment description. For instance for the case of game-based learning assessment engine, the following components should be considered according to IMS LTI:

- *Game-based learning assessment Runtime:* which is used by the tool consumer - i.e. SOFIA - to communicate with the tool provider -the game based assessment engine. This component forms as proxy tool which is meant to be environment-independent where it does not require specialized code. The proxy tool is entirely a descriptor-based package that describes the deployment, configuration, and runtime context.

- *Web Services:* a set of services that have to be implemented in the hosting environment - i.e. SOFIA middleware. These web services facilitate the deployment, configuration, and launching of the proxy tool through its main services of, d*eployment Service, configuration Service, launch Service,* and o*utcome Service.*

In order to tackle the problem of learning tools interoperability, SOFIA middleware has been used to support target users - e.g. LMS - to access the third-party tools web services. On the level of the third-party tools it is required that the tools provide an application programming interfaces (API) or web services interface to facilitate the access and transport of the tool data to the target user (see Section 6.4.1). Moreover, it is recommended that the third-party tools developers describe their web services using WSDL documents that fosters on the fly integration of these web services (AL-Smadi & Gütl, 2010). The services of the middleware handle the machine-to-machine interaction with both of the target user services and the third-party tool services. Moreover they maintain security and privacy aspects as well as control level of access, and handle multiple instances of services and data transfer.

To this end, the use of interoperability standards - i.e. content and tools - fosters SOFIA to author, and deliver QTI-based items using platform-independent web services as well as browser-independent QTI player based on HTML5 emerging technology. Nevertheless, enriching SOFIA with a SOA-based middleware with LTI specification compliance makes aggregating assessment forms for domain specific fields – e.g. game-based learning assessment - more easy, flexible, and thus interoperable.

Learning Design and Workflow

As discussed in Chapter 4, learning *workflow* in e-Education refers to the automation of the provision of learning activities controlled by a set of rules that defines the pre-requisite, sequence, and consequence of each learning activity. Learning workflow – which is also known as *learning design* – is known to the educational community by two main initiatives: the IMS Learning Design (LD)[57] and the Learning Activity Management System (LAMS)[58]. Moreover, some other initiatives such as, the Business Process Execution Language for Web Services (BPEL4WS)[59] and the Business Process Model and Notation (BPMN)[60] can be used to design the sequence of the business process. Despite the impact achieved in terms of flexibility and services orchestration and choreography using BPEL or BPMN, these initiatives lake pedagogical aspects - e.g. learning objectives - when it comes to use them to design workflows for e-education. Moreover, they require the e-education platform to be built on top of SOA.

In the context of SOFIA, the flexible architecture of SOFIA fosters it to provide easy access and use of the assessment services as well as the third-party tools. Thus, LMSs as target users can design a learning activity - using any of the aforementioned initiatives - in which they can specify any of the SOFIA third-party tools or assessment services, and SOFIA through its flexible architecture and middleware can provide the access required to the tool or e-assessment services and facilitates data transformation.

For instance, the CLR provided in the context of ALICE project for the application contexts of automated and integrated assessment for self-directed learning represents a composite didactic resource for self-directed learning that consists of learning materials represented in SCORM, enriched with IMS QTI compliant test items automatically created from learning material. Moreover IEEE Learning Object Metadata (LOM) is used to annotate the CLR. The CLR has been provided by the Intelligent Web Teacher (IWT). IWT is a learning management system allowing the definition and execution of personalized e-learning experience tailored on the basis of learners' cognitive status and learning preferences based on fully integrated tools and services (Capuano, Miranda, & Orciuoli, 2009). Moreover, IWT uses ontological approach to design learning workflow and provide personalized learning path considering learners and knowledge models. Nevertheless, the design of the scenario learning workflow took place on IWT and SOFIA middleware provided SSO based flexible access and data transport to and from the AQC third-party tool to automatically create QTI-based test items.

6.4.3. Security and Privacy

Security and privacy are crucial aspects educational institutes have to consider in order to provide quality e-assessment. The level of security highly depends on the stake of e-assessment. Educational institutes should pay more attention to security aspects when it comes to provide high stakes e-assessment - i.e. often summative assessment. The International Test Commission (ITC) has published guidelines on computer-based and

---

[57] http://www.imsglobal.org/learningdesign/

[58] http://www.lamsinternational.com/

[59] http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsbpel

[60] http://www.bpmn.org/

internet delivered testing (ITC, 2006), in which they stressed on considering security and privacy aspects in e-assessment – i.e. security of test materials, security of test-takers data transferred over the Internet, and the confidentiality of test-taker results (see Chapter 4).

Enterprise LMS that provide e-assessment such as Blackboard™[61] and Questionmark Perception™[62] use secure tools to deliver high stakes assessment that prevents students from browsing the internet or accessing the LMS to get the course material once they started the summative assessment. For instance, the Respondus LockDown Browser™[63] is an example of such tool that customizes students browsing during high stakes assessment. This tool locks down the browser, prevents internet or learning material access, and disables minimize button, mouse-right-click, and page-source view.

Providing authenticated access to SOFIA target users - i.e. users and systems - is another concern for security. As discussed earlier (*cf.* AL-Smadi & Guetl, 2011), different delegation permits initiatives such as OAuth (OAuth specification, 2009), Delegation Permits (Hasan et al., 2008), Shibboleth (Shibboleth, 2009), and OpenSSO (OpenSSO, 2009) are used to control authentication and access of online resources and e-assessments. In order to explain the security aspects within SOFIA, a use case of a student target user who is using a LMS - e.g. Moodle - and trying to access a third-party tool - e.g. Co-Writing Wiki - through a SSO approach is considered. Figure 6.14 shows the sequence diagram of how SOFIA Middleware is used to help the user to access the third-party tool through its services.

1. After the user is successfully logged into the LMS, he requests access to the tool by clicking on the link designed for the Tool or based on automatic delivery within a course.

2. The Tool web page within LMS platform requests access to the tool from SOFIA middleware.

3. The ‘service manager’ in the middleware checks if the user is authenticated by calling the ‘security service’.

4. The ‘security service’ checks if the LMS is authorized to use the requested Tool through the ‘context manager’.

5. If the LMS is authorized to use the Tool the middleware checks if the Tool is still available using the tool proxy information – i.e. adapter (see Section 6.3.3).

6. If the Tool is available and accessible, then LMS is provided access to the Tool which is rendered to the student right after. The user role from the LMS user management is used to define access level within the tool.

[61] http://blackboard.com/

[62] http://www.questionmark.com

[63] http://www.respondus.com/products/lockdown.shtml

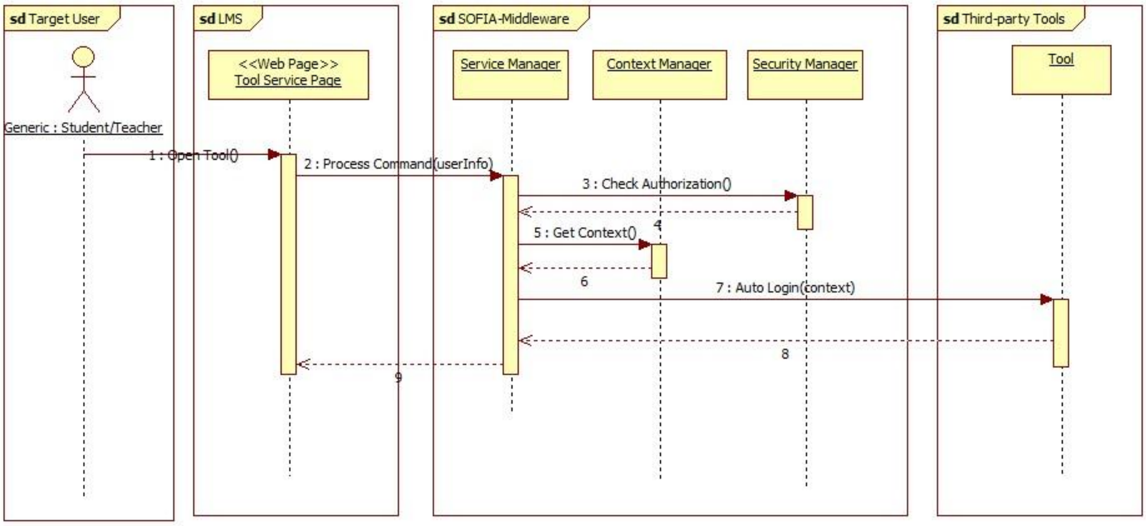


FIGURE 6.14. Sequance Diagram for the User Case of accessing Thrid-party tool via SOFIA from LMS traget user.

The user in this scenario uses the third-party tool – e.g. Co-Writing Wiki – as part of a learning activity. The user's data is stored as protected resources on the side of Co-Writing Wiki and cannot be accessed from other service without being given the required permission. Using SOFIA middleware addresses this problem through using OAuth protocol (OAuth specification, 2009). OAuth protocol is a secure approach through which the user data (Protected Resources) of Co-Writing Wiki (Service Provider) can be accesses from Moodle (Service Consumer) without Co-Writing Wiki gives the user credentials to Moodle. The sequence followed to provide authenticated access using OAuth protocol is depicted in Figure 6.15.

OAuth is criticized by the week security aspects as it relies on redirected requests. Despite the use of signing methods - i.e. HMAC-SHA1 and RSA-SHA1 - to provide digital signature used in OAuth protocol, the second version of the protocol recommends the use of transport layer security (TLS) between Service Provider and Consumer to maintain security over transported data and tokens as follows, "*If a redirection request will result in the transmission of an authorization code or access token over an open network (between the resource owner's user-agent and the client), the client SHOULD require the use of a transport-layer security mechanism. Lack of transport-layer security can have a severe impact on the security of the client and the protected resources it is authorized to access. The use of transport-layer security is particularly critical when the authorization process is used as a form of delegated end-user authentication by the client (e.g. third-party sign-in service).* " (OAuth V2.0 draft 22)[64]

[64] http://tools.ietf.org/html/draft-ietf-oauth-v2-25#section-3.1.2

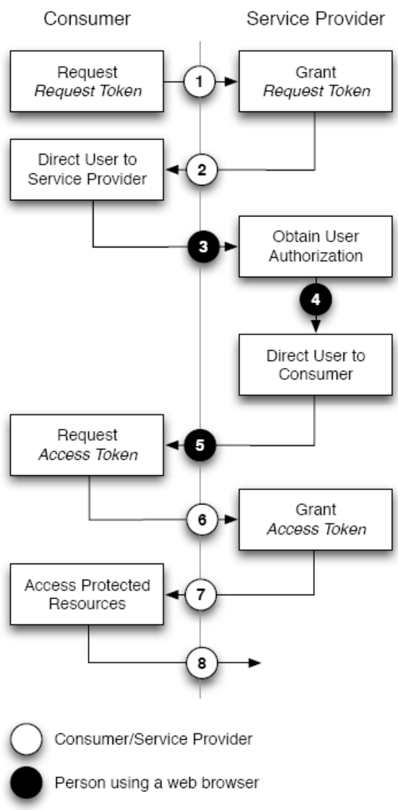


FIGURE 6.15. OAuth Authentication Sequence (Kaila, 2008)

In case of the user is trying to access the same tool several times - i.e. multiple instances - the service 'instance management' form the middleware is used to manage the instances through the 'session management' and 'process management' services. The sequence diagram depicted in Figure 6.14 assumes that both the LMS and the Tool are using the same generalized user roles - e.g. student, teacher, and admin. In case that they use different representation of user roles then the service 'control management' from the SOFIA middleware is used to check the tool access level information through the tool proxy and manages the mappings between user roles by defining a set of permissions for the user role on the Tool resources, see Figure 6.9 for the middleware services.

Privacy is an important aspect which must be considered in e-assessment. Privacy can be defined as "*Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to other*" (Westin, 1970). Based on this definition, e-assessment systems should be capable to protect collected data from users, and to provide services that enable user control on their data access and sharing. Moreover, they should maintain confidentiality of individual information and assessment results. As anonymous users are often prevented in e-assessment systems especially when it comes to provide high stakes assessment, pseudonymous identification - e.g. nickname - can be used to hide the real identity of the user and to link the user profile with the user collected data without revealing the user identity. For instance in some assessment forms the user identity should be hidden to maintain assessment quality and reliability. A good example is peer-assessment in which the assessor does not know who provided the answer otherwise aspects such as friendship may affect the quality of their evaluation.

In SOFIA environment privacy is maintained through preventing anonymous access to user's information and by using pseudonymous identities such as *user1* to show progress charts or contribution charts - as in Co-Writing Wiki. Moreover the users' passwords are encrypted and saved to the database to prevent teachers, and administrators from knowing them. On the level of the SOFIA middleware a service specifically for privacy handling have been added to the architecture as part of the security management services. Moreover, for the application contexts that do not follow the role-based security - authorized access using SSO - a technology provided in .NET framework called Code signing allows the developer to sign an assembly with a private key, and distribute the corresponding public key to each application that references the assembly. This approach is used in the application context of integrated assessment for game-based learning (see Section 7.5) in which the assessment engine web services is signed with a public key and the game engine web player can only invoke this web service using that key.

6.4.4. Standalone SOFIA

One of the main goals of SOFIA is to have a flexible design to work as a standalone system or to be used with other systems to provide integrated assessment to CLRs (see Section 6.3). This section shows SOFIA as a standalone assessment system and sheds the light on the main services developed to provide standard-conform e-assessment. As discussed in Section 5.3, in general e-assessment system has four main modules: Authoring, Scheduling, Delivering, and Reporting modules. Moreover, in order to provide a flexible and interoperable e-assessment system a Service-Oriented Framework for Assessment (SOFA) has been designed (see Section 5.3.2). The SOFA internal representation of e-assessment system is designed through a layered architecture which includes the following layers, (a) the application layer which refers to the user interfaces, (b) the application layer services which holds the e-assessment services - e.g. author, deliver, grade, (c) the common services which has the common services for e-assessment system - e.g. manage users, security, and data management, and (d) the infrastructure layer which covers system resources - e.g. databases and file system (see architecture in Figure 6.1 in Section 6.3.2).

With respect to the application layer services and the common services, Table 6.2 summarizes the SOFIA standalone e-assessment - i.e. MASS - provided services. The e-assessment services interacts with the 'QTI_Manager_Service' - developed as part of SOFIA middleware services - to interpret and validate imported/exported tests and question, to process the responses on questions, and to provide feedback and score. The common services handle aspects related to user management security management, and data management.

TABLE 6.2. SOFIA Standalone e-assessment services.

Category	SOFIA Standalone Service	Description
e-Assessment Services	Question_Author	Handles authoring QTI-based questions as part of the question module
	Answer_Author	Handles authoring answers for authored questions as part of the questions module
	Test_Author	Authors QTI-based tests as part of the assessment module
	Schedule	Manages the schedule meta-data in terms of when, where, and how to deliver authored test as part of the assessment module
	Answer_Mark	Handles the processing of users responses on answers and provides a score to users via the Question_Feedback service

	Test_Grade	Provides the test grade as part of the result module based on the grade scheme defined during authoring the test
	Question_Feedback	Provides the feedback designed during the question and test authoring
Common Services	User_Manager	Manages user data and roles
	Group_Manager	Manages user groups
	Security_Manager	Manages user access level and permissions
	Data_Manager	Manages access to user data and resources such as tests and questions and maintains data persistence

With respect to the application layer, the next part of this section depicts how users of teacher and student view and use the SOFIA standalone e-assessment system. Figure 6.16 depicts the home-page of SOFIA for a teacher user-role. As shown in the figure, different modules have been developed to provide standard-conform e-assessment namely, course, topic, question, assessment, results, and user modules. The course and topic module are used to link the authored questions and assessments to specific topics and courses. Focussing on assessment, the question module provides services of authoring and managing questions. Moreover, features such as question and test import/export are possible with the use of IMS QTI specifications and the support of 'QTI_Manager_Service' discussed earlier. Nevertheless, questions and tests can be created automatically with the use of AQC as one of the third-party tools, see section 6.3.

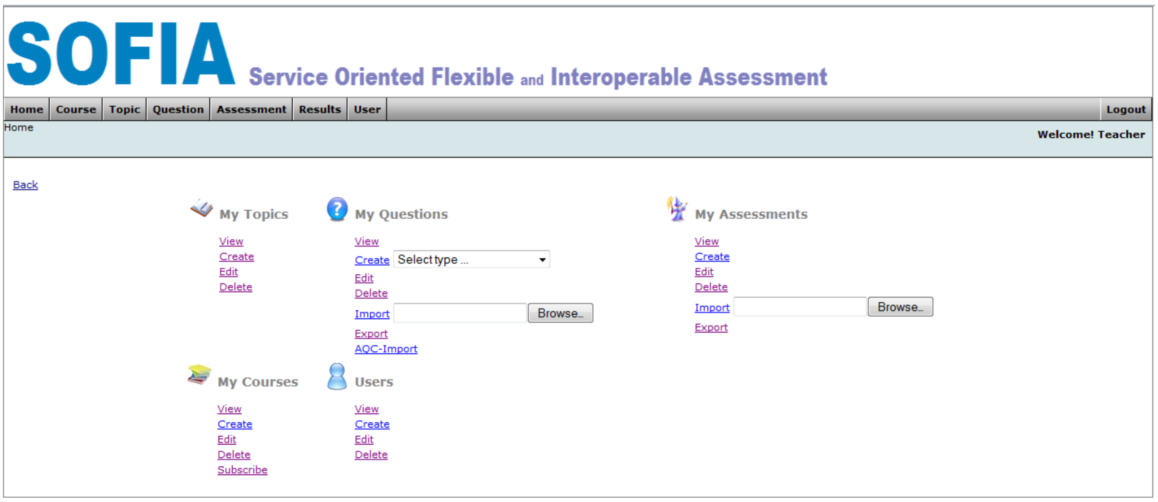


FIGURE 6.16. SOFIA Home-page for adminstrator user role (Teacher View).

In order to provide user friendly interfaces and to maintain an easy way of question and tests authoring, a wizard tabular view is used to author and manage tests step by step as depicted in Figure 6.17. The steps starts with providing meta-data for the test such as test title, test type (formative and summative), related course (only students registered to this course can access the test), and schedule dates (when this test be available), the next step is to create or select questions form the item bank, the next step is about assessment delivery preferences (maxim number of students attempts, time dependent or not, and allow skipping questions or not), whereas the next step is about the grading scheme (passing score threshold, and the range of grades based on marks scale), the next step shows the QTI-XML representation of the test in which specialists in IMS QTI can author and validate - with the support of 'QTI_Manager_Service' - the whole test using this xml editor, and the last step is to preview the test and to try it. The general meta-data in the first step and the preferences meta-data in the third step are based on IMS QTI meta-data for assessment.

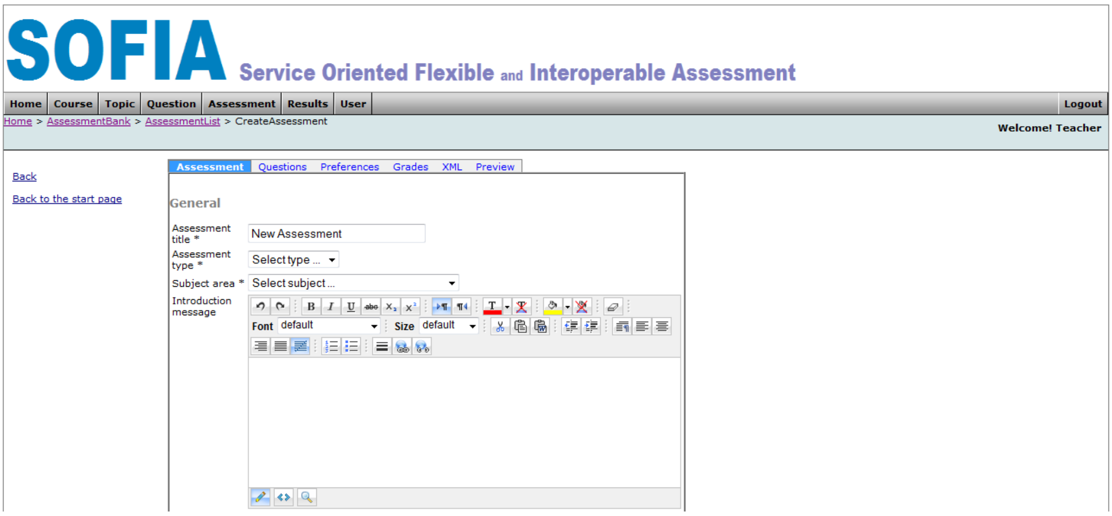


FIGURE 6.17. SOFIA assessment wizard (Teacher View).

Similar to assessment module, a wizard is used to author and manage questions, see Figure 6.18. The first step starts with selecting the question type. SOFIA supports different types of questions based on IMS QTI namely, single choice, true-false, multiple-choice, inline choice, text entry, hot spot, text order, graphic order, associate, and slider questions. Then, the question wizard is started for the selected question type (single choice in Figure 6.18) where, meta-data - based on IMS QTI for test item - can be provided, the answer list can be authored, and the similar to assessment wizard an XML editor for IMS QTI specialist and preview tab to show the question and test it. As discussed earlier the decoupling between QTI business logic developed as a web service - i.e. 'QTI_Manager_Service' - and its application in the context of SOFIA - e.g. support of question and test authoring - fosters SOFIA to share not only test items syntax but also their semantics which are flexibly accessed via the web service (*cf.* AL-Smadi & Guetl, 2011).

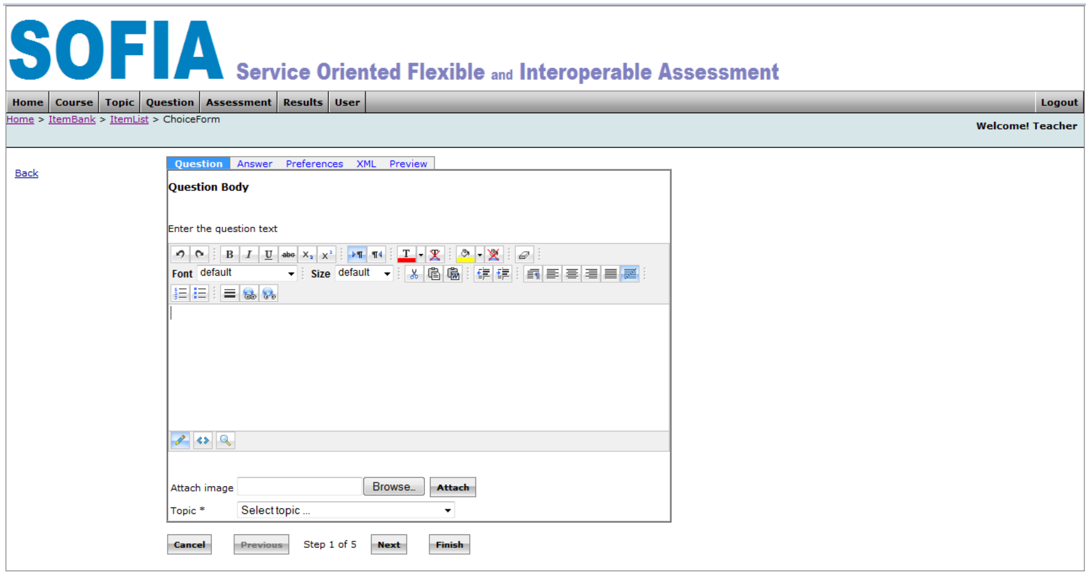


FIGURE 6.18. SOFIA question wizard (TeacherView).

The test and question player in SOFIA uses the HTML5-based QTI player discussed earlier in order to maintain browser-independent QTI delivery. Figure 6.19 shows the test player in SOFIA for a test automatically created using AQC. Once the user starts a test which is time dependent a timer is triggered to check the time consumed by the user against the time allowed to finish the test. Moreover, user responses on test questions are logged and analyzed either immediate to provide immediate feedback in case of formative assessment (see Figure 6.20), or at the end of the test to provide summative feedback (see Figure 6.21). The response processing is handled by the support of 'QTI_Manager_Service' which matches the user response using the response template of the question and the question IMS QTI representation to provide feedback and next step. For more information about IMS QTI response processing and feedback, see Chapter 4.

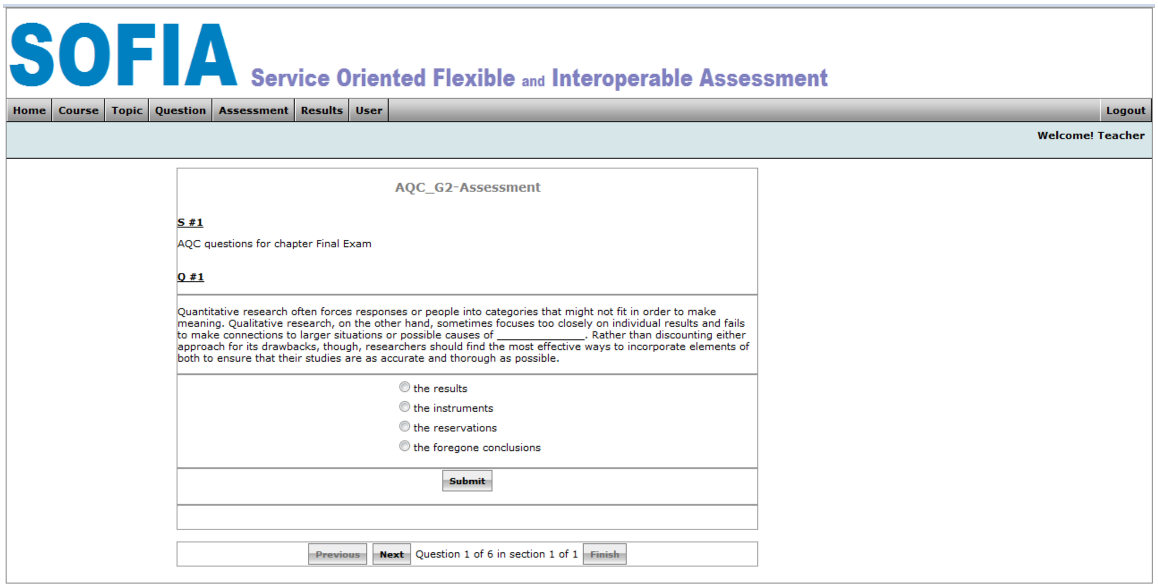


FIGURE 6.19. Playing a test automatically created using AQC in SOFIA (Teacher View).

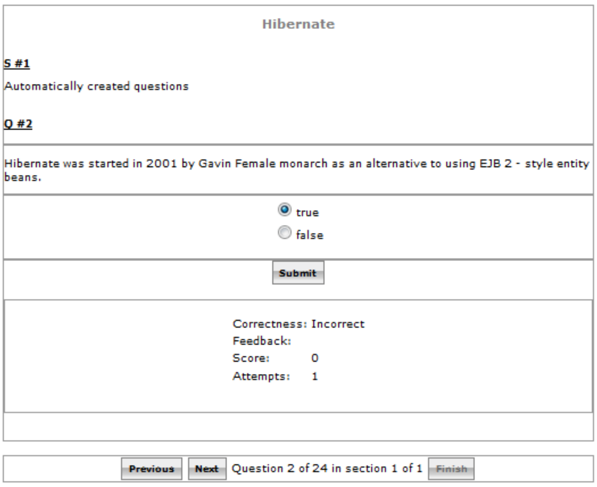


FIGURE 6.20. Question response and immediate feedback (Student View).

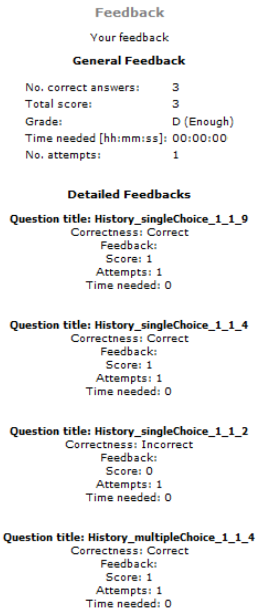


FIGURE 6.21. Post test summative feedback (Student View).

SOFIA provides a simple results module in which the student results based on their taken tests are provided. Aspects related to the test such as course, subject, topic, test title, test, type, number of student's attempts on the same test, number of correct answers, achieved score, final grade, time consumed to answer the test are feedback to the student using the results module, see Figure 6.22.

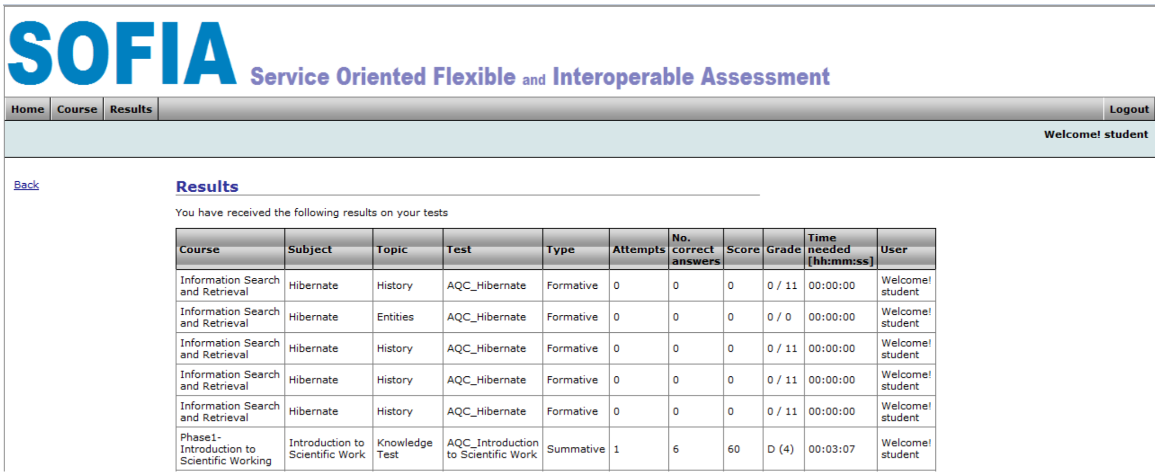


FIGURE 6.22. Part of studnet assessment results (Student View).

To this end, the technology used to develop the third-party tools used in SOFIA application scopes are as follows, PASS (AL-Smadi, Guetl, & Kappe, 2009) was developed using .Net 3.5 framework in particular C# for the code-behind business logic and ASP.Net for the front-end presentation layer, and MySQL database. The system applies the MVC (Model-View-Control) approach. Co-Writing Wiki (AL-Smadi, Hoefler, & Guetl, 2011) is an enhanced wiki for collaborative writing and peer-review has been developed based on ScrewTurn wiki[65]. ScrewTurn wiki is open source wiki developed using C# and ASP.Net for the front-end presentation layer. The engine is partitioned into two main blocks. The Core Assembly contains all the business logic, such as data management and caching, content formatting, provider's configuration and loading and system configuration. The Access control is directly performed by the ASP.Net pages, which also take care of the content presentation and user interaction. The assessment engine for game-based learning (AL-Smadi, Guetl, Dunwell, & Caballe, 2012) is developed using .Net 3.5 framework (C# for assessment engine components and web services). Proof-of-concept of these third-party tools is presented in Chapter 7 as part of CLRs. These CLRs were used in real learning settings to conduct studies in higher education (see the third part of this doctoral dissertation - i.e. experimentation and validation).

6.5. Summary

As discussed earlier in Chapter 5, the rationale for flexible and integrated e-assessment is to enrich complex learning resources - such as simulations, collaborative experiences, virtual experiences, and emotional elements - with integrated forms of alternative assessment capable to evaluate the results of those learning experiences and to support and scaffold students learning process. Moreover, it is important to have content and services interoperability and standard-conform tools when it comes to provide flexible e-assessment for CLR (see Chapter 4). Nevertheless, the necessity to have flexible software architecture to develop integrated and flexible e-assessment tools and services has been discussed in Chapter 5.

Despite the variety in educational standards - including e-assessment, there are some limitations and problems especially in the case of providing flexible and integrated standard-conform assessment forms for CLRs. The limitations are summarized in this chapter as follows:

- Limitations in e-assessment standards such as IMS QTI, (a) to provide test items for specific domains such as, mathematics assessment, programming assessment, or (b) to provide representation for alternative assessment such as performance assessment, behavioral assessment, or rubric-based assessment. Moreover, (c) no consideration of assessment referencing (see Section 3.1.4) - norm-related, criterion-based or ipsative - in the response processing and scoring, (d) there is no support for specific types of assessment such as – self, peer, group-assessment – on the level of test item authoring and scoring.

- Lack of pedagogical aspects in e-assessment standards - such as IMS QTI – as they do not consider, learner preferences, learning style, didactic objective, and learning objectives.

- Available e-assessment systems and tools lacks to some extent the alignment with learning theories and learning outcomes, questions such as what to assess? How to assess? Why to assess this way? When to assess? Who will assess? Who will be assessed? What is the learning type? What is the learning style? are often not considered in the design and the development of assessment tools. Thus, e-

[65] ScrewTurn Wiki - Free ASP.NET Wiki Software [http://www.screwturn.eu/]

assessment lacks the support and consideration of feedback in terms of type, frequency, format, and content.

- e-Assessment tools and systems usually adhere to e-assessment content standards such as IMS QTI and lacks learning tools interoperability standards conformation. Thus, extending e-assessment services with domain-based services through third-party tools requires extra effort in terms of redesign and redevelopment.

- e-Assessment standards and tools lack to some extent adaptive aspects neither on the level of content presentation nor on the level of the services and functions navigation.

By considering these limitations in e-assessment standards, and with the necessity to have integrated forms of e-assessment with the learning process, it is required to have flexible forms of assessment that is developed on top of interoperable software architecture and design in a consistent way to consider both instructional and learning (see Chapter 5). Therefore, the service-oriented approach has been used to develop what we called a service-oriented flexible and interoperable e-assessment environment (SOFIA). SOFIA has been designed and developed to address the aforementioned problems and to provide flexible e-assessment for several application contexts with a variety in learning and instructional outcomes. SOFIA uses the integrated model for e-assessment (IMA) and the service-oriented framework for e-assessment (SOFA) - discussed in Chapter 5 - to design and develop integrated and flexible assessment forms such as self, peer-assessment, automated assessment, rubric based assessment, and performance assessment to evaluate and support students' progress in their learning experiences (see Section 6.3). These learning experiences cover a variety of application scopes - such as, collaborative learning, self-directed learning, and game-based learning - which applies different learning theories and pedagogical approaches - such as, problem-based learning, self-regulation, reflective learning, active learning, and affective learning (see Section 6.1).

SOFIA has been developed to meet the goal of providing flexible, interoperable, and integrated forms of e-assessment for complex learning resources and in particular to achieve the following requirements: (a) *flexible design* to be used as a stand-alone service or to be easily integrated in existing systems. (b) *User-friendly interfaces* for both students and educators where a user interaction and online submission of solution and evaluation can be done. (c) Assessment environment for *various learning and assessment settings* which supports guided as well as self-directed learning. (d) *Management and (semi-)automatic support* over the entire assessment lifecycle (exercises creation, storage and compilation for assessments, as well as assessment performance, grading and feedback provision). (e) *Rubrics design* and implementation interfaces to allow the educators to design their own rubrics based on learning objectives to assess learners' performance against a set of criteria. (f) *Support of various educational objectives and subjects* by using various tools sets which for example enables automatic exercise generation or selection, automatic grading and feedback provision. (g) *Results analysis and feedback provision* (immediately or timely) of the current state of user knowledge and metacognitive skills for both educators and learners and also for adapting course activities and learning contents based on users' models. (h) *Standard-conform information and services* to be easily sharable, reusable and exchangeable. This will include the tests' questions, answers and students' results, rather than any other required services. And finally, (i) *Security and privacy* required mechanisms to ensure that confidential or private data is used or provided as the user wish, and considering organizational rules and ethical aspects. Moreover, secure user's access based on pre-defined roles and access levels is considered in the design and the development of the system.

In order to meet these requirements, SOFIA is built on top of a service-oriented middleware through which the target flexibility - in terms of technology and application domains - can be achieved (see Section 6.3.4). Moreover, SOFIA is standard-conform where IMS QTI has been used to represent the e-assessment content - items, test, and results - through the developed 'QTI_Manager_Service' which is used to interpret, annotate, and validate QTI-based items and tests. This decoupling between QTI business logic developed as a web service and its application in the context of SOFIA fosters SOFIA to share not only test items syntax but also their semantics which are flexibly accessed via the web service (*cf.* Dagger et al., 2007). Moreover, SOFIA is enriched with a browser-independent QTI Player (HTML5) which addresses the problem of having interoperability on the level of the GUI and provided a consistent, browser-independent QTI player (see Section 6.3). In addition to e-assessment content interoperability, SOFIA supports learning tools interoperability where the third-party tools developed in the context of this study adhere to the IMS LTI specifications to provide their services as web services with well-defined interfaces of how to access, launch, and transport services and data in a consistent manner.

Nevertheless, SOFIA middleware helps target users to access resources - i.e. e-assessment content and services and third-party tools - through a secure and authenticated access using SSO approach. The security service in the middleware requires the tools to adhere to the OAuth protocol and facilitates resources access without sharing user credentials.

To this end, in order to provide flexible, interoperable, and integrated e-assessment for CLRs a solution approach is proposed in the second part of this doctoral dissertation through which an integrated model for e-assessment (IMA) is designed to support designing aligned and integrated forms of e-assessment (see Chapter 5). Nevertheless, a SOA based flexible software architecture (SOFIA) has been used to develop those assessment forms and integrate them to complex learning resources (CLR) for three applications scopes namely collaborative learning, self-directed learning , and game-based learning (see Section 6.1). The next chapter sheds the light on the complex learning resources integrated with alternative assessment forms developed based on the SOFIA application scopes.

Chapter

7

7. Integrated Assessment in Complex Learning Resources

As discussed in the first part of this doctoral dissertation, quality education requires an alignment between learning, instruction and assessment (*cf.* Birenbaum, 2003; Biggs, 1999). Assessment has changed form being separated from the learning process to be more integrated and learner-centred. Therefore, the assessment paradigm has shifted towards advocating alternative forms of assessment such as performance assessment, behavioral assessment, portfolio assessment, self-, peer-assessment, and rubric assessment.

In order to address the challenges and problems which outcome from this shift in the assessment paradigm, new forms of learning experiences enriched with complex learning resources - designed based on instructional and learning objectives - integrated with new forms of assessment - e.g. performance assessment, self and peer-assessment, and behavioural assessment - should be considered. Therefore, a solution approach has been proposed in the second part of this doctoral dissertation through which an integrated model for e-assessment (IMA) has been designed to support designing aligned and integrated forms of e-assessment (see Chapter 5). Nevertheless, a SOA based flexible software architecture (SOFIA) has been used to develop those assessment forms and integrate them to complex learning resources (CLR) for three applications scopes namely collaborative learning, self-directed learning , and game-based learning (see Chapter 6).

To this end, this chapter discusses the developed CLRs for the application scopes mentioned earlier. The remaining parts of this chapter is organized as follows: section 7.1 discussed the CLR for enhanced approach for Peer-ASSessment (PASS), section 7.2 discusses Automated and Integrated Assessment in Self-directed Learning CLR, section 7.3 depicts a flexible and interactive tool for assessment rubrics, whereas the CLR for an Enhanced Approach for Collaborative Writing and Peer-review is covered in section 7.4, and section 7.5 explores the final CLR namely an Integrated Assessment Approach for Game-based Learning. Finally, section 7.6 concludes this chapter.

This chapter is based on (AL-Smadi, & Guetl, 2011a; AL-Smadi, & Guetl, 2011b; AL-Smadi & Guetl, 2010; AL-Smadi, Gütl, & Kannan, 2010; AL-Smadi, Guetl, & Helic, 2009a; AL-Smadi, Guetl, & Helic, 2009b; AL-Smadi, Guetl, & Kappe, 2009).

7.1. Enhanced Approach for Peer-Assessment (PASS)

The project PASS (AL-Smadi, Guetl, & Kappe, 2009) aims to develop a peer-assessment tool that can be used with SOFIA standalone e-assessment system (see Section 6.4.4) to support students in evaluating candidate answers using color-coded marks that analytically evaluates the candidate answer based on *correct, wrong,* or *irrelevant* parts. Thus, peer-assessment can be used as a tool for learning through which students learn as they assess and they learning using the feedback they receive from their peers' evaluations. In order to meet this goal, PASS has been developed based on the following requirements:

- Flexible design to be used with other e-assessment systems to support peer-assessment.
- User friendly interfaces to facilitate peer-assessment process and to maintain flexible and easy navigation among questions and candidate answers.
- Support users to prepare reference answers that can be used to maintain quality peer-assessment and should be displayed during the peer-assessment process per question.
- Maintain autonomous assessment as students should not know who provided this answer thus to maintain quality peer-assessment and to avoid biased grading, leniency in the marking process and paybacks by the peers.
- Provide students a color-coded tool to evaluate candidate answers as correct, wrong, or irrelevant.

In order to learn from other research and to identify problems related to peer-assessment, a literature survey has been conducted and related work has been collected as discussed in next section.

7.1.1. Related Work

Several computer-based tools to provide peer-assessment have emerged since the beginning of the 21st century. Some of these tools are part of computer-based assessment systems that implement peer-assessment methods (Davies, 2003). The earliest reported system to support peer-assessment developed at the University of Portsmouth, "*The software provided organizational and record-keeping functions, randomly allocating students to peer assessors, allowing peer assessors and instructors to enter grades, integrating peer- and staff-assessed grades, and generating feedback for students*" (Gehringer, 2001). One of the first systems with the peer-assessment methods was a tool for collaborative learning and nursing education based on multi-user database, which was called MUCH (Many Using and Creating Hypermedia) (Rada, Acquah, Baker, & Ramsey, 1993). In the late 1990s, NetPeas (Network Peer Assessment System) has been implemented, and Artificial Intelligence (AI) has been used to develop the tool of Peer ISM that combines human reviewing with artificial ones (Rada, Acquah, Baker, & Ramsey, 1993). Computer-assisted-peer-assessment systems have been also affected by the revolution of World Wide Web (WWW). An example of the first reported web-based system was a web-based tool for collaborative hypertext authoring and assessment via e-mail (Downing & Brown, 1997). Other systems such as, a web-based system for group contributions on engineering design projects (Eschenbach & Mesmer, 1998), the Calibrated Peer Review (CPR) which was introduced in 1999 (Carlson & Berry, 2005), the Peer Grader (PG) as a web-based peer evaluation system (Gehringer, 2000), The Self and Peer Assessment Resource Kit (SPARK) which is an open-source system designed to facilitate the self and peer assessment of groups (Freeman & McKenzie, 2002), The computerized Assessment by Peers (CAP) is another example (Davies, 2003). Further examples such as, OASIS which has automated handling for multiple-choice answers and peer assessment for free-text answers, The Online Peer Assessment System (OPAS), which has some abilities for assignment uploading and reviewing as well as groups management and discussions (Trahasch, 2004), An improvement for this system was introduced in Web-based Self and Peer Assessment (Web-SPA) system to avoid the lack in determining standards, methods of scoring and the workflow of the assessment process (Sung, Chang, Chiou, & Hou, 2005). Recent examples of peer-assessment developments are, the enhanced open-source implementation of WebPA system which was originally developed in 1998 (WebPA, 2009), as well as the Comprehensive Assessment of Team Member Effectiveness (CATME) system which assesses the effectiveness of team members contributions (Ohland, Pomeranz, & Feinstein, 2006).

7.1.2. PASS Development

With reference to PASS requirements discussed earlier and based on the architecture designed for this approach (see Section 6.3; Figure 6.2) PASS was developed using .Net 3.5 framework in particular C# for the code-behind business logic and ASP.Net for the front-end presentation layer, and MySQL database. The system applies the MVC (Model-View-Control) approach. PASS is designed to work with SOFIA standalone assessment thus uses assessment services namely *mark, grade, analyze, feedback* to provide its features. Nevertheless, services such as *manage users* and *security services* form SOFIA common services are used to maintain role based access and permissions during the peer-assessment process.

Focusing on peer-assessment process, online peer-assessments can be authored from SOFIA standalone (see Figure 7.1) by defining the grade schema of an online test as peer-assessment and selecting the test items that their answers will form as candidate answers during the peer-assessment process. Nevertheless, some configuration is required to group students if required and to design the feedback based on the peer-assessment process. Moreover, PASS provides a feature to define a set of calibrated answers for each question selected to be graded in the peer-assessment process (see Figure 7.2).

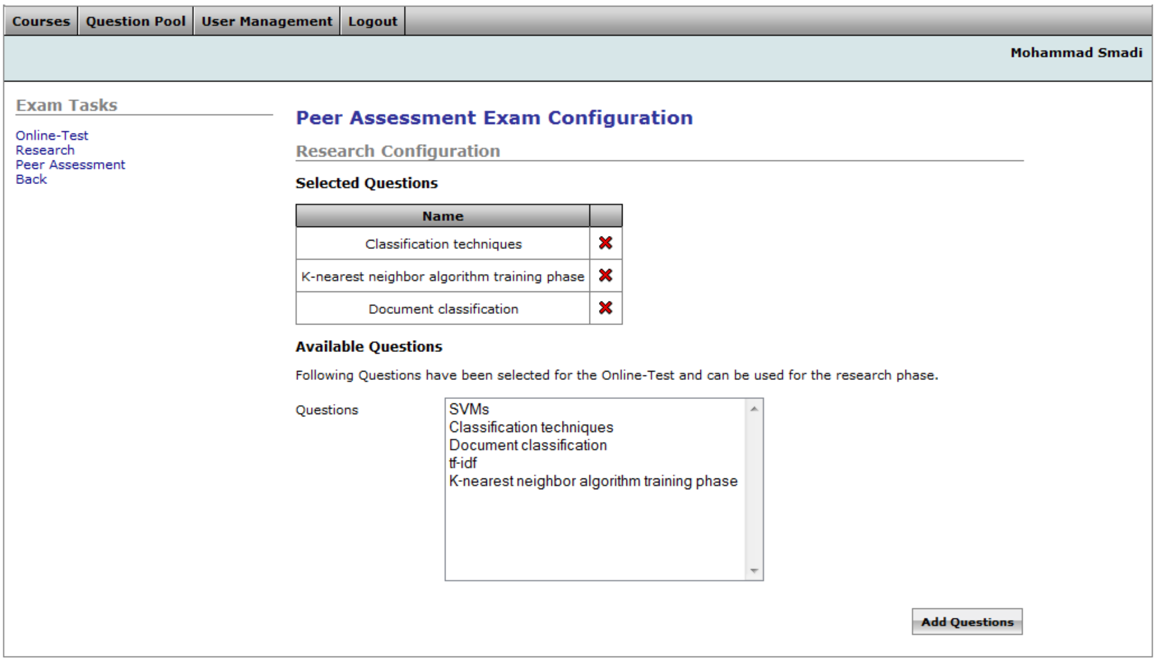


FIGURE 7.1. Selecting questions form the online test to be graded using peer-assessment (Teacher View)

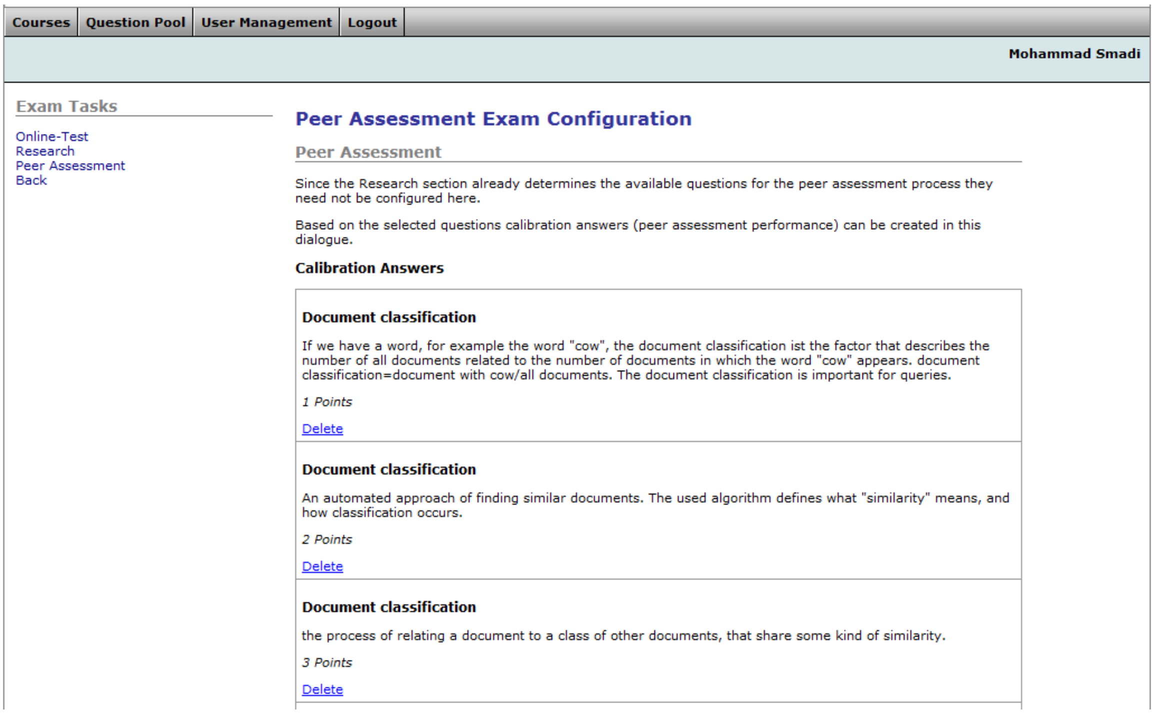


FIGURE 7.2. A set of claibrated answers provided during peer-assessment configuration (Teacher View).

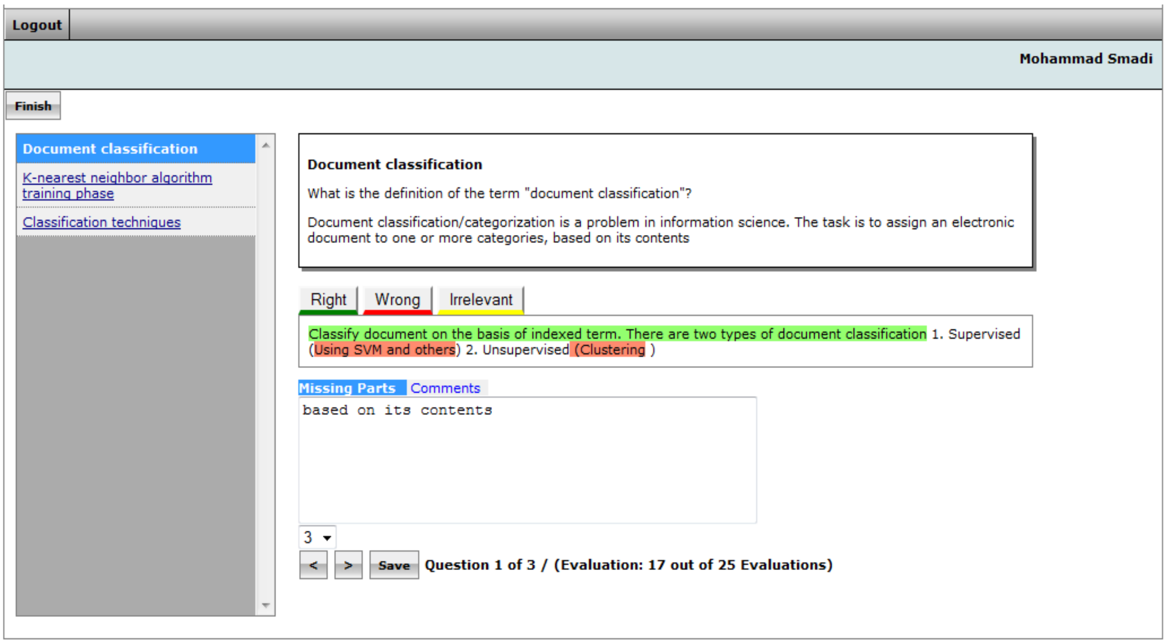


FIGURE 7.3. PASS tool running under SOFIA standalone assessment environment during the peer-assessment step (Student View).

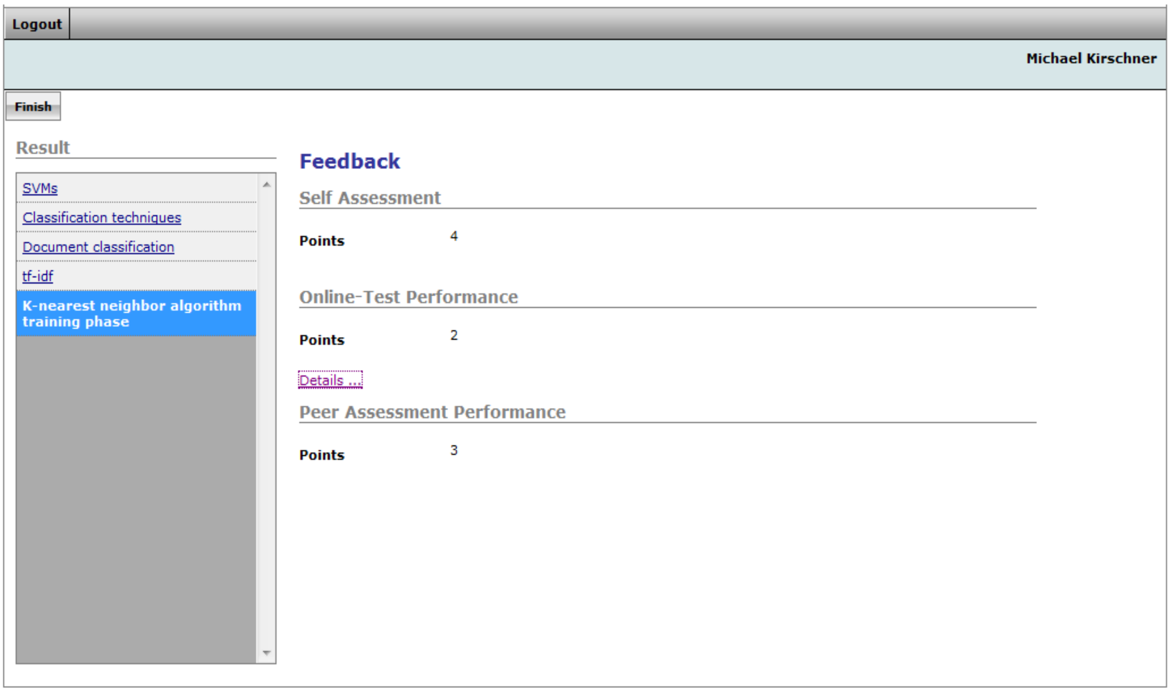


FIGURE 7.4. PASS Feedback tool (Student View)

Figure 7.3 depicts the features provided by PASS peer-assessment process. These features include question navigation menu (see left menu), question title, question body, and the student's prepared reference answer for the selected question, the marking tool which provides color-coded tagging based on correct, wrong, or irrelevant parts of the candidate answer, the candidate answer to be evaluated by the peer, and two textboxes to provide comments on the answer and missing parts of it, as well as score based on the scale defined during the online test authoring in SOFIA (e.g. 0 -5, 0 - 10, A - F), and finally navigation tool to navigate among the candidate answers for the selected question and updated feedback of how many answers have been evaluated and how many left.

One of the main benefits of peer-assessment is to provide feedback (see Chapter 3). In PASS a tool for providing feedback was developed. The feedback tool uses SOFIA feedback service to provide peers evaluations on candidate answer to the answer owner. Nevertheless, each student gets a grade computed based on the average of the absolute error between the students' average per-assessments and the tutors' reference rates per candidate answer (see Chapter 8). Figure 7.4 depicts a screenshot form PASS feedback tool. Feedback regarding student's self-assessment compared to tutors reference rates per candidate answer, online test performance based on tutors and peers grades, and each student peer-assessment performance using the average of the absolute error between his peer-assessments and the tutors' reference grades for the same candidate answers is provided.

PASS with the support of SOFIA standalone was used to conduct a study for peer-assessment in real learning settings. First findings shows that students liked PASS enhanced approach for peer-assessment. Nevertheless, the color-coded approach supported them to provide quality assessment and to learn from assessment, for more information about this study see Chapter 8.

7.2. Automated and Integrated Assessment

This scenario emerged from the ALICE project as discussed in Section 6.2.3. The scenario aims to support self-directed learners with automated assessment and feedback to scaffold their self-awareness about their knowledge state (AL-Smadi & Guetl, 2011a; AL-Smadi, Hoefler, & Guetl, 2011). In order to meet this goal SOFIA standalone assessment was extended with a third-party tool called Automatic Question Creator tool (AQC; (Gütl, Lankmayr, Weinhofer, & Höfler, 2011)). AQC creates automatically test items from textual learning material. With the support of SOFIA, AQC can be used as a third-party tool to automatically create test items. The automatically created test items can be used to automatically create test using SOFIA assessment author module.

SOFIA with AQC third-party tool support to main learning scenarios: (a) Fully automated scenario by which students and teachers have no control on the assessment author except selecting the learning material, (b) Interactive approach in which students and teachers are capable to select learning material, tag and select the main concepts, interact with AQC to create test items for these concepts, and get automated assessment using SOFIA after that.

The scenario was developed based on the following requirements: (a) advanced tool supporting the creation of four different test item types: multiple-choice questions, true/false, fill-in-the-blank, and open ended questions. (b) Learning setting dependent operating modes supports fully-automatic test item creation and interactive process types taking into account student or teacher input. (c) Domain knowledge independent methods allow test item creation of unseen textual content by applying statistical, semantic and structural analyses. (d) Language dependent data flow and process chain design provide multilingual test item creation, currently English and German, and support the easy extension to other languages. (e) Flexible design supports an easy integration or exchange of modules in the system to offer improved processing tasks or even new features. (f) Easy integration into other systems and service provision by a standard-conform web service interface. (g) Standard compliance enables an easy export and reuse of test items created by the tool.

7.2.1. Related Work

Automated test item creation from textual learning material has raised the interest of the community for quite a while (see Section 3.2.2). As discussed in (Höfler, AL-Smadi, & Gütl, 2012) There are many approaches to (semi-) automatically generate test items from a given learning content (e.g. Boyer & Piwek 2010; Liu & Calvo 2009; Stanescu, Spahiu, Ion, & Spahiu 2008). For instance, Chen, Aist, and Mostow (2009) developed a tool which generates open-ended test items out of informational texts. Liu and Calvo (2009) provided a tool that is able to support students in revising their own written essays through automatically generated open-ended questions form out of their essays. According to (Agarwal, Shah, & Mannem, 2011) there are two main challenges in test item generation, first, relevant content has to be identified for which test items are to be generated and second, a corresponding test item type has to be found. Usually, research concerning automatic test-item generation focuses on the generation of only one type of test item (mainly open-ended items or multiple-choice items) from one given sentence (e.g. Goto et al. 2010). Few studies deal with the generation of various types of test item. For instance, Myller (2007) investigated the generation of prediction questions for visualisations through test items of multiple-choice items, single-choice items, or open-ended items. Another example is the Automatic Question Creator tool (AQC; (Gütl, Lankmayr, Weinhofer, & Höfler, 2011)). AQC utilizes an automated process to create different types of test items out of textual learning content, more precisely to create single choice, multiple-choice, completion exercises and open ended questions. AQC is capable to process textual learning content stored in various file formats, extracts most important content and related concepts, creates different types of test items and reference answers, as well as exports the those items in QTI format.

In order to harness the potential of automated assessment in learning, the approaches of automatically create test items mentioned earlier should be utilized in real learning settings to provide assessment (AL-Smadi & Guetl, 2011; Höfler, AL-Smadi, & Gütl, 2012). An example on this potential use of automated assessment for learning is the scenario discussed in this sub-chapter. The scenario enhances the AQC tool and used in a self-directed learning approach to support learners with automated assessment thus to enhance and support their learning. The next section shows how this scenario was developed and gives an overview about student and teacher use cases.

7.2.2. Scenario Development

AQC uses a processing module to analyze textual learning material and extracts main concepts form it. The extracted concepts are then used to automatically create QTI-compliant test items namely open-ended, single-choice, multiple-choice, and fill-in-the-blank questions. By using the 'QTI_Manager_Service' SOFIA uses the created test items to automatically create test which can be used in a self-directed approach to support students test themselves based on their convenience (see Section 6.3.2 for scenario architecture). Nevertheless, AQC has been enhanced with the support of SOFIA middleware services to be used with other tools and systems in supporting self-directed learning settings with automated assessment. Moreover, students can tag the learning material with their manual concepts that they want to learn and SOFIA calls AQC with these concepts and automatically creates tests accordingly.

In order to extend SOFIA services with AQC a set of services was developed as part of SOFIA middleware. These services include 'AQC_Runtime' which is responsible for deploying AQC tool within the SOFIA environment runtimely, 'AQC_QuestionsManager' and 'AQC_ConceptsManager' which are responsible for calling AQC with learning material and concepts and handle AQC outcome - i.e. automatically created questions - to be used in the context of SOFIA assessment module, for more information about SOFIA middleware see Section 6.4.

Focusing on the two scenarios mentioned earlier, during the interactive approach the teacher can use AQC tool to create automatically test items during the authoring of the learning activity (see Figure 7.5), and then the resulted questions can be configured and used to create a test which can be delivered as part of the learning activity (see Figure 7.6). In this step the teacher is capable to define how many questions out of the created automatically per each question type can be used to create the automatic test using SOFIA assessment module. After the teacher saves the test configuration, SOFIA assessment module with the use of 'QTI_Manager_Service' creates the IMS QTI representation for this test. Once the learning material is delivered to students, a 'Test me' button is provided whereby students can receive an automated test created by selecting randomly questions from the automatically created ones based on the numbers provided in the test configuration step (see Figure 7.7).

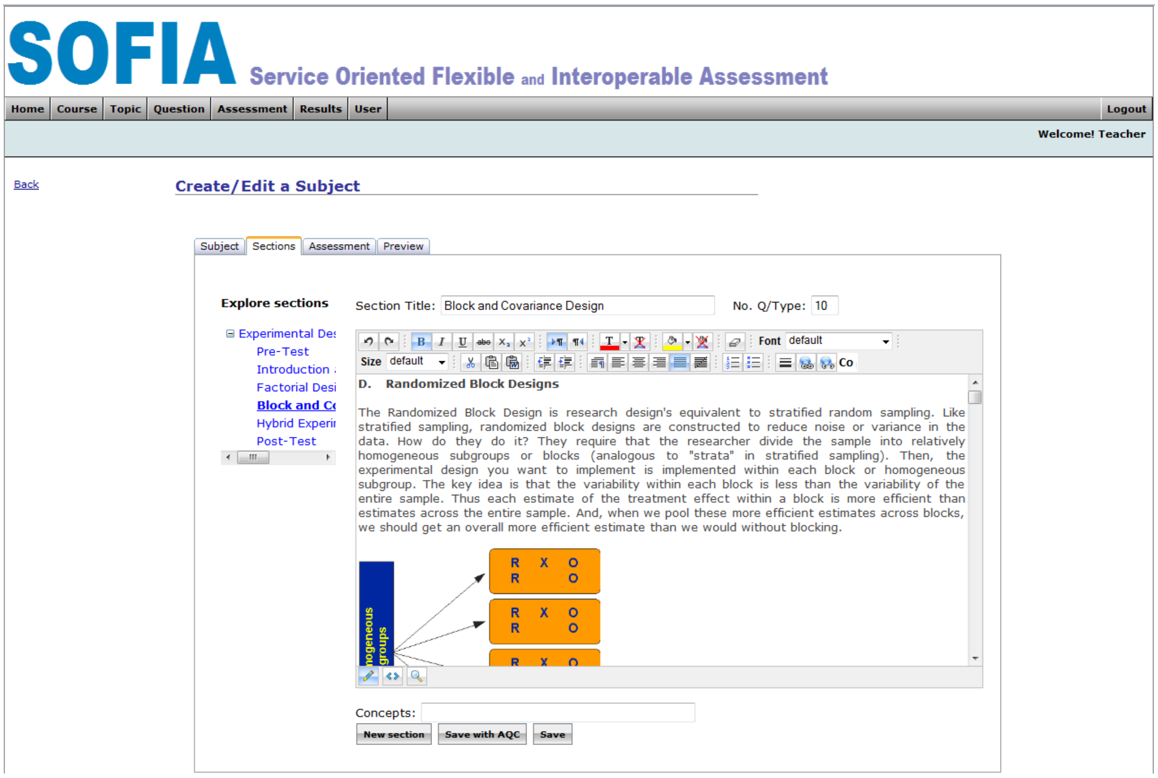


FIGURE 7.5. Authoring learning material with automated test using SOFIA and AQC (Teacher View)

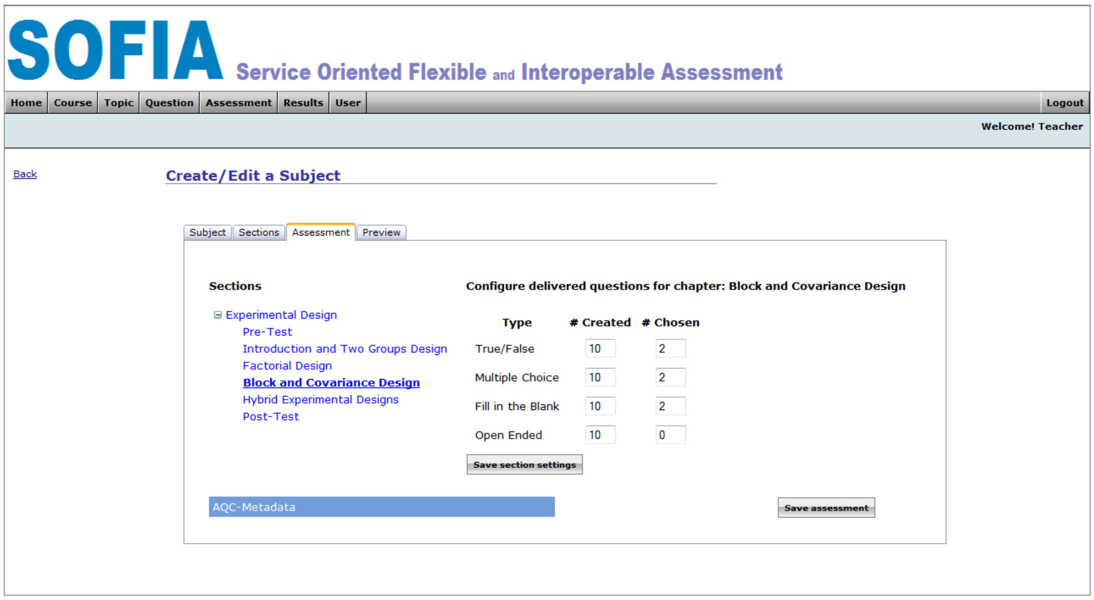


FIGURE 7.6. Configuring how many questions from the automatically created ones will be delivered randomly using SOFIA assessment deliver module (Teacher View)

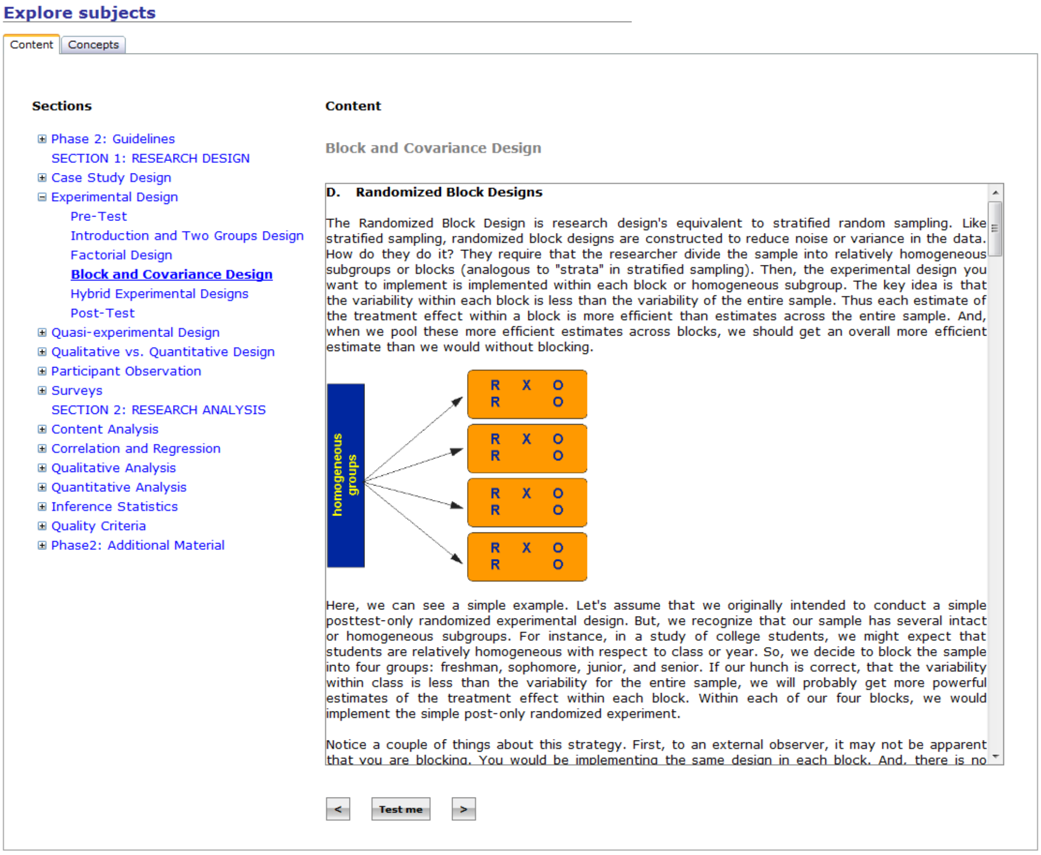


FIGURE 7.7. Online material provided using SOFIA with automated test called by 'Test me' Button (Student View)

During the learning phase students can select specific concepts from the learning content in the 'Concepts Tab' and click on the 'Test me' button to get automatically tests that are related to the selected concepts (see Figure 7.8). Using this approach, students only get questions related to the concepts they want to learn thus support them to regulate their learning based on their knowledge state.

To this end, this enhanced approach of automated assessment using SOFIA and AQC was used to support a study for self-directed learning CLR. Findings show that students liked the approach and they stated that testing themselves with questions often helped them for learning. Nevertheless, the students' motivation during the self-directed learning phase integrated with automated assessment was above average, for more information see Chapter 9.

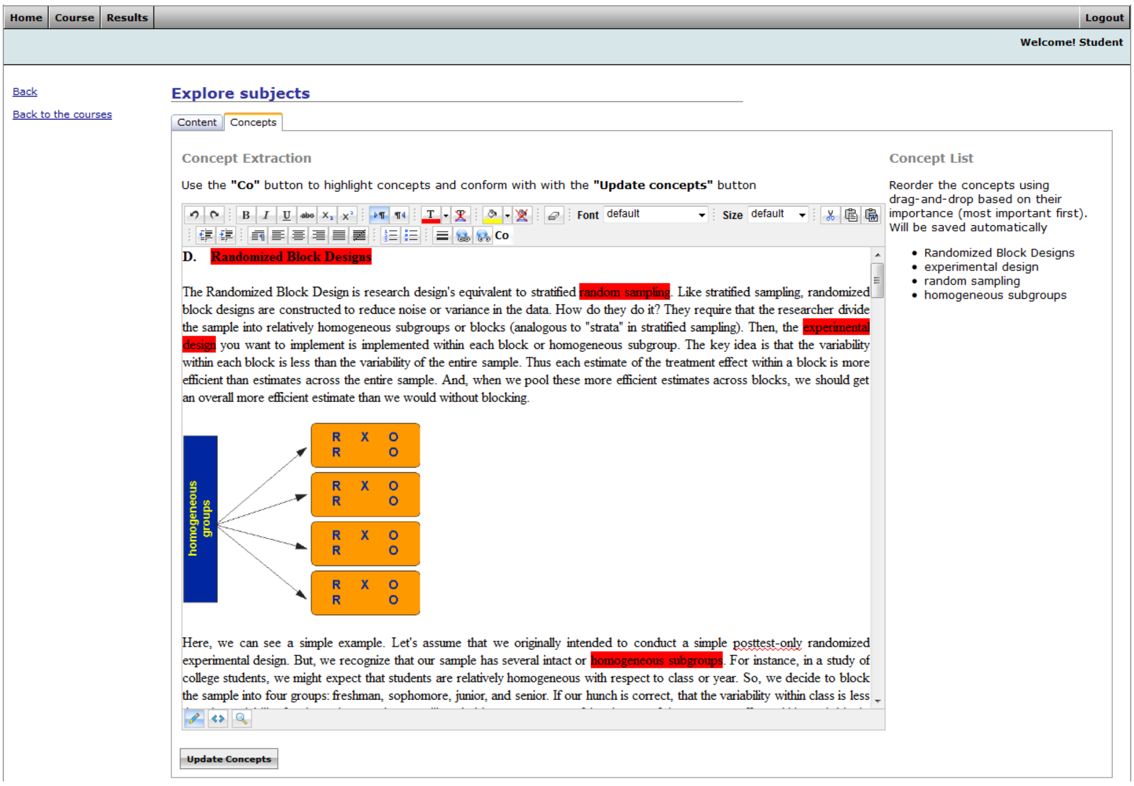


FIGURE 7.8. Selecting concepts from learning material during the interactive appraoch (Student View)

7.3. Flexible and Interactive Assessment Rubrics

One of main requirements for SOFIA is *rubrics design and implementation interfaces* to allow the educators to design their own rubrics based on learning objectives to assess learners' performance against a set of criteria (see Section 6.3.3). Assessment rubrics are a list of criteria each represented by levels of quality and their description in terms of mark scale or quality scale of poor to excellent (Andrade, 2000). In addition to that assessment rubrics help to figure the students strengths and weaknesses however its time consuming as they focus on variety of marking criteria. The use of rubrics in assessment enhances assessment transparency, reliability, and validity (Andrade, 2005; Tiereny & Marielle, 2004). Moreover, rubrics aim to support students with feedback based on clear levels of quality and marking criteria. Nevertheless, the use of rubrics can support students focus their effort and tasks to promote meta-cognitive skills such as, self-awareness and self-reflection as well as higher level of thinking (Andrade & Du, 2005; Tiereny & Marielle, 2004). However, creating an instructional rubric is a challenging task itself and teachers may find it time-consuming (see Section 3.1.3).

In order this requirement for SOFIA as well as to facilitate rubrics authoring and using a project with an aim of providing flexible and interactive assessment rubrics was started. The next sections shed the light on this project and in particular on the technology used to develop the tool for assessment rubrics.

7.3.1. Related Work

As discussed in (Weinberger, Dreher, AL-Smadi, & Guetl, 2011) examples of computer-based assessment rubrics include Writing Roadmap 2.0 is a tool designed for language training using a holistic rubric in which automatic results can be overridden by teachers (Rich & Wang, 2010). The system is limited as it cannot be used outside the language-training context and teachers cannot correct the additional analytical results. Other rubric-based systems such as iRubric[66] only provide tools to design rubrics and electronic grade handling but are not specific to essays and therefore offer no supporting essay analysis or other automatic features. In addition to iRubric other generalized assessment rubrics tools include Rubistar[67] which provides features to author and share rubrics for different application domains such as essay grading, reports grading, and oral presentations. Rubric Maker[68] provides an easy way to author rubrics for different topics such as behavioural, art, content, design, research, etc. store, and export in Excel and Text formats.

In the related domain of fully automated essay grading a small number of systems have come to prominence - Project Essay Grade, Intelligent Essay Assessor, MarkIT™, IntelliMetric, E-Rater and e-Examiner for example. Some of these, as Criterion, are mainly used in the context of language instruction. As the application of essay-style testing in higher education is broader, the content and concepts to be assessed in the submitted text are more varied than those for assessing basic writing skills making solutions that neglect the content unusable. MarkIT™ is an explicitly content centred approach employing semantic analysis but still needs around one hundred human scored essays in the training phase (Williams, 2006). Similarly, approaches based on Latent Semantic Analysis evaluate the content of essays to determine the similarity between documents (Landauer et al., 1998) as applied in the Intelligent Essay Assessor; it also requires 100-200 pre-graded essays (Palmer et al., 2002) (see Section 3.2.2).

The aim of this project is provide a flexible way to author and play assessment rubrics that can be used in different application scenarios and to be integrated to CLRs. Few specifications and standards are available to represent rubrics which make it difficult to share rubrics between different systems. For instance, IMS Rubrics specifications v1.0[69] is still in its early stages and not widely used in the domain or assessment rubrics tools. However, it is important to have a common information model to represent rubrics thus created rubrics can be reused in different scenarios and use cases. Authoring assessment rubrics require high level of interaction as for some rubrics you need a matrix of 30 controls representing 6 criteria and 5 mastery levels each. The next section shows the development of the assessment rubrics and how they are used integrated with CLRs to provide assessment.

7.3.2. Project Development

In order to facilitate the development of high level interaction and to support different internet browsers, the Rubric tool was developed using the jQuery 1.4 library. jQuery[70] is a cross-browser JavaScript library designed to simplify the client-side scripting of HTML traversing, and Ajax interactions for web development. Figure depicts the editor of the rubric tool while authoring an assessment rubric.

---

[66] www.rcampus.com/indexrubric.cfm

[67] http://rubistar.4teachers.org/

[68] http://myt4l.com/index.php

[69] http://www.imsglobal.org/ep/epv1p0/imsrubric_specv1p0.html

[70] http://jquery.com/

In order to use this tool with CLR to support different application contexts, the JavaScript code creates a JSON[71] (JavaScript Object Notation) representation in a hidden field before posting back to the web server. On the server side the JSON data are converted into the required storage representation. Alternatively, it is also possible to change the JavaScript code to generate XML instead of JSON or to adhere to rubrics specifications such as IMS Rubrics v1.0. Figure 7.9 shows an example of authoring a rubric using the Rubric Editor tool.

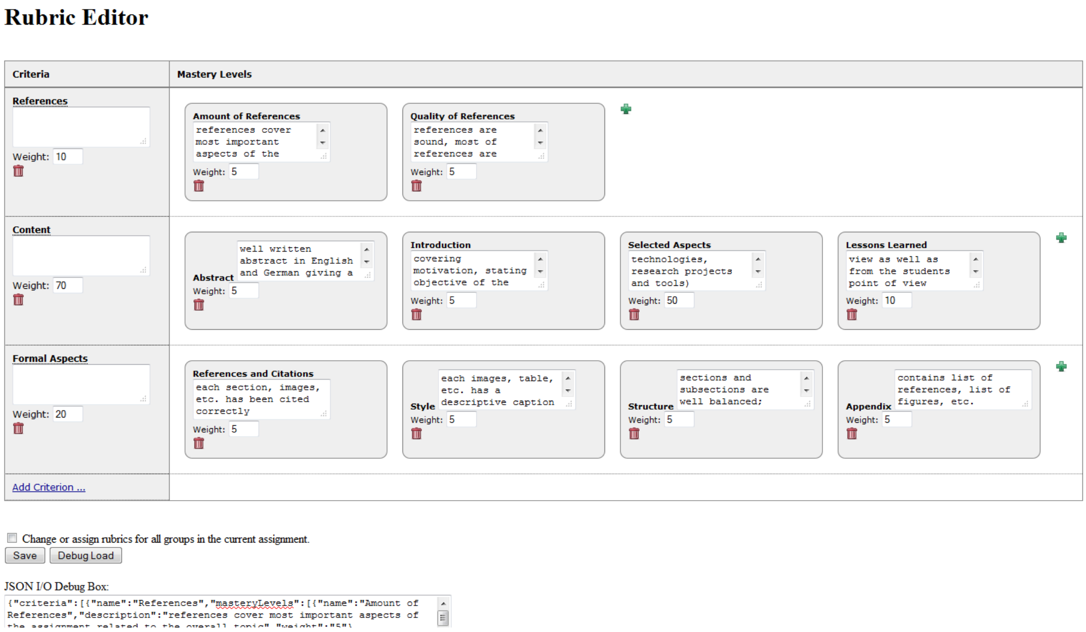


FIGURE 7.9. Authoring a rubric using the rubric editor (Teacher View).

The rubric tool editor is currently used with the latest version of Co-writing Wiki (AL-Smadi, Hoefler, & Guetl, 2011b) and used to design assessment rubrics for assignment group-assessment. The next sub-chapter explains in detail the Co-writing Wiki CLR development as part of an enhanced approach for collaborative writing and peer-review.

7.4. Enhanced Approach for Collaborative Writing and Peer-review

The scenario aims to provide a new form of assessment for collaborative writing using wiki, where self-, peer-assessment forms are used to evaluate students' contributions. In order to meet this goal a tool called Co-writing Wiki (AL-Smadi, Hoefler, & Guetl, 2011b) was developed to provide an enhanced approach for collaborative writing and peer-review.

Co-writing Wiki was developed based on the following requirements: (a) to provide enhanced tools to maintain task and social awareness and to support group well-being and production function during a collaborative writing assignment, (b) integrated self, peer, and group-assessment activities with the use of assessment rubrics designed for scientific writing, (c) continuous Feedback provision for learner learning scaffolding as well as for teachers to follow collaboration progress, (d) visualization tools to support both students and teachers to know who did what and when, and (e) motivational Charts in order to motivate peers to contribute and work in comparison with others in the same group as well as to motivate groups to contribute in comparison with other groups.

[71] http://www.json.org/

In order to support the instructor as well as the students with valuable feedback, an enhanced visualization tool was developed as part of Co-writing Wiki. The visualization tool retrieves the students' contributions and interactions and visualizes them in an enhanced way. The tool may support the instructor to mark the student's performance where the following questions could be answered. How much has each student contributed to the assignment product? How collaboration taking place? To what extent the students are collaborating within the group?

To this end, the next sub-chapters discuss related work which was considered during the design and development of Co-writing Wiki.

7.4.1. Related Work

Section 3.2.2 discusses in detail computer-based assessment in computer supported collaborative learning (CSCL) and in particular in wiki based co-writing. In addition to that section, this section focuses on problems that should be considered when it comes to provide CSCL and gives examples form literature about types of knowledge visualizations used in wikis to support co-writing.

As discussed in (AL-Smadi, Hoefler, & Guetl, 2011b; cited after Janssen *et. al.*, 2007), CSCL has some problems and challenges that include:

- *Lack of awareness:* awareness concerns useful information that group members need on what others are doing, what others know about the current task, what group members will do next. According to (Romero-Salcedo et al., 2004) awareness is a problem of perception and information. Group awareness information may reduce group members' effort to coordinate among them, may increase their efficiency, and may reduce the chance for errors (Gutwin & Greenberg, 2004). Moreover, awareness information is important to monitor group mutual performance, as group members are collaboratively working on shared tasks they need to monitor whether other members are performing well (e.g., Who is doing what? Is group member's performance on a sufficient level?). Nevertheless, Conversational awareness information is important to have quality discussions. Another important type of awareness information is social awareness. Social awareness is required to regulate social aspects of the collaboration, enhance group coordination (e.g., who is available for discussion and communication? Who needs help? Is collaboration going fine or should changes be made?) (Kreijns, 2004). Furthermore, social awareness may support group members to avoid the problem of free rider effect (Salomon & Globerson, 1989) by knowing who is doing what. Therefore, working in a CSCL environment requires group members to have not only task-related awareness information, but awareness information about social aspects as well.

- *Coordination problems:* CSCL is a difficult task to students as they are required to perform a variety of group activities while working on a collaborative learning task. During collaboration group members have to maintain communication and coordination among them regarding the collaborative tasks. They have to exchange ideas, ask questions, enter in arguments, and direct their effort and progress towards the group product. This process is called *production function* of groups where students involve in social interactions in order to maintain *group well-being* and share social space for *member-support* (McGrath, 1991). As collaboration involves different types of group activities, coordination among group members is required. Erkens (2004) identified three types of activities that affect group coordination: (a) *Activation of knowledge and* *skills:* this includes the initial communication and knowledge sharing among group members to define tasks and provide member support. Sharing knowledge and skills improvement are important activities for group's well-being. Moreover, they may foster collaboration with equal participation and contribution of group members so that each group member will have the opportunity to contribute to group production function, to engage in knowledge construction, and to utilize her/his skills during the production process (Barron, 2000). (b) *Grounding:* is another important activity that group members have to maintain. Group members have to have a common understanding of tasks and they have to ensure that they understand each other. In order to achieve grounding, the following strategies can be used: *tuning*, *joint attention*, *focusing*, and *checking*. (Janssen *et. al.*, 2007) (c) *Negotiation and coming to agreement:* despite the common understanding in the grounding processes and knowledge sharing strategies, group members have to negotiate the problem state and to come to an agreement about possible solutions and next steps.

- *Communication Problems:* research in CSCL has shown that communication problems mainly concern the communication media itself. Traditional CSCL communication media (e.g., e-mail or chat) lacks *media richness*. Media richness can be defined as "*a medium's ability to facilitate communication and the establishment of shared meaning. Factors such as the immediacy of feedback or the ability of the medium to transmit multiple cues (e.g., facial expressions, gestures, or intonation of voice) influence its richness*" (Janssen *et. al.*, 2007). Low media richness may prevent group members from understanding group discussions which affects CSCL process with coordination problems and lack of quality discussions. Therefore, rich CSCL communication media - in terms of facial expressions, gestures, or intonation of voice - such as video conferencing has been used to foster the group communication.

Knowledge visualization of behavioral data and interactions - using textual and graphical visualizations - have been recommended as a possible solution in order to support CSCL in both the collaborative learning process itself and group learning scaffolding (Janssen *et.al.*, 2007; Zumbach & Reimann, 2003; Reimann & Kay, 2010). Designing a suitable visualization highly depends on the following: (a) *what* information it will visualize: CSCL related information can be either task-related (e.g., How many problems have been solved by the group?) or social-related (e.g., How many messages have been sent by each group member, or how much each group member have contributed to the CSCL product?) or both. Moreover, selecting information related to the aforementioned production function, member-support, and group well-being functions (McGrath, 1991; Zumbach & Reimann, 2003); (b) *why* is it important to visualize those selected information; and finally (c) *how* those information will be visualized: regarding this question possible visualization can be textual representations (e.g. tables or hints) or graphical representations (e.g. graphs and charts) or a combination of both. However, visualizations have to be carefully selected and designed so that group members can easily perceive and interpret them correctly (Keller & Tergan, 2005). Furthermore, visualization aspects in CSCL can be used to scaffold task/social group activities in such a way to foster them to provide evidence for the assessment process (AL-Smadi, Hoefler, & Guetl, 2011b; Reimann & Kay, 2010).

Examples form literature regarding the use of knowledge visualization to support CSCL - in particular wiki based collaboration - with a focus on assessment aspects are summarized as follows. The work of (Zumbach & Reimann, 2003) discusses how knowledge extracted from social and task-related aspects of the collaborative process can be visualized in such a way to provide feedback to the online collaborators. The CSCL environment tracks and logs group members' interactions, analyzes these interactions, and feeds it back using a combination of textual and graph visualizations. The aim of this research is to investigate how the knowledge extracted from the interactions of small problem-based learning groups can be supported by means of visual feedback and used to scaffold group's function and well-being. However, they have analyzed firstly, parameters of interaction namely participation behavior, learners' motivation, and problem-solving capabilities by which they have investigated group coordination and enhanced group well-being. Secondly, they have tracked and analyzed interactions related to the task of problem solving design and provided feedback in form of problem-solving protocols. Furthermore, at regular intervals each group member had to rate his motivation using pre-defined forms. These data were aggregated over time and visualized using line graph showing all group members motivation. Nevertheless, group members' contributions were recorded by the CSCL environment and visualized as pie chart.

An example for a visualization of wiki site structure is WikiNavMap (Ullman & Kay, 2007). This tool enables the user to customize the view of the wiki in terms of time and in relation to the authorship of activity on the pages. Moreover, WikiNavMap shows a navigational role, and also increases member and task awareness (hence, affecting coordination), and helps to monitor coherence.

Another example that shows how visualizations aspects can be used to facilitate the assessment of wiki-based collaborative writing can be found in (Biuk-Aghai, Kelen, & Venkatesan, 2008). In this research the authors customized the "MediaWiki" to what they named "TransWiki" in order to be used in translation courses. Moreover, they developed visualizations in order to support the teacher answering the following research questions: How much has each student contributed to the final product? What is the process of collaboration? What is the depth of collaboration? Nevertheless, they used *color-coded textual visualization* to show individuals contribution to a wiki-page, the differences between two versions, as well as the depth of collaboration. They used the *analysis graph* (single/all users) to demonstrate the evolution of an article with all users or the evolution of a single user interaction per page. Moreover, they also used *Contribution summary graph* to demonstrate the amount of contribution per user.

In (Trentin, 2009), the author tested an approach for co-writing using wiki where the students used online discussion forum for co-planning and structuring the content for the co-writing phase. Moreover, they used online discussion forum for peer-review where they were required to peer-review their peers contributions and writings. 3D graphic projections had been used to visualize both the interaction among participants and among the links between the hypertext pages. Moreover, network analysis techniques had been used to represent the reticular relationships among those interactions. According to (Trentin, 2009) the use of 3D projections and the network analysis for the visualizing the reticular relationships among interactions has facilitated the evaluation of the level of group collaboration.

The work of (Larusson & Alterman, 2009) to visualize students' activities in a wiki-mediated co-blogging exercise is another example. Students as part of their participation may take three kinds of actions: blogging, commenting, and reading. In this research the authors developed visualizations to demonstrate student activity as: level and balance of participation; conversation locator, as well as interactions in a form of networked graph.

Another example can be found in (Khandaker & Soh, 2010). In this work the authors implemented what they called ClassroomWiki - an intelligent agent-based Wiki tool to assess the students' contributions toward their groups - and used it to assess students' contributions in group-based work for a wiki-writing assignment. As part of this wiki they implemented a tracking and modeling module (TAM) by which they track all the interactions and activities within the CalssroomWiki. Moreover, they provided a visualization of student activity counts over time by which teachers can assess group-members contributions and detect free-riding, scaffold group coordination and production function.

Another example is the research of (Reimann & Kay, 2010) in which they have investigated possible visualizations aspects of team performance and their ability to help in group production as well as team coordination i.e. to develop team skills. The research discusses the collaborative wiki writing and possible feedback strategies in order to scaffold group production function and well-being. According to their research they explain the challenges of collaborative wiki writing as wiki pages constitutes from semiotic perspectives of group members. This leads to two main challenges of group coordination on shared meaning of what is collaboratively written as well as wiki content coherence on both levels of text (sentences and paragraphs) and concepts (ideas and arguments). Therefore, in order to improve coordination of team members' activities and increase document coherence, their research is supporting using following forms: (a) by monitoring and visualizing group members' interactions and contributions, (b) by visualizing wiki site structure, and (c) by providing information on wiki page content based on a text-statistical analysis. However, the following visualizations are discussed in this research:

- *Wattle Trees* (Wattle tree is an Australian native plant with fluffy golden yellow round flowers): each member of the team is a single wattle tree, with its vertical green stem that grows up the page. Wiki-related activity is represented by yellow "flowers," the circles on the left of the trees. The size of the flower indicates the size of the contribution. After first experiences, the Wattle Trees were replaced by more interactive visualization of a set of "swim lanes" one for each group member. Color is used to represent the type of contribution (wiki, ticket, svn), per day (or other time units) and aggregated over the visualized time period. When the user clicks a point in one of the swim lanes that has an activity indicated (i.e., is colored), the underlying log data for that cell will be rendered on the screen.

- *Social networks diagrams* have been developed to visualize information regarding who contributes to the wiki-page. The authors used what they called "Interaction Network" (based on Social Network Analysis) to show the relationships and flows between entities. According to (Reimann & Kay, 2010) "*The network is modeled as a graph, with each node representing a team member, always shown in the same, fixed position. Lines between these nodes indicate interaction between these team members. We define interaction to occur when two people modify the same wiki page. The width of the edge is proportional to the number of interactions between them. For a given resource, the number of interactions is calculated as n = min (n1, n2) where n1 and n2 are the number of times user1 and user2 modified the resource*".

- *Visualizing wiki site structure*: while students are working on a wiki collaborative writing task they may need to know which parts have been changed since their last visit to the site. Or maybe which parts of the wiki have been changed by student "A". Therefore, the authors used WikiNavMap (Ullman & Kay, 2007) to support answering the following questions: Which are the pages that I have made contributions to? Which are the pages that another nominated person has made contributions to? Which are the pages associated with a certain task? Which are the pages with the most activity? Which pages changed in the last week? Which changed in a particular period of time, such as a particular month? What is the extent of the wiki?

- *Visualizing the Conceptual Structure of Wiki Page Content:* providing information regarding concepts contained in the wiki-page content and their semantic relations may help group member's collaborative writing. The authors presents an automatic concept analysis method based on "Carley's map analysis technique" and utilizes software called Glosser. Glosser uses text-mining techniques (based on Latent Semantic Analysis technique) to provide student writers with information about their text on a number of dimensions, including conceptual coherence. Glosser is capable to define concepts with hierarchical representation on multiple levels of generalization and abstraction. Moreover, it visualizes the concept map extracted from the wiki-page.

Based on the problems discussed earlier and the examples of wiki-based collaborative learning with knowledge visualization tools, wikis should be enhanced and enriched with new forms of assessment such as self, peer-assessment so that the processes of co-writing can be peer-reviewed. Moreover, enhanced visualization tools should be implemented to provide both students and teachers valuable feedback about the collaborative learning using wiki. The visualization tools and techniques should foster answering the following research questions: *How much has each student contributed to the assignment product? How collaboration is taking place? To what extent the students are collaborating within the group? Who did what and when?* (AL-Smadi, Hoefler, & Guetl, 2011). The next section explains how Co-writing Wiki was developed and enhanced with alternative forms of assessment namely self, peer-assessment, and group-assessment using assessment rubrics to support group well-being and production function, as well as to maintain task and social awareness during co-writing assignments.

7.4.2. Co-Writing Wiki Development

Co-Writing Wiki is developed on top of ScrewTurn wiki[72]. ScrewTurn wiki is an open source wiki developed using C# and ASP.Net for the front-end presentation layer. The engine is partitioned in two main blocks. The Core Assembly contains all the business logic, such as data management and caching, content formatting, provider's configuration and loading and system configuration. The Access control is directly performed by the ASP.Net pages, which also take care of the content presentation and user interaction. However, ScrewTurn wiki features have been extended with other components to handle integrated assessment as well as knowledge visualization based on user behaviour tracking and analysis. Nevertheless, Co-Writing Wiki is fully integrated with SOFIA where a single sign-on (SSO) mechanism is applied.

In order to extend SOFIA services with Co-writing Wiki to assess co-writing assignments, a set of services was developed as part of SOFIA middleware. These services include 'CoWritingWiki_Runtime' which is responsible for deploying Co-writing Wiki tool within the SOFIA environment runtimely, 'CoWriting_AssignmentAuthor' and 'CoWriting_GroupManager' which are responsible for authoring an assignment for co-writing

[72] ScrewTurn Wiki - Free ASP.NET Wiki Software [http://www.screwturn.eu/]

and peer-review and groups management on Co-writing Wiki, for more information about SOFIA middleware see Section 6.4.

Focusing on assessment, Co-writing Wiki is enhanced with the following integrated forms of assessment.

Self-assessment. During the edit of an assignment page students are required to select their edits intentions (e.g. add text, delete text, and change style) as well as to rate the importance and the added value of their edits. Moreover, they can provide some comments to be fed back to the assignment homepage as part of the actions feed section, see Figure 7.10.

FIGURE 7.10. Intention based self-assessment during contribution to Co-writing Wiki (Student View).

Peer-assessment. An internal peer-review follows this action as other group members can review this action and also rate it and provide feedback. The internal peer-review can be configured to be mandatory on each action, just for the final action on the page, or to be selective as in Figure 7.11.

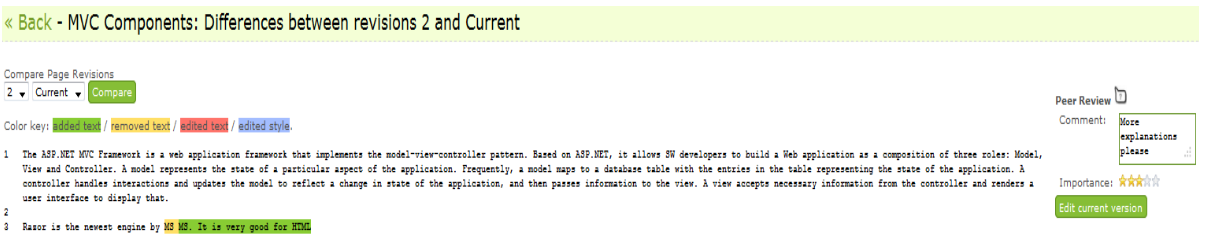


FIGURE 7.11. Internal peer-review for formative feedback and assessment (Student View).

Group-assessment. Co-writing wiki provides a tool for group's peer-assessment. By using this tool students and teachers can peer-assess other group's final product of the assignment and provide feedback. Moreover, they can assess a specific page from the assignment by clicking on 'Assess Page' from the action list. See Figure 7.12.

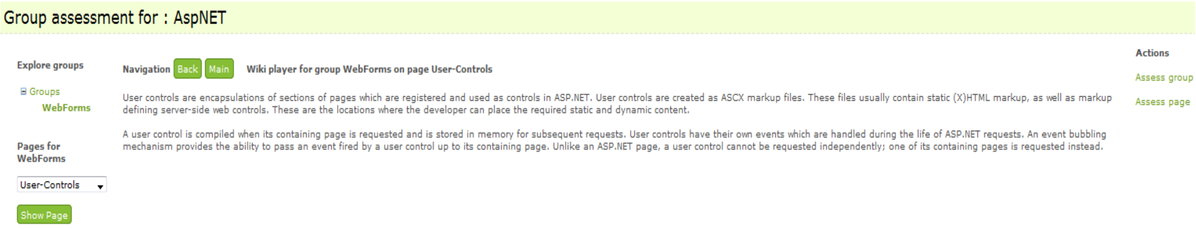


FIGURE 7.12. A tool for group-assessment- Student view for the group named AspNET (Student View).

Assessment Rubrics. The flexible and interactive Assessment Rubrics tool (see Section 7.3) is used to enhance the Group-assessment tool with assessment rubrics to facilitate the assessment process, provide feedback, and to maintain persistence and reliable assessment, see Figure 7.13.

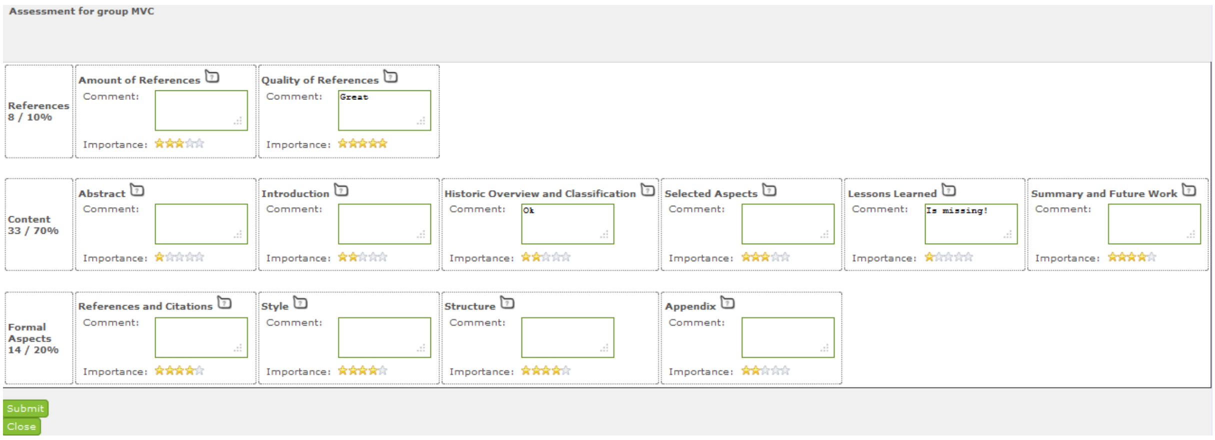


FIGURE 7.13. Rubric to facilitate group-assessment (Student View).

Focusing on visualization tools, Co-writing Wiki is enhanced with different forms of visualizations - textual and graphic- to overcome some of the CSCL problems discussed in related work section. As depicted in Figure 7.14, Co-writing Wiki has been enhanced with an assignment 'Homepage' by which students can get:

- Actions feed: the group members' actions on the assignment pages are fed back to the assignment homepage and grouped based on the page and ordered descending by action date within the same group. Nevertheless, the action record provides link "Preview" to the versions difference page by which the actions on the last version are visualized in colors to support the learners with suitable information about others actions (i.e. task-awareness). The actions are extracted automatically based on the interaction type (i.e. added text, removed text, edited text, and text changed style) of the learner with the wiki-page, see Figure 7.13. Moreover, a link "Review" to review others actions by which group peers can provide feedback based on others interactions. As well as a link “Edit” to edit the latest version of the page.

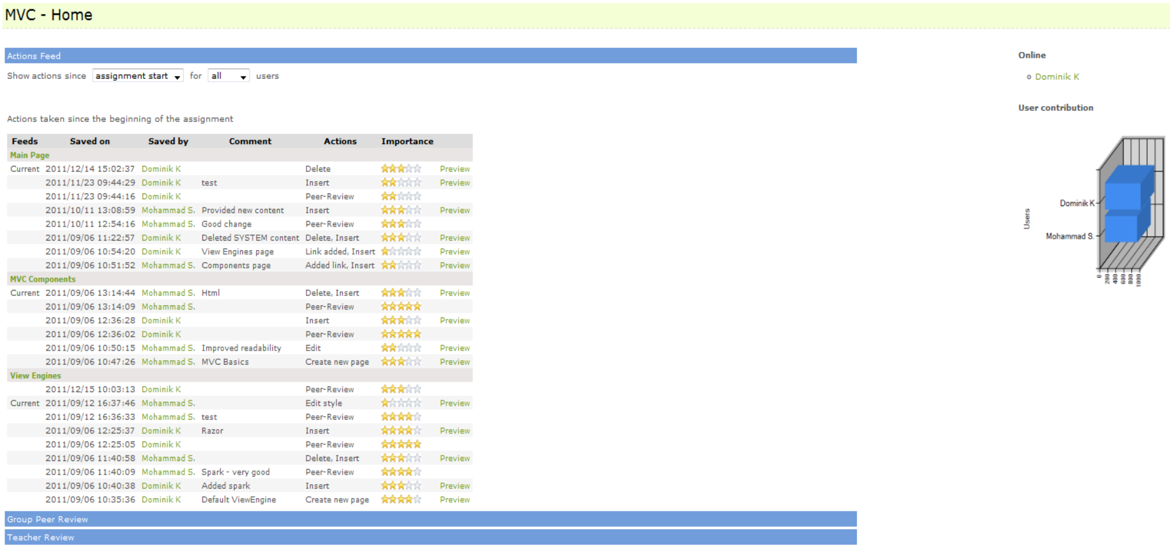


FIGURE 7.14. Actions Feed on the Assignment Homepage (Student View).

- Online peers: in order to provide social awareness and to maintain group production function and group well-being, the assignment homepage shows the currently online group members. This may motivate group members for further collaboration and contribution.

- Contribution chart: this graph represents the amount of letters each group member has contributed to the assignment wiki. In order to avoid meaningless and not related contribution an internal peer-review done by the group members is taking place during the collaboration process. Moreover, we used contribution rate which could be based on amount of letters, number of links within pages, feedback provision, and interaction time. However, until now the graphs still represents the amount of students contribution based on letters and provided to motivate group members to contribute more.

- Feedback: concerns the group/teacher assessment and feedback. ‘Group Peer Review’ section provides the average of the marks collected using the assessment rubric from groups assessment as well as the detailed feedback based on rubric mastery levels and criteria (see Figure 7.15). And the section ‘Teacher Review’ provides the teacher marks and comments based on mastery levels and criterion (see Figure 7.16)

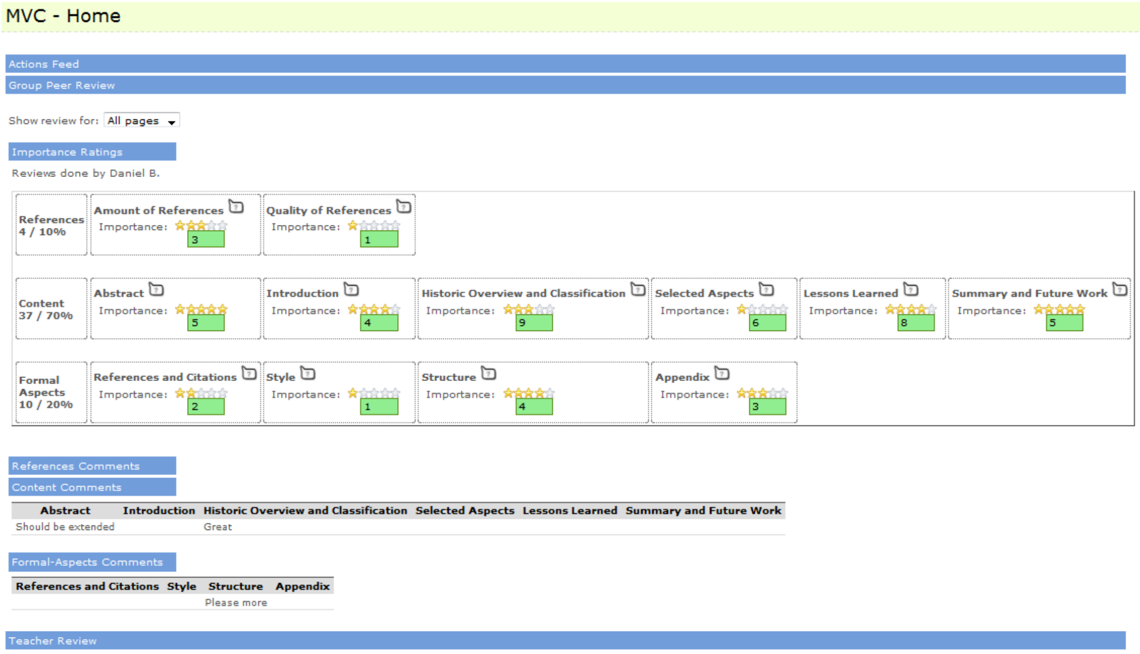


FIGURE 7.15. Feedback based on group-assessment in terms of AVG marks and detailed comments (Student View).

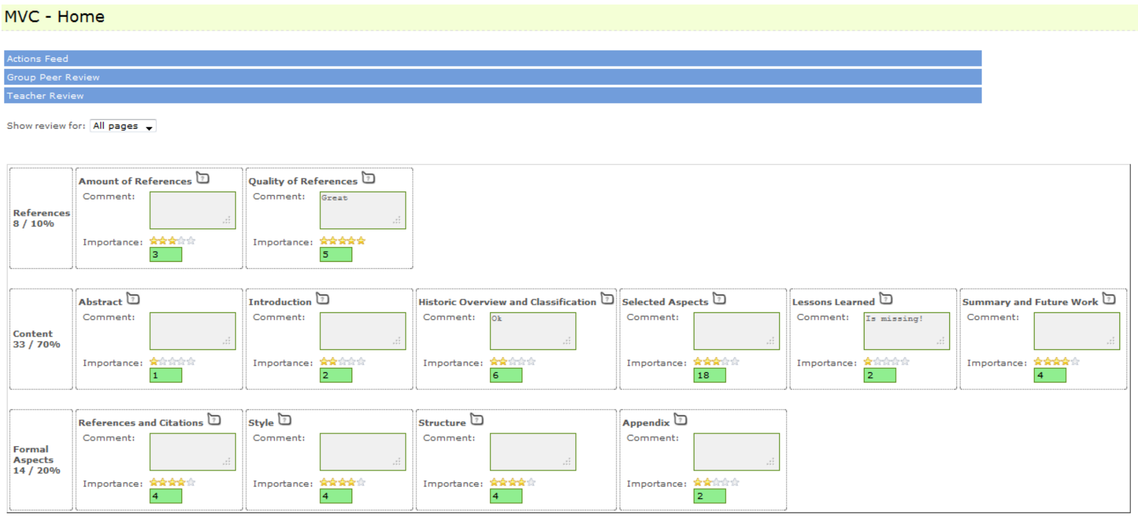


FIGURE 7.16. Teacher feedback based on teacher review using assessment rubric (Student View).

Nevertheless, Co-writing Wiki has been enhanced with visualization tools for the teacher as depicted in Figure 7.17. The 'Teacher view' page consists of the following:

- Group navigation: a tree-view has been provided to explore the assignment related groups of students. Each group member is assigned a unique colour which represents the colour of his own contribution to the assignment wiki document. Colour-based contributions may support the teacher with valuable information about who contributed what to the assignment document, see Figure 7.17.

- History player: the history player is tool that demonstrates the colour-based wiki document as a slide show. The player is flexible to be stopped and started on a specific version of the document. Moreover, it is enriched with navigation buttons to play forward or backward the wiki document, see Figure 7.17.

- Action list: this list contains the possible actions that the teacher may take to evaluate individuals and groups contributions. For instance, the teacher can provide feedback to a specific assignment which will appear in the action list of the assignment homepage. Moreover, the teacher can provide a score the individual/group product in the assignment see Figure 7.17.

- Useful information: in this part of the page the teacher gets some useful information about the collaboration process and students contribution. Examples of such information represents the assignment document number of pages, number of letters, number of links, how many text addition interactions, how many deletions and how many style changes. Moreover, the information panel is interactive and represents the selection from the group's navigation panel. This means that the information may represent the whole group or can be related to a specific member of the group, see Figure 7.17.

- Chart panel: in this panel, the information is visualized in different charts by which useful information is provided to the teacher. Possible charts can be contribution chart, wiki navigation graph, social network graph; Moreover charts may have different shapes such as column chart or pie charts, see Figure 7.17.

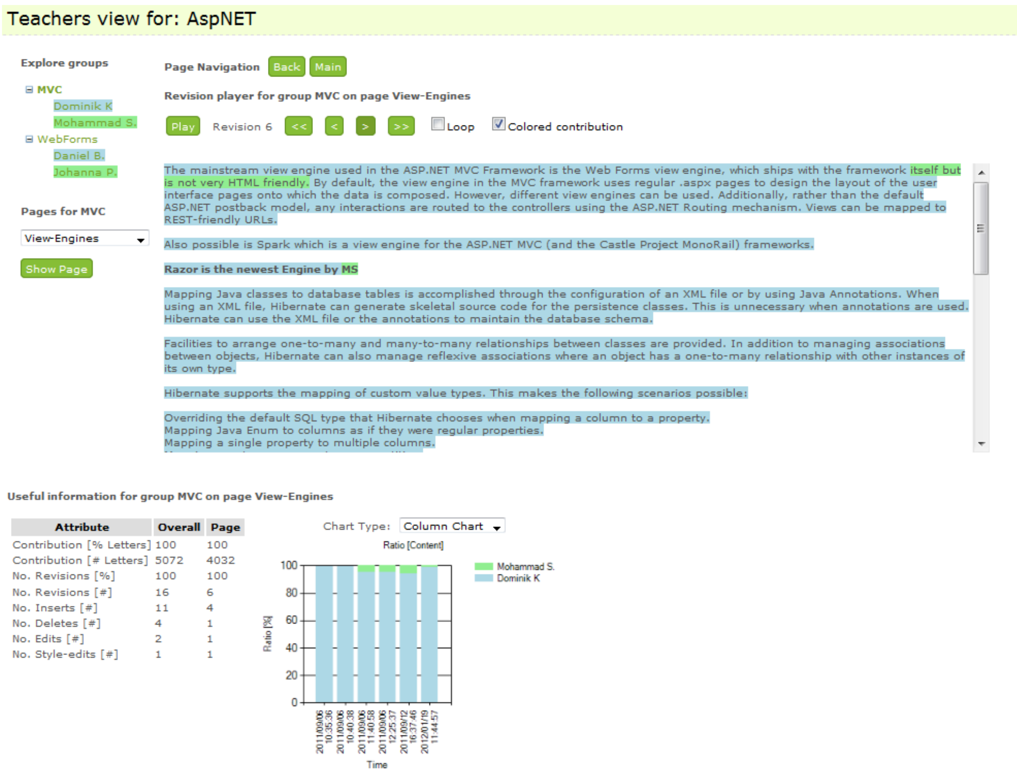


FIGURE 7.17. Teacher View and its tools to support evaluation and feedback provision (Teacher View).

Moreover, 'Co-writing Wiki' is enhanced with a tool to show the students / groups progress before and after group-assessment, and teacher assessment. As depicted in Figure 7.18, this type of visualization aims to show the student's perception of the feedback provided in Figures 7.15, 7.16 based on group-assessment and teacher assessment and the influence on their progress represented by the enhancement they have done on the assignment after receiving the assessment and feedback - i.e. 'Phase 3' in the graph. The graphs in this visualization are interactive as their values are updated based on the selection of group, individual, or specific page from the assignment. Moreover, once you click on the graph bar corresponding information are represented for instance, if you click on the red bar in the assessment graph the feedback provided from the teacher assessment of this assignment is shown as in Figure 7.16.

With the support of SOFIA middleware services, Co-writing Wiki is fully integrated through SSO approach with IWT LMS (see Figure 7.19). Nevertheless, the instructor using IWT can author an assignment and manage group to work on the assignment using Co-writing Wiki running under IWT LMS (see Figure 7.20).

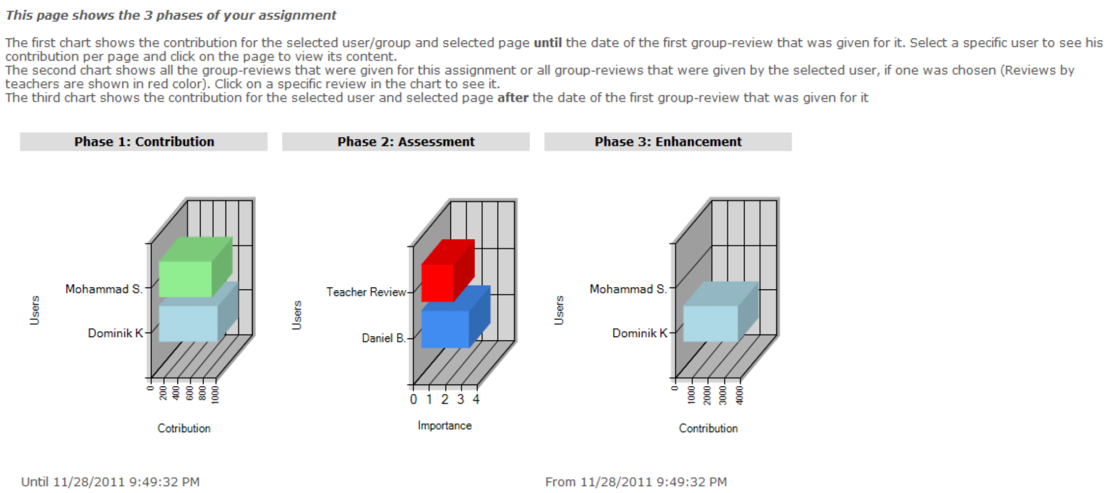


FIGURE 7.18. Individual/group progress before and after group-assessment (Teacher View).

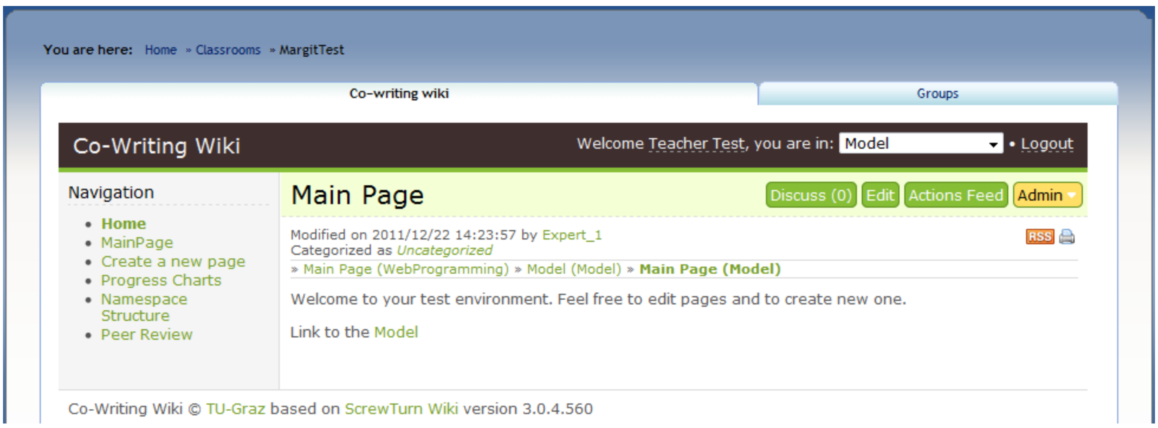


FIGURE 7.19. Co-writing Wiki running under IWT LMS using SSO appraoch (Student View).

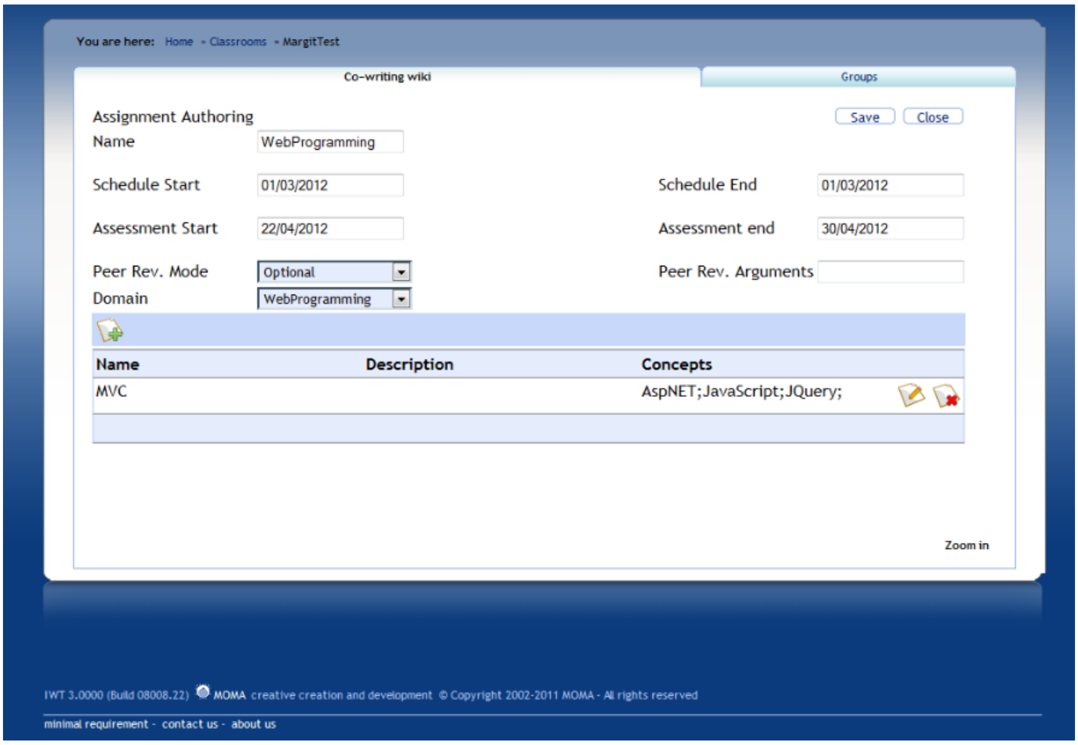


FIGURE 7.20. Authoring an assignment for co-writing using Co-writing Wiki from IWT LMS (Teacher View).

To this end, this enhanced approach of collaborative writing and peer-review using SOFIA and Co-writing Wiki was used to support a study for self-directed learning CLR. Findings show that students were in favor of the ability to discuss per topic, per page and creating and modifying pages. In addition, they mentioned that the tool was always available and consistent. However, some students complained about the usability of the Co-writing Wiki and its slowness. The students also mentioned that they were not aware of all available functions provided using the tool (see Chapter 9).

7.5. Integrated Assessment Approach for Game-based Learning

Providing assessment for serious games holds some challenges especially when it comes to provide dynamic evaluation and feedback for player's - i.e. students - based on their progress and interactions within the game (AL-Smadi, Guetl, Dunwell, & Caballe, 2012; AL-Smadi, Guetl, Chang, 2011; Shute, Ventura, Bauer, & Zapata-Rivera, 2009; Kickmeier-Rust, Steiner, & Albert, 2009. In order to tackle this problem an enhanced approach for integrated assessment in 'stealth' mode has been developed through the architecture discussed earlier in Section 6.3.2 (see Figure 6.6). The architecture has been designed to consider two main scenarios for assessment:

- **Post Evaluation:** in which a 'log file' has been designed to hold all the actions related to the assessment scenario for specific context – e.g. fire evacuation training – through tracking the players' interactions. Moreover, an 'assessment engine' is developed to interact with an 'assessment model' to evaluate the players progress – represented by log file actions – against a pre-authored assessment rules to assess specific learning objectives – e.g. crawling in Smokey areas during evacuation.

- **Dynamic Assessment and Feedback:** in which an 'assessment interface' is attached to the game engine in order to handle events coming from players interactions and calls the 'assessment engine' to evaluate those actions based on pre-defined assessment rules in the 'assessment model', and provides the pre-defined feedback associated to those assessment rules dynamically to the player.

This assessment approach was developed in the context of ALICE project to provide integrated assessment forms for serious games (AL-Smadi, Guetl, Dunwell, & Caballe, 2012). The architectural design of the proposed assessment approach is based on the Service-oriented flexible and interoperable assessment (SOFIA) framework (see Section 6.3.2). The assessment approach aims to provide feedback, thus to support students through a guided learning approach. In context of ALICE project, the assessment approach was used to provide dynamic feedback to the player thus to learn how to evacuate a school building in case of fire threat.

The next sections gives some insights form literature about game-based learning assessment and how those influenced the developemnt of this enhanced approach for game-based learning assessment and dynamic feedback.

7.5.1. Related Work

Serious games represent a challenging as well a rich domain for assessment practices. However, the efficacy of any assessment approach is highly related to the *target demographic*, *usage context*, *choice of technology*, and *underlying pedagogy* (de Freitas and Oliver, 2005). Hence an attempt to evaluate any assessment model typically results with lack applicability when transferred to other groups of learners, different context and educational situations. Examples of assessment trends in digital educational games can be found in the work of (Moreno-Ger, Burgos, Sierra, & Fernández-Manjón, 2008; Shute, Ventura, Bauer, & Zapata-Rivera, 2009; Kickmeier-Rust, Steiner, & Albert, 2009).

Digital games content is very interactive thus more engage students during learning. Nevertheless, this high level of interaction can be utilized for supporting assessment. When players interact with the game they eventually take possible actions pre-defined in the game model of actions. These interactions can be utilized to define assessment rules based on monitoring the player activities, logging all actions within the game session which can be used to assess the player activities within the game. Therefore, educational games should be integrated with LMSs thus to adapt their content, scenarios, and didactic objectives - e.g. the type of provided feedback - to fit with learners – i.e. players – preferences and skill and knowledge state (*cf.* Moreno-Ger, Burgos, Sierra, & Fernández-Manjón, 2008). For instance the extension of the evidence-centered design assessment model (ECD; see Section 3.2.3) with an *action model* instead of *task model* which has been used with Bayesian networks to track player actions within the game and to provide an evidence of progress within the game (Shute, Ventura, Bauer, & Zapata-Rivera, 2009). Another example is the so-called *micro-adaptivity* approach for assessment in educational games (Kickmeier-Rust, Steiner, & Albert, 2009; Kickmeier-Rust, & Albert, 2010). The approach has been developed in the context of the Learning Experience and Knowledge TRAnsfer (ELEKTRA)[73] project. The ELEKTRA framework uses the *Competence-based Knowledge Space Theory* (CbKST) to model the competencies required by the student to achieve a learning goal. The basic idea of CbKST is to associate problems in a domain with skills in order to provide a model of competencies for

[73] http://www.elektraproject.org

a specific domain which can be used to update the knowledge and skill state of the learner in a learning domain. However, according to (Kickmeier-Rust, & Albert, 2010) the approach demands extra load on authoring aspects to define all required information for the models as well as computational load as the game updates the CbKST based on each player action. Another example is the adventure game engine called <e-Adventure> (Moreno-Ger, Burgos, Sierra, & Fernández-Manjón, 2008), which authors claims that the <e-Adventure> is flexible to be used with educational modeling languages - i.e. IMS Learning Design (IMS LD) - to design the pedagogical impact of using assessment rules in the game engine to evaluate the progress of the players, and hence to provide personalized and adaptive digital educational games.

According to Burgos et al. (2008) in order to have an adaptive and personalized game-based learning, the game engine has to be integrated with a LMS. LMSs use the Log of player interactions within the game session to provide more personalized and adaptable content. The player flow within the game will form like a learning path where a third-party tool is needed to interact with the game engine, retrieves the player state, and communicate with LMS so the learner model can be updated as well as adaptive and personal content can be provided during the game next phases (AL-Smadi, Guetl, & Chang, 2011). Thus, when students play, they interact with the game by making decisions and taking right/wrong actions and paths. The game platform should have the possibility to define checkpoints (assessment rules) so that to assess players interactions and decisions. Moreover, it should provide valuable feedback.

7.5.2. Scenario Development

As discussed in Section 6.3.2, the developed scenario consists of (a) an assessment model through which XML based representation of assessment rules are provided, the assessment rules uses the game engine annotations of the pedagogical objects (e.g. bag, elevator, stairs, etc.) to design the assessment based on possible player behavior pattern and consequences (e.g. textual feedback, dialog with a teacher, etc.), (b) assessment engine represents the core component which evaluates the event received from the game engine against the assessment rules defined for the game application domain (e.g. fire evacuation training) and provides feedback, the feedback is provided through a virtual character within the game engine (the virtual character can be adapted to form as a teacher or a firefighter), (c) assessment interface handles the communication between the game engine and assessment engine, finally (d) a log file holds all tracked player interactions and environment changes, the log file is used for post evaluation to provide report based on player behaviour and performance within the game environment (AL-Smadi, Guetl, Dunwell, & Caballe, 2012).

The assessment engine is developed using C# in .Net 3.5 framework. The assessment engine is managed through 'GBL_Assessment_Runtime' developed as part of SOFIA middleware (see Section 6.4.1). For this service an interface is provided and descried using the Web Services Description language (WDSL) and uses the Simple Object Access Protocol (SOAP) for messages communication and transport. This service is used to call the assessment engine methods during the two scenarios for assessment (i.e. post evaluation and dynamic assessment and feedback). Table 7.1 summarized the assemblies developed as part of this scenario - an assembly is similar to a JAR file in a Java application.

TABLE 7.1. A list of assemblies developed as part of the GBL-Assessment approach.

SOFIA GBL-Assessment Assembly	**Description**
Sofia.Gbl.Assessment.Model.dll	This assembly contains objects to represent the XML assessment model structure. It is used to load, browse and manipulate models.
Sofia.Gbl.Assessment.Evaluation.dll	This assembly defines all communication objects and interfaces needed to mediate between an evaluation engine instance and the game. It could be seen as the essential part of a reference implementation (communication contract) for an assessment interface.
Sofia.Gbl.Assessment.Evaluation.Engine.dll	This assembly contains the actual evaluation engine. It implements all interfaces from the common Evaluation assembly (Sofia.Gbl.Assessment.Evaluation.dll) and performs the evaluation process based on a given model described by the objects of the Model assembly (Sofia.Gbl.Assessment.Model.dll).
Sofia.Gbl.Assessment.Evaluation.Service.dll	This assembly provides an instance management facility for using multiple evaluation engines for different clients. It prepares a service like interface for using the evaluation engine without focusing on the actual integration aspects.
Sofia.Gbl.Assessment.Evaluation.WebService.asmx	Offers a WSDL/SOAP compliant web service based on the generic evaluation service (Sofia.Gbl.Assessment.Evaluation.Service.dll**)**.

The game – developed at the Serious Games Institute (SGI) at Coventry University - adopts a freely navigable 3D environment, created within the Unity Engine[74]. The game contains elements of crowd simulation within fire evacuation scenarios, effectively placing the player within the building and monitoring their actions as they evacuate. Hence, provide effective feedback and assessment, it is essential that the game monitors and correctly identifies key actions which may indicate correct and incorrect behaviours. The principal means through which it is proposed is achieved through the implementation of virtual 'checkpoints' within each scenario, recording players' time and state as they pass within a radius of a single point within the virtual space. Nevertheless, the game designer annotates the pedagogical objects and share these annotations as XML file with the LMS - SOFIA in this scenario. The annotations are used by the assessment designer to provide assessment rules thus to provide feedback to the player once an event regarding one of these pedagogical objects is sent to the Assessment Engine.

[74] http://unity3d.com/

The Game engine tracks the player behaviour and environment changes and save them to a log file designed for this GBL-assessment. Moreover, the actions of the players on tagged pedagogical objects fires an event to save a record to the log file. A JavaScript function is developed to send this event to through the 'Sofia.Gbl.Assessment.Evaluation.WebService' to the assessment engine. For instance, a scenario of teaching students that they should not collect their possessions during fire evacuation was tested using this approach. The instructor used the GBL-Assessment Editor and designed an assessment rule for the game object 'bag', a feedback message is added to the consequence section of the assessment rule of "*You took your bag during fire evacuation! You should not collect your possessions before evacuating*". During playing, the player collected his bag before leaving the class room. This action fired an event (see Figure 7.21) and saved to the log file. The JavaScript function in the Web-based game player called the 'Sofia.Gbl.Assessment.Evaluation.WebService' with that event. The Assessment Engine evaluated that event against the assessment rule defined by the instructor and replied with the feedback message. This assessment scenario was used to guide the player not to collect their possessions during fire evacuation using dynamic feedback provision (see Figure 7.21).

```
<logfile>
    <event time="00:00:19.6811257" player-id="noname">
        <action name="UseObject">
            <set property-name="object_type">bag</set>
        </action>
    </event>
</logfile>
```

FIGURE 7.21. Part of the log file relatd to the event of a player collected their bag during fire evacuation.

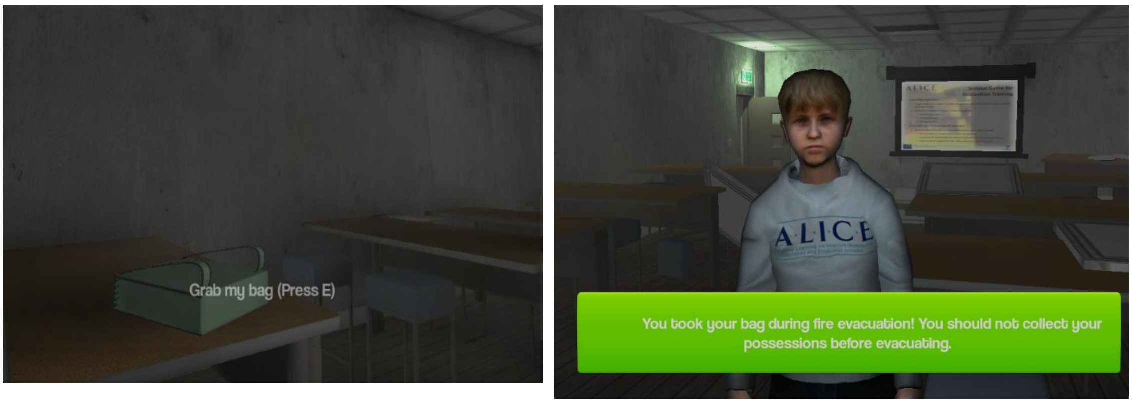


FIGURE 7.22. The player action to collect his possessions triggered an assessment rule which has a consequence of providing feedback using a virtual character.

The decoupling of the game engine - as a complex learning resource - and the assessment engine - utilized via web services - fosters the accommodation of various learning contexts and pedagogical approaches. For instance, the same approach is used to provide guidance in other scenarios such as, using elevator during evacuation as prohibited action, and crawling in a Smokey area as correct action. Focusing on post evaluation assessment approach, Figure 7.23 depicts a proof-of-concept of the output of the post evaluation on three scenarios during fire evacuation. This post evaluation output can be used to update the student competence level as well as can be provided at the end of the game as a final feedback regarding the player level.

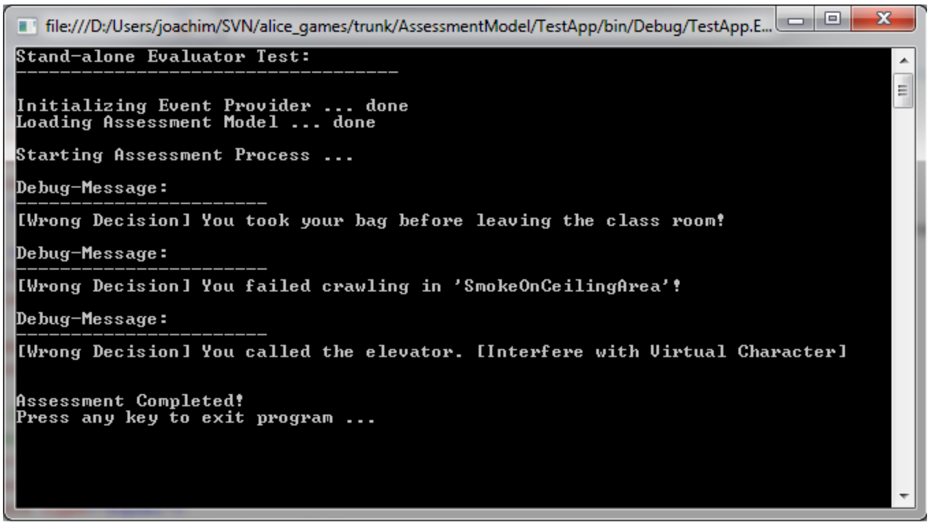


FIGURE 7.23. Post Evaluation feedback on the scenarios (a) collecting possessions, (b) crawling in smokey areas , and (c)using elevator

7.6. Summary

Assessment has changed form being separated from the learning process to be more integrated and learner-centred. Therefore, the assessment paradigm has shifted towards advocating alternative forms of assessment such as performance assessment, behavioural assessment, portfolio assessment, self-, peer-assessment, and rubric assessment (see Chapter 3).

In order to address the challenges and problems which outcome from this shift in the assessment paradigm, new forms of learning experiences enriched with complex learning resources - designed based on instructional and learning objectives - integrated with new forms of assessment - e.g. performance assessment, self and peer-assessment, and behavioural assessment - should be considered. Therefore, a solution approach has been proposed in the second part of this doctoral dissertation through which an integrated model for e-assessment (IMA) has been designed to support designing aligned and integrated forms of e-assessment (see Chapter 5). Nevertheless, a SOA based flexible software architecture (SOFIA) has been used to develop those assessment forms and integrate them to complex learning resources (CLR) for three applications scopes namely collaborative learning, self-directed learning , and game-based learning (see Chapter 6).

To this end, this chapter discusses a set of CLRs with integrated forms of e-assessment. Nevertheless, this chapter sheds the light on how SOFIA has been extended with the services provided by these CLRs through the flexible support of the middleware layer. More precisely, the developed CLRs include: (a) an enhanced approach for peer-assessment (PASS), (b) automated and integrated assessment in self-directed learning CLR, (c) a flexible and interactive tool for assessment rubrics, (d) an enhanced approach for collaborative writing and peer-review, and finally, (e) an integrated assessment approach for game-based learning.

The developed CLRs served as third-party tools to SOFIA assessment environment and used to support conducting studies in real learning settings. First findings show that SOFIA with the CLRs supported students for better and deep learning, and empowered students with learner-centred assessment forms such as self, peer-assessment, and rubric assessment. More detailed information about the studies is provided in next chapters of the third part of this doctoral dissertation.

Part

C

Experimentation & Validation

Chapter

8

8. Enhanced Approach for Peer-assessment

Chapter 7 presented a set of complex learning resources (CLR) integrated with assessment forms. However, in order to show how these CLRs can be used in real learning settings, SOFIA and the developed CLRs have been used to conduct studies in the context of science education in higher education. To this end, this chapter aims to show the applicability of using the solution approach discussed in the second part o f this dissertation. More precisely, this chapter shows the use of SOFIA-standalone assessment enhanced with the peer-assessment tool (PASS) to conduct a study regarding an enhanced approach for peer-assessment.

This Chapter discusses peer-assessment as an alternative form of assessment and shows how it can be used to support students learning. Nevertheless, concerns and aspects related to peer-assessment - e.g. students motivation, students accountability to provide peer-assessment, and peer-assessment reliability - are discussed in this chapter and used to design the study and define the main goals. Despite these problems, the study shows that the enhanced approach provided with the support of SOFIA motivated the students to learn and to provide pee-assessment. Moreover, students had positive attitude during the study phases and provided reliable and quality peer-assessment. The remaining parts of this chapter are organized as follows: Section 8.1 discusses peer-assessment benefits and problems and reflects on the goals of the study. Section 8.2 discusses the study conducted to investigate the enhanced approach for peer-assessment. Finally, Section 8.3 concludes this chapter and summarizes rooms for improvements.

This chapter is based on (AL-Smadi, Guetl, & Helic, 2009a; AL-Smadi, Guetl, & Kappe, 2009; AL-Smadi, Gütl, & Kappe, 2010).

8.1. Purpose

As discussed in Chapter 3, Peer-assessment is defined as "*an arrangement for the peers to consider the level, value, worth, quality or successfulness of the products or outcomes of learning of others of similar status*" (Topping, 1998 p.250). Based on this definition, peer-assessment is not a method for measurement but it is a source of assessment that can be used to support students learning. Peer-assessment has gained more importance from its emphasis on the major role students' play in the assessment process not only as assessee but also as assessor where students and tutors collaboratively work together in the assessment criteria (Orsmond, Merry, & Reiling, 2000). In addition to advocating the learner-centred model, peer-assessment may decrease staff workload and time consumed on the assessment process as well as it may develop certain skills for the students such as, communication skills, self-evaluation skills, observation skills and self-criticism. (Dochy& McDowell, 1997; Sullivan, Hitchcock, & Dunnington, 1999; McLaughlin & Simpson, 2004; Prins, Sluijsmans, Kirschner, & Strijbos, 2005).

On the other hand literature shows that there are some aspects and concerns to be considered when it comes to provide peer-assessment. Such aspects and concerns include student's motivation to participate in peer-assessment, students' accountability to provide fair and consistent peer-assessment. Moreover, the reliability of peer grading is a crucial issue which includes biased grading, leniency in the marking process and paybacks by the peers. Related to this is the problem whether or not students have the required competences and skills to provide peer-assessment or whether or not they understand the assessment criteria and are capable to apply them consistently and fairly. Furthermore, the issue of how to grade the students' peer-assessment performance with respect to the experts reference grades. Findings from literature show that there is a problem of reliability in terms of experts' reference grades as they vary based on the complexity of the assessment task and criteria. (Dochy & McDowell, 1997; Orsmond, Merry, & Reiling, 2000; Divaharan & Atputhasamy, 2002; Ward, Gruppen, & Regehr, 2002; Topping, 2003; McLaughlin & Simpson, 2004; Hamer, Ma, & Kwong, 2005)

This Chapter focuses on how a computer-assisted peer assessment can motivate students to participate in the learning process as well as to provide them with added value assessment. Moreover, discussing the following aspects related to peer-assessment: (A1) Reliability of peer-assessment results, (A2) Appropriate measurement for peer-assessment performance, (A3) Motivation and attitudes, (A4) Knowledge acquisition, And (A5) Usability aspects. Therefore, a PASS as an enhanced peer assessment tool with the support of SOFIA standalone (see Chapter 6) has been used to conduct this study. To this end, the rest of this chapter is structured as follows: Section 2 describes the study, and provides results and discussion. Section 3 concludes this paper.

8.2. Study

Based on the study aspects and goals discussed so far, this sub-chapter discusses this study results with respect to the hypotheses, evaluation criteria and metrics presented in Table 8.1. Nevertheless, this study aims to investigate a suitable form for grading peer-assessment process performance by considering the students' peer-assessments and the tutors' reference rates for the same candidate answers.

TABLE 8.1. Study hypotheses, and their evaluation criteria and metrices

Hypotheses	Evaluation Criteria	Metrics
H1: the enhanced approach for peer-assessment supported the students to gain more knowledge.	**C1.1**: to evaluate the students self-estimation of their knowledge acquisition during the study phases	**M1.1**: ratings for students' self-estimation on their knowledge acquisition (students questionnaire)
H2: the enhanced approach for peer-assessment motivated students to participate in peer-assessment.	**C2.1**: to evaluate the students attitudes towards peer-assessment in modern learning settings	**M2.1**: ratings for students' votes on their preferability to participate in peer-assessment or to use peer-assessment as part of their learning settings (students questionnaire)
H3: the use of the tools is easy even if the user is a non-expert.	**C3.1**: to evaluate the user's level of satisfaction towards the tools	**M3.1**: ratings for functionality/usability of the tool itself, and frequency of use (students questionnaire)
	C3.2: to identify possible	**M3.2:** suggestions and comments

	improvements for the tool based on comments and suggestions	based on open questions. (students questionnaire)
H4: the peer-assessment results using the enhanced approach are reliable on both levels of tutors and students.	**C4.1**: to evaluate the grades of students peer-assessment performance with respect to the tutors reference rates	**M4.1**: students peer-assessment scores with respect to tutors reference rates for same candidate answers

8.2.1. Method

The experiment was performed as an e-learning activity for the course of "Information Search & Retrieval (ISR)" at Graz University of Technology in the winter term 2009/2010. The experiment was conducted in a controlled environment in the computer lab with a supervision of the course lecturer.

Participants

A group of 25 students enrolled at the course of ISR. All of them participated in the experiment. 13 (52%) of the students were taking part in the course as a bachelor program, where 10 (40%) were master students, and 2 (8%) were doctoral students. 3 (12%) were females and 22 (88%) were males. The average age of the students was 26.7 years old with a minimum age of 22 and a maximum one of 36. The tutors were a group of 5 PhD students at the IICM (Institute for Information Technology and Computer Media) of Graz university of Technology. All of them were males and have a master degree of computer science.

Apparatus and Stimuli

The web-based peer-assessment tool (PASS) with the support of SOFIA environment was used provide this study. Using SOFIA students and tutors were capable to participate in the study. SOFIA standalone assessment environment was used to author the online test and the peer-assessment phase based on selected questions out of the online tests. The questions are authored in compliance to IMS QTI specifications (see Chapter 4) and PASS features were used to support students and tutors to peer-assess the candidate answers (see Procedure section). Moreover, a survey tool based on the LimeSurvey[75] deployed on the university campus server has been used to deliver the student questionnaire. The questionnaire was provided to the students and investigated information on demographic data, student's previous knowledge about the study topic - i.e. Document Classification, the student's knowledge acquisition during the study phases, PASS usability, and what students liked and disliked during the study as well as comments for future improvements. Table 8.2 shows the questionnaire questions and their response scale.

TABLE 8.2. Part of Students Questionniare

Question	Description	Scale
Q1	I was already familiar with the selected topic for the online learning phase	5-point Likert scale
Q2	I had a good level of understanding for the selected topic after the online learning phase	5-point Likert scale
Q3	the preparation of the reference answers helped me to acquire a better understanding of the selected topic	5-point Likert scale

[75] [http://www.limesurvey.org/]

Q4	The evaluation of the candidate (peer-assessment phase) answers helped me to better understand the online learning selected topic	5-point Likert scale
Q5	The time given for the online learning phase was	very short, appropriate, very long
Q6	I used additional learning material in the learning phase	Yes, No
Q7	I used additional learning content during the phase of preparing the reference answers	Yes, No
Q8	I used additional learning content to evaluate the candidate answers	Yes, No
Q9	In the evaluation process, it was enough use the reference answers to evaluate the candidate answers	5-point Likert scale
Q10	The evaluation of candidate answers required me to reread part of the course materials with more concentration	5-point Likert scale
Q11	The understanding and evaluation of candidate answers helped me to notice more details in the course materials	5-point Likert scale
Q12	There were many answers to evaluate for each question	5-point Likert scale
Q13	The time of the peer-assessment phase was too long	5-point Likert scale
Q14	I like it that the assessment activities are part of the learning process	5-point Likert scale
Q15	I think it's good to consider the quality of my reviews as part of the mark	Yes, No
Q16	I would like to have online assessment activities as part of my future courses	Yes, No
Q17	I think that annotating the candidate answers with (Wrong, Irrelevant, and Right) helped me to better mark the answer	5-point Likert scale
Q18	I like to have peer-assessment as part of learning activity	5-point Likert scale
Q19	I like to have peer-assessment as part of my performance grading	5-point Likert scale
Q20	I like to have peer-assessment as part of future learning settings	5-point Likert scale
Q21	Overall impression of the Assessment Tool	5-point Likert scale

Procedure

The experiment procedure is as follows:

- Introductory talk (10 minutes): at the beginning of the experiment a short introduction was given by the ISR course lecturer about the subject domain as well as the assessment in general and the peer-assessment as an emerging form of assessment. The importance of knowledge acquisition and knowledge assessment in modern learning settings was discussed briefly. The learning objectives behind this experiment were mentioned. The lecturer also stressed on the importance of the students performance during the experiment and clarified that the performance will be given 10 points as part of the overall grade for both the online test and the online peer assessment session of 5 points each.
- Online learning session (45 minutes): "Document Classification" as one of the main topics of ISR course was chosen to formulate the online learning material of the experiment. The material language is English and it has been extracted from Wikipedia. The material is formulated out of four web-pages and an introduction one, where the students were allowed to access and navigate between them as well as a set of further readings hyperlinks related to the subject domain.
- Online testing session (15 minutes): The knowledge that was gained by the student from the last session is assessed in this session. An online test of five questions was provided using SOFIA (see Table 8.3). During this session the students were not allowed to access any course materials. The test items were variable, where the first questions was a definition one, the second was an enumeration, the third and the

fourth were asking for a concept explanation while the fifth was an abbreviation. For each of the fifth questions an answer and a confidence value out of 5 (i.e. 0 as very poor to 5 as very good) had to be provided. The confidence value is used to evaluate the student self-assessment on the provided answer.

- Break (15 minutes).
- Online reference answers preparation (15 minutes): During this session, the students were asked to prepare reference answers for the questions Q1, Q2 and Q5 (see Table 7.1). Differently from the last session the students were asked to access the course content and other useful materials to help them in preparing the reference answers.
- Online peer assessment session (45 minutes): in this session the students used the reference answers from the last session to evaluate and to peer-assess their answers from the online test session. Every student had to evaluate around 21 randomly selected answers for 3 different questions as well as 15 pre-prepared calibrated answers. For each answer, the students were capable to select parts from the candidate answer and mark them as *correct, wrong, or irrelevant.* Special colors are used to mark the selected part of the candidate answer based on its correctness (i.e. correct as green, wrong as red and irrelevant as yellow). A score should also be provided by the student for the answer from "0" (very poor) to "5" (very good). Using colored marks for the candidate answer supports the students for scoring the answer and to provide a reasonable score based on his colored marks. Moreover, the colored marks will be provided as a valuable feedback to the student who wrote this answer. Input-boxes for missing parts of the answer and additional notes were provided for the students to write into them as in Figure 8.1.
- Experiment questionnaire (10 minutes): the students were asked to fill in a questionnaire that diagnoses their attitudes about the assessment activity of its three parts self-directed, online test and the peer-assessment one, as well as the usability of the web-based assessment prototype and their suggestions for further enhancements and notes, see Table 8.3.
- Results delivery: the students peers marks of their candidate answers have been used to compute the online test performance grade and provided as feedback as in Figure 8.2.

TABLE 8.3. Questions used in the Online test phase

Question	Description	Type
Q1	What is the definition of the term "document classification"?	Definition question - Short free-text answer
Q2	What are the four most commonly used classification techniques?	Enumeration question
Q3	What does the abbreviation tf-idf stand for?	Abbreviation question
Q4	What are SVMs?	Short free-text answer
Q5	How are training and classification with the k-nearest neighbor algorithm performed?	Open ended question

In order to compare the students' peer-assessment results with a reference grading values, a group of 5 tutors had participated in the experiment. The tutors' peer-assessment process was as follows:

- Experiment Introduction: an e-mail was sent to all the tutors, in which a brief introduction about the experiment goals and procedures were outlined.

- Reference answer preparation: the tutors were asked to use the course content and other related materials to prepare a set o f reference answers that they will use later on in the evaluation process.
- Online peer-assessment: in this step, all the candidate answers from the students were evaluated by the tutors. The same colored marking facilities of some parts of the candidate answers were used. As well as the possibility of adding notes and missing parts of the candidate answers.

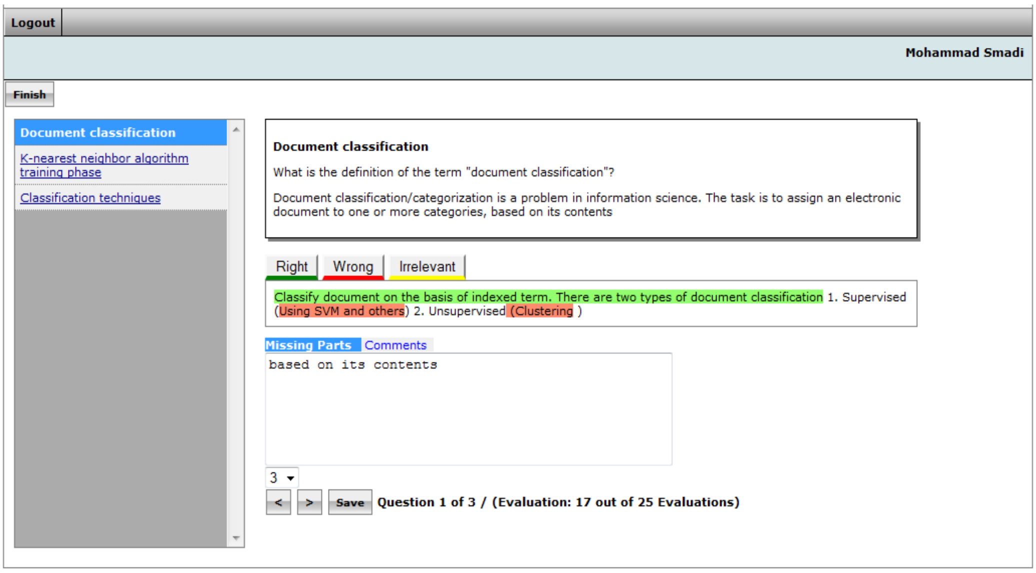


FIGURE 8.1. Screenshot from PASS tool running under SOFIA assessment environment during the peer-assessment step.

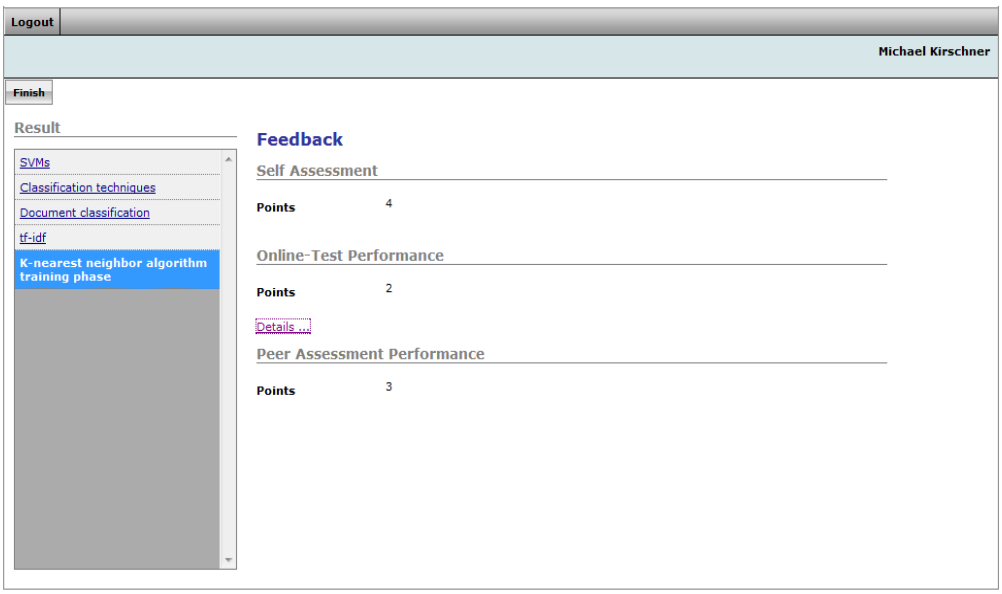


FIGURE 8.2. Screenshot from the feedback step.

8.2.2. Results Evaluation and Discussion

Students Questionnaire

As part of the students phase, students were asked to fill in a questionnaire regarding their attitudes and comments on the experiment. The questionnaire diagnosed student's knowledge acquisition, learning attitudes, and the usability of the tool.

With respect to metric M1.1

The self estimation of students' knowledge acquisition has been discussed in several researches (*cf.* Magin & Churches, 1988; Sluijsmans, 2002; McLaughlin & Simpson, 2004). Figure 8.3 shows the results for the students' self estimation of their knowledge acquisition from the overall study. From the students' point of view, their basic knowledge in the subject before the experiment was with a mean value of 3.84 (σ = 1.31) where (0 represents fully disagreement and 5 represents fully agreement). The knowledge gained from the online learning phase was with a mean value of 4.12 (σ = 1.2). Preparation of reference answers has supported the students to get better knowledge in the subject domain with a mean value of 4.64 (σ = 1.04), where the knowledge that they had gained from the peer assessment task was with a mean value of 4.40 (σ = 1.35). Furthermore, students had used the course content during the peer assessment task with a mean value of 3.40 (σ = 1.80), moreover for them it was appropriate to use only the reference answers for evaluating the candidate answers with a mean value of 4.64 (σ = 1.15). Obviously, students had positive attitude towards the study phases and interestingly the findings from the online learning phase as students gained knowledge probably because of the incentive of having an online test after this step which more motivated students to focus on this step. The same attitude is clear in the preparation of the reference answers which they will use later on in the peer-assessment. This positive attitude and self estimation of knowledge acquisition is in line with findings from other research (Magin & Churches, 1988; Sluijsmans, 2002; McLaughlin & Simpson, 2004).

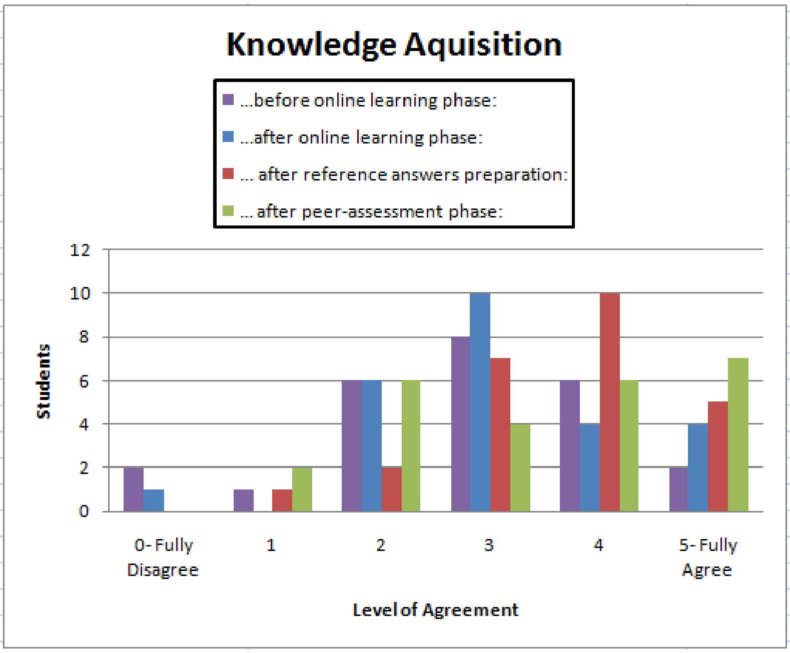


FIGURE 8.3. Student's self estimation of knowledge acquisition.

With respect to metric M2.1

By analyzing the students' attitudes on the e-assessment as part of modern learning settings, 15 (60%) students like to have e-assessment as part of their future learning activities whereas 16 (64%) they think it is good to consider their peer-assessments in the questions final score. Marking the candidate answers (right, wrong, and irrelevant) helped the students to better mark and score the answers with a mean value of 3.80 ($\sigma = 1.47$). The students argued that the time of the peer-assessment phase was too long with a mean value of 3.40 ($\sigma = 1.38$) where the suggested time for this phase was with an average value of 56.4 minutes (i.e. 11.4 extra minutes than the given time). They also argued that the required candidate answers per question to be evaluated were too many with a mean value of 4.08 ($\sigma = 1.44$) where they prefer the number per question to be with a mean value of 11.28 ($\sigma = 6.65$) candidate answer (i.e. 0.62 less than the required number which was 12 answers/question). 16 (64%) students think that it is a good idea to consider the quality of their peer-assessments as part of the final mark.

By analyzing the students' impressions on the peer-assessment as part of a modern learning settings, students like peer-assessment as part of the learning activity with a mean value of 2.74 ($\sigma = 1.51$), where they recommend it to be part of computing their performance grading with a low mean value of 1.56 ($\sigma = 1.45$), as well as to be considered as part of the future learning settings with a low mean value of 1.85 ($\sigma = 1.32$). These results are presented in Figure 8.4.

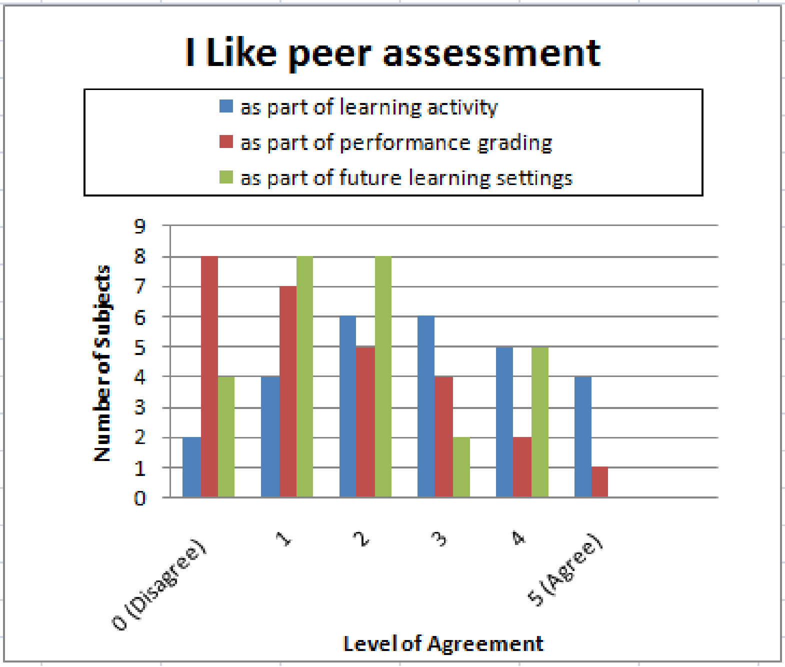


FIGURE 8.4. Students' Impressions on peer-assessment in modern learning settings.

With respect to metric M3.1 & M3.2

To get better idea about the usability of the tool, students were asked in the questionnaire about the tool functionalities and usability. According to the questionnaire, the overall usability of the tool was with a mean value of 3.96 ($\sigma = 1.31$) where (0 represents fully unusable and 5 represents fully usable). In addition to this satisfactory impression about the tool usability, students were asked to explain shortly what they liked, disliked, and what can be improved regarding the tool in the study phases. Findings can be summarized as follows: focusing on the online test functionality, students argued that they liked the design of the UI as test representation was clear and precise. Moreover, they liked the left menu indicating the test items and the easy navigation among questions and answerers and the possibility to edit an answer after you save it before finishing the test. On the other hand the students asked for a timer that indicates the time left for the online test in the UI, thus they recommended adding a timer to the online test UI and some of them asked to add the answer provided to the left menu of the questions for faster navigation. Focusing on peer-assessment functionality, students liked the features provided in terms of easy navigation among questions left menu, the provision of the question body and their provided reference answer, the ability to mark the candidate answer with colors representing (correct, wrong, irrelevant), as well as the ability to provide a score and the feedback about how many evaluations finished and how many left (see Figure 8.1). Whereas some of the students faced problems with browsers and the colored marking feature, and some students complained about having marking candidate answers as mandatory to navigate to next answer. In a question about what can be improved for the peer-assessment functionality, students recommended to enhance the color-coded marking functionality, indicating that score 5 is the best as this can be different in some universities based on the scoring schema.

Tutors Phase

Because of the diversity in tutors experience the weighted mean has been chosen to compute the reference marks for the candidate answers. Table 8.4 shows the tutors experience represented in weights. The weights given to the tutors have been decided based on the tutors experience as well as the arithmetic mean of tutors grading from table 8.4 where a grade value of 2.5 represents the reasonable mean of a scale between 0 and 5. All of the tutors are PhD students in computer science (CS) some of them have advanced knowledge in information retrieval (IR) as well as in assessment activities (AS).

TABLE 8.4. Tutors weights based on their experiences and grading.

	Experience				**Grading**	
	CS	***IR***	***AS***	***Weight***	***Mean***	***σ***
T1	☑	☑	☑ ☑	2	2.53	1.58
T2	☑	☑ ☑	☑ ☑	3	2.81	1.59
T3	☑	☑	☑	1	2.99	1.86
T4	☑	☑	☑	1	3.63	1.75
T5	☑	☑	☑	1	2.85	1.78

The five tutors had to score 125 candidate answers for the five test items (see Table 8.2). However, in order to compare the scores from students' peer-assessment with the tutors scores a subset of the entire students candidate answers have been considered namely the tutors scores for only three test items - i.e. Q1, Q2, and Q5. For this subset the tutors had to score 75 candidate answers with weighted mean value of 2.87 ($\sigma = 1.39$) which is some half point lower compared to the arithmetic mean value of 2.96 ($\sigma = 1.40$) and indicates that the more experienced tutors in average assign lower grades.

Table 8.5 outlines the cross-correlations of the tutors' assessment results as well as the comparison with the weighted mean values of the candidate answers. For all of the test items the cross-correlation values vary between 0.507 (T1, T4) and 0.717 (T2, T3) by a mean value of 0.61 (σ = 0.15). For test Item 1 which asks for a definition, it has the best cross-correlation values between 0.483 (T2, T5) and 0.765 (T1, T2) by a mean value of 0.62 (σ = 0.20). Test Item 2 which asks for an enumeration, the cross-correlation values are between 0.324 (T1, T5) and 0.833 (T2, T3) by a mean value of 0.58 (σ = 0.36). For test Item 3 the cross-correlation values are the worst because it asks for an explanation of a concept which is more complex than definition and enumeration types, they are between 0.291 (T1, T4) and 0.656 (T2, T3) by a mean value of 0.47 (σ = 0.26). The same findings can be found in the literature where the variance between the tutor's correlation values ranges between 0.2 and 0.9, and depends on their experience as well as on the complexity of the assessment task. However, in order to tackle this problem and to maintain quality reference scores, clearly defined assessment criteria should be provided - e.g. using assessment rubrics - and by using multiple reference ratings. (Magin & Churches, 1988; Flotz, Laham, & Landauer, 1999; Sullivan, Hitchcock, & Dunnington, 1999; Ward, Gruppen, & Regehr, 2002).

TABLE 8.5. Cross-correlations for tutors' assessment results.

		T1	T2	T3	T4	T5	WMW
All Test Items	**T1**	1.000	0.610	0.551	0.507	0.565	0.811
	T2		1.000	0.717	0.604	0.576	0.912
	T3			1.000	0.567	0.608	0.820
	T4				1.000	0.531	0.743
	T5					1.000	0.755
	WMW						1.000
Test Item 1	**T1**	1.000	0.765	0.741	0.664	0.727	0.938
	T2		1.000	0.687	0.560	0.483	0.892
	T3			1.000	0.552	0.590	0.829
	T4				1.000	0.612	0.754
	T5					1.000	0.759
	WMW						1.000
Test Item 2	**T1**	1.000	0.497	0.426	0.628	0.324	0.712
	T2		1.000	0.833	0.536	0.565	0.926
	T3			1.000	0.559	0.721	0.875
	T4				1.000	0.383	0.699
	T5					1.000	0.718
	WMW						1.000
Test Item 3	**T1**	1.000	0.496	0.578	0.291	0.552	0.776
	T2		1.000	0.656	0.390	0.652	0.882
	T3			1.000	0.314	0.573	0.793
	T4				1.000	0.496	0.569
	T5					1.000	0.810
	WMW						1.000

TABLE 8.6. The absolute errors for tutor's assessment performance.

	All Test Items		**Test Item 1**		**Test Item 2**		**Test Item 3**	
	Mean	***σ***	***Mean***	***σ***	***Mean***	***σ***	***Mean***	***σ***
T1	0.78	0.61	0.52	0.47	1.05	0.57	0.76	0.68
T2	0.51	0.42	0.48	0.52	0.43	0.30	0.61	0.40
T3	0.87	0.63	0.87	0.79	0.84	0.34	0.89	0.65
T4	1.12	0.83	1.10	0.91	1.12	0.65	1.14	0.95
T5	0.88	0.77	0.97	0.92	0.85	0.67	0.82	0.71

In order to investigate the results, the absolute error of the tutors' individual score values is computed as the difference between the weighted average and the tutor score per candidate answer. As in Table 8.6, the absolute error for all of the test items is between 1.12 ($\sigma = 0.83$) as worst result and 0.51 ($\sigma = 0.42$) as best result. For test item 1 the absolute error varies between 1.10 ($\sigma = 0.91$) as worst result and 0.48 ($\sigma = 0.52$) as the best one. The best case can be seen in test item 2 which reflects the simplicity of the assessment activity done by this item as an enumeration item where the absolute error is between 1.12 ($\sigma = 0.65$) and 0.43 ($\sigma = 0.30$). Test item 3 as the most complex item has not only lower cross-correlation but also higher absolute error values between 1.14 ($\sigma = 0.95$) and 0.61 ($\sigma = 0.40$). Moreover, all the best results are achieved by Tutor 2 which shows that the more experience the tutor is the lower absolute errors she/he has. Less information can be found in literature about the absolute error of tutors' grading. Palmer, Williams and Dreher (2002) outline the distribution of given marks of 3 tutors, and MARKIT (2008) reports about the average error rate of 9.67 % between two tutors' marking results. In contrast to findings from literature, in Table 7.5 results outline more details for each of the question types, which provide a useful base to define a measure for assessment performance.

Students Phase

With respect to metric M4.1

In order to compare the student's peer-assessment performance with the tutor's reference scores, the arithmetic mean of peer's individual results per candidate answer has been used and the absolute error as the difference between the student's arithmetic mean and the tutor's reference marks has been computed. For all the three test items the arithmetic mean of absolute error is quite low with 0.60 ($\sigma = 0.48$). For the three test items individually, test item 1 the arithmetic mean of the absolute error is 0.62 ($\sigma = 0.41$). Test item 2 has the lowest value of 0.47 ($\sigma = 0.38$) since it is easier to score an enumerated question than scoring a short-free answer. Test item 3 has a higher value with 0.72 ($\sigma = 0.61$) which reflects the complexity of the assessment activity done by this item as a concept explanation one (open ended question). The correlation between the arithmetic mean of the student's evaluations and the tutor's reference marks for each candidate answer is quite strong with 0.84 for all the three test items, 0.88 for test item 1; 0.78 for test item 2; and 0.82 for test item 3. Figure 8.5 represents a scatter plot for the tutor's reference grading in comparison with the students peer assessments for the three test items sorted in ascending order by the tutor's reference grading values.

The comparison between the students average peer-assessment scores ($M = 3.03$, $\sigma = 1.15$) and the tutors average reference scores ($M = 2.94$, $\sigma = 1.37$) for the same candidate answers of the three test items shows that there is no significant difference among them (t (148) = 0.44, p = 0.66). Nevertheless, the results show that there is no significant difference between the students average peer-assessment scores and the tutors average reference scores for all test items as follows: for test item 1 the students average peer-assessment scores ($M = 2.81$, $\sigma = 1.20$) and the tutors average reference scores ($M = 3.20$, $\sigma = 1.43$; t (48) = 1.03, p = 0.31), for test item 2 the students average peer-assessment scores ($M = 3.63$, $\sigma = 0.95$) and the tutors average reference scores ($M = 3.57$ $\sigma = 0.89$; t (48) = 0.23, p = 0.82), and for test item 3 the students average peer-assessment scores ($M = 2.66$, $\sigma = 1.07$) and the tutors average reference scores ($M = 2.06$, $\sigma = 1.29$; t (48) = 1.78, p = 0.08). Having a closer look on p-values in comparison to a threshold value of 0.05 it can be assumed that the quality of students' peer-assessment is close to the tutors' assessment. Moreover, the difference between student's peer-assessment scores and the tutors' reference scores depends on the test item type and difficulty. For instance, for test item 2 of enumeration question type, the p-value is higher than the p-values of test items 1 and 2. Nevertheless, for test item 3 - i.e. open ended question - the p-value ($p = 0.08$) is quite close to the threshold value of 0.05 which indicates to some extent that in case of difficult test items - such as open-ended questions - there is a tendency to have a difference between the students peer-assessment and the tutors one. These findings is in line with the findings from the tutors phase, moreover findings from literature show that the quality and reliability of students' peer-assessment is influenced by motivational aspects as well as experience (Sullivan, Hitchcock & Dunnington, 1999; McLaughlin & Simpson, 2004; Hamer, Ma, & Kwong, 2005). It can be argued that this problem has been tackled in our approach as the students had an online learning session with incentive of an online test after that phase which to some extent motivated the students to learn better (knowledge gained $M = 4.12$, $\sigma = 1.2$), as well as the reference answers preparation session which also enhanced the students knowledge state about the topic ($M = 4.64$, $\sigma = 1.04$) (see findings from the student questionnaire).

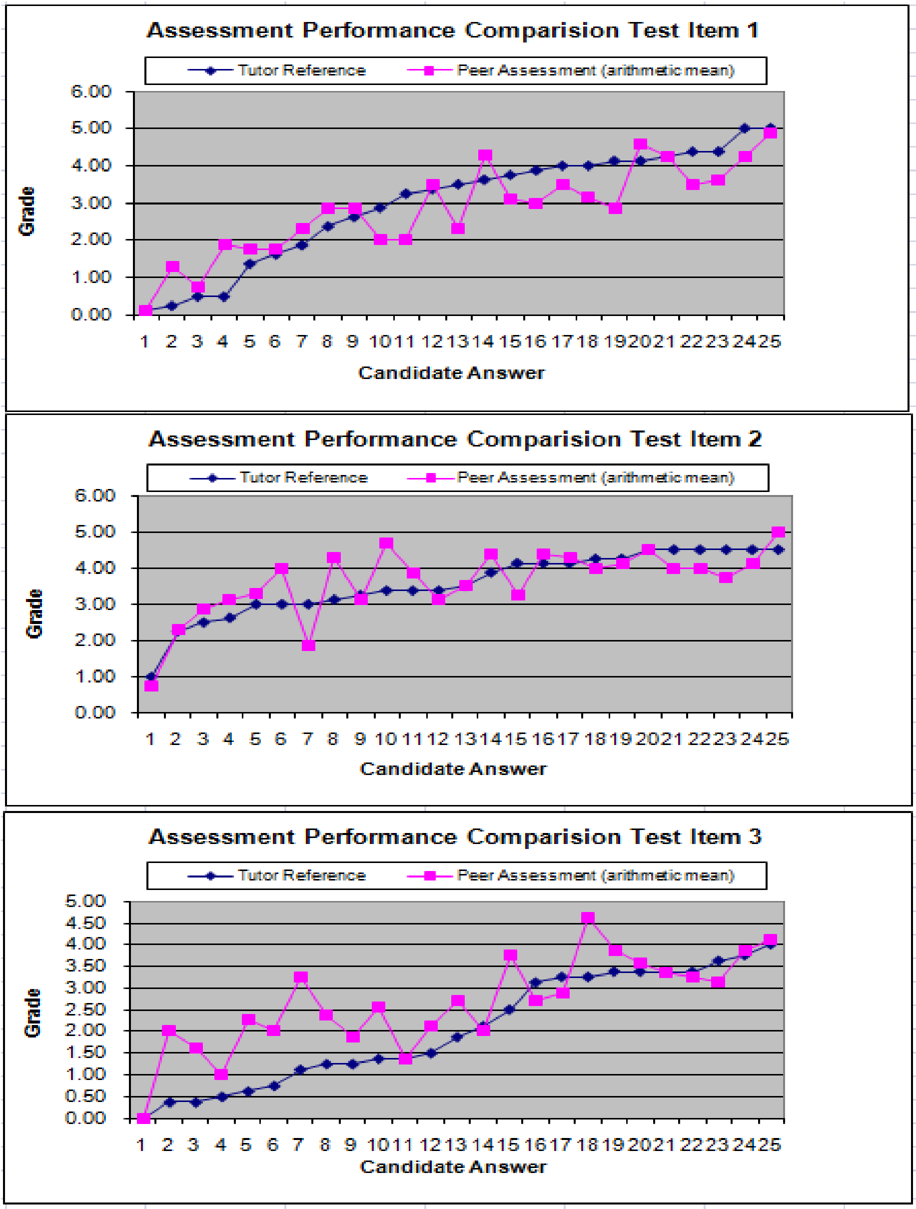


FIGURE 8.5. Comparison between students peer-assessment performance and the tutores reference scores.

In order to grade the students peer-assessment performance the absolute error (e) as the difference between the student's arithmetic mean and the tutor's reference marks has been. In order to provide grades based on the absolute error, five grades ranges from 5 as the best to 0 as the worst are used based on the following grades: grade 5 ($e < 1.1$), grade 4 ($1.1 \leq e < 1.9$), grade 3 ($1.9 \leq e < 2.6$), grade 2 ($2.6 \leq e < 3.1$), grade 1 ($3.1 \leq e < 3.5$) and grade 0 ($3.5 \leq e$). This grading scheme has been applied and the results are as follows: 75 (74%) for grade 5, 19 (24%) for grade 4, 1(2%) for grade 3, and 0 for grades 2, 1. This can be interpreted as 74% of the students' peer-assessment has a quality of grade 5. Moreover the same grading scheme can be used to grade the online test based on peer-assessment as the candidate answer that belongs to grade 5 from the peer-assessment can be graded as 5 as well for the online test. The use of the absolute error as a peer-assessment performance measure has been used in other researches such as (Palmer, Williams, & Dreher, 2002; Hamer, Ma, & Kwong, 2005). As the experts reference grades vary according to their experience and based on the complexity of assessment (question type in this study), the grades scheme can be adapted to each question type.

8.3. Conclusion and Outlook

Peer-assessment as an alternative form of assessment has proven to provide added value to learning and education in particular through advocating learner-centred approach and providing timely feedback. Participating in peer-assessment students provide assessment for learning thus the assessment practices align with learning and instruction goals and objectives. However, literature research has shown several aspects and concerns regarding peer-assessment which mainly include students motivation to participate in peer-assessment, students' accountability to provide fair and consistent peer-assessment, the reliability of peer grading, whether or not students have the required competences and skills to provide peer-assessment or whether or not they understand the assessment criteria and are capable to apply them consistently and fairly. By considering these aspects, this chapter discusses a study which is designed to investigate these aspects and in particular focuses on: *(A1) Reliability of peer-assessment results, (A2) Appropriate measurement for peer-assessment performance, (A3) Motivation and attitudes, (A4) Knowledge acquisition, And (A5) Usability aspects.* In order to investigate these aspects a set of hypotheses, evaluation criteria, and metrics have been provided.

After analyzing and validating the research hypotheses using the evaluation criteria and metrics designed for this study, the following findings were resulted. Regarding the reliability of the peer-assessment results (A1), the level of agreement between the student's peer evaluations and the tutor's reference grading varies according to the complexity of the assessment task (represented by the test items), the experience of the individuals, as well as the motivation and attitudes. Experiment results showed for students as well as for tutors the highest level of agreement was for simple assessment tasks such as definitions and enumeration answers, where the level of agreement was fair with more complex assessment activities such as concept explanation answers. In contrast to findings from literature, it can be argued that these findings were influenced by the enhanced approach - procedural and technical - designed for this study. As part of the study procedure, students had an online learning session with incentive of an online test after that phase which to some extent motivated the students to learn better (knowledge gained $M = 4.12$, $\sigma = 1.2$), as well as the reference answers preparation session which also enhanced the students knowledge state about the topic ($M = 4.64$, $\sigma = 1.04$) (see findings from the student questionnaire). This explains the findings of no significant difference from the t-test analysis on the students' peer-assessments and the tutors' reference rates, as well as the findings from the average of the absolute error between student's peer-assessment and their correspondence tutors reference scores. A weighted average has been used to enhance the tutor's assessment values as they have different levels of experience.

The average of the absolute error between the tutors' weighted average grades and the students' average marks for each candidate answer has been used to evaluate the performance of the students in the peer-assessment task (A2). By focusing on motivation and attitudes aspects (A3), overall students argued that the peer assessment task is an interesting alternative and they have gained new knowledge form it. Moreover, students and tutors acquired assessment skills and more detailed knowledge about the subject domain (A4). By focusing on usability aspects (A5), students in general liked the experiment procedure and they provided us with comments that can be considered as rooms of future improvements. Moreover, the students overall votes on the tools usability and functionality was above average with positive attitudes regarding online test and peer-assessment functions and usability. On the other hand, students reported some technical and procedural problems which includes lack of time feedback as the SOFIA assessment player used for this study didn't has a timer control, problems with the color-coded marker on some browsers, lack of information about scoring scale as students were confused whether 5 is the best score or the worst. However, reported problems have been considered as rooms for tools improvement.

As future work, more focus on functionality and usability should be considered. Nevertheless, more research on reliability of peer-assessment results should be conducted. For instance more focus on sharing a clear understanding of assessment and grading criteria with students should be considered. This can be achieved by using SOFIA Rubrics tool for peer-assessment through which students can get clear assessment criteria and mastery levels underpinned with exemplars to support them in providing consistent and fair assessment (see Section 2.5).

Chapter

9

9. Automated and Integrated Assessment in Self-directed and Collaborative Learning

Recently, information and communication technology plays a main role in education and learning. As a result, modern learning settings of learner-centred practices have become more dominant. A new culture of assessment of integrating assessment to CLRs to address requirements of assessing skills such as cognitive (e.g. problem solving, critical thinking), meta-cognitive (e.g. self-reflection and self-evaluation), social (e.g. leading discussions and working within groups), and affective aspects (e.g. internal motivation and self-efficiency) have arisen. In this new culture of assessment, students play major roles in the assessment where new forms of assessment have been adapted to suit the learning styles of the modern learners. Such forms include performance assessment, process and product assessment, directed assessment, authentic assessment, alternative assessment, collaborative assessment and self- and peer-assessment. (see Chapters 2, 3 for more information).

This chapter aims to discuss alternative assessment practices integrated with CLRs (see Chapter 7) to support self-directed learning and collaborative learning. More precisely, this chapter discusses an empirical study about emerging forms of assessment namely automated assessment, peer-assessment, rubric assessment, and group-assessment integrated with CLRs in self-directed and collaborative learning. The first findings show that students were intrinsically motivated towards this approach. Moreover, automatic and peer-assessment supported the students to achieve their learning goals. Nevertheless, the study discussed in this chapter has been conducted online in distance learning mode with the support of SOFIA assessment environment and CLRs of enhanced AQC and Co-writing Wiki (see Chapter 7). Moreover, the results of the study show that the enhanced AQC and Co-writing Wiki integrated with alternative forms of assessment supported students in their learning process.

This chapter is based on (AL-Smadi, Hoefler, Wesiak, & Guetl, 2012; AL-Smadi, Wesiak, Guetl, & Holzinger, 2012).

9.1. Purpose

As discussed in Chapter 3, assessment has become a useful tool for learning. Assessment is no more considered to be isolated from the learning process and thus is provided embedded with the learning activity. Nevertheless, students have more responsibility in the learning process in general and in assessment activities in particular. They become more engaged in developing assessment criteria, participating in self, peer-assessments, reflecting on their own learning, monitoring their performance, and utilizing feedback to adapt their knowledge, skills, and behavior. Consequently, assessment tools have emerged from being stand-alone represented by monolithic systems through modular assessment tools to more flexible and interoperable generation by adopting the service-oriented architecture and modern learning specifications and standards (see Chapter 6). In this chapter, integrated automated assessment forms provided through flexible and SOA-based tools are discussed. Moreover, it presents a show case of how these forms of assessment have been integrated with a Complex Learning Resource (CLR) and used for self-directed and collaborative learning.

In self-directed learning, learners set their learning goals and plans, self assess their progress of learning, and self reflect on their used plans and goals. In order to support self-regulators alternative assessment forms that mainly focus on feedback provision and do not require teacher involvement are required. Automated assessment is a suitable assessment form that can be used to support self-directed learners and informal learning as well. This chapter shows how the enhanced AQC integrated with CLR for self-directed learning supports learner's progress and self-regulation.

In collaborative learning students consider learning goals and divide the work among the group and interacts in a social and learning environment to achieve the learning goals. The use of computers to support collaborative learning has faced some problems and challenges. For instance, lack of awareness which concerns useful information that group members need on what others are doing, what others know about the current task, and what group members will do next, and lack of coordination and communication. During collaboration group members have to maintain communication and coordination among them regarding the collaborative tasks. They have to exchange ideas, ask questions, enter in arguments, and direct their effort and progress towards the group product. This process is called *production function* of groups where students involve in social interactions in order to maintain *group well-being* and share social space for *member-support*. In Chapter 3 the use of computers to support assessment in CSCL has been discussed. Findings from literature show that CSCL activities should be linked to assessment practices thus to attract and more engage students. Moreover, providing feedback using visualization aspects - textual and graphical - have been recommended as a possible solution in order to support CSCL in both the collaborative learning process itself and group learning scaffolding. (Janssen *et al.*, 2007; Zumbach & Reimann, 2003; Reimann & Kay, 2010) this chapter shows how the Co-writing Wiki CLR integrated with self, peer-assessment activities underpinned by assessment rubrics has supported students to maintain task and social awareness as well as provided them timely feedback through enhanced visualizations (see Chapter 7 for more information about Co-writing Wiki).

To this end, this research aims to investigate the following goals: (G1) the students perception towards the use of CLRs integrated with emerging forms of e-Assessment during self-directed learning activities, and the applicability of using flexible and interoperable education tools in one complex learning resource, moreover (G2) students motivation and attitudes concerning assessment forms such as, automated assessment, self, peer-assessment, and assessment rubrics. Finally, (G3) the students preferred learning style when it comes to use SOFIA CLRs. The rest of this chapter is organized as follows: section 2 explains the study design and analysis, and section 3 concludes the results and reflects on the research goals and hypotheses.

9.2. Study

Based on the study aspects and goals discussed so far, this sub-chapter discusses the study results with respect to the hypotheses, evaluation criteria and metrics presented in Table 9.1.

TABLE 9.1. Study hypotheses, and their evaluation criteria and metrics

Hypotheses	Evaluation Criteria	Metrics
H1: the use of the tools is easy even if the user is a non-expert.	**C1.1**: to evaluate the user's level of satisfaction towards the tools	**M1.1**: ratings for functionality/usability of the tool itself, frequency of use. (students questionnaire)
		M1.2: ratings for emotional aspects while using the tools. (students questionnaire)
	C1.2: To identify possible improvements for the tool based on comments and suggestions,	**M1.3**: suggestions and comments based on open questions. (students questionnaire)
H2: using the tools has a positive impact on the students' motivation concerning their learning activities	**C2.1**: to evaluate students' motivation concerning their learning activities	**M2.1**: ratings of students' extrinsic and intrinsic motivation regarding peer-assessment activity before using the tool (students questionnaire)
		M2.2: ratings of students' extrinsic and intrinsic motivation regarding the course and its tasks after using the tool. (students questionnaire)
		M2.3: ratings of students' group-assessment activities using Co-writing Wiki (students questionnaire)
	C2.2: to identify preferable learning style of the students	**M2.4:** ratings regarding the learning styles. (students questionnaire)

9.2.1. Method

The study was conducted through providing a course of Scientific Research online. The study has three phases where taken by participants and ran along the entire course. The course was delivered in distance learning settings and participants got to know their partners for group work within the study activities.

Participants

In this study 12 students had participated, for 5 of them the course was mandatory and the rest participated as life-long learners. Participants are multicultural and from three continents - i.e. South America, Europe, and Australia. Eight participants are males and four females with an age range of 22 and 41 years old (M = 32, SD = 6.53). With respect to education level, 3 of them hold a Bachelor degree and 8 holds a Master degree, and 1 has a PhD. degree.

Only 6 students finished the entire study as the course was mandatory for 5 of them. One student participated in all the three phases but s/he did not finish the requirements of phase 3. Two students finished phases 1, and 2 and three students only participated in phase 1.

Apparatus and Stimuli

The course material and tests have been provided online using SOFIA environment. The enhanced AQC approach (see Section 7.2) has been used to support self-directed learning with automatically created tests - using automatic question creator - and based on the e-assessment framework discussed in Chapter 6. Moreover, the tool named “Co-writing Wiki” (see Section 7.4) integrated to SOFIA environment based on Single Sign-On (SSO) approach was used by participants in the third phase of the study to collaboratively solve a problem. Moreover, a survey tool based on the LimeSurvey[76] deployed on the university campus server has been used to deliver three questionnaires - one for each phase of the study - to investigate aspects such as, motivation and attitudes, emotions, preferable learning style, and usability. The three questionnaires are explained in next sections in more details.

Pre-questionnaire

This questionnaire was provided at the beginning of the study and investigated information on demographic data, pervious experience in group work and collaborative learning, general attitudes on self, peer-assessment after (Pintrich, Smith, Garcia, & McKeachie, 1991), and motivational aspects towards using CLR enriched with automated assessment for self-directed learning.

The section of attitudes concerning self-, peer-assessment has been adapted from the work of (Pintrich, Smith, Garcia, & McKeachie, 1991) to investigate the following four scales of motivation: *intrinsic motivation* scale measures the student’s motivation doing the peer-assessment activity for its own sake, just out of pleasure, e.g. “In a peer-assessment activity I liked opinions from peers because I got more ideas.”, *extrinsic motivation* scale measures the student’s motivation doing the peer-assessment activity in order to get approval from the teacher and a good grade, e.g. “In a peer-assessment activity I think the opinions of my work from teachers were more important than those from peers.”, *evaluating scale* measures the confidence of the students in evaluating their peer’s work, e.g. “In a peer-assessment activity I found the strengths of my peer's work when I reviewed it.”, and *receiving scale* measures how students can handle the peer’s assessment in order to recognize their own weaknesses, e.g. “In a peer-assessment activity I recognized my weakness when I got comments from peers.”. Moreover, Answers were given on a 5-point Likert scale (“strongly disagree” - “strongly agree”), so that students could state their level of agreement or disagreement.

In order to investigate the participants motivation towards the course in general and the study phases in particular, a section adapted from (Tseng & Tsai, 2010) has been added based on the following three motivation scales: *Intrinsic Goal-Oriented* scale measures the students’ intrinsic motivation regarding the course, for instance: “I prefer course material that arouses my curiosity, even if it is difficult to learn” A high value on this scale would mean that the students are doing the course for reasons such as challenges and curiosity, *Extrinsic Goal-Oriented* scale deals with the extrinsic motivation of the students, e.g. “Getting a good grade is the most satisfying thing for me right now” A student is extrinsically motivated when s/he is rather interested in rewards or good grade than in the task itself, and *Task Value* scale is about the learning task itself, i.e. how important, interesting, and useful the task and the task material are for the students. More interest in the task should lead to more involvement in one’s learning. To give an example, one item out of this scale is: “I think I will be able to use what I learn in this course in other courses”. Answers were given based on a 5-point Likert scale as described above.

[76] [http://www.limesurvey.org/]

Intermediate questionnaire

This questionnaire was provided after the second phase of the study – self-directed with automated formative assessment - (see procedure section for more details) to investigate aspects such as, quality of learning material and tests, preferred learning style, emotional aspects, and tools usability. Regarding the learning material quality a scale of ("very bad" (1) - "very good" (5)) was used. Students were asked how often they had taken a test based on a scale of "never" (1), "seldom" (2), "sometimes" (3), "often" (4).

Regarding the "usability of the learning scenario" we used the System Usability Scale (SUS) (Brooke, 1996) which contains 10 items and a 5-point Likert scale to state the level of agreement or disagreement (e.g. "I think that I would like to use this system frequently").

The learning style of 'elaborating' or 'repeating' has been investigated in order to find out if the students' learning process is rather superficial or aims at a deeper understanding. For this section, items developed by (Wild, 2000) have been translated into English (e.g. item regarding the elaborating learning style: "In my mind I try to connect what I have learned with already known issues concerning the same topic", item regarding the repeating learning style: "I try to learn the content of scripts or other notes by heart"). The answers were also given based on a 5-point Likert scale.

To figure the participants emotional state during the second phase, the Computer Emotion Scale (CES) (Kay, & Loverock, 2008) has been used. This scale includes 12 items and measures emotions related to learning new computer software as follows:

- *Happiness:* ("When I used the tool, I felt satisfied/excited/curious."),
- *Sadness:* ("When I used the tool, I felt disheartened/dispirited"),
- *Anxiety:* ("When I used the tool, I felt anxious/insecure/helpless/nervous"), and
- *Anger*: ("When I used the tool, I felt irritable/frustrated/angry").

For this section answers followed a 4-point scale of: "None of the time", "Some of the time", "Most of the time" or "All of the time".

Post-questionnaire

At the end of the third phase, this questionnaire was provided to participants. Aspects such as, task difficulty and learning effort - in terms of hours, attitudes towards the group-assessment with rubrics as part of Co-writing Wiki, Co-writing Wiki usability and participant's emotional state when they had used it, and further comments and suggestions.

Moreover this questionnaire investigates the participants motivation during the study three phases and their perception about their peer's motivation as well. For instance students were asked "How motivated were you according to the following tasks?": reading the contents, working with the self-directed tool, testing myself with questions, writing the essays, working with the Co-writing wiki, planning a study, group assessment activity, and filling in the evaluation questionnaires. A scale of ("absolutely unmotivated" (1) - "very motivated" (4)) has been used to get the participants answers.

Procedure

As mentioned before the study procedure has three phases as follows.

Phase 1: Introduction to Scientific Research

At the beginning of this phase, students were asked to answer the pre-questionnaire. Moreover, they were provided an introductory learning material about scientific research course in general, how to plan a study, and experimentation design and analysis using SOFIA. Nevertheless, information about assessment scheme as well as description of the study phases and requirements they need to achieve. Moreover, they were asked to take a summative test based on automatically created questions from the provided learning content using the enhanced AQC approach integrated to SOFIA (see Figure 9.1).

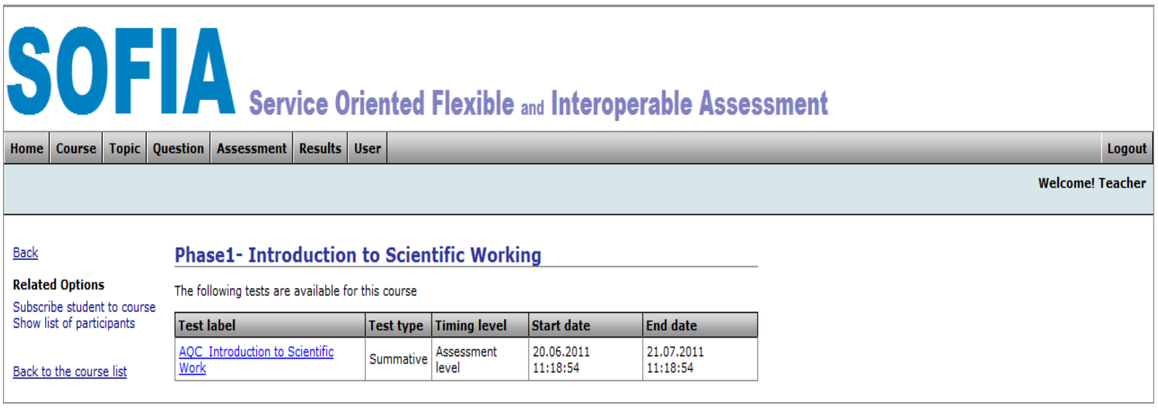


FIGURE 9.1. Teacher view for the summative test created automatically using AQC and provided for phase 1.

Phase 2: Selected Topics on Experimentation Planning

Students have been grouped by the instructor into 6 groups - two members each - based on their interest in the course (i.e. mandatory of 3 groups and volunteer of 3 groups). After that an online learning material covers scientific research has been provided using the developed system. The content has been divided into two main categories experimentation design and experimentation analysis. For each of them, 6 articles have been delivered. Each group member was requested to select one article from both categories different than the ones selected by his peer within the same group. In order to avoid members from the same group selecting similar articles they have been asked to use the discussion forum from the Co-writing Wiki to agree on their selections, and to edit the group main page on the Co-writing Wiki with their selections. Moreover, participants introduced each other using the forum, and selected their articles based on their interest.

Furthermore, the self-directed learning system supported students with the ability to test themselves before reading the article, during reading the article, and after finishing the article. A "TestMe" button has been added to the course player by which the provided learning content is used to automatically create tests based on the student preference. Those created tests could be taken several times in a formative way to get formative feedback about their current knowledge state with respect to the learning material (see Figure 9.2).

FIGURE 9.2. Student view for phase 2 learning material and the feature of "TestMe" for pre-test, section-test, and post-test.

After that, students were asked to write two essays - 1000 words per article - summarizing the topics in his/her selected articles using the Co-writing Wiki. Using the peer-review features provided in the Co-writing Wiki, group members could provide feedback on their peers' essays and learn their topics consequently.

Finally, the essays content-per group - have been used to create automatically a test - for each group - using the self-directed system. Taken this test from group members require them to be aware of peers' topics. Moreover, students were asked to answer the intermediate questionnaire after this phase.

Phase 3: Experimentation Planning

In this phase, groups have been given a problem based on this research question "Is there a difference between 'Facebook' users and 'non-Facebook' users concerning their sport activities?" Then they were requested to collaboratively plan a study using the Co-writing Wiki to solve the problem, peer-assess other groups studies using online assessment rubric has been designed for this purpose, and provide feedback. Accordingly, each group will receive feedback from other groups based on peer-assessment as well as they learn from others' ideas

The students had to write a method section by which they describe how they would investigate this research question. The students were asked to write maximally 4-5 pages in total (max. 2500 words). Furthermore, they did not have to provide any introduction with related research (although this would be mandatory in a real scientific paper). Instead, they only focused on the design of the study and gave some ideas how the analysis could be performed. Group's final products after peer-assessment and enhancement phase were evaluated by the instructor and detailed feedback has been provided for each group. After all phases of this study have been finished the students were asked to answer the post-questionnaire.

9.2.2. Results Evaluation and Discussion

This section discusses the results analyzed from students' answers on the three questionnaires and tests the study hypotheses using the evaluation criteria and metrics discussed earlier in Table 9.1 as follows.

H1: the use of the tools is easy even if the user is a non-expert

With respect to metric M1.1

Results have shown that 7 out of 8 have taken formative tests during the self-directed learning in phase 2, and one student said that s/he has never took a test because s/he did not have time. Counting the tests which the students took optionally during phase 2, 30 tests were taken in total. Regarding the three different types of tests the students stated on a 4-point rating scale that they seldom took a test before, during, or after reading the topic (pre-test: $M = 2.13$, $SD = 0.64$; sub-sections test: $M = 2.25$, $SD = 0.71$; and post-test: $M = 2.25$, $SD = 0.87$). However, looking at the actual data, the students had 6 times pre-test and post-test (maximal twice per person), and 18 times for the sub-sections tests (between 0 and 8 times per person).

Moreover, the students were asked about "what they like about the three types of tests". Results have shown that the different types of questions helped them getting an overview about the topics. Furthermore, some students also stated that the sub-section and post-tests supported them in observing their learning progress. However, the tests were criticized as they focus factual knowledge.

With respect to SOFIA self-directed learning tool usability, the average SUS score based on students responses is 66.88, where the SUS scale gives a score within a range of 0 and 100 (see Figure 9.3). According to (Sauro, 2012) "*The average SUS score from all 500 studies is a 68. A SUS score above a 68 would be considered above average and anything below 68 is below average*". The reference provides a calculator to convert the SUS score into a percentile rank through a process called normalizing. A score above than 80.3 is considered as A (the top 10% of scores). Scoring at the mean score of 68 gets you a C and anything below a 51 is an F (putting you in the bottom 15%). The calculator "*takes raw SUS scores and generates percentile ranks and letter-grades (from A+ to F) for eight different application types*". The score 66.88 the SOFIA self-directed learning CLR achieved indicates that the tool has higher perceived usability than (40% - 50%) of all products that have been tested, and it can be interpreted as a grade C.

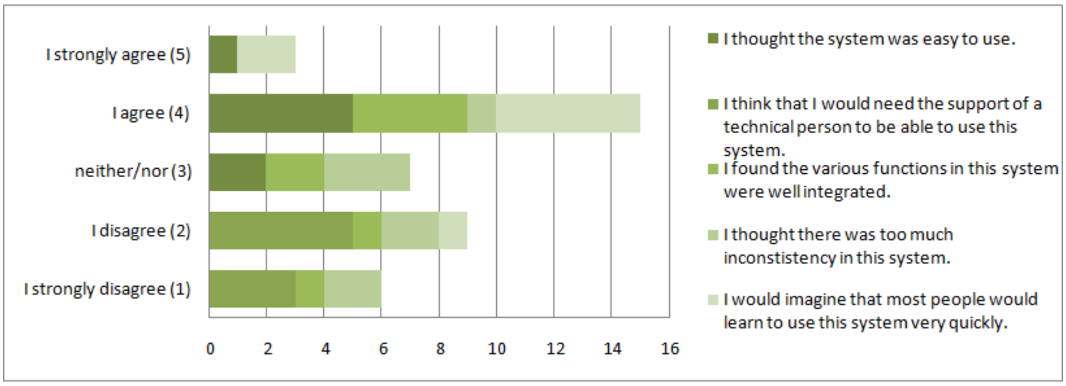


FIGURE 9.3. Students findings on a subset of SUS scale regarding SOFIA self-directed learning tool usability (n =8)

According to (Bangor, Kortum, & Miller 2008) this score can be considered as "OK" having that the complexity of the learning scenario and the use of multiple tools in a flexible and interoperable way within the same learning scenario. Moreover, with respect to what the students liked about the tool, students stated that they were in favor of the simplicity of the tool and the division of the content into meaningful modules. Furthermore the students liked the consistency and the possibility to have an overview of the learning progress and their own test results. On the other hand, students mentioned that session time-out was short. Some also complained about the slow interface. Regarding the Test Module within the self-directed tool, some students criticized the difficulty to navigate to different questions. Regarding comments and suggestions for improvement, some students would prefer a faster responding system and a faster navigation.

With respect to usability of the Co-writing Wiki itself, an average SUS score of 52.08 has been computed (see Figure 9.4). Moreover, almost all students stated that the Co-writing Wiki is easy to use. They also were in favor of the ability to discuss per topic, per page and creating and modifying pages. In addition, they mentioned that the tool was always available and consistent. However, some students complained about the usability of the Co-writing Wiki and its slowness. The students also mentioned that they were not aware of all available functions. It was also annoying for some of them self-assess their contributions. They also mentioned some editing problems, especially when this content has been copied from 'Microsoft Word' where special style tags are attached to the text and conflict with the wiki Markup and syntax. Besides, for some of them it was a little bit confusing to find the pages. Regarding comments and suggestions for improvement of Co-writing Wiki in particular, the students would also like to receive notifications on content changes or new discussion posts as to keep the user up to date. Another suggestion was to include all created pages in the tool main menu.

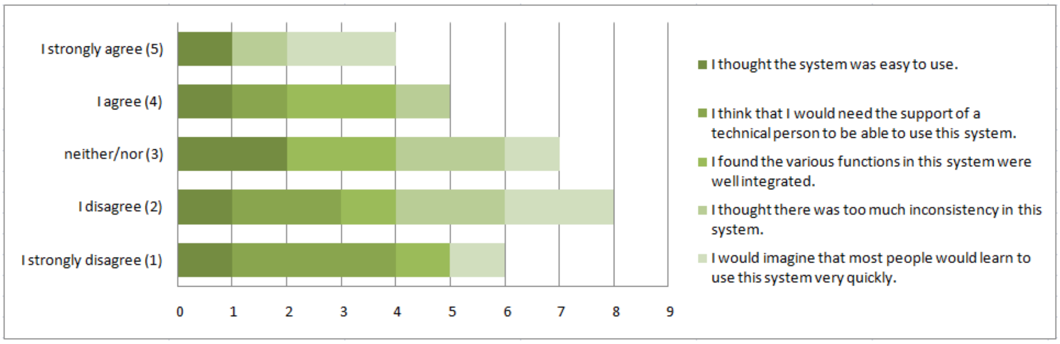
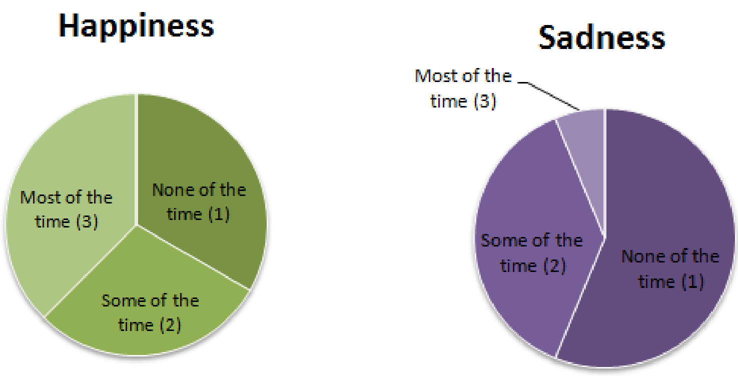


FIGURE 9.4. Students findings on a subset of SUS scale regarding Co-Writing Wiki tool usability (n =6)

With respect to metric M1.2

Concerning students' emotions during working with the SOFIA self-directed learning tool (see Figure 9.5), the comparison of the mean values indicates that the students felt equally happy (M = 1.88, SD = 0.80), sad (M = 1.5, SD = 0.60), anxious (M = 1.41, SD = 0.65), and angry (M = 1.54, SD = 0.31). Since a one-sample Kolmogorov-Smirnov Test (KS-Test) showed that the data for all four emotions are distributed normally (p-values range between 0.257 and 0.69), a one-way ANOVA (N = 8) for repeated measures was performed. With F = 0.874, df = 3, and p = 0.47 the results show no significant difference among the four types of emotion. By interpreting the mean values, it can be assumed that the students seldom felt consciously happy, sad, anxious, or angry. Linking this emotional state with the tool frequency of use form M1.1, It can be assumed that despite the unclear emotional state during the self-directed learning activity, they frequently requested an automatic test with a rate of (twice per student) on pre, and post-tests, as well as (between 0 - 8 times per student) on sub-sections tests.



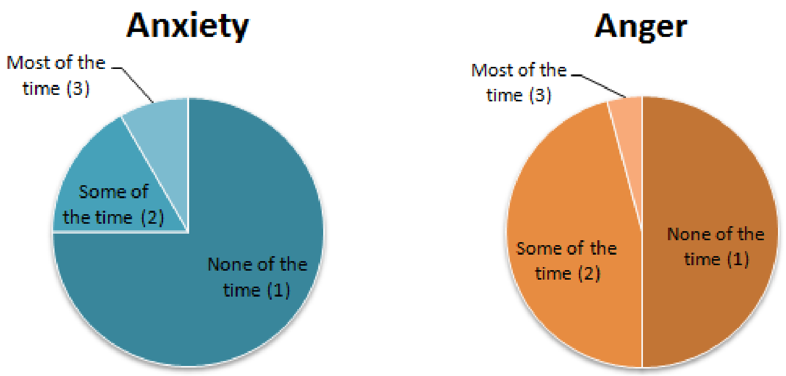


FIGURE 9.5. Students emotional states while using SOFIA self-directed learning tool (*n* =8).

The results are similar to Co-writing Wiki, the results from a 4-point rating scale showed that the students felt equally happy (M = 1.72, SD = 0.65), sad (M = 1.33, SD = 0.41), anxious (M = 1.42, SD = 0.34), and angry (M = 1.61, SD = 0.53). Since a one-sample Kolmogorov-Smirnov Test (KS-Test) showed that the data for all four emotions are distributed normally (p-values range between 0. 682 and 0. 957), a one-way ANOVA (N = 6) for repeated measures was performed. With F = 0. 619, df = 1.296, and p = 0. 500, the results show no significant difference among the four types of emotion (see Figure 9.6).

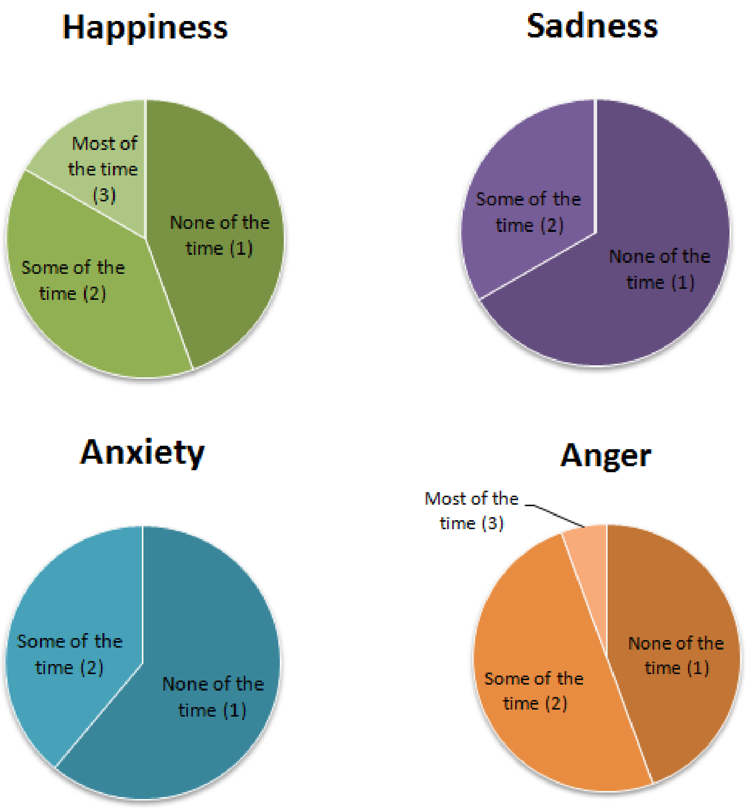


FIGURE 9.6. Students emotional states while using Co-writing Wiki tool (*n* =6).

H2: Using the tools has a positive impact on the users' motivation concerning their learning activities

With respect to metric M2.1

The student's motivation concerning the peer-assessment, a comparison of the mean values (t (11) = 5.99, p<.01) shows that the student's intrinsic motivation (M = 3.75, SD = 0.51) is significantly higher than their extrinsic motivation (M = 2.65, SD = 0.48).Thus, the students would participate in assessment for its own sake and out of pleasure and not just for getting a good grade or approval from the teacher. It can be assumed that the student's first aim was to learn something out of the course and that getting a grade does not play such an important role for them. This result stands in accordance with the fact that half of the students participated in the course voluntarily. For instance, students stated that they liked opinions from peers in order to get more ideas (M = 4.08, SD = 0.67). In contrast, they would not feel that they have learned nothing if they get a low peer score on their work (M = 1.75, SD = 0.75).

With respect to metric M2.2

The results of the student's motivation regarding the course and its tasks shows that the intrinsic motivation (M = 3.94, SD = 0.53) is significantly higher than the extrinsic motivation (M = 2.83, SD = 0.79; $t_{(11)}$ = 3.43, p<.01). This means that they are interested in the course for reasons such as curiosity and challenge, whereas high grade or rewards were not so important for them. These findings are supported by the results of the task value scale. A mean value of 3.83 (SD = 0.74) shows that the students were really interested in the task itself. The task material was also very useful and important for them. Due to their high interest, it can be assumed that this also leads to more involvement in their learning activities.

In general, questions regarding students' motivation concerning their learning activities during the three phases revealed that they were motivated up to very motivated over the course of the study. Table 9.2 shows the mean ratings as well as the respective medians in order to take account of extreme values.

TABLE 9.2. Mean ratings of motivation during the course

Motivation while:	***M (SD)***	***Md***
reading the content	3.50 (0.55)	3.5
working with the tool	2.67 (0.52)	3.0
testing themselves with questions	2,50 (0.84)	3.0
writing essays	3.50 (0.55)	3.5
planning a study	3.67 (0.52)	4.0
using the Co-writing Wiki	2.67 (1.03)	3.0
performing group-assessment	3.00 (0.0)	3.0
filling in the questionnaire	3.00 (0.0)	3.5

Note: ratings were given on a 4-pt. scale

Focusing on assessment practices and as depicted in Table 9.2, the students motivation to test themselves during phases 1 using AQC and SOFIA was above average (M = 2.5, SD = 0.84; 4-point scale). Linking this finding with the students tool frequency of use form M1.1, It can be assumed that students were motivated during the self-directed learning activity as they frequently requested an automatic test with a rate of (twice per student) on pre, and post-tests, as well as (between 0 - 8 times per student) on sub-sections tests. In addition, the students stated that testing themselves with questions often helped them for learning (M = 3.63, SD =

1.50). Nevertheless, students were motivated to participate in group-assessment (*M* = 3.0, *SD* = 0.0) using Co-writing Wiki assessment rubrics.

With respect to metric M2.3

In phase 3 students were asked to evaluate the work of other groups. Regarding the assessment rubric provided for the group review, the students stated that the assessment rubric was easy to use (*M* = 3.67, *SD* = 1.51). In addition 50 % of the students agreed on the statement that the assessment rubric supported them to effectively review the product of the other groups (*M* = 3.17, *SD* = 0.98).

The students neither agreed nor disagreed on the statements "The assessment rubric provided for the group review supported me to learn more about other group's topic." and "Using the rate control (stars) was very helpful to assess the student's level of mastery based on the rubric criteria." In addition, the students were asked what they liked regarding this group-assessment. All of the students mentioned that they liked the group-assessment because of the opportunity to see how other groups approached the problem and solved it in order to improve their own products. In the other hand, some students answered that they would have preferred to give textual detailed feedback to state suggestions and improvements instead of providing short feedback by using the assessment rubric.

With respect to metric M2.4

As depicted in Figure 9.7 a comparison of the mean values shows that there is a significant difference between the elaborating learning style (*M* = 4.05, *SD* = 0.56) and the repeating learning style (*M* = 3.04, *SD* = 0.82; t (7) = 2.71, p<.05). The students prefer the elaborating learning style, which means that their learning process aims at deeper understanding and is less superficial. Concerning elaborating, for instance the students stated that they try to link new terms or new theories to familiar terms and theories (*M* = 4.38, *SD* = 0.52). In contrast to that, the students said that they do not learn the content of scripts or other notes by heart (*M* = 2, *SD* = 1.07) which would indicate a repeating learning style.

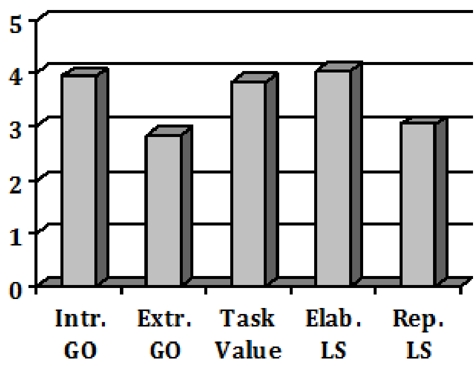


FIGURE 9.7. Mean ratings (5-pt. scales) for intrinsic and extrinsic goal orientation (GO), task value, elaborating and repeating learning styles (LS).

From M2.2 and M2.4 results we can figure the relation between elaborating learning style and deep learning based on intrinsic motivation to participate in the learning activity. The results from M2.2 show that students were intrinsically motivated after the first Phase of the course.

So due to their learning style preference, it can be assumed that the students were still intrinsically motivated in the second Phase, where they received the questions during the self-directed learning phase. Thus, the students answered the questions out of pleasure with the aim to deepen their knowledge. These findings go in line with other research such as (Kellaghan, Madaus, & Raczek, 1996) which shows that there is an evidence of the influence of intrinsic motivation on learners' engagement that leads to 'deep' learning through higher level thinking skills and the conceptual understanding. Moreover, Crooks (1998) highlights the problems associated with extrinsic motivation as it leads to 'shallow' instead of 'deep' learning.

In addition, the students stated that testing themselves with questions often helped them for learning (M = 3.63, SD = 1.50). This result is in line with the results discussed above. Therefore, it can be assumed that providing self-directed learning courses with the ability to create automatic tests supported the students to achieve their learning goals.

9.3. Conclusion and Outlook

In learner-centred learning environments active learning approaches such as self-directed learning and collaborative learning are highly common (see Chapter 2). In order to maintain quality education these learning approaches should be enriched with integrated forms of assessment thus to achieve learning goals. As discussed in the second part of this doctoral dissertation, SOFIA as a flexible e-assessment environment provides enhanced approaches for assessment that meet the potential goals of applying active learning (see Chapters 5,6, & 7). SOFIA is enhanced with CLRs with integrated alternative forms of assessment which have been developed based on an integrated model for e-assessment (IMA; see Chapter 5). Thus the designed forms of assessment are aligned to learning and instruction.

In this chapter a study with the support of SOFIA and two of its CLRs namely enhanced AQC (see Section 7.2) and Co-writing Wiki (see section 7.4) was conducted. In addition to show a proof-of-concept for the solution approach discussed in the second part of this doctoral dissertation, other aspects that include the impact of using SOFIA on student learning and motivation are discussed. The study goals include: (G1) the students perception towards the use of CLRs integrated with emerging forms of e-Assessment during self-directed and collaborative learning activities, moreover (G2) students motivation and attitudes concerning assessment forms such as, automated assessment, self, peer-assessment, and assessment rubrics. Finally, (G3) the students preferred learning style when it comes to use SOFIA CLRs.

Despite the small sample of subjects of maximum 12 participants during the study three phases, findings can be summarized as follows. With respect to the study goals, summarizing (G1) findings, it can be assumed that the tools developed to integrate assessment forms to CLRs are user friendly, and usable because of the satisfactory SUS score the tools have reached. Moreover, the students were in favor of the various functions of the tools and their simplicity. Moreover, they stated that the tools gave them a good overview of their learning progress. For further improvement, a closer look on the questions quality enhancement and on a faster interface should be considered. Moreover, the study shows the applicability of combining interoperable and flexible learning tools in one complex learning scenario.

Regarding students' motivation (G2), the results show that the students were intrinsically motivated at the beginning of the course. So they were really interested in the course and its tasks, which lead also to more involvement in their learning activities. Moreover, students' motivation was high during reading content, writing essays, doing the peer and group assessment, working with the Co-writing Wiki in a problem-based learning scenario, and filling in the questionnaires. In addition, testing themselves with automatically created tests and working with the self-directed learning tool also motivated them.

By investigating students' learning styles, we found out that the students' learning process aims at deeper understanding and is less superficial. This result is in line with the results discussed above, because intrinsic motivation is an important condition for this learning style. Thus, it can be assumed that students answered took tests out of pleasure with the aim to deepen their knowledge. Besides, students also stated that testing themselves often supported them in their learning process (G3).

For future work, more focus on functionality and usability should be considered. Nevertheless, the students' interactions and behavior using SOFIA and its CLRs should be considered in the analysis of the students' perception on SOFIA integrated forms of e-assessment. Moreover, a larger sample of participants should be used to investigate the study goals.

Chapter

10

10. Conclusions and Outlook

As discussed in Chapter 1, this doctoral dissertation aims at providing a flexible and integrated e-assessment in complex learning resources (CLR). Thus, in addition to traditional assessment, the proposed solution provides alternative forms of e-assessment - e.g. peer-assessment, rubric assessment, and performance assessment - integrated to CLRs that can be used in different application scopes - e.g. collaborative learning, self-directed learning, and game-based learning. In order to meet this goal a research methodology of three phases namely survey phase, development phase and evaluation phase has been followed.

This chapter aims to discuss research results, research outcomes, and experiences gained so far concerning the three parts of this doctoral dissertation. Thus, this chapter summarizes this doctoral research in Section10.1 and provides the research outcomes with a reflection on research questions in Section 10.2. Moreover, experiences gained from literature survey phase, development phase, and evaluation phase are illustrated in Section 10.3 and used to identify open issues and further work. Section 10.4 gives insights for further improvements of SOFIA system and SOFIA research in general.

10.1. Summary

Findings from theoretical background research in Chapter 2 show that education systems have been influenced by a new age of information where information and communication technology (ICT) plays a major role in education and learning society. Over the last decades education has evolved to be administered and provided using technology. However, this shift of education paradigm to e-education has been dominated by the technology with shy attention to pedagogy and theories of learning (Watson, 2001). As a result, educators are faced with the challenge to select among a variety of resulting educational technology and tools to meet their educational goals (Ravenscroft, 2001). Therefore, experts in education and educational professional organizations recommended designing frameworks and models to foster education through: (a) standards and guidelines to administer, develop, and provide quality e-education, (b) an emphasis on pedagogy and theories of learning as well as types of learning such as, collaborative learning, self-directed learning, intuitive learning, and social learning, (c) adaptive and personalized educational tools and services through which learners are provided learning environments that maintain their social identity, their learning progress, and life-long learning skills.

In Chapter 3, findings from literature research show that assessment forms provided in e-education should be aligned with instruction and learning (Biggs, 1999) thus to meet the education goals. Therefore, it is required to consider teaching strategies, learning objectives, learning theories and pedagogy when it comes to design assessment. Assessment has evolved to advocate alternative forms of assessment - such as, performance assessment, self and peer-assessment, behavioural assessment, portfolio-based assessment, and rubric-based assessment - through which high level of metacognitive skills are evaluated, on-going feedback is provided, and students are more engaged in the learning process (Birenbaum, 2003). As a result, assessment is considered as a tool for learning. Moreover, it becomes part of the learning process and represented as integrated assessment forms. Nevertheless, students have more responsibility in the learning process in general and in the assessment activities in particular. They become more engaged in developing assessment criteria, participating in self, peer-assessments, reflecting on their own learning, monitoring their performance, and utilizing feedback to adapt their knowledge, skills, and behaviour (see Chapter 3). Consequently, educators are faced with the challenge of having to develop appropriate, authentic, reliable, and ethical e-assessment that is integrated with the learning process, evaluates learning, engages students, appraises the students' learning process, and promotes further learning (*cf.* Bartley, 2006).

Chapter 4 discusses educational standards and specifications with emphasis on e-assessment. Findings show that learning content reusability and interoperability, learner's information accessibility and share ability, are main maters of quality for any LMS and thus for e-assessment. Therefore, LMS should be designed and implemented to be standard-conform. Interoperability standards and specifications are important aspects to be considered when it comes to provide flexible and interoperable e-assessment. Therefore, not only e-assessment content should be standard-conform but also e-assessment tools and services. Learning tools interoperability fosters e-assessment platforms with assessment third-party tools that acts as service, thus can be easily used to extend the main system tools and services. This flexible extension enables e-assessment platforms to provide assessment services for different application contexts - such as, collaborative learning and game-based learning - in a way to support different learning theories, learning types, and pedagogical approaches, as well as provide alternative forms of assessment as mentioned earlier in Chapter 3.

In the second part of this doctoral dissertation the solution approach is provided. The solution approach aims at addressing the challenges of having appropriate assessment practice to provide flexible and integrated e-assessment forms. To achieve this solution approach, an integrated model for e-assessment (IMA) is proposed. By using IMA, aligned assessment with instruction and learning can be designed. Moreover, the solution approach recommends using e-assessment standards and specifications in a way to assure flexible integration with complex learning resources (see Chapter 5). In designing assessment and feedback, IMA assures designing quality assessment through considering several aspects which include: (a) assessment domain: cognitive, affective, and psychomotor (*cf.* Bloom, 1956), (b) assessment type: diagnostic, formative, and summative assessment (*cf.* AL-Smadi & Guetl, 2008; Crisp, 2007; Bransford, Brown, & Cocking, 2004), (c) assessment strategy: traditional assessment, individual assessment, group assessment, self-assessment, peer-assessment, instructor-based assessment, and system-based assessment (*cf.* Dochy & McDowell, 1997), (d) assessment referencing: norm-related, criterion-related, or ipsative (*cf.* McAlpine, 2002), (e) assessment practice: behavioral assessment, performance assessment, portfolio assessment, and rubric-based assessment (*cf.* Buzzetto-More & Alade, 2006), (f) assessment adaptation: micro-adaptive assessment, or macro-adaptive assessment (*cf.* Kickmeier-Rust, Steiner, & Albert, 2009), (g) assessment method: quantitative or qualitative (*cf.* AL-Smadi & Guetl, 2008; Crisp, 2007; Culwin, 1998; Bloom, 1956), (h) assessment feedback: feedback type, format, frequency, and content (*cf.* Nicol, Milligan, 2006; Wiggins, 2001; Black & Wiliam, 1998; Charman, & Elms, 1998).

e-Assessment standards - see Chapter 4 - lack the representation of alternative forms of assessment as well as the integration with complex learning resources (CLR) - represented by complex learning objects. Therefore, the solution approach argues the importance of having standard-conform e-assessment system which not only adheres to content standards - e.g. IMS QTI - but also considers learning tools interoperability specifications - e.g. IMS LTI - when it comes to provide integrated assessment to CLRs. In addition, providing assessment that is aligned to instruction and learning requires flexible technology which leads to flexible pedagogy. Adopting flexible and accessible software architecture fosters assessment tools to be flexible and to be used in several application domains thus, meeting different pedagogical requirements (AL-Smadi & Guetl, 2011; Dagger et al, 2007). Therefore, by considering these limitations in e-assessment standards, and with the necessity to have integrated forms of e-assessment with the learning process, it is required to have flexible forms of assessment that is developed on top of interoperable software architecture and design in a consistent way to consider both instructional and learning (see Chapter 5). Therefore, the service-oriented approach has been used to develop what we called a service-oriented flexible and interoperable e-assessment environment (SOFIA).

SOFIA is designed and developed to address the aforementioned problems and to provide flexible e-assessment for several application contexts with a variety in learning and instructional outcomes. SOFIA uses the integrated model for e-assessment (IMA) and the service-oriented framework for e-assessment (SOFA) - discussed in Chapter 5 - to design and develop integrated and flexible assessment forms such as self, peer-assessment, automated assessment, rubric based assessment, and performance assessment to evaluate and support students' progress in their learning experiences (see Chapter 7). These learning experiences cover a variety of application scopes - such as, collaborative learning, self-directed learning, and game-based learning - which applies different learning theories and pedagogical approaches - such as, problem-based learning, self-regulation, reflective learning, active learning, and affective learning.

Nevertheless, SOFIA is integrated to complex learning resources (CLR) to provide integrated assessment for three applications scopes namely collaborative learning, self-directed learning , and game-based learning (see Chapter 7). SOFIA is extended with the services provided by these CLRs through the flexible support of the middleware layer. More precisely, the developed CLRs include: (a) an enhanced approach for peer-assessment (PASS), (b) automated and integrated assessment in self-directed learning CLR, (c) a flexible and interactive tool for assessment rubrics, (d) an enhanced approach for collaborative writing and peer-review, and finally, (e) an integrated assessment approach for game-based learning.

The solution approach discussed earlier was evaluated in the final part of this doctoral dissertation (see Chapters 8 and 9).The developed CLRs acts as third-party tools to SOFIA assessment environment and used to support conducting studies in real learning settings. First findings show that SOFIA with the CLRs supported students for better and deep learning, and empowered students with learner-centred assessment forms such as self, peer-assessment, and rubric assessment. More detailed information about the studies is provided in chapters 8 and 9 of the third part of this doctoral dissertation.

After this brief summary of this doctoral dissertation, a reflection on research motivations and questions - discussed so far in Chapter 1 - is provided in next section.

10.2. Research Results and Outcomes

The research conducted for this doctoral dissertation investigates issues of flexible and integrated educational assessment, and designing and developing flexible and integrated e-assessment in complex learning resources. Thereby, in accordance to the findings discussed so far in previous section, special focus is set on the following main research goals:

To what extent e-assessment practices fulfill multi-purpose e-Education.

A rich and comprehensive literature survey has been conducted using terminological and fictional e-assessment aspects, through which the theoretical and technological background of educational assessment provided online has been investigated (see Chapters 2, 3, & 4). Findings from this survey revealed the following arguments.

Over the last decades e-assessment has emerged with the influence of ICT. The so-called 'e-Assessment 2.0' (Elliott, 2008) is a result of this influence as well as the use of web 2.0 technology in e-learning. Rather than pedagogy and student support, the evolution of using ICT in education assessment has been led by technology (*cf.* Watson, 2001). Consequently, a variety of e-assessment tools have been developed for different contexts and application domains (see Section 3.2.2). Thus, selecting the most appropriate one for a specific learning activity has become a complex task (AL-Smadi, Guetl, & Helic, 2009b; Ravenscroft, 2001). Despite the richness in technological capabilities used in e-assessment, developed assessment tools lack to some extent the alignment with theories of learning and pedagogy. Moreover, minor group of e-assessment tools are standard-conform thus supports content and services reusability, share ability and interoperability (see Chapter 4).

The paradigm shift for online learning and assessment has caused researchers to rethink assessment practices. Traditional assessment practices - often based on objective testing - are neither adequate for testing meta-cognitive skills such as critical thinking, creativity, and self-reflection nor to test authentic learning and to support life-long learning (*cf.* Haken, 2006). Thus, rethinking e-assessment practices towards advocating alternative assessment has emerged. Alternative assessment practices - including self and peer-assessment, portfolio-assessment, behavioral assessment, and performance assessment (*cf.* Buzzetto-More & Alade, 2006) - address the lack of considering theories of learning and pedagogy in online assessment through advocating constructive, authentic, contextualized, and deep learning assessment. Consequently, educators are faced with the challenge of having to develop appropriate, authentic, reliable, and ethical e-assessment that is integrated with the learning process, evaluates learning, engages students, appraises the students' learning process, and promotes further learning (Bartley, 2006). (see Chapters 2, & 3)

As a result, a new culture of assessment of integrating assessment to complex learning resources (CLR) in order to meet requirements of assessing skills such as cognitive (e.g. problem solving, critical thinking), meta-cognitive (e.g. self-reflection and self-evaluation), social (e.g. leading discussions and working within groups), and affective aspects (e.g. internal motivation and self-efficiency) have arisen. In this new culture of assessment, students play major roles in the assessment where new forms of assessment have been adapted to suit the learning styles of the modern learners. Such forms include performance assessment, process and product assessment, directed assessment, authentic assessment, alternative assessment, collaborative assessment and self- and peer-assessment. (see Chapters 2, 3 for more information).

How to achieve high level of e-assessment flexibility in terms of: technology - secure, reusable, and accessible assessment content, tools and services - and pedagogy - assessment that is aligned with instruction and learning.

Findings from the comprehensive survey were used to propose solution approach in which flexibility - in terms of pedagogy and technology - is considered. The solution approach aims at addressing the challenges of having appropriate assessment practice to provide flexible and integrated e-assessment forms. As part of the solution approach, the goal of flexible and interoperable e-assessment is achieved through developing the e-assessment system that is (a) standard-conform (see Section 5.3) and (b) underpinned by a rich and comprehensive model for e-assessment covering essential aspects such as educational and psychological, technology and standards, and a clear guidance of how to design integrated forms of assessment - including feedback - with CLRs (see Section 5.2), thus to be (c) pedagogically flexible to support different application contexts such as collaborative writing, game-based learning and self-directed learning. Moreover, this e-assessment system (d) should be developed based on flexible software architecture that provides a technical flexibility through accessible services interfaces and transparent data transportation and communication.

Based on that, the solution approach developed in this doctoral dissertation consists of: (a) an integrated model for e-assessment (IMA) supported with a (b) bottom-up layered framework that implements and iterative approach of software development, in order to develop the integrated forms of e-assessment in CLRs (see Section 5.2), (c) a framework to design standard-conform e-assessment services and tools (see Section 5.3), (d) a service-oriented flexible and interoperable e-assessment system (SOFIA) implementing a service-oriented software approach to develop integrated and flexible standard-conform e-assessment forms integrated to complex learning resources, and (e) complex learning resources (CLR) to provide integrated assessment for three application scopes namely collaborative learning, self-directed learning , and game-based learning (see Chapter 7).

More precisely, the developed SOFIA software consists of, a modular assessment system for modern learning settings (MASS), MASS represents SOFIA standalone e-assessment system, SOFIA middleware which is developed using a service-oriented approach thus provides on the one hand the required usability, flexibility, and interoperability of learning tools and information. On the other hand the middleware provides platform-independent access to SOFIA assessment services and third-party tools. Moreover, SOFIA is enhanced with CLRs which provides alternative forms of e-assessment such as self-assessment, peer-assessment, rubric assessment, and game-based learning assessment. The developed CLRs include: (a) an enhanced approach for peer-assessment (PASS), (b) automated and integrated assessment in self-directed learning CLR, (c) a flexible and interactive tool for assessment rubrics, (d) an enhanced approach for collaborative writing and peer-review, and finally, (e) an integrated assessment approach for game-based learning. SOFIA is extended with the services provided by these CLRs through the flexible support of the middleware layer.

To this end, education goals can be met through an alignment between instruction, learning, and assessment - which is achieved by IMA - during the design of the assessment forms. This alignment requires a clear guidance and understanding for e-assessment process and services which is achieved by the framework. This framework is used to develop flexible and standard-conform e-assessment on top of a service-oriented architecture that facilitates content and services accessibility and interoperability. The solution approach has proven to achieve the required flexibility - in terms of pedagogy and technology - and SOFIA and its CLRs were used to conduct studies in real learning settings.

To what extent flexible and integrated e-assessment with complex learning resources supports students' learning.

The developed solution approach within the context of this study was used to conduct studies in different learning settings - i.e. self-directed learning and collaborative learning - and aspects such as tools usability, students' motivation, their emotional states, and their knowledge acquisition level were investigated. First findings show that SOFIA with the CLRs supported students for better and deep learning, and empowered students with learner-centred assessment forms such as self, peer-assessment, and rubric assessment. More detailed information about the studies is provided in the third part of this doctoral dissertation (Chapters 8 and 9).

By focusing on student's perception of integrated and assessment forms provided in CLRs, overall students argued that the peer assessment task is an interesting alternative and they have gained new knowledge form it. Moreover, students and tutors acquired assessment skills and more detailed knowledge about the subject domain (see Chapter 8). Nevertheless, the results show that the students were intrinsically motivated to use SOFIA and its CLRs at the beginning of the course. So they were really interested in the course and its tasks provided online using SOFIA, which lead also to more involvement in their learning activities. Moreover, students' motivation was above average during reading content, writing essays, doing the peer and group assessment, working with the Co-writing Wiki in a problem-based learning scenario. In addition, testing themselves with automatically created tests and working with the self-directed learning tool also motivated them (see Chapter 9).

10.3. Lessons Learned

This sub-chapter discusses experiences gained so far concerning the three parts of this doctoral dissertation. Providing flexible e-assessment system that can be used as standalone system or integrated with other systems is a challenging research from a software architect viewpoint. Nevertheless, providing integrated and interoperable assessment forms in complex learning resources holds great promises when it comes to develop appropriate, authentic, reliable, and ethical e-assessment that is integrated with the learning process, evaluates learning, engages students, appraises the students' learning process, and promotes further learning. However, as a researcher with computer science background, I started my research with aspects related to flexible software architectures but with false thoughts about educational assessment of providing a tool that imitates conventional classroom assessment. However, after conducting rich literature survey and having insightful discussions with our research group we learned that designing assessment is not a linear-process which starts from learning objectives and ends with learning outcome evaluation. Nevertheless, assessment is dynamic continuous process in which students are provided quality feedback which they can use to reflect on their learning and to scaffold their learning progress. According to (Martell & Calderon, 2005) "*assessment is an ongoing process that involves planning, discussion, consensus building, reflection, measuring, analyzing, and improving based on the data and artifacts gathered about a learning objective*".

In addition to that we learned that providing e-assessment is more than developing e-assessment tool that imitates conventional classroom assessment (*cf.* Dochy & McDowell, 1997; Biggs, 1999; Orsmond, Merry, & Reiling, 2000; Birenbaum, 2003; Buzzetto-More & Alade, 2006; Martell & Calderon, 2005; Elliott, 2008). According to (Buzzetto-More & Alade, 2006) "*The use of information technologies and e-learning to augment the assessment process may include: pre and post testing, diagnostic analysis, student tracking, rubric use, the support and delivery of authentic assessment through project based learning, artifact collection, and data aggregation and analysis*".

In order to design aligned assessment to learning and instruction and to provide quality assessment, assessment models are required (see Section 3.2.3). However, after analyzing the available assessment models the following findings were reached. Assessment models are either general and discuss key elements for assessment in general (e.g. Chudowsky, & Glaser, 2001) or specialized and emphasize on specific aspects of the assessment process (e.g., Almond, Steinberg, & Mislevy, 2002). However, the available assessment models lack to some extent aspects such as: (a) pedagogical flexibility and the alignment with theories of learning, (b) the suitable assessment form for the learning activity or task, (c) available technology - in terms of systems, tools, and services, (d) standards, specifications, and guidelines of how to design, and develop assessment for the target learning practice, (e) feedback as a crucial component for quality assessment practice, and (f) guidelines or frameworks of how to use these models to support developing learning tools with integrated assessment.

Based on that, an integrated model for e-assessment (IMA) has been designed as part of the solution approach we used to provide flexible and integrated e-assessment (see Chapter 5). However, using an assessment model to support developing integrated forms of e-assessment in CLR raised some challenges especially to the tools developers. Therefore, a framework based on an iterative approach has been designed to address these challenges and used as a methodology to design the assessment forms based on IMA and the application scope - e.g. co-writing in Co-writing Wiki CLR, to identify the requirements, and to develop the CLR - through an agile software development - to provide a running version of the CLR after each iteration which was evaluated by the final step of the framework namely experimentation and validation, and then the findings were used to update the requirements and to develop a new running version for new experiment. This approach were used in developing Co-writing Wiki CLR (see Section 7.4) and supported in enhancing the CLR features and performance as well as students perception on using it in real learning settings.

Focusing on experiences gained from development phase, practical insights are gained through laboratory experiments and software prototypes testing of SOFIA assessment environment and its developed third-party tools integration (see Chapters 5, 6, & 7). Moreover, experiences coming out from conducted studies using SOFIA assessment environment in real learning settings (see Chapters 8 & 9). These practical insights are summarized as follows.

Providing standard-conform e-assessment form a general viewpoint requires the developed assessment tools to adhere to available assessment specifications - often IMS QTI as widely used. Based on this research aim of providing flexible and integrated e-assessment in CLR, the developed assessment forms should be standard-conform thus to achieve flexibility in terms of accessibility, reusability, and interoperability (see Chapter 4). However, available assessment specifications such as IMS QTI lack a representation of alternative forms of assessment such as peer-assessment, rubric assessment, and performance assessment (see Chapter 4). Thus, utilizing IMS QTI in representing assessment forms developed in the CLRs (see Chapter 7) was only possible in PASS and the automated assessment based on AQC scenarios. Moreover, from our practical point of view, developing QTI-compliant assessment forms is complex as the specification information model is complex and requires a long time to be comprehended. Nevertheless, IMS QTI lacks a common representation of items and tests on the level of user interface. As discussed in Chapter 6, QTI lacks a platform-independent test items player, or even guidelines of how to design controls to play QTI items. For instance, in the context of QTI players you can find tools that utilize browser applets such as Java Applets, Silverlight, or Flash controls to design the front-end QTI players. In the context of SOFIA, a QTI player has been developed within MASS author, and deliver modules using ASP.Net controls for simple items - e.g. MCQ, True/False, and FIB, AJAX controls for specific items such as Slider, whereas for highly interactive test items such as Hot Spot, Graphic Order the player functionality development was complex and has low level of usability. Therefore, HTML5 based player to handle highly interactive items such Hot Spot, Graphic Order, and their combinations is developed.

Iintegrated assessment forms in CLRs have been developed as part of this doctoral research to support students learning in different application scopes (see Chapter 7). However, utilizing these CLR in a learning activity requires a clear understanding of their inputs, outputs, dependencies, and processes. From the instructional designer viewpoint, these CLRs represent a black box. Therefore, in the solution approach proposed in the development phase of this doctoral dissertation, CLRs are required to expose their data and services in an accessible way to facilitate sequencing them in learning activities. For this purpose a middleware layer of service-based approach is added to the SOFIA architecture. Thereby, the CLRs expose their data and services as web services within this middleware layer thus to facilitate accessing the CLRs resources during learning (see Section 6.4). However, we faced a problem of describing web services interfaces using WSDL lacks semantics of the service and information on non-functional features. For instance, aspects related to performance and dependability or only covered by the programmer documentation and this is represented by natural language. A promising solution is to use ontology-based description. For example using The Semantic Markup for Web Services (OWL-S[77]) to represent SOFIA services as further work, Thus to improve accessibility and reusability of SOFIA.

Nevertheless, providing standard-conform e-assessment considers in addition to content reusability and share ability, learning tools interoperability (see Chapter 4). The approach proposed where CLRs expose their data and services as web services within the SOFIA middleware follows the architecture solution for learning tools interoperability provided by IMS LTI specifications (see Chapter 4). Through web services, CLRs provide well-defined interfaces through which other tools and systems such as LMS and assessment tools can access CLRs data and services and utilize them in learning activities. In contrast to IMS QTI practical findings, IMS LTI is simple and in its early stages. However, further research should be used to enhance and test IMS LTI specifications.

From the viewpoint of SOFIA overall architecture (see Section 6.3), the decision for a layered architecture implementing a service-oriented approach was due to the main requirement of SOFIA e-assessment of having a flexible design thus to be used as a standalone system or integrated with other systems (see Section 6.3.3). The service-oriented approach is promising when it comes to have open, flexible, reusable, accessible, and scalable solution. On the other

[77] http://www.w3.org/Submission/OWL-S

hand, these advantages are opposed by complex communication and low performance threat. Nevertheless, challenges raised by security, privacy and reliability.

Form the SOFIA practical point of view, the service oriented approach is an important part of the solution framework, and achieves the desired flexibility of SOFIA. Examples on the achieved flexibility can be found in Chapters 8, 9 where SOFIA and the CLRs were used to conduct studies with students in learning activities as part of university courses. Moreover, SOFIA middleware has supported in integrating the CLRs (see Chapter 7) in the SOFIA e-assessment environment. Nevertheless, CLRs of Co-writing Wiki (see Section 7.2) and the automated assessment in self-directed learning (see Section7.4) are integrated to IWT LMS (Capuano, Miranda, & Orciuoli, 2009) and used for providing learning activities.

By focusing on security and privacy at the user level, a role-based security access based on OAuth protocol has been developed. Moreover, a single sign-on (SSO) approach has been used to access the CLRs from SOFIA environment. At the platform level, we have built on top of security mechanisms provided by the .NET framework. For instance, a security feature called Code signing - allows the developer to sign an assembly with a private key, and distribute the corresponding public key to each application that references the assembly - have been used to address problems of context based access of the CLRs from SOFIA environment. An example is the application context of integrated assessment for game-based learning (see Section 7.5) in which the assessment engine web services is signed with a public key and the game engine web player in SOFIA or IWT LMS can only invoke this web service using that key.

Regarding performance aspects, in the SSO approach the user data and group management are done of the level of the LMS, whereas in the CLR these data are required to define levels of access, permissions, and resources ownership. Therefore, accessing the CLR through the web services provided to access data as well as the services related to user data and groups management may cause a communication problem especially if the CLR gets user data from the LMS for each process it needs them. For instance, in the case of Co-writing Wiki the group's contribution is color-coded and when the teacher opens the contribution page of the assignment, the page is calling the user management service on SOFIA to get each user data, as well as the group's management services to get the user group data. If the assignment has different groups, then for each group member these two actions are done. And when the teacher selects a group to get its contribution on the assignment, the whole operation is repeated. This performance problem can be solved by caching the user data and group data once it is called the first time.

Another related problem is when a CLR is providing a service and different user agents are asking for this service in the same time, then a problem of instance management may rise. We have faced this problem with the CLRs of automated assessment using AQC and the assessment for game-based learning. In the scenario of automated assessment, AQC is desktop application developed using Java, every time the students are calling AQC to get automated assessment, the AQC runtime service developed in SOFIA middleware creates a new instance based on the calling context - i.e. IWT LMS, SOFIA, and a specific session to get the created question. Then the LMS uses the contextID and the sessionID internally to assign the test to the user. Thereby, the problem of multiple instances is solved through the middleware services related to context management and CLRs runtime services. However, another problem of risk management should be considered. For instance, the recovery of services, data, and communication should be developed as part of further research.

Focusing on experiences gained from experimentation and validation phase, we learned that providing alternative integrated assessment forms is not the same for all application scopes. For instance in the enhanced approach for peer-assessment, from a development point of view meeting the requirements of this scenario can be limited to providing a user control with features to select part of the answer and clicking on a button to change the selected text background color thus to be marked based on a color-code. What we learned from this scenario that some aspects should be considered when it comes to provide computer-based peer-assessment. These aspects include students' motivation to participate in peer-assessment, students' accountability to provide fair and consistent peer-assessment, the reliability of peer grading, whether or not students have the required competences and skills to provide peer-assessment or whether or not they understand the assessment criteria and are capable to apply them consistently and fairly (*cf.* Dochy & McDowell, 1997; Orsmond, Merry, & Reiling, 2000; Divaharan & Atputhasamy, 2002; Ward, Gruppen, & Regehr, 2002; Topping, 2003; McLaughlin & Simpson, 2004; Hamer, Ma, & Kwong, 2005). Therefore, these aspects were considered in designing the enhanced approach and thus an online learning as well as reference answers preparation phases were added to facilitate peer-assessment and to enhance peer-assessment quality (see Chapter 8).

Nevertheless, in order to enhance education outcome, provided assessment should be aligned to learning and instruction goals (*cf.* Birenbaum, 2003; Biggs, 1999). For instance, the developed integrated peer-assessment in PASS (see Section 7.1) and Co-writing Wiki (see Section 7.4) were based on different learning goals. Peer-assessment in PASS is used to reduce the teacher assessment workload and thus used to grade the students' candidate answers based on peer-assessment scores. However, quality peer-assessment is a crucial concern which includes aspects such as biased grading, leniency in the marking process and paybacks by the peers. Therefore, in PASS peer-assessment is provided autonomously thereby students do not know who provided this answer as well as who provided this evaluation or feedback (see Chapter 8). In contrast, peer-assessment provided in Co-writing Wiki aims at enhancing group production function through maintaining task awareness and feedback provision thus group members know who did the action on the assignment before they provide their peer-review and so they stay aware about: actions done by their group members, the assignment and task progress, as well as social awareness by providing feedback and online discussion (see Chapter 9).

To this end, the insights of experiences gained from this doctoral research were used to provide the solution approach and to conduct studies in real learning settings as discussed so far in this doctoral dissertation. Moreover, they helped to identify further research and improvements on the SOFIA e-assessment environment as summarized in next sub-chapter.

10.4. Open Issues and Further work

Providing flexible, interoperable, and integrated e-assessment in CLRs gained more interest from researcher due to the ability of providing not only traditional assessment but also alternative forms of assessment. Nevertheless, provided assessment forms in CLR cover a variety of pedagogical approaches through advocating different learning theories and teaching strategies. Due to these benefits, the research holds between its folds several trends of providing integrated assessment for different contexts.

Among these trends of research, providing alternative assessment forms in CLRs for social learning and affective/emotional learning holds great promises. Theories of social learning - such as Social Learning Theory (Bandura, 1977), Social Development Theory (Vygotsky, 1978), and Situated Learning Theory (Lave, 1990) - are discussed in Chapter 2. These theories of learning build on constructive theories on learning and define learning as learning by interacting with the learning environment and the social atmosphere to solve authenticated problems. However, providing assessment in social learning scenarios requires more attention on the learner's behavior and interactions. Nevertheless, on the level of SOFIA, services for tracking learners behavior should be implemented thus to analyze behavior and interaction to extract knowledge that is useful to scaffold students learning and encourage them for further learning. This research has already started in the SOFIA environment in particular in MASS core assessment system, and Co-writing Wiki for collaborative writing and peer-review. However, further work should be done to enhance and evaluate the provided services.

On the other hand, affective/emotional learning is an emerging research which attracts researcher in the domain of e-education. Theories of emotional learning and their relation to motivation are also discussed in Chapter 2. Hascher (2010) stated that there is *"rarely any learning process without emotion"* (p.13). Nevertheless, with respect to the relation between emotion and affect, research provides an evidence for the claim that emotion, together with cognition and motivation are the key components of learning (D'Mello et al., 2005). Moreover, according to Bransford *et al.* (2004) motivation affects the time and effort learners plan or consume to learn or to solve problems. Nevertheless, motivation is considered as an important outcome of education (Harlen, 2006, p. 61), thus teaching and learning activities should be carefully designed to promote motivation. For instance assessment is considered as one of these factors that affect motivation, however Stiggins (2001, p.36, cited after Harlen, 2006, p. 62) argues that assessment practices that are provided within a course can enhance or destroy students' desires to learn more quickly and more deep than any other tools. Therefore, assessment practices provided by SOFIA should pay more attention on emotion and affect. However, this requires further research on this domain and developing assessment services dedicated for this purpose. A possible trend of research is based on pedagogical software agents and their application in detecting the emotional state of the learner.

Focusing on the application contexts of SOFIA, providing integrated assessment and dynamic feedback for immersive education is one of the future research trends for SOFIA. This research has already started in the context of SOFIA where an enhanced approach for providing integrated assessment and dynamic feedback is developed as part of SOFIA provided assessment services (see Section 7.5). The success of this scenario in terms of achieved requirements and flexibility of providing assessment and feedback has opened the road ahead for further research to use the same approach in 3D virtual environments and augmented reality. However, further research is required to enhance the approach and to apply it in different contexts of immersive education.

Focusing on further work on SOFIA development, as discussed earlier in last section we gained experiences through practical insights which led to identify further work on SOFIA development level. For instance, on the level of SOFIA middleware further research to enhance services and recourses accessibility and flexibility can be considered. As an example we mentioned using ontologies to represent the semantics of services and their non-functional requirements such as dependencies and performance thus to further enhance their accessibility and reusability. Moreover, further research on learning tools interoperability, services workflow, and learning design specifications and standards should be considered. In the current state of SOFIA, middleware services related to application contexts and CLRs interacts with each other via direct invocation. However, this is applicable in small cases but lacks to some extent dynamic assessment design internally in SOFIA. Therefore, further work can be considered in the future to implement learning design services - e.g. IMS LD and LAMS - internally within SOFIA on both levels of the CLRs and on the level of their services as well. This will increase the level of flexibility in terms of services accessibility and reusability and thus enrich SOFIA with dynamic design of assessment practices using CLRs assessment services. However, this requires SOFIA services and CLRs to be described using ontologies such as The Semantic Markup for Web Services (OWL-S) and further testing on performance aspects.

By focusing on performance, some further work can be taken to enhance SOFIA with caching mechanisms on different levels, internally between the data layer and the presentation layer, and externally between SOFIA CLRs and application layer. Example on internal level of caching can be found on the level of knowledge visualization in co-writing wiki and their associated data in the data layer. On the other hand, another example for caching can be found in the application of automated assessment for self-directed learning where the questions created automatically for specific learning material are cached on SOFIA for further calls on the same learning content.

For further work on security and privacy, research on data and communication security can be conducted. For instance, data encryption can be used to maintain data security and users privacy. Moreover, the solution of code signing used in the game-based learning assessment to prevent unauthorized access to the assessment engine can be extended and used on all services of SOFIA. Thus, only target users - services, tools, and systems - can access SOFIA services using a ticketing mechanism and security keys.

Further work on experimentation can be directed towards analyzing the students' interactions and behavior using SOFIA and its CLRs, and use the results in the analysis of the students' perception on SOFIA integrated forms of e-assessment. Moreover, different application domains in conducting studies such as teaching foreign languages and programming courses can be considered.

List of Figures

List of Tables

References

- ADL (2008). The Advanced Distributed Learning. Last retrieved October 7th 2008 from http://www.adlnet.gov/about/index.aspx.
- Agarwal, M., Shah, R., and Mannem P. (2011). Automatic question generation using discourse cues. I*n Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications*: 1–9. Oregon: Portland.
- AICC (2008). The Aviation Industry CBT (Computer Based Learning) Committee. Last retrieved October 7th 2008 from http://www.aicc.org/index.html.
- Aist, G., Kort B., Reilly R., Mostow J. & Picard, R. (2002). Experimentally augmenting an intelligent tutoring system with human-supplied capabilities: Adding human-provided emotional scaffolding to an automated reading tutor that listens. *ITS 2002-Workshop on Empirical Methods for Tutorial Dialogue Systems*, San Sebastian, Spain.
- Aldabe, I., Lopez de Lacalle, M., Maritxalar, M., Martinez, E., & Uria, L. (2006). ArikIturri: An Automatic Question Generator Based on Corpora and NLP Techniques*, Proceeding of the 8th International Conference on Intelligent Tutoring Systems,* p.584-594.
- Alessi, S.M. and Trollip, S.R. (1991). Computer-based Instruction: Methods and Development. Second Edition, Chapter 1. Prentice Hall, Englewood Cliffs, NJ. Chapter 13; laste retrieved Nov. 22nd, 2007 from http://www.quasar.ualberta.ca/edmedia/readingsnc/Drefale.html
- Allen, M. (2003). *Guide to E-learning*. John Wiley & Sons Inc, United States of America.
- Ally, M. (2004). Foundations of Educational Theory for Online Learning. In T. Anderson & F. Elloumi (Eds.), *Theory and Practice of Online Learning* (pp. 3-31): Athabasca University & Creative Commons.
- Almond, R., Steinberg, L. & Mislevy, R. (2002). Enhancing the design and delivery of assessment systems: A four process architecture. *The Journal of Technology, Learning, and Assessment, 1(5) 1-64*.
- AL-Smadi M. et al. (2011). ALICE (Adaptive Learning Via Intuitive/Interactive, Collaborative And Emotional Systems) project (2011), Deliverable D5.1.2 "Integrated Model for e-Assessment v2" (revision), project co-funded by the European Commission within the 7th Framework Programme (2007-2013), n. 257639 (2010).
- AL-Smadi M., & Guetl C. (2010). SOA-based Architecture for a Generic and Flexible E-assessment System, Proceedings of *IEEE EDUCON Education Engineering 2010 - The Future of Global Learning Engineering Education*, April 14 - 16, 2010, Madrid, SPAIN.
- AL-Smadi M., and Guetl C. (2011a). Supporting Self-Regulated Learners with Formative Assessments using Automatically created QTI-Questions, Proceedings of *IEEE EDUCON Education Engineering 2011 - The Future of Global Learning Engineering Education*, April 2 - 6, 2011, Amman, Jordan.
- AL-Smadi, M. and Guetl, C. (2011b) 'Service-oriented flexible and interoperable assessment: towards a standardised e-assessment system', *Int. J. Continuing Engineering Education and Life-Long Learning*, Vol. 21, No. 4, pp.289–307.
- AL-Smadi M., Guetl, C., & Chang V. (2011). Addressing e-Assessment practices in e-Learning Activities: A Review. *Proceedings of Global Learn Asia Pacific 2011-Global Conference on Learning and Technology to be held in Melbourne*, Australia, March 28-April 1, 2011.
- AL-Smadi M., Hoefler M. and Guetl, C. (2011a), An Integrated Model for e-Assessment of Learning Experiences Enriched with Complex Learning Resources, Proceedings of *International Workshop on Adaptive Learning via Interactive, Collaborative and Emotional Approaches (ALICE 2011)*, 3d IEEE INCoS-2011 conference: Third International Conference on Intelligent Networking and Collaborative Systems, November 30 - December 2, 2011, Fukuoka, JAPAN
- AL-Smadi M., Hoefler M. and Guetl, C. (2011b), Enhancing Wikis with Visualization Tools to Support Groups Production Function and to Maintain Task and Social Awareness, *Proceedings of 4th International Conference on Interactive Computer-aided Blended Learning,* November 2-4, 2011, Antigua Guatemala, Guatemala
- AL-Smadi M., Höfler M., Wesiak G., & Gütl, C. (2012). Complex Learning Resources Integrated with Emerging Forms of e-Assessment: an Empirical Study. *IEEE International Conference on Advanced Learning Technologies (ICALT 2012)*, Rome 7-10 July 2012, ACCEPTED
- AL-Smadi M., Wesiak G., Gütl, C., Holzinger, A. (2012). Assessment for/as Learning: Embedded Assessment Forms in Complex Learning Resources. Proceedings of *the 2nd International Workshop on Adaptive Learning via Interactive, Collaborative and Emotional Approaches (ALICE 2011)*, The Sixth

International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2012), 4-6 July, 2012, Sanpaolo Palace Hotel, Palermo, Italy. ACCEPTED

- Al-Smadi, M., Guetl, C., Dunwell, I., & Caballe, S. (2012). Enriched Learning Experience V2, ALICE (Adaptive Learning Via Intuitive/Interactive, Collaborative And Emotional Systems) project co-funded by the European Commission within the 7th Framework Programme (2007-2013), n. 257639 (2010 -2012).
- AL-Smadi, M., & Gütl, C. (2008). Past, Present and Future of e-Assessment: Towards a Flexible e-Assessment System. *Special Track on Computer-based Knowledge & Skill Assessment and Feedback in Learning Settings (CAF 2008)*, ICL 2008, Villach, Austria, September 2008.
- AL-Smadi, M., Gütl, C. & Helic, D. (2009a). Using Computers to Enhance Peer-assessment Activities: Towards a Flexible e-Assessment System. *Special Track "Computer-based Knowledge & Skill Assessment and Feedback in Learning Settings (CAF)", ICL 2009*, Villach, Austria, September 2009, 1098-1105.
- AL-Smadi, M., Guetl, C. & Helic, D. (2009b). Towards a Standardized e-Assessment System: Motivations, Challenges and First Findings. *iJET, Vol 4 (2009) Special Issue: IMCL2009*, 6-12.
- AL-Smadi, M., Guetl, C., Kappe, F. (2009). Peer Assessment System for Modern Learning Settings: Towards a Flexible E-Assessment System. *iJET, Special Issue 2: "MIPRO 2009"* 5, 5–11
- AL-Smadi, M., Gütl, C. & Kannan R. (2010). Modular Assessment System for Modern Learning Settings: MASS. *International Conference on Futuristic Computer Applications*, IISc Bangalore, India, March 20-21, 2010.
- Anane, R., Crowther, S., Beadle, J., & Theodoropoulos, G. (2004) eLearning Content Provision, *The 15th IEEE International Workshop on Databases and Expert Systems (DEXA04*), Zaragoza, Spain, pp420-425.
- Anderson, J. & McCormick, R. (2006). A common framework for e-learning quality. In: A. McCluskey (ed.). *Policy and Innovation in Education. Quality Criteria.* Brussels: European Schoolnet, pp. 4–9.
- Anderson, O. R. (2009). Neurocognitive theory and constructivism in science education: A review of neurobiological, cognitive and cultural perspectives. *Brunei International Journal of Sciences &* Mathematical Education, 1, 1-32.
- Anderson, T. (2004). Toward a Theory of Online Learning. In T. Anderson & F. Elloumi (Eds.), Theory and Practice of Online Learning (pp. 33-60): Athabasca University & Creative Commons. Retrieved February 10, 2006 from http://cde.athabascau.ca/online_book/pdf/TPOL_book.pdf
- Andrade, H and Du, Y. (2005) Student Perspectives on Rubric-Referenced Assessment. Practical Assessment, Research and Evaluation, 10(3). ISSN 1531-7714.
- Andrade, H. G. (2000). Using rubrics to promote thinking and learning. Educational Leadership, 57, 13-18.
- Andrade, H. G. (2005). Teaching with Rubrics: The Good, the Bad, and the Ugly. College Teaching, 53(1), 27-30.
- ARIADNE (2008). *ARIADNE Foundation.* Last retrieved October 7th 2008 from http://www.ariadne-eu.org/.
- Atkins, M. J. (1993). Theories of Learning and Multimedia Applications: An Overview. *Research Papers in Education*, 8(2), 251–271.
- ATP Association of Test Publishers (2002). *Guidelines for computer-based testing.* ATP.
- Bandura, A. (1977). *Social learning theory*. New York: General Learning Press.
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008): An Empirical Evaluation of the System Usability Scale, *International Journal of Human-Computer Interaction*, 24:6, 574-594.
- Barker, T. and Lee, S. Y. (2007). The verification of identity in online assessment: A comparison of methods, Proceedings of 11th Computer Aided Assessment Conference, Loughborough, July, 2007.
- Baron, J. & Keller, M. (2003). Use of rubrics in online assessment In the *Evaluations and Assessment Conference.* Adelaide, South Australia: University of South Australia.
- Barr, N. (2006). *Assessment Provision through Interoperable Segments.* Last retrieved October 12nd 2010 from the website of the APIS project: http://sourceforge.net/projects/apis/
- Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. Journal of the Learning Sciences, 9, 403-436.
- Bartley, J. (2006). Assessment is as assessment does: A conceptual framework for understanding online assessment and measurement. In M. Hricko & S. Howell (Eds.), *Online Assessment and Measurement: Foundations and Challenges.* Hershey, PA: ISP.

- Bennett, R. E. (2002). Inexorable and inevitable: The continuing story of technology and assessment. Journal of Technology, Learning, and Assessment, 1 (1).
- Biggs, J. B. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 1-18.
- Biggs, J. B. (1999) *Teaching for quality learning at university.* Buckingham: Open University.
- Billett, S. (1996). Situated learning: Bridging sociocultural and cognitive theorising. *Learning and Instruction*, 6(3), 263-280.
- Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of quality and standards* (pp. 13–36). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Biuk-Aghai, R.P., Kelen, C. & Venkatesan, H. (2008). Visualization of interactions in collaborative writing. In Proceedings of the 2008 Second IEEE International Conference on Digital Ecosystems and Technologies (pp. 97-102), Phitsanulok, Thailand.
- Bizonova, Z., & Ranc, D., (2008) Interoperability and Reuse Between Systems in eLearning. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2008* (pp. 1700-1705). Chesapeake, VA: AACE.
- Blackboard Inc. Retrieved on 7 July 2008 from www.blackboard.com.
- Bloom, B. S. (Ed.), Engelhart, M. D., Furst, E. J., Hill, W. H. & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain.* New York: David McKay.
- Bone, A. (1999). *Ensuring Successful Assessment*. University of Warwick, Coventry, UK: The National Centre for Legal Education.
- Boston, C. (2002). The concept of formative assessment. Practical Assessment, Research & Evaluation, 8(9). Last retrieved August 9th 2010 from http://pareonline.net/getvn.asp?v=8&n=9.
- Boyer, K.E., and P. Piwek, eds. (2010). Proceedings of QG2010: The third workshop on question generation. Pittsburgh: questiongeneration.org.
- BPS (2002). *Guidelines for the Development and Use of Computer-based Assessments*. Leicester: British Psychological Society.
- Bracey, P. (2010). Self-directed Learning vs Self-regulated Learning: Twins or Just Friends?. In *J. Sanchez & K. Zhang (Eds.)*, Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2010 (pp. 1600-1607). Chesapeake, VA: AACE. Retrieved from http://www.editlib.org/p/35780.
- Bransford, J. D., Brown, A. L. & Cocking, R. R. (2004). *How people learn.* Washington D.C.: National Academy Press.
- Bransford, J.D., Brown, A.L., Cocking, R.R. (Eds.) (2000). *How People Learn: Brain, Mind, Experience, and School.* Expanded Edition. Washington DC: National Academies Press.
- Bridgeman, S., Goodrich, M.T., Kobourov, S.G., Tamassia, R., (2000). PILOT: An Interactive Tool for Learning and Grading, In *Proc. of SIGCSE 3/00 Austin, TX, USA, ACM Press*, pp. 139-143.
- Brinke, D., Bruggen, J., Hermans, H., Burgers, J., Giesbers, B. & Koper, R. (2007). Modeling assessment for re-use of traditional and new types of assessment. *Computers in Human Behavior, 23*(6), 2721-2741.
- Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In Usability evaluation in industry. London: Taylor & Francis.
- Brown, J., Frishkoff, G., Eskenazi, M., (2005). Automatic question generation for vocabulary assessment. In: *Proceedings of the Annual Human Language Technology Meeting*, October 6–8, Vancouver, BC.
- Brown, S., Race, P. (1996). *500 Tips on assessment*, Cogan Page, London, UK.
- Brusilovsky, B., & Peylo, C. (2003) Adaptive and Intelligent Web-based Educational Systems. International Journal of Artificial Intelligence in Education 13, 156–169
- Brusilovsky, P. (2000). Adaptive Hypermedia: From Intelligent Tutoring Systems to Web-Based Education. Lecture Notes in Computer Science, *Intelligent Tutoring Systems*, Volume 1839/2000 (pp. 1-7). Berlin/Heidelberg: Springer.
- Bull, J. & McKenna, C. (2001). *Blueprint for Computer Assisted Assessment.* London: RoutledgeFalmer.

- Bull, S., Brna, P., Critchley, S., Davie, K. & Holzherr, C. (1999). The Missing Peer, Artificial Peers and the Enhancement of Human-Human Collaborative Student Modelling. In *Proceedings of AIED 99,* 269-276.
- Burgos, D., Moreno-Ger, P., Sierra, J. L., Fernández-Manjón, B., Specht, M., and Koper, R. (2008). Building adaptive game-based learning resources: The marriage of IMS learning design and <e-Adventure>. Simulation & Gaming, 39, 414-431.
- Burstein, J., Kukich, K., Wolff, S., Lu, C., Chodorow, M., Bradenharder, L. & Harris, M. D. (1998). Automated scoring using a hybrid feature identification technique, in '*Proceedings of the Annual Meeting of the Association of Computational Linguistics*'.
- Burstein, J., Leacock, C. & Swartz, R. (2001). Automated evaluation of essays and short answers, in *Proceedings of the International CAA Conference*.
- Buzzetto-More, N. A. & Alade, A. J. (2006). Best Practices in e-Assessment. *Journal of Information Technology Education, 5*, 251-269.
- Caballé, S., Daradoumis, T. & Xhafa, F. (2008). *Providing an Effective Structured and Context-aware Asynchronous Discussion Forum for Collaborative Knowledge Building.* In Proc. of ED-MEDIA 2008. Vienna, Austria. AACE Press.
- Cabanac, M. (2002). What is emotion? *Behavioural Processes, 60*, 69-83.
- Callear, D., Jerrams-Smith, J. & Soh, V. (2001). Caa of short non-mcq answers. *In Proccedings of the 5th International Computer Assissted Assessment conference*.
- Campbell, J., De Blois, P.B. and Oblinger, D. (2007). Academic analytics: a new tool for a new era, *EDUCAUSE Review*, 42(4), pp. 42-57.
- Capuano, N., Miranda, S., & Orciuoli, F., (2009). IWT: A Semantic Web-based Educational System, *IV Workshop of the Working Group on "AI & E-Learning"* 2009, pp. 11–16.
- Carlson, P. A. & Berry, F. C. (2005). Calibrated Peer Review: A tool for assessing the process as well as the product in learning outcomes," *Proceedings*, ASEE Conference.
- Carlson, P., & Frederick Berry, F. (2007). A Web-Based Tool for Implementing Peer-Review. In Proceeding of ASEE.
- Chang, W. C., Hsu, H. H., Smith, T. K. & Wang, C. C. (2004). Enhancing SCORM metadata for assessment authoring in e-Learning. *Journal of Computer Assisted Learning, 20*, 305-316.
- Charman, D., Elms, A. (1998). *Computer Based Assessment: A guide to good practice*, Volume I, University of Playmouth.
- Chen, W., G. Aist and J. Mostow (2009). Generating questions automatically from informational text. In *Proceedings of the 2nd Workshop on Question Generation*: 17- 24. UK: Brighton.
- Christie, J. R. (1999). Automated essay marking-for both style and content. In M. Danson (Ed.), *Proceedings of the Third Annual Computer Assisted Assessment Conference*, Loughborough University, Loughborough, UK.
- Christie, J. R. (2003). Email communication with author. 14th April.
- Chun, M. (2002). Looking where the light is better: A review of the literature on assessing higher education quality. *Peer Review. Winter/ Spring.*
- Chung, G. & O'Neill, H. (1997). *Methodological approaches to online scoring of essays,* Technical Report 461, UCLA, National Center for Research on Evaluation, Student Standards, and Testing.
- Cole, M. (2009). Using wiki technology to support student engagement: Lessons from the trenches. *Computers & Education, 52*(1), 141-146.
- Collins, A., Brown, J. S., &Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics. In L. B. Resnick (Ed.), Knowing, learning and instruction: Essays in honour of Robert Glaser (pp. 453-494). Hillsdale, NJ: Erlbaum.
- Conlon, G. (1986), Objective measures of writing ability, New York: Longman, chapter Writing assessment: issues and strategies, pp. 109–125.
- Costagliola, G., Ferrucci, F., & Fuccella, V., (2006). Scorm run-time environment as a service, *in ICWE '06: Proceedings of the 6th international conference on Web engineering*. New York, NY, USA: ACM, pp. 103-110.
- Crisp, G. (2007). *The e-Assessment Handbook*. Continuum International Publishing Group, New York.
- Crooks, T.J. (1988) The impact of classroom evaluation practices on students, *Review of Educational Research*, vol. 58, no. 4, pp. 438-481.
- Cross, J. (2004) "An informal history of eLearning", On the Horizon, Vol. 12 Iss: 3, pp.103 – 110

- Cubric , M.; & Tosic , M. (2010). Towards automatic generation of e-assessment using semantic web technologies. In: Proceedings of the 2010 International Computer Assisted Assessment Conference, Jul , vol 2010 .
- Cubric, M. (2007). Wiki-based Process Framework for Blended Learning. In *WikiSym'07*, Montréal, Québec, Canada, 21-23 October 2007, 11-22.
- Culwin, F. (1998). Web hosted assessment: possibilities and policy, *Proceedings of the 6th annual Conference on the Teaching of Computing/3rd Annual ITiCSE Conference on Changing the Delivery of Computer Science Education*, 55-58.
- Dagger D., O'Connor A., Lawless S., Walsh E., and Wade V.P. (2007). Service-Oriented E-Learning Platforms from Monolithic Systems to Flexible Services, *Published by the IEEE Computer Society 2007*.
- Daradoumis, T., Martínez Monés, A. & Xhafa, F. (2006). A layered framework for evaluating on-line collaborative learning interactions. *International Journal of Human-Computer Studies*. *64*(7), 622-635.
- Darus, S. & Stapa, S. (2001). Lecturers' expectations of a computer-based essay marking systems. *Journal of the Malaysian English Language Teachers' Association (MELTA)* 30, 47-56.
- Davies, P. (2003). Peer-Assessment: No marks required just feedback? – Evaluating the Quality of Computerized Peer-Feedback compared with Computerized Peer-Marking. In J. Cook & D. McConnell (Eds.), *Communities of Practice*, Research Proceedings of the 10th Association for Learning Technology Conference (ALT-C 2003), 8-10 Sept 2003, Sheffield, UK.
- Davies, W. M. and Davis, , H. C. (2005). Designing Assessment Tools in a Service Oriented Architecture*, In Proceedings of 1st International ELeGI Conference on Advanced Technology for Enhanced Learning (in press)*, Napoli, Italy, 2005.
- Dawson, S., Heathcote, L. and Poole, G. (2010). Harnessing ICT potential: The adoption and analysis of ICT systems for enhancing the student learning experience, International Journal of Educational Management 24(2) pp. 116-128.
- DC (2008). *The Dublin Core Metadata Initiative*, Last retrieved October 7th 2008 from http://dublincore.org/.
- Dessus, P. Lemaire, B. & Vernier, A. (2000). Free text assessment in a virtual campus, in *Proceedings of the 3rd International Conference on Human System Learning*, 61-75.
- Devedžic, V. (2006). *Semantic Web and Education*, Springer.
- Devedžic, V., Jovanovic, J., Gaševic, D. (2007). The Pragmatics of Current e-Learning Standards, *IEEE Internet Computing, Special Issue on Distance Learning*, *11*(2), 16-24.
- Di Nitto, E., Mainetti, L., Monga, M., Sbattella, L. & Tedesco, R. Supporting Interoperability and Reusability of Learning Objects: The Virtual Campus Approach. *Educational Technology & Society*, 2006, 9 (2), 33-50.
- Dietal, R. J., Herman, J. L., & Knuth, R. A. (1991). What does research say about assessmen*t?, North Central Regional Educational Laboratory*.
- Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.) *Collaborative –learning: Cognitive and Computational Approaches* (pp. 1-19). Oxford: Elsevier.
- Divaharan, S., & Atputhasamy, L. (2002). An attempt to enhance the quality of cooperative learning through peer assessment. *Journal of Educational Inquiry*, 3 (2), 72–83.
- D'Mello, S., Craig, S., Gholson, B., Franklin, S., Picard, R. & Graesser, A. (2005). *Integrating affect sensors in an intelligent tutoring system.* International conference on intelligent user interfaces, New York, AMC Press, 7-13.
- Dochy, F. (2001). A new assessment era: Different needs, new challenges. *Research Dialogue in Learning and Instruction, 2*, 11-20.
- Dochy, F. J. & McDowell, L. (1997). Introduction. Assessment as a tool for learning. *Studies in Educational Evaluation, 23*, 279-298.
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and coassessment in higher education: A review. *Studies in Higher Education*, 24 (3), 331–350.
- Douce, C., Livingstone, D. & Orwell, J. (2005). Automatic test-based assessment of programming: A review, *Journal on Educational Resources in Computing, ACM.*
- Downing, T. & Brown, I. (1997). Learning by cooperative publishing on the World-Wide Web. *Active Learning 7,* 14-16.
- Dron, J. and Anderson, T. (2009). On the design of collective applications, Proceedings of the 2009 International Conference on Computational Science and Engineering , Volume 04, pp. 368-374.

- Dror, I.E. (2008). Technology enhanced learning: The good, the bad, and the ugly. *Pragmatics & Cognition*, 2(16), pp.215-223.
- Dunwell, I., Jarvis, S. & de Freitas, S. (2010). *Four-dimensional consideration of feedback in serious games.* In Digital Games and Learning, P. Maharg & S. de Freitas (Eds.) Continuum Publishing.
- Dynan, L., Cate, T. & Rhee, K. (2008). The impact of learning structure on student's readiness for self-directed learning. *Journal of Education for Business*, v84 n2 p96-100.
- Ebner, M., Kickmeier-Rust, M. & Holzinger, A. (2008). Utilizing wiki-systems in higher education classes: A chance for universal access? *Universal Access in the Information Society,* 7, 199-207.
- Elias, T. (2011). Learning Analytics: Definitions, Processes and Potential. Retrieved January 15, 2012 from http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf
- Elliot B., (2008). Assessment 2.0: Modernising assessment in the age of web 2.0. Scottish Qualifications Authority, Retrieved on 28 May 2008 from: http://www.scribd.com.
- Elliott B. (2008). Online Collaborative Assessment, Scottish Qualifications Authority.
- Elliott, B. (2008). *Online Collaborative Assessment*, Scottish Qualifications Authority.
- Erkens, G. (2004). Dynamics of coordination in collaboration. In J. van der Linden & P. Renshaw (Eds.), Dialogic learning: Shifting perspectives to learning, instruction, and teaching (pp. 191-216). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Eschenbach, E. A. & Mesmer, M. A. (1998). *Web-based forms for design team peer evaluations*. American Society for Engineering Education 1998, Annual Conference and Exposition, Session 2630.
- Fielding R. T. (2000), Architectural Styles and the Design of Network-based Software Architectures, *Doctoral Dissertation (University of California, Irvine, CA, 2000)* 180p.
- Flotz, P.W., Laham, D., Landauer, T.K. (1999). The Intelligent Essay Assessor: Applications to Educational Technology. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning (IMEJ)*, 1, No. 2, last reviewed April 28th, 2012 from http://imej.wfu.edu/articles/1999/2/04/printver.asp
- Forsythe, G. E. & Wirth, N. (1965). Automatic grading programs, *Communications of the ACM, 8*(5), 275-278.
- Fraser, G. A., Crook, A. C. & Park, J. R. (2007). A Tool for Mapping Research Skills in Undergraduate Curricula, Bioscience Education e-journal, (9), Last retrieved October, 19th 2010 from http://www.bioscience.heacademy.ac.uk/journal/vol9/beej-9-1.pdf.
- Freeman, M. & McKenzie, J. (2002). SPARK, a confidential web-based template for self and peer assessment of student teamwork: benefits of evaluating across different subjects. *British Journal of Educational Technology, 33*(5), 551-569.
- Gamma, E., Helm, R., Johnson, R. & Vlissides, J. (1995). *Design Patterns, Elements of Reusable Object-Oriented Software.* Addison-Wesley.
- Gardner, H. (2006). *Multiple Intelligences: New Horizons in Theory and Practice*. Paperback.
- Gardner, J. (2006). *Assessment and Learning*. London: SAGE.
- Garris, R., Ahlers, R. & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*, 441-467.
- Garrison, D. R., Anderson, T. & Archer, W. (2001). Critical Thinking, Cognitive Presence, and Computer Conferencing in Distance Education. *The American Journal of Distance Education*, *15*, 7-23.
- Gehringer, E. F. (2000). Strategies and mechanisms for electronic peer review. In *Proceedings, Frontiers in Education Conference, 2000, 1*, F1B/2 - F1B/7.
- Gestalt (1999). GESTALT: Getting Educational Systems Talking Across Leading-Edge Technologies; Work Package 3, D0301 Design and Specification of the RDS (Resource Discovery Service), 1999. Retrieved on 7 July 2008 from http://www.fdgroup.co.uk/gestalt.
- Gibbs, G. & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education, 1,* 3-31.
- Gibbs, G. (1999) Using assessment strategically to change the way students learn. IN BROWN, S. & GLASNER, A. (Eds.) Assessment Matters in Higher Education: Choosing and Using Diverse Approaches. Buckingham, SRHE/Open University Press.
- Glaser, B. (2005). The Grounded Theory Perspective III: Theoretical Coding. Mill Valley, CA: Sociology Press.
- Goldstein, P. J. and Katz, R. N. (2005). Academic Analytics: The Uses of Management Information and Technology in Higher Education, ECAR Research Study Volume 8. Retrieved January 15, 2012 from http://www.educause.edu/ers0508

- Goleman, D. (1995). *Emotional Intelligence*. New York: Bantam Books.
- Goto, T.; Kojiri, T.; Watanabe, T.; Iwata, T.; & Yamada, T. (2010). Automatic Generation System of Multiple-Choice Cloze Questions and its Evaluation. Knowledge Management & E-Learning: An International Journal, Vol.2, No.3., 210-224.
- Gouli, E., Gogoulou, A., & Grigoriadou, M. (2008). Supporting Self-, Peer- and Collaborative-Assessment in ELearning: the case of the PECASSE environment. Journal of Interactive Learning Research, 19*(4)*, 615-647.
- Graesser, A.C., VanLehn, K., Rose, C., Jordan, P., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. AI Magazine, 22, 39-51.
- Grondlund, N. E. (1985), *Measurement and evaluation in teaching*. New York: Macmillan.
- Guetl C. (2007). Moving towards a Fully Automatic Knowledge Assessment Tool, *extended paper of the IMCL 2007 paper, iJET International Journal of Engineering Technologies in Learning*.
- Gütl, C. (2008). Moving towards a Fully-Automatic Knowledge Assessment Tool. *iJET International Journal of Emerging Technologies in Learning, 3(1) and* http://online-journals.org/i-jet/article/view/172/240.
- Gütl, C. (2010).The Support of Virtual 3D Worlds for enhancing Collaboration in Learning Settings, in *Francesca Pozzi and Donatella Persico (Eds.)* ,Techniques for Fostering Collaboration in Online Learning Communities: Theoretical and Practical Perspectives, IGI Global, 2011, 278-299.
- Gütl, C., Chang, V. (2008).Ecosystem-based Theoretical Models for Learning in Environments of the 21st Century. *International. Journal of Emerging Technologies in Learning (iJET)*, submitted Nov. 2008, December 2008
- Gütl, C.; Lankmayr, K.; Weinhofer, J.; & Höfler, M. (2011). Enhanced Approach of Automatic Creation of Test Items to foster Modern Learning Setting. Electronic Journal of e-Learning, 9, 1, ECEL 2010 special issue Apr 2011, 23‑38.
- Gutwin, C., & Greenberg, S. (2004). The importance of awareness for team cognition in distributed collaboration. In E. Salas & S. M. Fiore (Eds.), Team cognition: Understanding the factors that drive processes and performance (pp. 177-201). Washington: APA Press.
- Haken, M. (2006). Closing the loop-learning from assessment. Presentation made at the *University of Maryland Eastern Shore Assessment Workshop*. Princess Anne: MD.
- Hall, K. (1995). *Co-assessment: participation of students with staff in the assessment process: a report of work in progress*. Invited paper presented at the Second European Electronic Conference on Assessment and Evaluation. EARLIAE list European Academic & Research Network (EARN) (EARLI_AE on Listserv@nic.surfnet.nl).
- Hamer, J., Ma, K. T. & Kwong, H. H. (2005). A method of automatic grade calibration in peer assessment. In Proceedings, *7th Australasian conference on Computing education (ACE '05)*, 2005, 67-72.
- Hara, N., Bonk, C. J. & Angeli, C. (2000). Content analysis of online discussion in an applied educational psychology course. *Instructional Science, 28*, 115-152.
- Harlen, W. (2006). The role of assessment in developing motivation for learning. In J. Gardner (Ed.). *Assessment and Learning* (pp. 61-80). London: Sage Publications.
- Hascher, T. (2010). Learning and Emotion: Perspectives for theory and research. *European Educational Research Journal, 9*, 13-28.
- Hayes, B. G. (1999). Where's the data? Is multimedia instruction effective in training counselors? Journal of Technology in Counseling 1.1 [On-line]. Retrieved on 7 July 2008 from: http://jtc.colstate.edu/vol1_1/multimedia.htm.
- Hazari, S. (2004). Strategy for assessment of online course discussions. Journal of Information Systems Education, 15(4), 349-355.
- Hearst, M. (2000). The debate on automated essay grading. *IEEE Intelligent Systems, 15*(5), 22-37, IEEE CS Press.
- Heift, T., & Nicholson, D. (2001). Web delivery of adaptive and interactive language tutoring. International Journal of Artificial Intelligence in Education, 12(4), 310-324.
- Heilman, M. (2011). Automatic Factual Question Generation from Text. PhD thesis, Carnegie Mellon University.
- Helic, D. (2006). *Template-based Approach to Development of Interactive Tests with IMS Question and Test Interoperability*. In Proceedings of ED-MEDIA 2006, 2075-2081, AACE, Charlottesville, USA.
- Henze, N., & Nejdl, W. (2001). Adaptation in open corpus hypermedia. International Journal of Artificial Intelligence in Education, 12(4), 325-350.

- Hirsch, C., Hosking, J.G., Grundy, J.C., Chaffe, T., MacDonald, D., and Halytskyy, Y (2009). The Visual Wiki: A New Metaphor for Knowledge Access and Management, In Proceedings of the 42nd Hawaii International Conference on System Sciences (HICSS 2009), Big Island, Hawaii, Jan 5-8 2009, IEEE CS Press.
- Ho, C. H. (2004). *Assessing Electronic Discourse: A Case Study in Developing Evaluation Rubrics*. Paper presented at the 14th Annual Meeting of the Society for Text and Discourse (ST&D). Chicago, August 1-4, 2004.
- Höfler, M., AL-Smadi, M., & Gütl, C. (2012). Investigating the suitability of automatically generated test items for real tests. *The International Journal of eAssessment, 2*(1) [Online]. Available: Doc. No. 26. http://journals.sfu.ca/ijea/index.php/journal/article/viewFile/27/26.
- Hollingsworth, J. (1960). Automatic graders for programming classes. *Commun. ACM*, 3(10):528–529.
- Honey, P., & Mumford, A. (1992). *The manual of learning styles*. Maidenhead: Peter Honey.
- Huang, H. (2002). Toward constructivism for adult learners in online learning environments. *British Journal of Educational Technology*, *33*(1), 27-37.
- IEEE LOM (2008). *IEEE Standard for Learning Object Metadata*. Last retrieved October 7th 2008 from http://www.ieeeltsc.org/standards/1484-12-1-2002/.
- IEEE LTSA (2008). *IEEE Standard for Learning Technology-Learning Technology Systems Architecture*. Last retrieved October 7th 2008 from http://www.ieeeltsc.org/standards/1484-1-2003/.
- IEEE LTSC (2008). *The IEEE Learning Technology Standards Committee*. Last retrieved October 7th 2008 from http://ieeeltsc.org/.
- IEEE PAPI (2003). *IEEE P1484.2 Learner Model Working Group: PAPI Learner, Draft 7* Specification, IEEE Standards Committee on Learning Technology, 2003.
- IMS (2005). IMS Question and Test Interoperability Overview. Version 2.0 Final Specification, Unpublished, 2005, Retrieved 19 June 2008 from: http://www.imsglobal.org/question/qti_v2p0/imsqti_mdudv2p0.html.
- IMS CP (2008). *Content Packaging Specification v1.1.4*. Last retrieved October 7th 2008 from http://www.imsglobal.org/content/packaging/index.html.
- IMS GLC (2008). *The IMS Global Learning Consortium*. Last retrieved October 7th 2008 from http://www.imsglobal.org/background.html.
- IMS GWS (2009), "The IMS general web services v1.0", http://www.imsglobal.org/gws/index.html, last retrieved October 7th 2009.
- IMS LD (2008). *IMS Learning Design Specification*. Last retrieved October 7th 2008 from http://www.imsglobal.org/learningdesign/.
- IMS LIP (2005). IMS Learner Information Package (IMS LIP); Version 1.0.1 IMS Learner Information Package Specification, IMS Global Learning Consortium Inc., (January,17), Retrieved on 7 July 2008 from: http://www.imsglobal.org/profiles/index.html, last visit: 2004-11-05.
- IMS LIP (2008). *IMS Learner Information Package Specification*. Last retrieved October 7th 2008 from http://www.imsglobal.org/profiles/index.html.
- IMS LRM (2008). *Learning Resource Meta-data Specification Version 1.3 - Final Specification – HTML*. Last retrieved October 7th 2008 from http://www.imsglobal.org/metadata/index.html.
- IMS LTI (2009), "The IMS Tools Interoperability guidelines", http://www.imsglobal.org/toolsinteroperability2.cfm, last retrieved April 1st 2009.
- IMS QTI (2008). *IMS Question & Test Interoperability Specification, Version 2.0 - Final Specification*. Last retrieved October 7th 2008 from http://www.imsglobal.org/question/index.html.
- Issroff, K. & del Soldato, T. (1996). Incorporating motivation into computer-supported collaborative learning. In *Proceedings of European conference on artificial intelligence in education*. Ficha Tecnica, Lisbon.
- ITC: International Test Commission. (2006). International guidelines of computer-based and Internet-delivered testing. International Journal of Testing, 6, 143–171.
- J., Sauro, "Measuring Usability with the System Usability Scale (SUS)", retrieved on March 3rd, 2012 from http://www.measuringusability.com/sus.php
- Jackson, D., and Usher, M. (1997). Grading student programs using ASSYST. In *Proc. 28th SIGCSE Tech.Syrup.*, pp. 335--339.
- James, M. (2006). Assessment, Teaching and Theories of Learning. In J. Gardner (Ed.). *Assessment and Learning.* London: Sage Publications, 47-60.

- Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning? Computers & Education, 49, 1037-1065.
- Jegede O. (2005). Towards a New Paradigm for Assessment in Lifelong Learning, *Presented at the International Association for Educational Assessment (IAEA) Conference* held at Abuja from 5-9 September 2005.
- JISC (2007). *Effective Practice with e-Assessment*, Last retrieved October 19th 2010 from http://www.jisc.ac.uk/media/documents/themes/elearning/effpraceassess.pdf.
- JISC (2008). *Effective Practice with e-Portfolios*, Last retrieved October 19th 2010 from http://www.jisc.ac.uk/media/documents/publications/effectivepracticeeportfolios.pdf
- JISC (2009). *Effective Practice in a Digital Age*, Last retrieved October 19th 2010 from http://www.jisc.ac.uk/media/documents/publications/effectivepracticedigitalage.pdf.
- JISC (2010). *Effective Assessment in a Digital Age: A guide to technology-enhanced assessment and feedback*, Last retrieved October 19th 2010 from http://www. jisc.ac.uk/media/documents/programmes/ elearning/digiassass_eada.pdf.
- John H. Wolfe. (1976). Automatic question generation from text - an aid to independent study. *ACM SIGCUE Bulletin*, 2(1), 104-112.
- Jonassen, D. H. (1994). Thinking technology: Toward a Constructivist design model. *Educational Technology*, 34-37.
- Jones, A. & Issroff, K. (2005). Learning technologies: Affective and social issues in computer-supported collaborative learning. *Computers & Education, 44*, 395-408.
- Judd, T., Kennedy, G., & Cropper S., (2010). Using wikis for collaborative learning: Assessing collaboration through contribution. Australasian Journal of Educational Technology, 26(3), 341-354.
- Kaila, P. (2008). OAuth and OpenID 2.0. *The Seminar on network security* pp. 18 - 22.
- Kalyuga, S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction, 19*, 402-410.
- Karanikolas, N.N. (2010). Computer Assisted Assessment (CAA) of Free-Text: Literature Review and the Specification of an Alternative CAA System. *In Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE)*, 2010 19th IEEE International Workshop on. 116 – 118.
- Kay, R.H., & Loverock, S. (2008). Assessing emotions related to learning new software: The computer emotion scale. Computers in Human Behavior. 24, 1605-1623.
- Kellaghan, T., Madaus, G. F., Raczek, A (1996) *The Use of External Examinations to Improve Student Motivation.* Washington, DC:AERA
- Kellaghan, T., Madaus, G. F., Raczek, A (1996) *The Use of External Examinations to Improve Student Motivation.* Washington, DC:AERA
- Keller, T., & Tergan, S. O. (2005). Visualizing knowledge and information: An introduction. In S. O. Tergan & T. Keller (Eds.), Knowledge and information visualization: Searching for synergies (Vol. 3426, pp. 1-23). Berlin: Springer-Verlag Berlin.
- Kellough, R.D. & Kellough, N.G. (1999). *Secondary school teaching: A guide to methods and resources; planning for competence*. Upper Saddle River, New Jersey Prentice Hall.
- Khandaker, N.; Leen-Kiat Soh (2010). Assessing individual contributions to groupwork by a summary of tracked computer-supported collaborative activities. In Proceedings of the Frontiers in Education Conference (FIE), 27-30 Oct. 2010, (pp. S1E-1 - S1E-6), IEEE CS Press.
- Kickmeier-Rust, M. and Albert, D. (2010), Micro-adaptivity: protecting immersion in didactically adaptive digital educational games. Journal of Computer Assisted Learning, 26: 95–105.
- Kickmeier-Rust, M.D., Steiner, C.M., and Albert, D. (2009). Non-invasive Assessment and Adaptive Interventions in Learning Games, International Conference on Intelligent Networking and Collaborative Systems INCOS 09, pp. 301 – 305.
- Kim, M. (2004). *Why online peer assessment is needed?* Paper presented at the AECT, Chicago, IL.
- Kim, M. (2008). *Peer assessment as a learning method. Saarbrücken*, Saarland: VDM Verlag Dr. *Müller*.
- Klassen, J (2001). Pedagogical Support for the use of Information technology in teaching, Conference Proceedings for Informing Science, Krakow, Poland.
- Klenowski, V., (1995). Student self-evaluation process in student-centred teaching and learning contexts of Australia and England. Assessment in Education: Principles, Policy and Practice, 2 (2), 145–163.
- Knight, P. (1995). *Assessment for learning in Higher Education*. London: Kogan Page.

- Knight, P. (1996). Quality in higher education and the assessment of student learning. Invited paper at the Third European Electronic Conference on Assessment and Evaluation, March 4-8, EARLIAE list European Academic & Research Network (EARN) (EARLI_AE on Listserv@nic.surfnet.nl).
- Kolb, A. Y. & Kolb, D. A. (2005). *The Kolb learning style inventory – Version 3.1 Technical Specifications*. Technical Report.
- Kolb, A. Y. (1984). *Experiential Learning: Experience as the source of learning and development*. New Jersey: Prentice Hall.
- Koper, E. J. R. (2001*). Modelling units of study from a pedagogical perspective: The pedagogical metamodel behind EML*. Open University of the Netherlands.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 41*, 212-218.
- Krathwohl, D. R., Bloom, B. S. & Bertram, B. M. (1973). *Taxonomy of educational objectives, the classification of educational goals. Handbook II: Affective domain*. New York: David McKay Co. Inc.
- Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. Computers in Human Behavior, 19, 335-353.
- Kreijns, K., Kirschner, P. A., Jochems, W., & Van Buuren, H. (2004). Measuring perceived quality of social space in distributed learning groups. Computers in Human Behavior, 20, 607-632.
- Kriz, W. C. (2003). Creating Effective Learning Environments and Learning Organizations through Gaming and Simulation Design, Simulation and Gaming, Vol. 34, Iss. 4, pp 495-511.
- Kukich, K. (2000). Beyond Automated Essay Scoring. *IEEE Intelligent Systems, 15*(5), 22-27.
- Kumar, V. S., Shakya Gress, C. L. Z., Hadwin, A. F., & Winne, P. H. (2010). Assessing process in CSCL: An ontological approach. Computers in Human Behavior, 26, 825–834.
- Larkey, L. S. (1998). Automatic essay grading using text categorization techniques, in '*Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in InformationRetrieval*', 90-95.
- Larusson, J.A. & Alterman, R. (2009). Visualizing Student Activity in a Wiki-mediated Co-blogging Exercise. In CHI EA '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems (pp. 4093-4098). New York, NY: ACM.
- Lave, J. & Wenger, E. (1991). *Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Lave, J. (1990). The culture of acquisition and the practice of understanding. In J. *W. Stigler, R. A. Schweder, & G. Herdt (Eds.)*, Culturalpsychology (pp. 259-86). Cambridge: Cambridge University Press.
- Lave, J. (1993). The practice of learning. In S. Chaiklin & J. Lave (Eds.), Understanding practice: perspectives on activity and context (pp. 3-32). Cambridge: Cambridge University Press.
- LeClerc, D. & Poumay (2005). *The 8 learning events model and its principles.* Last retrieved October, 21st 2010 from http://www.labset.net/media/prod/8LEM.pdf.
- Lei He. (2006). A novel web-based educational assessment system with Bloom's taxonomy. *Current Developments in Technology-Assisted Education, Published by FORMATEX, Badajoz, Spain,*Volume III, 1861-1865.
- Leuf, B. & Cunningham, W. (2001). *The Wiki Way Quick Collaboration on the Web.* Addison Wesley.
- Lin, C.Y. (2004). Rouge: A package for automatic evaluation of summaries. In *Proceedings of ACL 2004 workshop on Text Summarization Branches Out*, Barcelona, Spain, pp. 74-81.
- Liu, M., and R.A. Calvo. 2009. An automatic question generation tool for supporting sourcing and integration in students' essays. Poster presented at the 14th Australasian Document Computing Symposium, Sydney, Australia.
- Lütticke R. (2006). Using Semantic Networks for Assessment of Learners' Answers, *Proceedings of ICALT 2006*, 1070-1072.
- Macdonald, J. (2003). Assessing Online Collaborative Learning: Process and Product. Computers & Education, 40 (4), 377-391.
- Magin, D., & Churches, A. (1988). What do students learn from self and peer assessment? *In Proceedings, EdTech'88 Conference, Australian Society for Educational Technology*, 27-29 September 1988, last reviewed April 28th, 2012 from http://www.ascilite.org.au/aset-archives/confs/edtech88/magin.html

- Marais, E., Argles, D. and von Solms, B. (2006). *Security Issues Specific to e-Assessments.* In: 8th Annual Conference on WWW Applications, 6-8th September, 2006, Bloemfontein.
- MARKIT (2008). Project MarkIT, official Website, Curtin University of Technology, last retrieved April 28th, 2012 from http://www.essaygrading.com/index.jspx
- Martell, K. & Calderon, T. (2005). Assessment of student learning in business schools: What it is, where we are, and where we need to go next. In K. Martell & T. Calderon, *Assessment of student learning in business schools: Best practices each step of the way* (Vol. 1, No. 1, pp. 1-22). Tallahassee, Florida: Association for Institutional Research.
- Martell, K., & Calderon, T. (2005). Assessment of student learning in business schools: What it is, where we are, and where we need to go next. In K. Martell & T. Calderon, *Assessment of student learning in business schools: Best practices each step of the way* (Vol. 1, No. 1, pp. 1-22). Tallahassee, Florida: Association for Institutional Research.
- Mason, O. & Grove-Stephenson, I. (2002). Automated free text marking with paperless school. In M. Danson (Ed.), *Proceedings of the Sixth International Computer Assisted Assessment Conference*, Loughborough University, Loughborough, UK.
- Mayer, R. E. & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43-52.
- McAlpine, M. (2002). *Principles of assessment*. Bluepaper No. 1. CAA Centre, University of Glasgow. Last retrieved August 18th 2010 from http://www.caacentre.ac.uk/dldocs/Bluepaper1.pdf.
- McConnell, D. (2000). *Implementing computer supported cooperative learning* (2nd ed.). London: Kogan Page.
- McGrath, J. E. (1991). Time, interaction, and performance (TIP). Small Group Research, 22, 147-174.
- McLaughlin, P., & Simpson, N. (2004). Peer assessment in first year university: How the students feel. S*tudies in Educational Evaluation*, 30 (2), 135-149.
- Melis, E., Andrès, E., Büdenbender, J., Frishauf, A., Goguadse, G., Libbrecht, P., Pollet, M., & Ullrich, C. (2001). ActiveMath: A web-based learning environment. International Journal of Artificial Intelligence in Education, 12(4), 385-407.
- Merceron, A., & Yacef, K. (2003). A Web-based tutoring tool with mining facilities to improve learning and teaching. In U. Hoppe, F. Verdejo, & J. Kay (Eds.), AI-Ed'2003 (pp. 201-208). Amsterdam: IOS Press.
- Merriman, J. (2008). *Redefining interoperability (Or why the IEEE and Oxford English Dictionary have it Wrong).* The Open Knowledge Initiative (OKI), Last retrieved October 19th 2010 from http://www.okiproject.org/view/html/site/oki/node/2916.
- Mertler, Craig A. (2001). Designing scoring rubrics for your classroom. Practical Assessment, Research & Evaluation, 7(25). Retrieved February 16, 2012 from http://PAREonline.net/getvn.asp?v=7&n=25 .
- Mikhailov, A. (1998). 'Indextron', Intelligent Engineering Systems Through Artificial Neural Networks 8, 57.
- Millard, D., Howard, Y., Bailey, C., Davis, H., Gilbert, L., Jeyes, S., Price, J., Sclater, N., Sherratt, R., Tulloch, I., Wills, G. & Young, R. (2005). Mapping the e-Learning Assessment Domain: Concept Maps for Orientation and Navigation. *In Proceedings of e-Learn 2005*, Vancouver, Canada.
- Milligan, C. (2003). Question and Test Interoperability (QTI): Extending the specification for Mathematics and Numerical Disciplines*, Maths CAA Series, The Maths, Stats & OR Network, University of Birmingham*, UK, Nov. 2003.
- Mine, T.; Suganuma, A., Shoudai, T. (2000).The Design and Implementation of Automatic Exercise Generator with Tagged Documents based on the Intelligence of Students:AEGIS,Proceedings of the ICCE/ICCAI 2000,pp.651-658.
- Ming, Y., Mikhailov, A. & Kuan, T. (2000). '*Intelligent essay marking system'*, Learners Together.
- Mislevy, R. & Riconscente, M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology, PADI Technical Report 9.* Last retrieved 20th October 2010 from http://padi.sri.com/downloads/TR9_ECD.pdf.
- Mislevy, R. J., Almond, R. G. & Lukas, J. F. (2004). *A brief introduction to evidence-centered design* (CSE Technical Report 632). Last retrieved 20th October 2010 from http://www.eric.ed.gov/PDFS/ED483399.pdf.
- Mislevy, R. J., Steinberg, L., & Almond, R. G. (1999). Evidence-centered assessment design. Available from http://www.education.umd.edu/EDMS/mislevy/papers/ECD_overview.html

- Mitchell T., Aldridge N., and Broomhead P., (2003).Computerised Marking of Short-Answer Free-Text Responses, *Manchester IAEA conference*.
- Mitchell, T., Russel, T., Broomhead, P. & Aldridge N. (2002). Towards robust computerized marking of free-text responses. In M. Danson (Ed.), *Proceedings of the Sixth International Computer Assisted Assessment Conference*, Loughborouh, UK: Loughboroug University.
- Monfardini, E., Brovelli, A., Boussaoud, D., Takerkart, S., Wicker, B., (2008). I learned from what you did: retrieving visuomotor associations learned by observation. *Neuro-Image* 42, 1207–1213.
- Moreno-Ger, P., Burgos, D., Sierra, J. L., & Fernández-Manjón, B. (2008). Educational game design for online education. Computers in Human Behavior, 24, 2530-2540.
- Moskal, Barbara M. (2003). Recommendations for developing classroom performance assessments and scoring rubrics. Practical Assessment, Research & Evaluation, 8(14). Retrieved February 16, 2012 from http://PAREonline.net/getvn.asp?v=8&n=14 .
- Muehlen, M., Nickerson, J.V., Swenson, K.D.,(2005). Developing Web Services Choreography Standards-The Case of REST vs. SOAP; *in Decision Suppport Systems*, Vol. 40 Nr 1; pp. 9-29.
- Murphy, S. (1994). Portfolios and curriculum reform: Patterns in practice. *Assessing Writing*, *1*, 175-206.
- Myers, R. (1986). Computerized Grading of Freshman Chemistry Laboratory Experiments. *Journal of Chemical Education, 63*, 507-509.
- Newbold, W. W. (1990). Computers and writing assessment: A preliminary view. *Computers and Composition* 8 (1990), pp. 7–22.
- Nichols, M. (2003). A theory for eLearning. *Educational Technology & Society*, 6(2), 1-10
- Nicol, D. J. & Milligan, C. (2006), Rethinking technology-supported assessment in terms of the seven principles of good feedback practice. In *C. Bryan and K. Clegg (Eds), Innovative Assessment in Higher Education*, Taylor and Francis Group Ltd, London
- 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.
- Norris, D., Baer, L., Leonard, J., Pugliese, L. and Lefrere, P. (2008). Action Analytics: Measuring and Improving Performance That Matters in Higher Education, EDUCAUSE Review 43(1).
- O' Regan (2003). Emotion and e-learning. *Journal of Asynchronous Learning Networks, 7*(3), 78-92.
- Oda, T., Satoh, H., & Watanabe, S. (1998). Searching deadlocked Web learners by measuring similarity of learning activities. Proceedings of Workshop "WWW-Based Tutoring" at 4th International Conference on Intelligent Tutoring Systems (ITS'98), August 16-19, 1998.
- Ohland, M. W., Loughry, M. L., Carter, R. L., Bullard, L. G., Felder, R. M., Finelli, C. J., Layton, R. A. & Schmucker, D. G. (2009). *The Comprehensive Assessment of Team Member Effectiveness (CATME): A New Peer Evaluation Instrument.* Proceedings of the 2006 ASEE Annual Conference, Chicago, Illinois, June 2006. Information about CATME may be found at http://www.catme.org.
- Olsen, J. B. (2000). Guidelines for Computer-Based Testing. Last retrieved October 19th 2010 from http://www.isoc.org/oti/articles/0500/olsen.html.
- Orrell, J. (2005). *Assessment literacy: A precursor to improving the quality of assessment.* Paper presented at ATN Evaluations and Assessment Conference , Sydney, Australia.
- Orsmond, P., (2004): Self- and peer-assessment: guidance on practice in the biosciences. *In Teaching Bioscience Enhancing Learning Series*, eds S. Maw, J. Wilson, and H. Sears, pp. 1-47 Leeds, UK: The Higher Education Academy Centre for Bioscience.
- Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education, 25*(1), 23-38.
- Orsmond, P., Merry, S., & Reiling, K. (2002). The use of examplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education, 27*(4), 309-323.
- Page, E. B. (1994). Computer grading of student prose, using modern concepts and software. *Journal of Experimental Education, 2*(62), 127-142.
- Page, E. B. (1996). Grading essay by computer: Why the controversy? Handout for *NCME Invited Symposium* 34.
- Palmer, J. Williams, R. & Dreher, H. (2002). Automated Essay Grading System Applied to a First Year University Subject – How Can We do it Better? In *Proceedings of the Informing Science and IT Education (InSITE) Conference, Cork, Ireland*, 1221-1229.

- Palmer, K. and May, C. (2004). Using Learning Styles Theory to Improve On-line Learning through Computer Assisted Diagnosis. *Proceedings for 8th CAA Conference 2004*, Loughborough University, Myles Danson (Editor), July 2004.
- Papasalouros A., Kotis K., Kanaris K. (2008). Automatic generation of multiple-choice questions from domain ontologies. IADIS e-Learning 2008 conference, Amsterdam.
- Papasalouros A., Kotis K., Kanaris K. (2011). Automatic generation of tests from domain and multimedia ontologiesInteractive Learning Environments, Volume 19, Issue 1, 2011, 5-23.
- Papazoglou, M. P. (2008). *Web Services: Principles and Technology*. Printice Hall, pp. 22-26.
- PAPI (2000). Public and Private Information; IEEE 2000, Draft Standard for Learning Technology - Public and Private Information (PAPI) for Learner, IEEE P1484.2/D6, 2000. Retrieved on 7 July 2008 from: http://ieeeltsc.org
- Park, O., & Lee, J. (2003) Adaptive Instructional Systems. Educational Technology Research and Development, 2003(25). p. 651-684, 2003.
- Pellegrino, J., Chudowsky, N. & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.
- Pelz, W. (2004). Three principles of effective online pedagogy. *Journal of Asynchronous Learning Networks, 8*(3), 33-46.
- Pérez-Marín D., Pascual-Nieto I., Rodríguez P., (2009). Computer-assisted assessment of free-text answers. *The Knowledge Engineering Review*, 24: 353-374 Cambridge University Press.
- Pintrich, P.R., Smith, D.A.F., Garcia, T., & McKeachie, W.J. (1991). A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ). Technical Report, 91, 7-17.
- Popham, W.J. (1997). What's wrong - and what's right - with rubrics. *Educational Leadership*, 55, 72-75.
- Prensky, M. (2001). Digital Natives, Digital Immigrants. In *On the Horizon*, October 2001, 9(5)NCB University Press.
- Prins, F. J., Sluijsmans, D. M., Kirschner, P. A. & Strijbos, J.-W. (2005). Formative Peer Assessment in a CSCL Environment: A Case Study. *Assessment & Evaluation in Higher Education, 30*, 417-444.
- Race, P. (1998). Practical pointers in peer assessment. In S. Brown (Ed.), *Peer Assessment in Practice*. Brimingham: SEDA.
- Race, P. (2001). *The Lecturer's Toolkit: a practical guide to learning, teaching and assessment* (2nd ed.). London: Kogan Page.
- Rada, R., Acquah, S., Baker, B. & Ramsey, P. (1993). Collaborative Learning and the MUCH System. *Computers and Education, 20*(3), 225-233.
- Ravenscroft, A. (2001). Designing E-learning Interactions in the 21st Century: revisiting and rethinking the role of theory. *European Journal of Education*, 36 (2), 133-156.
- Recker, M. & Wiley, D. (2001)"A Non-authoritative Educational Metadata Ontology for Filtering and Recommending Learning Objects," Interactive Learning Environments, vol. 9, no. 3, , pp. 255–271.
- Reimann, P. & Zumbach, J. (2003). Supporting virtual learning teams with dynamic feedback. In K. T. Lee & K. Mitchell (eds.), T*he "Second Wave" of ICT in Education: from Facilitating Teaching and Learning to Engendering Education Reform* (pp. 424-430). Hong Kong: AACE.
- Reimann, P., & Kay, J. (2010). Learning to learn and work in net-based teams: Supporting emergent collaboration with visualization tools. In *M. J. Jacobson & P. Reimann (Eds.), Designs for learning environments of the future*, (pp. 143-188) New York: Springer.
- Reimann, P., Frerejean, J., & Thompson, K. (2009). Using process mining to identify models of group decision making processes in chat data. In C. O'Malley, D. Suthers, P. Reimann & A. Dimitracopoulou (Eds.), Computer-supported collaborative learning practices: CSCL2009 conference proceedings (pp. 98-107). International Society for the Learning Sciences.
- Reiser, R. A. (2001). A History of Instructional Design and Technology: Part I: A History of Instructional Design. *ETR&D, Vol. 49*, No. 1, 53-64.
- Ridgway, J., McCusker, S., and Pead, D. (2004). *Literature Review of e-Assessment. Bristol*, UK: Nesta Future Lab.
- Roberts, T. S. (2006). *Self, peer and group assessment in e-learning.* Hershey: Information Science Publishing.
- Rogoff, B. (1990). *Apprenticeship in Thinking*. Cognitive Development in Social Context. New York: Oxford University Press.

- Rogoff, B. (1995). Observing sociocultural activities on three planes: participatory appropriation, guided appropriation and apprenticeship. In *J. V. Wertsch, P. Del Rio & A. Alverez (Eds.)*, Socioculturul studies of the mind (pp. 139-164). Cambridge: Cambridge University Press.
- Rohrmann, B. (2002). Verbal qualifiers for rating scales: Sociolinguistic considerations and psychometric data. Retrieved February 16, 2012 from http://www.rohrmannresearch.net/pdfs/rohrmann-vqs-report.pdf.
- Romero-Salcedo, M., Osuna-Gómez, C., A., Sheremetov, L., Villa, L., Morales, C., Rocha, L., et al. (2004). Study and analysis of workspace awareness in CDebate: A groupware application for collaborative debates. Interacting with Computers, 16, 657-681.
- Rose, C., Roque, A., D., D. B. & VanLehn, K. (2003). *'A hybrid text classification approach for analysis of student essays'*, Build Educational Applications Using Natural Language Processing, pp. 68-75.
- Rottmann R. M. and Hudson H. T. (1983). Computer Grading As an Instructional Tool, *Journal of College Science Teaching*, Volume 12, Pages 152-156.
- Rovai, A. P. (2000). Online and traditional assessments: What is the difference? *Internet and Higher Education, 3*, 141-151.
- Rudner, L. M. & Liang, T. (2002). Automated essay scoring using Bayes' Theorem. *The Journal of Technology, Learning and Assessment, 1*(2), 3-21.
- Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment and Evaluation in Higher Education, 28*(2), 147-164.
- Sadat, H., & Ghorbani; A.A. (2004). On The Evaluation of AdaptiveWeb Systems. In: Proc. of the Second International Workshop on Web-based Support Systems, pp. 127-136.
- Salmon, G. (2000). *E-moderating: The Key to Teaching and Learning Online*. London: Kogan-Page.
- Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. International Journal of Educational Research, 13, 89-99.
- Salovey, P. & Mayer, J. (1990). *Emotional Intelligence. Imagination, Cognition, and Personality*.
- Sanz-Lobera, A.; González Roig, A.; González Requena, I. (2011). Methodology for automated generation of multiple choice questions in self assessment. REES 2011.
- Sanz-Santamaria, S., Gutierrez Serrano, J. & Vadillo Zorita, J.A. (2007). e-Assessment Tools in the Open. A Review from a Different View: Standards & Pedagogy. In T. Bastiaens & S. Carliner (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2007* (pp. 7350-7357).
- Sclater, N. and Howie, K. (2003). User requirements of the "ultimate" online assessment engine, *Computers & Education*, 40, 285–306.
- SCORM (2008). *ADL Sharable Courseware Object Reference Model. Last r*etrieved October 7th 2008 from http://www.adlnet.gov/scorm/index.aspx.
- Screen Digest. (2009). There's life beyond World of Warcraft. Subscription MMOGs: Life beyond World of Warcraft, Retrieved on 8 December, 2009 from http://www.screendigest.com/press/releases/pdf/PR-LifeBeyondWorldOfWarcraft-240309.pdf
- Segers, M., Dochy, F., & Cascallar, E. (2003). The era of assessment engineering: Changing perspectives on teaching and learning and the role of new modes of assessment. In *M. Segers, F. Dochy, & E. Cascallar (Eds.), Optimising new modes of assessment: In search of qualities and standards* (pp. 1–12). The Neatherlands: Kluwer Academic Publishers.
- Shepherd, E. (2006). Definitions, Uses, and Benefits of Standards, Questionmark Corporation, USA, In *Online Assessment and Measurement Foundations and Challenges, Information Science Publisiong* , USA , 64-84.
- Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), Serious games: Mechanisms and effects (pp. 295–321). Mahwah, NJ: Routledge, Taylor and Francis.
- Skinner, B. F. (1974). *About Behaviorism.* London: Jonathan Cape.
- Sluijsmans, D. A., Prins, F. J. & Martens, R. L. (2006). The design of competency-based performance assessment in e-learning. *Learning Environments Research, 9*, 45-66.
- Sluijsmans, D.M.A. (2002). Student involvement in assessment.The training of peer assessment skills. *Unpublished doctoral dissertation*, Open University of the Netherlands, The Netherlands.

- Smith, A. S. G., & Blandford, A. (2003). MLTutor: An Application of Machine Learning Algorithms for an Adaptive Web-based Information System. International Journal of Artificial Intelligence in Education. 13(2-4), 233-260.
- Smythe, C. & Roberts, P. (2000). An Overview of the IMS Question & Test Interoperability Specification, *Computer Aided Assessment (CAA'2000)*, Leicestershire, UK, 2000.
- Smythe, C. (2003). The IMS Abstract Framework: White Paper. IMS Technical Report. 01 July 2003, http://www.imsglobal.org/af/afv1p0/imsafwhitepaperv1p0.html, last retrieved October 13th 2008.
- Städler, T. (2003). *Lexikon der Psychologie* [Encyclopaedia of Psychology]. Stuttgart: Kröner.
- Stahl, G., Koschmann, T. & Suthers, D. (2006). "Computer-supported collaborative learning: An historical perspective". In R. K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 409–426).
- Stanescu, L., Spahiu, C.S., Ion, A., & Spahiu, A. (2008). Question generation for learning evaluation. Proceedings of the *International Multiconference on Computer Science and Information Technology*, 509-513.
- Stemler, S. (2001). An overview of content analysis. *Practical Assessment, Research and Evaluation, 7*(17).
- Stiggins, R. J. (2001). *Student-involved assessment for learning* (4th ed). Upper Saddle River, NJ: Prentice-Hall.
- Sukkarieh, J., Pulmand, S. & Raikes, N. (2003). Auto-marking: Using computational linguistics to score short, free text responses.
- Sullivan, M, Hitchcock, M, & Dunnington, G.L. (1999). Peer and Self Assessment during Problem-Based Tutorials. *The American Journal of Surgery*, 177 (March 1999), 266-269.
- Sung, Y.-T., Chang, K.-E., Chiou, S.-K. & Hou, H.-T. (2005). The design and application of a web-based self- and peer-assessment system. *Computer & Education, 45*(2), 187-202.
- Swan, K., Shen, J. & Hiltz, R. (2006). Assessment and collaboration in online learning. Journal of Asynchronous Learning Networks, 10 (1), 45-62.
- Thorne S. (2004). Conceptual Architectural Framework, *O.K.I. Project Site. MIT*, retrieved from http://www.okiproject.org/view/html/site/oki/node/1402, last visited October, 19th , 2010.
- Tierney, Robin & Marielle Simon (2004). What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels. Practical Assessment, Research & Evaluation, 9(2). Retrieved February 14, 2012 from http://PAREonline.net/getvn.asp?v=9&n=2
- Tinoco, L., Fox, E., and Barnette, D. (1997). Online evaluation in WWW-based courseware. In *Proe. 28th SIGCSE Tech. Syrup.*, pp. 194-198.
- Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research,* 68(3), 249-276.
- Topping, K. J.; Smith, E. F.; Swanson, I.; Elliot, A., (2000): Formative Peer Assessment of Academic Writing Between Postgraduate Students. *Assessment & Evaluation in Higher Education*, Vol. 25, No. 2, p. 150-169.
- Topping. K. (2003): Self and Peer Assessment in School and University: Reliability, Validity and Utility. In Mien Segers, Filip Dochy and Eduardo Cascallar (Eds.), Optimising New Modes of Assessment: In Search of Qualities and Standards, Springer Netherlands, 55-87.
- Trahasch, S. (2004). Towards a flexible peer assessment system. In Proceeding, Information Technology Based Higher Education and Training (ITHET 2004), 516-520.
- Trentin, G. (2009). Using a wiki to evaluate individual contribution to a collaborative learning project. Journal of Computer Assisted Learning, 25, 43-55.
- Tsai, C. C., Lin, S. S. & Yuan, S.-M. (2002). Developing science activities through a networked peer assessment system. *Computers & Education, 38(1-3)*, 241-252.
- Tseng, S.-C., & Tsai, C. C. (2010). Taiwan college students' self-efficacy and motivation of learning in online peer-assessment environments. Internet and Higher Education, 13, 164-169.
- Ullman, A. & Kay, J. (2007). WikiNavMap: A visualisation to supplement team based wikis Conference on Human Factors in Computing Systems, CHI '07 extended abstracts on Human factors in computing systems (pp. 2711–2716). San Jose, CA: ACM Press.
- Valenti, S., Neri, F. & Cucchiarelli, A. (2003). An overview of current research on automated essay grading, *Journal of Information Technology Education 2*.
- Vantage Learning Technology (2000). A Study of Expert Scoring and Intellimetric Scoring Accuracy for Dimensional Scoring of Grade 11 Student Writing Responses. *Technical Report RB-397*, Vantage, USA.

- Vogten, H., Martens, H., Nadolski, R., Tattersall, C., Van Rosmalen, P. & Koper, R. (2006). Integrating IMS Learning Design and IMS Question and Test Interoperability using CopperCore Service Integration. *Proceedings of the International Workshop in Learning Networks for Lifelong Competence Development* at the TENCompetence Conference in March 30th-31st, 2006, Sofia, Bulgaria: TENCompetence.
- Vygotsky, L. S. (1978) *Mind in Society: The development of higher order psychological processes*. Cambridge and London: Harvard University Press.
- Wang, W.M., Hao, T.Y., Liu, W.Y. (2008). Automatic question generation for learning evaluation in medicine. In: *Leung, H., Li, F., Lau, R., Li, Q. (eds.) ICWL 2007. LNCS*, vol. 4823, pp. 242–251. Springer, Heidelberg (2008).
- Ward, M., Gruppen, L., & Regehr, G. (2002). Measuring Self-assessment: Current State of the Art. *Advances in Health Sciences Education*, 7 (1), 63-80.
- Watson, D. (2001). Pedagogy before technology: Re-thinking the relationship between ICT and teaching. *Education and Information Technologies*, 6 (4), 251-266.
- WebCT Inc. Retrieved on 7 July 2008 from: www.webCT. com.
- WebPA (2009). http://webpaproject.lboro.ac.uk. Last retrieved 5th February 2009.
- Weinberger, A., Dreher, H., AL-Smadi, M. & Guetl, C. (2011). Analytical Assessment Rubrics to facilitate Semi-Automated Essay Grading and Feedback Provision. Proceedings of the ATN Assessment Conference 2011, 20-21 October, Curtin University Perth, Western Australia, pp. 170 – 177. http://otl.curtin.edu.au/atna2011/files/ATNA_2011_Proceedings.pdf
- Westin, A.F. (1970).*Privacy and Freedom*; Atheneum, Bodley Head.
- Whitelock, D. & Brasher, A. (2006). *Roadmap for e-Assessment* Report for JISC.
- Whittingdon, D. & Hunt, H. (1999). Approaches to the computerised assessment of free-text responses. *In Proceedings of the 3rd International Computer Assisted Assessment Conference*, Loughborough, UK.
- Wiggins, G. (1990). The case for authentic assessment. [online]. Last retrieved August 16th 2010 from http://eric.ed.gov/PDFS/ED328611.pdf.
- Wiggins, G. (2001), Educative Assessment, San Francisco: Jossey-Bass.
- Wild, K.P. (2000). *Lernstrategien im Studium. Strukturen und Bedingungen*. Münster: Waxmann.
- Williams, R. & Dreher, H. (2004). Automatically Grading Essays with Markit, In *proceedings of Informing Science*, Rockhampton, Australia.
- Williams, R. (2001). Automated essay grading: an evaluation of four conceptual models. *In Proceedings of the 10th Annual Teaching and Learning Forum: Expanding Horizons in Teaching and Learning*, Curtin University of Technology, Perth, Australia.
- Williamson, D., Bejar, I. and Hone, A. (1999), Mental model comparison of automated and human scoring, *Journal of Educational Measurement* 36, 158–184.
- Woolley, D. R. (1994). PLATO: The emergence of on-line community. *Computer-Mediated Communication Magazine*, 1(3), 5.
- Wresch, W. (1993). The imminence of grading essays by computer–25 years later. *Computers and composition, 2*(10), 45–58.
- WS-I (2009), "Web services interoperability basic profile v1.1", http://www.ws-i.org/, last retrieved October 7th 2009.
- Yeh, S. W., & Lo, J. J. (2009). Using online annotations to support error correction and corrective feedback. *Computers & Education*, 52(4), 882–892.
- Zimmerman, B. J., (2002). Achieving self-regulation: the trial and triumph of adolescence. In *F. Pajares & T. Urdan (Eds.)*, Academic motivation of adolescents (vol. 2, pp. 1 – 27). Greenwich, CT: Information Age.
- Zimmerman, B. J., Bonner, S., & Kovach, R. (1996). Developing self-regulated learners: Beyond achievement to self-efficacy. *Washington, DC: American Psychological Association.*
- Zumbach, J., & Reimann, P. (2003). Influence of feedback on distributed problem based learning: Enhancing online communities in group functioning and well-being. In *B. Wasson, S. Ludvigsen, & U. Hoppe (Eds.), Designing for change in networked learning environments* (pp. 219-228).