A·L·I·C·E

Adaptive Learning via Intuitive/Interactive Collaborative and Emotional systems

Project Number:	**257639**
Project Title:	ALICE: ADAPTIVE LEARNING VIA INTUITIVE/INTERACTIVE, COLLABORATIVE AND EMOTIONAL SYSTEMS
Instrument:	Specific Targeted Research Projects
Thematic Priority:	ICT-2009.4.2: Technology-Enhanced Learning
Project Start Date:	June 1st, 2010
Duration of Project:	24 Months
Deliverable:	**D.5.1.2 Integrated Model for E-Assessment v2**
Revision:	2.8
Workpackage:	WP5: New Forms of Assessment
Dissemination Level:	Public
Due date:	30/11/2011
Submission Date:	30/11/2011
Responsible:	TUG: Mohammad AL-Smadi
Contributors:	TUG, COVUNI, UOC

PROJECT CO-FUNDED BY THE EUROPEAN COMMISSION WITHIN THE SEVENTH FRAMEWORK PROGRAMME (2007-2013)

SEVENTH FRAMEWORK PROGRAMME

Version History			
Version	**Date**	**Changes**	**Contributors**
v0.1	17/08/2010	Initial version	TUG
v0.2	30/08/2010	Restructuring	TUG
v0.3	14/10/2010	Changes in structure after internal revision	TUG
v0.4	29/10/2010	Input from Ian Dunwell	TUG, COVUNI
v0.5	25/11/2010	Changes after incorporation from input provided by Michael Feidakis	TUG, UOC
v0.6	29/11/2010	Changes after incorporation from input provided by Michael Feidakis	TUG, UOC
v0.7	30/11/2010	Formatting and minor changes	TUG
v0.8	27/01/2011	Revision after internal feedback from project partners	TUG
v0.9	02/02/2011	Insertion of Chapter 7 (Model of assessment)	TUG
v1.0	24/02/2011	Revision (formatting and minor changes)	TUG
v1.1	25/04/2011	Development of bottom-up framework	TUG
v1.2	15/05/2011	Validation of IMA by cognitive science expert	TUG
v1.3	15/06/2011	Revison of IMA after expert validation	TUG
v1.4	30/06/2011	Development of sub-modell for assessment (Section 7.4)	TUG
v1.5	02/07/2011	Literature review for AQC	TUG
v1.6	15/07/2011	Input of AQC survey (Section 4.4.3)	TUG
v1.7	25/09/2011	Application of IMA in CLR (Section 7.6)	TUG
v1.8	20/10/2011	Application of assessment model for CLR (Section 7.6.1.3)	TUG
v1.9	15/11/2011	Application of bottom-up framework (Section 7.6.4)	TUG

v2.0[1]	12/12/2011	Formatting and minor changes based on consortium review	TUG
v2.1	23/12/2011	Preparing expert validation (questionnaires, tool guide)	TUG
v2.2	31/01/2012	Validation of model and tools by experts	TUG
v2.3	15/02/2012	Revision of model after expert validation	TUG
v2.4	18/02/2012	Application of assessment model for storytelling (Section 7.10)	TUG
v2.5	21/02/2012	Insertion of Section 7.8 (Assessment for collaborative and social learning)	UOC
v2.6	22/02/2012	Insertion of Section 7.9 (Assessment for experiential learning and serious games)	COVUNI
v2.7	29/02/2012	Integration of contributions	TUG
V2.8[2]	10/03/2012	Formatting and minor changes based on consortium review	TUG

[1] D5.1.2: has been delivered on two phases

- **a:** has been delivered on time based on WP5 contribution (v2.0)
- **b:**after experts validation see section 7.7 (v2.8)

[2] D5.1.2: has been enhanced based on experts validation (see section 7.7) in order to provide a validated integrated assessment model.

A·L·I·C·E

Table of Contents

1 Introduction 8
1.1 Purpose and overall summary 8
1.2 Methodology 9
2 Theoretical Foundations of Learning 10
2.1 Learning 10
2.1.1 E-learning 12
2.2 Behaviorist, constructivist, and cognitive theories of learning 13
2.2.1 Behaviorist Theories 13
2.2.2 Constructivist Theories 13
2.2.3 Cognitive Theories 13
2.3 Collaborative learning 17
2.3.1 Bandura's Social Learning Theory 17
2.3.2 Social Development Theory 18
2.3.3 Co-operative vs. Collaborative Learning (Stahl, 2000) 19
2.3.3.1 Theories of collaborative learning 20
2.4 Motivation and learning 20
2.4.1 Maslow's Hierarchy of Needs 21
2.4.2 Self-efficacy Theory 21
2.4.3 ARCS Model 21
2.4.4 Attribution Theory 23
2.4.5 Flow Theory 23
2.4.6 Effectance Motivation 24
2.5 Emotion and learning 24
2.5.1 The emotion circuitry 25
2.5.2 Educational Perspectives 26
2.5.3 Emotional Intelligence (EQ) 27
2.5.4 Flow 27
2.5.5 The Learning Cycle 28
2.5.6 Theories of emotion 29
2.6 Learning styles 31
2.7 Summary of the Chapter 33
3 Assessment 34
3.1 Definition and purpose of assessment 34
3.1.1 Referencing of assessment 35
3.1.2 Types of assessment 36
3.1.3 Strategies of assessment 36
3.2 Feedback 38
3.3 Assessment from the viewpoint of cognitive and social theories 39

3.4 Assessment from the viewpoint of motivational and emotional aspects 40
3.4.1 Assessment of motivation 40
3.4.2 Assessment of emotion 41
3.4.3 Emotion assessment systems 42
3.5 Adaptive Assessment 43
3.6 Quality of Assessment from a psychological viewpoint 44
3.6.1 Objectivity 44
3.6.2 Reliability 44
3.6.3 Validity 45
3.7 Summary of the chapter 45
4 Assessment - Technology and Education 46
4.1 Overview of projects and initiatives addressed to assessment 46
4.2 Assessment Models 48
4.3 E-Assessment Software 56
4.4 Other types of e-assessment 63
4.4.1 Peer-assessment 63
4.4.2 Computer Assisted/Based Assessment of Essays and Free Text Answers 64
4.4.3 Automatic Test Item Creation 66
4.4.4 Assessment of Computer-Supported Collaborative Learning 75
5 Standards in Modern Learning Settings 80
5.1 National organization standards and guidelines 80
5.2 Professional association standards and guidelines 81
5.3 Content standards, specifications and guidelines 84
5.3.1 E-Assessment Content standards, specifications and guidelines 85
5.4 From Specifications to Standards 92
5.5 Benefits and added values of standards 92
5.6 Challenges of Releasing Standards 93
5.7 Interoperability 93
5.7.1 OKI Open Service Interface Definitions (OSID) 94
5.7.2 CopperCore Service Integration (CCSI) 97
5.7.3 IMS Tools Interoperability 98
5.8 Flexible Integration 100
6 Game-based & Simulations 101
6.1 Assessment and Feedback in Game-based Learning 101
6.2 Key considerations in game-based learning 104
6.2.1 Leaderboards 105
6.2.2 Achievements 105
6.2.3 Competition 106
6.3 Summary and conclusions 106
7 An Integrated Model for E-assessment 107

7.1 Purpose 107
7.2 Objectives of ALICE 107
7.3 An integrated model for enriched learning experiences 108
7.3.1 General Model 109
7.3.2 IMA-Core Methodology 111
7.3.3 Inputs to the enriched learning experience 112
7.3.4 Interaction with other models 114
7.4 New forms of e-assessment 114
7.5 Framework for the use of the IMA-Model 119
7.5.1 Show Case 121
7.6 First application of the model in selected complex learning scenarios 123
7.6.1 Enriched learning experience for a self-directed learning course with collaborative writing assignment 123
7.6.2 Inputs to the enriched learning experience 127
7.7 Validation of the IMA and assessment model by experts 130
7.7.1 Validation of the model 130
7.7.2 Validation of the tools 134
7.8 Assessment for collaborative and social learning 138
7.8.1 Knowledge assessment 139
7.8.2 Emotional assessment 141
7.9 Assessment for experiential learning and serious games 144
7.9.1 Assessing social learning in serious games 145
7.9.2 Adaptivity in serious games 147
7.10 Assessment for storytelling learning resources 148
7.11 Summary of the chapter 150

References 152

Annex A 170
A.1 First validation of model by querying experts 170
A.1.1 Questionnaire (Validation of the Model) 170
A.2 Second validation of the model and the tools by experts 173
A Demographic data 173
B Integrated Model for e-Assessment 173
C Automatic Question Creator (AQC) 176
D Co-writing WIKI 177
User Guide for Co-writing wiki 181
You can enter the Co-writing wiki with your test accounts, which we provided for you. **181**
You received three accounts: **181**
This User Guide should help you to find the most relevant components in the system. **181**
Student View 181
Group-, Self- and Peer-assessment 184
Teacher View 185

A.3 Example of Use regarding the AQC **188**
The AQC is integrated in the Intelligent Web Teacher (IWT). To enter the IWT, please register yourself at first. The following example gives you an overview of the AQC used for self regulated learning.......... **188**
AQC for Self Regulated Learning **188**

1 Introduction

1.1 Purpose and overall summary

This report describes activities of Work package 5, Task 5.1 of the ALICE project. The aim of ALICE is to build an adaptive and innovative environment for e-learning. To this end, personalization, collaboration, and simulation aspects are combined and also affective and emotional aspects are considered. In particular, two specific contexts will be considered in ALICE: university instruction and training about emergency and civil defence.

As the task of WP 5 is to develop new forms of assessment, the main objective of Subtask 5.1 is to design an integrated framework for e-assessment that shows the state of the art of e-assessment. Therefore, this document (D5.1.2) has two aims: First, we will outline and summarize findings of research with respect to assessment in context of e-learning. In particular, we will present the state-of-the-art research based on a broad literature review of recent research in Psychology, Pedagogy, and Computer Science. Second, based on the broad literature review, we will design an integrated Model for e-Assessment that addresses the needs of the ALICE Project.

During the last decades, educational goals and learning activities changed. Nowadays, students are no longer seen as passive recipients of information but as rather active learners that should be involved in the development of their learning activities. This change was also increased because information and communications technology became more and more important in context learning. Due to these so called e-learning activities also assessment activities changed. E-assessment, i.e., assessment in context of e-learning activities is a challenging field of research for Computational Science, Pedagogy, and Psychology. In this document we not only give a review of the state of the art of e-assessment but also provide theoretical aspects that influence e-assessment. We present theories and models of learning and also consider emotional and motivational aspects of learning and assessment. Furthermore, we present an overview of assessment models and assessment software and describe standards and guidelines addressed to e-assessment and give an overview about assessment in context of game-based learning. Based on this state-of-the-art, we describe an integrated model for e-assessment that considers the issues of the ALICE-project. Additionally a bottom-up framework has been proposed in order to use the IMA in real learning scenarios. The model and bottom-up framework have both been evaluated by experts and based on this, the original IMA model was slightly revised and a separate model for assessment has been developed to explain this part of the IMA model in more detail. Furthermore, concrete examples for assessment in collaborative and social learning, experiential learning and serious games, as well as in storytelling are given.

Therefore, we organized this paper as follows: We will first give a broad overview about basic theories of learning and will also outline aspects of motivation and emotion in context of learning (Chapter 2). In Chapter 3, we will introduce assessment in a more general term and will define and describe forms and types of assessment. In Chapter 4, we then will define assessment more precisely in context of e-learning, describing state-of-the art e-assessment software and techniques. Also some assessment models are presented in this Chapter. In Chapter 5 we then present standards and guidelines in modern learning settings and discuss the challenges in this topic. In Chapter 6 we provide an overview about e-assessment and feedback in context of game based learning. Finally, in Chapter 7 we describe the integrated model of e-assessment (Sections 7.3 and 7.4), a framework for its use (Section 7.5), and the results of an expert validation for the model (Section 7.7). The Chapter closes with three sections on concrete applications of assessment in different complex learning scenarios, namely collaborative learning, serious games, and storytelling (Sections 7.8 through 7.10).

1.2 Methodology

As the aim of this document is to provide a state-of-the-art overview on e-assessment, an extensive literature research was conducted. In a first step, we searched in data bases and search engines for general terms (e.g., "learning theories", "learning theories & review", "emotion & assessment") to become a basic idea and common knowledge of the terms and definitions. In a second step, we then refined the search, using more specific terms (e.g., "cognitive load theory" within "learning theories"). In general, our search terms remained broadly during the first steps to cover technical, psychological, and pedagogical approaches in learning and assessment. The following data bases and search engines were used to find relevant literature:

- ISI Web of Knowledge
- PsycINFO
- ScienceDirect
- Google

Further resources were technical and psychological journals and special issues of these journals (e.g., "Learning and Instruction", Special Issue: Unravelling assessment, August 2010), books (e.g, Bransford, Brown, & Cocking, 2004; J. Gardner, 2006), and conference proceedings. Access to those media was either provided via the libraries of Graz University of Technology or Karl-Franzens University of Graz or the media were freely accessible via internet. Moreover, references from relevant articles were checked for other studies and projects. We also considered information provided from special track papers, workshops, and working groups. For other projects of the European Community and other funding organizations relevant for this review, we searched through the respective websites (e.g. CORDIS website http://cordis.europa.eu/home_en.html).

2 Theoretical Foundations of Learning

In this section we will present a broad overview of theoretical foundations of learning in general. In particular, we will provide input from cognitive and social theories as well as from motivational and emotional theories of learning. It has to be mentioned that learning cannot be covered solely by one of the dimensions discussed here. 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 emotional or motivational aspects. Our aim is also to take into account the latter aspects because of the significant impact of motivation and attitudes on learning and learning outcomes.

2.1 Learning

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, et al., 2004). But how can learning be described in general? As described later on in this section, there are several theories on how learning occurs. For instance, whereas behaviorist theories suggest that a learner begins as a clean slate (*tabula rasa*, see for instance http://www.learning-theories.com/) 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

- Community-entered 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). Therefore, using the 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.

The levels are arranged in a hierarchical order with increasing difficulty (see also Table 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.

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.

Table 1. Taxonomy of Bloom, cognitive domain (revised by Krathwohl, 2002).

2.1.1 E-learning

With the development of information technologies, a rather new form of learning came important during the last decades, namely e-learning. E-learning is an educational process due to which information and communications technologies are used to mediate both asynchronous and synchronous learning (Naidu, 2006). Furthermore, Naidu defined e-learning as

"the intentional use of networked information and communications technology in teaching and learning" and moreover, "as the letter 'e' in e-learning stands for the word 'electronic', e-learning would incorporate all educational activities that are carried out by individuals or groups working online or offline, and synchronously or asynchronously via networked or standalone computers and other electronic devices". (p.1)

The growing interest in e-learning may be because of several factors (Naidu, 2006). For instance, online learning is a logical extension for distance education programs to enhance their education activities whereas in the corporate section, e-learning decreases the costs of in-house training activities. Due to the changes in technologies in the last years also learning patterns are changing, however. This changing is mostly caused by the fact that the latest generation of learners has grown up with digital media such as blogs, wikis, instant messaging etc. A broad overview regarding learning in the so-called web 2.0 can be found in Redecker (2009). However, before coming back to e-learning (and moreover, e-assessment) activities in the next chapters of this document, we will first introduce and discuss the most traditional theories of learning that still have important impact on recent research.

2.2 Behaviorist, constructivist, and cognitive theories of learning

2.2.1 Behaviorist Theories

Behaviorist theories of learning were the most dominant perspectives in the 1960s and 1970s and still influence on recent research. One of the most important representatives of behaviorism is B. F. Skinner (e.g., Skinner, 1954). Important behaviorist theories are *classical conditioning* and *operational conditioning*. Classical conditioning is an automatic type of learning in which a stimulus evokes a response that was originally evoked by another stimulus (e.g. see http://www.learning-theories.com/). In opposite, in operational conditioning, a behavior is followed by either positive or negative reinforcement that changes the possibility that the behavior will happen again. According to these theories, learning is defined as the response to external stimuli (James, 2006). The consequences of such responses, namely rewards and penalties, are powerful tools to acquire or extinguish behavior (see for instance Siang & Rao, 2003, who gave an overview about the influence of behaviorist theories from a computer game perspective). Furthermore, it is also posited that the most effective way of learning is to deconstruct complex performance into small elements and to practice each of these elements separately. However, behaviorist theories do not consider inner states of learners but rather suggest that behavior is considered to be occasioned by stimuli from the environment. The learner is seen as a “Black box” whose motivation, creativity, thinking, remembering etc. are neither accessible nor necessary to explain behavior.

2.2.2 Constructivist Theories

Contrariwise to behaviorist theories, constructivist theories do not propose that our knowledge and behavior are determined by our environment but that we are actively create knowledge by building explanations of ourselves and our environments (see e.g. Anderson, 2009, for a review). There is an interaction of knowledge building and sensory experience. Knowledge is created actively by comparing new information with already existing information stored in memory. Accordingly, J. Piaget (see Atherton, 2010) proposed that *assimilation* and *accommodation* are two important processes in knowledge building. Assimilation is a process due to which new knowledge is merged to the network of already existing knowledge. Accommodation is a process due to which already existing knowledge is restructured when it is no longer compatible with the new knowledge. The construction of knowledge is facilitated when multiple channels of communications are used (Verhoeven, Schnotz, & Paas, 2009). Moreover, constructivist theories propose that social dialogues convey knowledge building (see also Vygotsky, 1996). Therefore, the importance of other persons regarding the learning of the individual is emphasized in these theories.

2.2.3 Cognitive Theories

First cognitive theories of learning arose in the 1960, often in reaction to behaviorist theories (James, 2006). According to these theories, learning requires the active engagement of learners. The focus lies on the question how people construct meaning, organize structures and concepts, and hence, form mental models of the world. Understanding and problem solving as much as inductive and deductive reasoning are proposed to be important factors in acquiring new knowledge. The differences between novices and experts are due to the way experts are organizing their knowledge. Hence, learners should be supported in acquiring “expert” understanding of structures and strategies to solve problems effectively. In this review, we will discuss the Cognitive Load Theory by John Sweller (see e.g., Kalyuga, 2009), Theory of Multimedia Learning (Mayer & Moreno, 2000) and the Experiential Learning Theory of Kolb (1984). Note that these theories are not independent but should rather be seen as complements of each other. Furthermore, we will also review a more recent theory of neuro-cognition (Anderson, 2009) that synthesizes different approaches from cognitive sciences, learning theory and neurophysiology.

Cognitive Load Theory

Cognitive load theory (CLT) was first described by John Sweller and focuses on the limitations of processing of our cognitive system (see Kalyuga, 2009). It is proposed that there are two key components of our cognitive system: The first key component is a large and permanent store which is related to long-term memory (LTM) and in which information is organized. The second component is related to working memory (WM). The main attributes of WM are that it has a limited capacity and that it is time-limited (e.g., Baddeley, 2005). Moreover, while capacity of working memory was suggested to be seven ± 2 items for a long time (Miller, 1956), recent research suggests a capacity of about four items (Cowan, 2001). Because of these limitations of WM, new information is relatively slow transferred into long-term memory. Knowledge structures stored in LTM, however, are essential to prevent working memory from being overloaded.

According to CLT, learning is hindered when the provided instructional materials overwhelm the limited resources of the learner (e.g., Artino, 2008). Therefore, three types of contributions to cognitive load can be distinguished (de Jong, 2010): (a) *intrinsic cognitive load* caused by the difficulty of the content to be learned, (b) *extraneous cognitive load* caused by the instructional material, and (c) *germane cognitive load* caused by the learning process. As other theories, CLT does not consider motivational aspects, however. But note that Paas, Tuovinen, Merriënboer, and Darabi (2005) suggested that mental effort and performance, as important constructs in CLT, have both cognitive and motivational components. Recently, the assumptions of CLT became more and more important in context of e-learning. Van Merriënboer and Ayres (2005) stated that the shift from written materials to working on online learning task may increase cognitive load and thus, may overstrain learners that are not familiar with such tasks. To avoid high cognitive load that perhaps precludes efficient learning, the authors suggested that not all information should be presented at once but rather sequentially, specifically when element interactivity of the provided material is high and learners are novices (see also Clarke, Ayres, & Sweller, 2005).

Cognitive Theory of Multimedia Learning

Mayer and Moreno (2000; see also Moreno & Mayer, 2000) presented a cognitive theory of multimedia learning (CTML) which also considers assumptions from cognitive load theory. The model bases on the assumption that working memory includes independent capacity-limited auditory and visual working memories and that human have separate systems for representing verbal and non-verbal information. Learning therefore occurs in three steps: *selection*, *organization*, and *integration*. The learner first selects relevant information in each of the working memory stores, organizes it in each store into a coherent representation, and finally, integrates the representations to make connections between each of the stores and previous knowledge from long-term memory (see Figure 1).

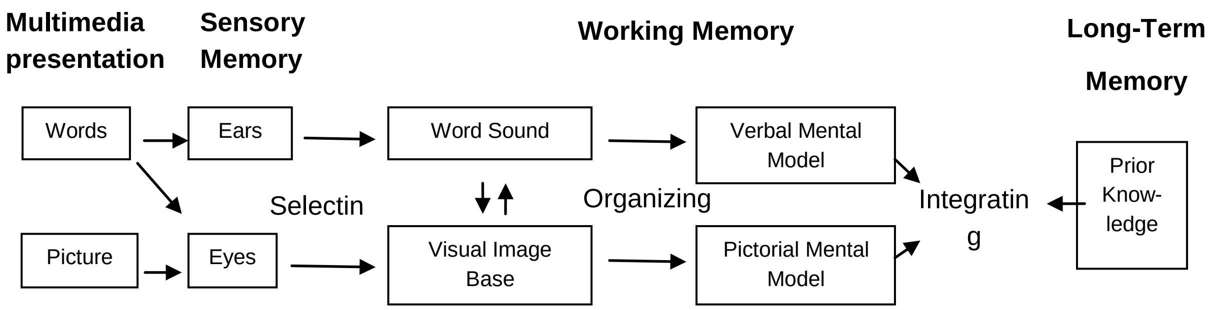


Figure 1. Cognitive theory of multimedia learning (adapted from Moreno & Mayer, 2000)

Based on CTML, Moreno and Mayer (2000, see also Mayer & Moreno, 2000) presented six principles of instructional design:

- Split-Attention Principle
- Modality Principle
- Redundancy Principle
- Spatial-Contiguity Principle
- Temporal-Contiguity Principle
- Coherence Principle

According to the split-attention principle, attention should not be split between multiple sources of mutually referring information. The modality principle suggests that verbal information should be presented auditory rather than visually. The redundancy principle suggests that redundant material should not be used as learners do not profit when they both hear and see the same verbal information. Following the spatial-contiguity principle, on-screen text and visual materials should be physically integrated, following the temporal-contiguity principle, verbal and visual materials should be temporally synchronized. Finally, the coherence principle suggests that extraneous material should be excluded in multimedia explanations.

Experiential Learning Theory

One further important theory in context of learning is the Experiential Learning theory (ELT; Kolb; 1984; see also Kolb & Kolb, 2005). From this theory, experiential learning is described as

"a process of constructing knowledge … [and] *this process is portrayed as an idealized learning cycle or spiral where the learner 'touches all the bases' – experiencing, reflecting, thinking, and acting – in a recursive process that is responsive to the learning situation and what is being learned"* (Kolb & Kolb, 2005; p. 2).

In this cycle of learning, concrete experiences lead to observations and reflections about those experiences (see Figure 2). The observations and reflections then lead to abstract forms of concepts and a moreover, to a generalization. These concepts and generalizations in turn are tested in new situations and finally are the bases for new experiences.

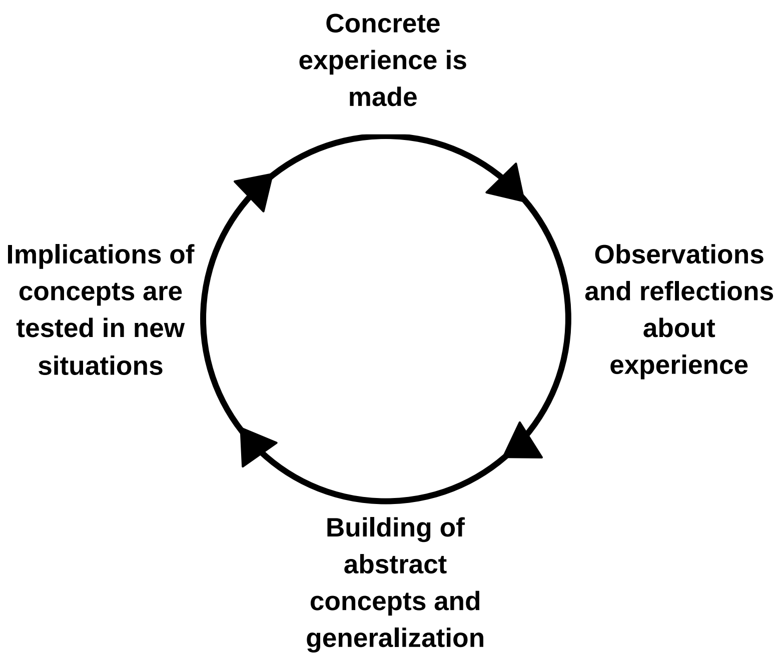


Figure 2.The experiential learning cycle (adapted from Kolb & Kolb, 2005).

Neurocognitive Theory of Learning

Recently, neurocognitive theories (NCT) of learning become more important. While behaviorist theories neglected the inner states of the mind (see above), NCT emphasize the importance of brain functions and internal processing in cognition. Moreover, neurocognitive theories usually base on or can be verified by findings of neurosciences. For instance, several neurobiological correlates have been found that support assumptions made in Anderson's neurocognitive learning theory (2009). According to this theory, the suggestions of three different areas are synthesized: Neurophysiology (focuses on brain and neural activities), Cognitive Science (focuses on information processing and internal representation), and Learning Theory (focuses on the interaction between the learner and his or her environment). Furthermore, also emotional aspects are considered in this theory.

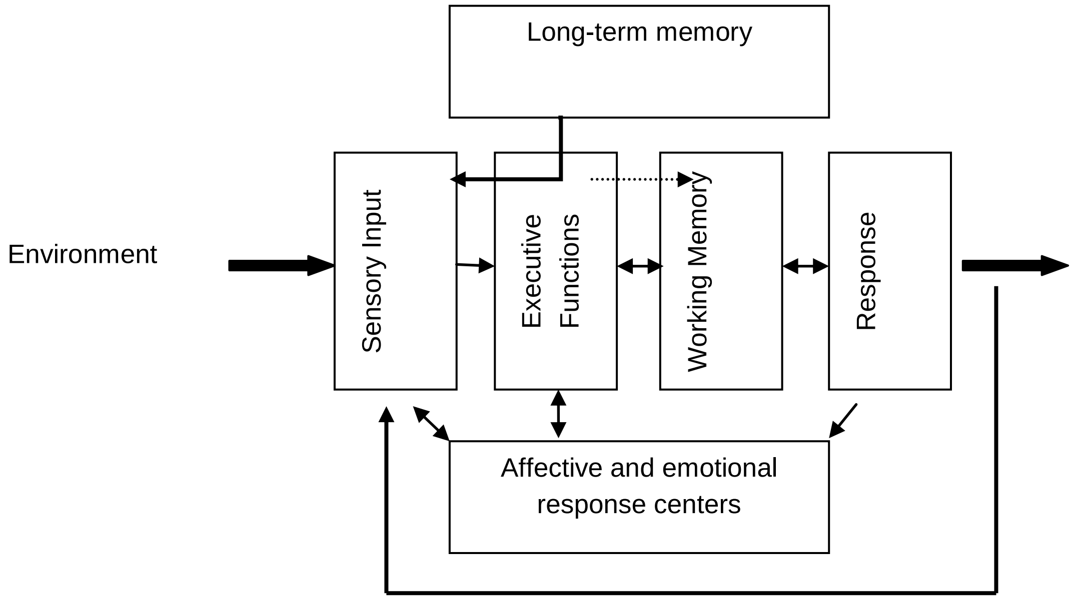


Figure 3. Neurocognitive model of information processing (adapted from Anderson, 2009)

Figure 3 shows the neurocognitive theory of Anderson (2009). Incoming sensory input is coordinated with existing knowledge in an executive functions module. This executive function module interacts with the working memory and centers of affective and emotional responses. In particular, information from prior memory is stored in working memory (dotted arrow line). In working memory, information is processed flexibly and response pattern are selected. The (usually motor) response changes the environment. Those changes are perceived by the individual through sensory feedback.

Summarizing the theories presented so far in this section, there is obviously a different viewpoint in the understanding of the role of the learner. Therefore, the role of the learner changed from a passive learner who simply learns to respond to a stimuli and whose inner states are neglected (behaviorism) to a rather active learner who explores his or her environment in order to acquire new skills and knowledge and integrate them to already existing knowledge (constructivism) though the cognitive resources are generally limited (cognitivism). However, a further aspect has to be considered at this point. As described before in context of constructivist theories, Vygotsky (1996) emphasized that social dialogues convey knowledge building. Therefore, learning often takes place in group setting; i.e., several persons work on the same problem and have the same learning objectives. Thus, in the next section we will introduce so-called collaborative learning before discussing social theories of learning. Note that collaborative learning will also be described more detailed later on in this document (see Chapter 4).

2.3 Collaborative learning

Dillenbourg (1999) defined collaborative learning as "a situation in which two or more people learn or attempt to learn something together" (p.1). Though such a definition of collaborative learning is quite intuitive, Dillenbourg mentioned that this definition is imprecise: Each of the elements of the definition can be interpreted differently. For instance, "two or more people" can be a small group up to five subjects as well as a class, a community or even a society of several thousands of people. Learning "together" can be interpreted as face-to-face learning or e-learning or synchronous or not synchronous. A further issue is that CL is neither one single mechanism which causes learning nor a method which ensures efficient learning. The former one means that someone may not simply learn from collaboration but rather from the extra activities involved (such as explanation, disagreement, etc.). The latter one suggests that the interactions observed in CL may have a low predictability: It is not guaranteed that some types of interactions will occur during CL because the underlying conditions like instructions, institutional constraints, etc., may vary. Some intervention to increase such predictability of interactions can be, for instance, specifying interaction rules (e.g., "everybody has to give his or her opinion, respectively") or to monitor and regulate the interactions (e.g. giving some hints).

Recently, computer-based or computer-supported collaborative learning (CSCL) became more and more important. There is evidence that computers support social interactions and cooperation (e.g., Hoyles, Healy, & Pozzi, 1992). Another benefit of CSCL is that learning can occur in distance. The enhancement in communication technology also supported a rapid growth of learning communities due to which learning occurs "through participation in communities of common purpose" (Kilpatrick, Barrett, & Jones, 2010, p.2). According to Kester et al. (2006), cooperation in such learning communities is influenced by the characteristics of its members. Thus, Kester et al described three profiles of members, namely *veterans* (as opposite to "newbies"), *trendsetters*, and *posters*. Veterans support and encourage peers. As opposite to "newbies", they show good community behavior by sharing knowledge and experiences. Trendsetters can either be "connectors", "mavens", or "salesmen". Connectors are sociable and rapidly make friends. Mavens are information experts and salesmen are persuaders. Finally, posters support and advance the enhancement of the community.

Moreover, for emerging a social space, Kester et al (2006) presented three factors that should be considered. First, there should be *continuity*, meaning that there should be a continuity of contact, a recognizability of members, and a historical record of actions. Second, a community should be *populated heterogeneously* with all types of members (as presented before) to ensure liveliness of the community. Third, *clear boundaries* and set of rules are required that can be monitored and sanctioned. Such boundaries facilitate cooperation. This is in line with Daradoumis, Xhafa, and Pérez (2006) who suggested that "CSCL technology and methods need guidance and support in order to collaborate effectively and achieve their learning goals successfully" (p. 278). Note that CSCL will be picked up again more detailed in Chapter 4.

In order to present the theoretical aspects of social learning, we discuss social theories of learning in the next section. According to these theories, social relationships are fundamental for learning. Learning occurs usually in an interaction between the individual and his or her social environment (James, 2006); it is a social and collaborative activity in which people develop their understanding together and in which cultural aspects such as language play a crucial role. Moreover, it is assumed that group knowledge generated from individual knowledge is greater than the sum of the knowledge of the individuals. With respect to teaching, social theories suggest that an environment has to be created in which students are involved in both generating the problems and the solutions. Problems should be solved collaboratively in order to enhance the individual skills of the students and their understanding.

2.3.1 Bandura's Social Learning Theory

According to Bandura's social learning theory (SLT, 1977), learning occurs due to interaction with others, i.e., in a social context. Behavior, but also attitudes, and emotional reactions are developed by observing, imitating, and modeling the behavior of other people. In particular, behavior is more likely to be acquired when the result of this behavior is desirable. Therefore, there are four processes that underlie social learning: *attention*, *retention*, *motor reproduction*, and *motivation* (e.g., http://www.learning-theories.com/). Therefore, one prerequisite of learning is that attention has to be paid to an object or task. Attention is varied by several factors like individual's characteristics (e.g. sensory capacities or arousal level) or complexity. Retention means that it is necessary to remember for what attention was paid. Reproduction means that the image has to be reproduced, and motivation means that there must be a good reason to imitate the image.

In their paper, Smith and Berge (2009) investigated the influence of SLT in *SecondLife* (http://secondlife.com/). *SecondLife* is a three-dimensional, virtual world where users are represented by avatars. Smith and Berge suggested that the proposed components of SLT (observing, imitating, and modeling) can also be observed in virtual worlds and that SecondLife is "a great example of social learning theory in action, although there are some components that cannot be satisfied in-world". For instance, it is not possible to observe attitudes and emotional actions in *SecondLife*.

2.3.2 Social Development Theory[3]

The major theme of Vygotsky's (1978) theoretical framework is that social interaction plays a fundamental role in the development of cognition. Vygotsky states: "Every function in the child's cultural development appears twice: first, on the social level, and later, on the individual level; first, between people (interpsychological) and then inside the child (intrapsychological). This applies equally to voluntary attention, to logical memory, and to the formation of concepts. All the higher functions originate as actual relationships between individuals."

A second aspect of Vygotsky's theory is the idea that the potential for cognitive development depends upon the "zone of proximal development" (ZPD): a level of development attained when children engage in social behaviour. Full development of the ZPD depends upon full social interaction. The range of skill that can be developed with adult guidance or peer collaboration exceeds what can be attained alone.

Vygotsky's theory was an attempt to explain consciousness as the end product of socialization. For example, in the learning of language, our first utterances with peers or adults, are for the purpose of communication, but once mastered, they become internalized and allow "inner speech". Vygotsky's theory is complementary to the work of Bandura on social learning and a key component of situated learning theory.

Scope/Application:

This is a general theory of cognitive development. Most of the original work was done in the context of language learning in children (Vygotsky, 1962), although later applications of the framework have been broader (Wertsch, 1985).

Example:

Vygotsky (1978) provides the example of pointing a finger. Initially, this behaviour begins as a meaningless grasping motion; however, as people react to the gesture, it becomes a movement that has meaning. In particular, the pointing gesture represents an interpersonal connection between individuals.

Principles:

[3] http://tip.psychology.org/vygotsky.html

(1) Cognitive development is limited to a certain range at any given age.
(2) Full cognitive development requires social interaction.

2.3.3 Co-operative vs. Collaborative Learning (Stahl, 2000)

Both cooperative and collaborative learning theories oppose the view that knowledge consists of facts told by teachers for students to repeat back. They may advocate a student-centered, constructivist approach in which students construct their own meaning using the ways in which they personally learn best. Social aspects of learning are considered theoretically important and the use of small group processes is emphasized in practice.
The difference may be defined in terms of the “unit of analysis.” Cooperative learning still privileges the teacher as the orchestrator of the educational process and still looks to the assessment of individual student knowledge as the sign of learning. Collaborative learning – for instance in versions like Lave and Wenger (1991) – analyzes things at the level of the community. Here, the teacher is just another participant within the changing roles of the community, and learning consists of evolution of the group and the abilities of its members to participate within it. The classroom may be reconceptualized as a knowledge-building community (Scardamalia & Bereiter, 1996) or a learning organization (Brown, Collins & Duguid, 1989), where the essential outcomes are measured at the group level not the individual. Thus, collaborative learning constitutes a distinct educational paradigm with a very different approach to defining and assessing learning.

Whereas cooperative learning is still measured by post-test evaluations of individual student learning based on teacher-defined goals, collaborative learning is concerned with evidence of social cognition (Crook, 1994, pp. 132f; Koschmann, 1996, p. 15). Social cognition may involve the creation of new socially-shared meanings, the increasingly skilled enactment of social practices by students, or the evolution of the learning community as such.

Given this distinction, one can see cooperative learning as a halfway stage to collaborative learning in the sense that the dissemination of the former provides an important basis for the implementation of the latter. Collaborative learning – whether supported by computer technology or not – must adopt many of the classroom practices of cooperative learning, such as its refined use of small group processes. The differences between individual, cooperative and collaborative learning theories are illustrated in Figure 4.

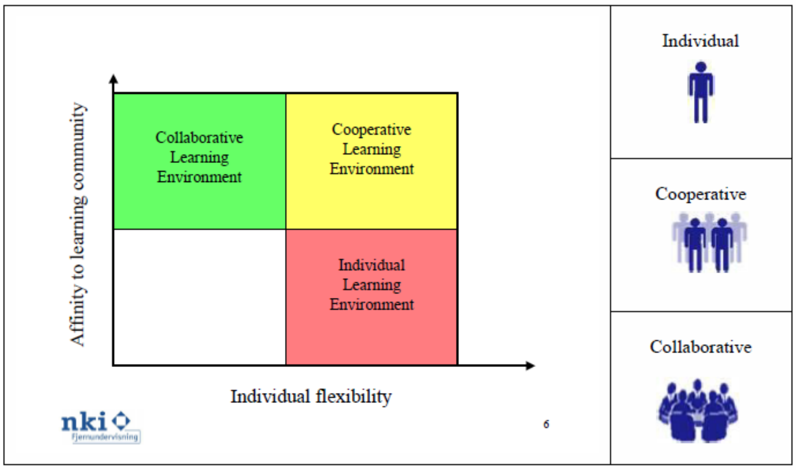


Figure 4. Individual, cooperative and collaborative learning theories (Paulsen, 2007)

2.3.3.1 Theories of collaborative learning

There are several theoretical approaches regarding collaborative learning (e.g., Dillenbourg, 1999; Daradoumis, Xhafa, & Marques, 2002). In his theory of collaborative learning, Dillenbourg mentioned that four items have to be considered with respect to collaborative learning, namely *situation*, *interactions*, *processes*, and *effects*. In particular, a collaborative situation is characterized by the symmetry of actions (i.e., the extent to which the same range of action is allowed to be shown by each subject), symmetry of knowledge or skills (i.e., the extent to which individuals possess the same level of knowledge or skill), and symmetry of status (i.e., the extent to which individuals have a similar status). The interactions characterized as collaborative are interactivity, synchrony in communication, and negotiation. Processes which can mainly be found in collaborative settings are for instance internalization (i.e., the transfer of group tools from the social level to an inner level), and mutual modeling (i.e., the comparing of individual knowledge with the knowledge of another person may lead to a discrepancy and thus, leads to awareness of one's own knowledge). However, the nature of interaction also changes with the context in which it occurs (Woo & Reeves, 2007). Interaction in face-to-face situation certainly differs from interaction at a distance because the latter one is mediated by a medium (Hillman, Willis, & Gunawardena, 1994).

2.4 Motivation and learning

Motivation is central to learning but also central in the development of learning. For instance, Garris, Ahlers, and Driskell (2002) suggested that motivated learners are enthusiastic, focused, engaged, and interested in what they are doing. Theories of motivation are designed to explain why people behave in a particular way. Historically, mechanistic theories dominated the field, viewing humans as passive and driven to act by biological disequilibrium toward homeostatic restoration. Nevertheless, compared to cognitive aspects, it has not received much attention in research in the past. For instance, from a behaviorist viewpoint, motivation does simply not exist although there might be an interconnection between motivation and reinforcement.

However, by the 1960s, with the onset of the cognitive revolution, theories now viewed humans as active explorers with goals, interests, perceptions, values and choices--all of which play a role in understanding behaviour. Theories of intrinsic motivation developed during this time and focused specifically on understanding why people do activities for their own sake rather than for instrumental reasons.

In general, motivation is defined as a psychological process in which, based on the needs of an individual, behavior is directed towards a goal (Maslow, 1970). Typically, we differentiate between intrinsic and extrinsic motivation. Factors which make a task intrinsically motivating are for instance challenge, curiosity, and fantasy (Malone, 1981). Extrinsic factors could be for instance grading, money or the chance to get a job. It is suggested that extrinsic rewards can be less effective than intrinsic behaviors (Garris, Ahlers, & Driskell, 2002) though both are important for learning. According to the self-determination theory (Deci & Ryan, 1975), people have three innate psychological needs: a need for competence, autonomy and relatedness. Intrinsic motivation develops out of these needs. When people feel competent (challenged and able to conquer challenge), autonomous (free to set goals and choose behaviours) and self-determined (internal locus of causality), they will freely seek what interests them.

Current research continues to investigate the conditions that support and undermine intrinsic motivation and the consequences that follow. Most of the new approaches to motivation developed from these theories – especially the ones from constructivist theories (i.e. self-efficacy, self-regulation, meta-cognition etc.) were explored in classroom context – both to verify theories and to build interventions to increase motivation.

According to Ford (1992), human have six goals that represent the consequences to be achieved with respect to motivation:

1. *Affective goals* including entertainment, happiness and physical well-being goals
2. *Cognitive Goals* including exploration, understanding, and creativity
3. *Subjective organization goals* such as unity
4. *Self-assertive social relationship goals* including individuality and self-determination
5. *Integrative social relationship goals* such as social responsibility, equity and belongingness
6. *Task goals* including mastery, management and safety goals.

2.4.1 Maslow's Hierarchy of Needs

Maslow's Hierarchy of Needs (often represented as a pyramid with five levels of needs) is a motivational theory in psychology that argues that while people aim to meet basic needs, they seek to meet successively higher needs in the form of a hierarchy (http://www.learning-theories.com/maslows-hierarchy-of-needs.html).

Maslow's Hierarchy of Needs has often been represented in a hierarchial pyramid with five levels. The four levels (lower-order needs) are considered *physiological needs,* while the top level is considered *growth needs*. The lower level needs need to be satisfied before higher-order needs can influence behaviour. The levels are as follows.

- *Self-actualization* – morality, creativity, problem solving, etc.
- *Esteem* – includes confidence, self-esteem, achievement, respect, etc.
- *Belongingness* – includes love, friendship, intimacy, family, etc.
- *Safety* – includes security of environment, employment, resources, health, property, etc.
- *Physiological* – includes air, food, water, sex, sleep, other factors towards homeostasis, etc.
-

2.4.2 Self-efficacy Theory

According to the theory of Bandura (1989), self-efficacy is important with respect to reaching goals successfully. Self-efficacy can be defined as the belief in one's own abilities. High self-efficacy supports individuals in coping with test situations. Low self-efficacy sometimes leads to *learned helplessness* (Seligman, 1975) which can be defined as a "low level of motivation, attributed to the belief that nothing [learners] could do will make a difference" (Svinicki, 1999, p. 21).

According to Bandura, perceived efficacy determines how much effort a person is willing to put into an activity as well as how long they will persevere in the face of obstacles. Many studies have documented the adaptive consequences of high self-efficacy. For example, it is known that high self-efficacy and improved performance result when students: (1) adopt short-term over long-term goals, inasmuch as progress is easier to judge in the former case; (2) are taught to use specific learning strategies, such as outlining and summarizing, both of which increase attention to the task; and (3) receive performance-contingent rewards as opposed to reinforcement for just engaging in a task, because only in the former case does reward signal task mastery. All these instructional manipulations are assumed to increase the belief that "I can do it," which then increases both effort and achievement. Efficacy beliefs have been related to the acquisition of new skills and to the performance of previously learned skills at a level of specificity not found in any other contemporary theory of motivation.

2.4.3 ARCS Model

One of the most influential models regarding motivation is the ARCS-Model of Keller (1983, 1987; see also Shellnut, 1998) which has also be shown to be relevant for e-learning. The model defines four motivational categories, namely *attention*, *relevancy*, *confidence*, and *satisfaction*, which represent the components of motivation (see Table 2). Attention includes perceptual arousal, inquiry arousal, and variability. Relevance includes goal orientation, motive matching, and familiarity. Confidence involves strategies to provide learning requirements, success, and to provide personal control over the learning. Finally, satisfaction includes strategies to increase the intrinsic reinforcement to provide extrinsic rewards and to assure equity of rewards. Furthermore, sets of strategies and a systematic design process are given in the ARCS-Model. The design process supports profiling the motivation of a person in a given environment.

Category	Subcategory	Process Tasks[a]
Attention	A1 Perceptual Arousal	Use novelty, surprise or uncertainty to gain interest
	A2 Inquiry Arousal	Stimulate interest by use of problem-solving and questioning
	A3 Variability	Provide a variety of methods and media for a change of pace
Relevance	R1 Goal Orientation	Present the objectives or the mean outcome of learning
	R2 Motive Matching	Match objectives to students' needs
	R3 Familiarity	Present contents understandably and related to already existing experiences
Confidence	C1 Learning Requirements	Provide learning requirements in form of clear objectives
	C2 Success Opportunities	Provide success opportunities for successful learning early and often enough
	C3 Personal Control	Provide personal control with choices of content, objectives, and activities.
Satisfaction	S1 Intrinsic Reinforcement	Enhance and support intrinsic enjoyment
	S2 Extrinsic Rewards	Provide positive consequences and motivational feedback
	S3 Equity	Provide rewards so that they match achievements

Table 2. Categories of the ARCS-Model of Keller (1983, 1987).

[a] *Note.* Adapted from Shellnut (1998)

However, because of the different characteristics, Chen and Jang (2010) noticed that motivational theories established in context of face-to-face learning cannot directly be transplanted to online learning. However, there is evidence that ARCS model is valid in context of e-learning (Keller & Suzuki, 2004). Lin and Gregor (2006) interviewed experts in order to find significant features that encourage students to learn from a museum's e-learning website. According to this, six suggestions for motivating in learning were found: *appearance*, *interactivity*, *ease of use*, *accessibility*, *simplicity*, and *partnerships*. Furthermore, five development guidelines for designing websites for enjoyment and learning were extracted:

- Adoption of multimedia and interactive technology
- Considering the characteristics of self-directed learning
- Have qualified staff and adequate financial support
- Consider the targeted audience
- Make the information more shareable

2.4.4 Attribution Theory

Another important theory of motivation is the Attribution Theory of Heider (1958). Attribution theory states that people try to understand what causes events and behaviours in the world by considering personal and environmental forces. It contends that personal causality comes from intentionality. It also states that there are many paths to achieve an intended behaviour, but that personal causality implies that the individual must set a goal and choose the specific path that he thinks will lead to its accomplishment (Deci, 1975).

Building on Heider's work, Weiner (1986) specified three dimensions of causality:

- Locus (internal/external),
- stability (stable/unstable),
- and controllability (controllable/uncontrollable).

The locus dimension influences pride and self-esteem experienced after an event if there is an internal attribution, the stability dimension influences future expectancies and the controllability dimension can influence future volition.

1. Attribution is a three stage process: (1) behaviour is observed, (2) behaviour is determined to be deliberate, and (3) behaviour is attributed to internal or external causes.
2. Achievement can be attributed to (1) effort, (2) ability, (3) level of task difficulty, or (4) luck.
3. Causal dimensions of behaviour are (1) locus of control, (2) stability, and (3) controllability.

Attribution theory has been used to explain the difference in motivation between high and low achievers. According to attribution theory, high achievers will approach rather than avoid tasks related to succeeding because they believe success is due to high ability and effort which they are confident of. Failure is thought to be caused by bad luck or a poor exam, i.e. not their fault. Thus, failure doesn't affect their self-esteem but success builds pride and confidence. On the other hand, low achievers avoid success-related chores because they tend to (a) doubt their ability and/or (b) assume success is related to luck or to "who you know" or to other factors beyond their control. Thus, even when successful, it isn't as rewarding to the low achiever because he/she doesn't feel responsible, i.e., it doesn't increase his/her pride and confidence.

2.4.5 Flow Theory

Flow theory was developed from Csikszentmihalyi's (1979) interest in how an intrinsically rewarding experience feels. From his research and interviews, he has concluded that pure intrinsically motivated behaviours involve enjoyment, complete immersion in the activity, detailed focus, feelings of competence and loss of conception of time. He stated that the enjoyment from the flow experience further motivates the individual to seek additional challenges (1988). This experience or 'flow' can only result from a situation where high challenges are matched with high skills. A skill/challenge imbalance leads to less than ideal emotional states: when challenge is higher than skill, anxiety will be experienced; when challenge is low and skills are high, boredom will result; when both skill and challenge are low, apathy will be experienced. Csikszentmihalyi goes on to say that although leisure activities typically lead to flow, any activity has the potential for creating it (i.e., people may increase challenge in more mundane tasks by trying to do them more creatively or more efficiently).

2.4.6 Effectance Motivation

Harter's effectance theory (1978) builds on White's (1959) ideas of a need for, and inherent enjoyment in, mastery. When mastery of challenging tasks is successful, the person experiences feelings of enjoyment, and internal rewards of competence and control. When he is unsuccessful, intrinsic motivation decreases, with a rise in need for external reinforcement to restore a sense of self and alleviate anxiety. Harter developed a self-report mastery scale to measure children's intrinsic versus extrinsic motivation toward learning and school mastery. It is divided into 5 subscales with intrinsic/extrinsic poles: learning motivated by curiosity verses learning in order to please the teacher, incentive to work for its own satisfaction verses working to please the teacher and to get good grades, preference for challenging work verses preference for easy work, desire to work independently verses dependence on the teacher for help, and internal criteria for success verses external criteria.

2.5 Emotion and learning

Just as motivation, also emotional aspects of learning have been neglected for a long time (e.g. Pekrun, Elliot & Maier, 2006; Kay & Loverock, 2008). Emotions have wittingly neglected from scientific research, and there is one significant reason for that; "everyone knows what emotion is until they are asked to define it" (LeDoux, 1999). Emotions are believed to be inherently non-scientific and susceptible to the risk to be focused on subjective emotional experience (Picard, 1997). When Cognitive Psychology was established by Ulric Neisser's homonymous book (1967), emphasis began shifting from the construction of meaning to the processing of information (Davou, 2000). Human mind could be understood and studied as complex information processing system (Bruner, 1990), leaving Behavioural theory (Watson, Thorndike, Skinner) behind. Since then, Cognitive Psychology has dominated for more than three decades, endowed by the rapid evolution that was attained in Computer Science and Neuroscience. This trend has been acclaimed as the 'cognitive revolution' in psychology. Within the framework of mind dominance, emotions were one particular form of cognitive processing that was determined by cognition. This approach to emotions however has been proven inefficient to explain many emotional phenomena of everyday life (Davou, 2000).

Psychology turned to Cognitive Psychology, ignoring creative thinking and consciousness, and Cognitive Psychology has been dominated by Cognitivism, leaving human emotions out of question. By keeping terms like emotion and subliminal processing in the background, Cognitive Psychology reached its deadlocks (Costall, & Still, 1987). As Norman (1980) has acknowledged, "the organism we are analysing is conceived as pure intellect, communicating with one another in logical dialogue, perceiving, remembering, thinking when appropriate, and reasoning its way through well formed problems that are encountered during the day. Alas that description does not fit actual behaviour".

One reason for this might be the difficulty to measure emotional states of a person. Providing questionnaires, for instance, usually allows testing the emotions of a person before or after the event and often leads to biases with respect to the individual's responses: A person might give rather desired answers or perhaps might be not able to reflect or express his or her sensation anyway. It is also possible to measure emotions through physiological measurements like heart rate, respiration or skin conductance measurement (see e.g., Janssen & van den Broek, 2009, for a review). Although those measures may reflect emotions more objectively than questionnaires, their appliance outside of the laboratory is often simply not feasible. Moreover, from a methodological viewpoint, the appliance of physiological tools itself might affect the resulting measurement: Putting on some physiological sensors might change the emotional state so that the measurements are noisy. Recent neurological studies, however, have proven the essential role of human's emotional centers not only in perception and learning, but also in decision making (Damasio, 1996). Minds are not either cognitive or emotional; they are both and more (Le Doux, 1999). Emotions are present in any form of education: learners worry, hope, become bored, embarrassed, envy, get anxious, feel proud, and become frustrated, and so on (Ochs & Frasson, 2004).

In general, emotions are essential for learning and memory. Hascher (2010) stated that there is "rarely any learning process without emotion" (p.13). However, there is a general problem the term "emotion" has to be defined (e.g., Cabanac, 2002). For instance, Kleinginna and Kleinginna (1981; cited after Cabanac) 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). Davou (2007) offers an enlightening discrimination:

1. *Emotion* (derives from the Latin prefix emot=moving away) refers to a "shaking" of the organism as a response to a particular stimulus (person, situation or event), which is generalized and occupies the person as a whole.
2. *Affect* is the effect of emotion in the organism; a synthesis of all likely effects of emotion (cognitive, organic, etc) and includes their dynamic interaction, but is not evened individually with any of them.
3. *Feeling* is always experienced in relation to a particular object of which the person is aware. It may have various levels of intensity, and its duration depends on the length of time that the representation of the object remains active in the mind of the individual.

In her review, Hascher (2010) described three widely accepted characteristics of emotions: First, an emotion is an affective reaction. It can be determined and described relatively precisely and can be attributed to a cause. Second, the experience of an emotion is related to for situations that are important for a person. Third, when an emotion is experienced, it becomes the centre of the individual's awareness.

Affective elements in the analysis of the interaction with the student, has become an increasingly prominent theme in recent years. This is due to a clear evidence of correlation between affect and learningThe 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. They are learning theories that although acknowledge 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'.

2.5.1 The emotion circuitry

The neurological evidence indicates emotions are not a luxury; they are essential for rational human performance (Picard, 1997). Advances in Neurobiology, unfolded the impact of the Emotional Brain (Limbic System) in reasoning, learning and behavior.

Whenever an external stimulus is perceived by a sensory modality, it travels inside the human neuro-network by triggering specific neurotransmitters, following the pathway of the human brain evolution, before being redistributed to the cortex for analysis. On this pathway, Goleman (1995) discriminate the following important brain parts that play a less or more crucial role in the stimulus' decryption process:

The old/emotional brain that featured by:

- The Brain Stem or Primitive Brain (MacLean's Reptilian Brain) that regulates Cardiac and respiratory function.
- The limbic system (Hypothalamus, Amygdala, Hippocampus), the seat of emotion memory and attention.

The thinking/cognitive brain (Neocortex or Cortex) that is involved in higher functions, such as

- sensory perception
- generation of motor commands
- spatial reasoning
- conscious thought
- language

In other words, there are two brains: the emotional and the cognitive, residing in different locations in the cerebrum. The above three distinct brain areas (Brain Stem, Limbic, Neocortex) emerged successively in the course of evolution and now co-inhabit the human skull (Maclean, 1990). The basic circuitry that aims to satisfy the primary human needs (survival, feed, protection from natural threats, breeding, sociality, etc.) was named as Regulatory System (Norman, 1980). This system's requirements are the roots of the Cognitive System that is going to be developed, later on. In other words, the Cognitive System tents to serve the Regulatory System, and not the opposite, as commonly believed by humans.

2.5.2 Educational Perspectives

By decrypting what lies beneath the emotion and affect expression, Cognitive Sciences provided evidence for the claim that emotion, together with cognition and motivation are the key components of learning (D'Mello et al., 2005). It is important to help students know how and when their "emotional intelligence" works to help or hinder their success. **Expert teachers** are very adept at recognizing and addressing the emotional states of students and, based upon impressions, taking some action that positively impacts learning. But what these expert teachers see, and how they decide on a course of action is an open question (Kort & Reilly, 2002).

Daniel Goleman (1995) suggests that a student with a positive disposition would see an F on a math test as evidence that he needs to work harder, while another may see it as evidence that he is stupid. When negative emotions create a pessimistic perceptual attitude they divert the learner's attention to aspects irrelevant to the task which activate intrusive thoughts that give priority to a concern for a well-being rather than for learning (Boekaerts, 1993).

Howard H. Gardner (2006) has referred to the 'personal intelligences', and more specifically:

Interpersonal intelligence is concerned with the capacity to understand the intentions, motivations and desires of other people. It allows people to work effectively with others. Educators, salespeople, religious and political leaders and counsellors all need a well-developed interpersonal intelligence.

Intrapersonal intelligence entails the capacity to understand oneself, to appreciate one's feelings, fears and motivations. In Howard Gardner's view it involves having an effective working model of ourselves, and to be able to use such information to regulate our lives.

2.5.3 Emotional Intelligence (EQ)

The early Emotional Intelligence theory was originally developed during the 1970s and 80s by the work and writings of psychologists Howard Gardner, Peter Salovey and John Mayer. According to Salovey and Mayer (1990), Emotional Intelligence is:

- “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”

Pugh (2008) has given a more simple definition of emotional intelligence “Tuning into emotions and taking appropriate action”.

Daniel Goleman (1995) presents convincing evidence that the emotional intelligence quotient (EQ) is just as important in academic success as cognitive intelligence, as measured by IQ or SAT scores. Salovey and Mayer (1990) combined the work of several researchers to define the following measures of effective use of emotion. He used five attitudes to determine Emotional Intelligence:

1. Knowing one’s emotions: Self-awareness
2. Managing emotions: Handling feelings
3. Motivating Oneself: Self-motivation
4. Recognizing emotions in others: Empathy
5. Handling Relationships: The art of relationships

In his homonymous book, Goleman (1995) illustrates a variety of self-help programs (Transaction Analysis, Self Science) that have been developed to assist people in gaining control of their emotions. This is different from suppressing the emotions. Rather, the goal is to get in touch with feelings, to know what causes them and to take appropriate action in response to them (Culver, 1998).

2.5.4 Flow

Most of us have had that "involved" moment happen, when we concentrated our attention so intensely on solving a problem, reading a book, climbing a mountain, on some task, that we lost track of time and when we became aware of our surroundings, a few hours or more had passed by as if they were minutes. Such "flow", according to Csikszentmihalyi (1990) is "optimal experience" that leads to happiness and creativity.

Flow is the state in which people are so involved in an activity that nothing else seems to matter; the experience is so enjoyable that people will do it even at great cost, for the sheer sake of doing it (see Figure 5). It is a single-minded immersion and represents perhaps the ultimate in harnessing the emotions in the service of performing and learning. In flow the emotions are not just contained and channelled, but positive, energized, and aligned with the task at hand (Goleman, 1995). Since 70s, William O Perry (1968) has said that people learn best in their pleasure zone – between panic and boredom.

If a task is not challenging enough, boredom sets in, while too great a challenge results in anxiety, and both cases result in task, and thus learning, avoidance. As one's skills increase, then the challenge must also increase for one to remain in a state of flow. Because flow is an enjoyable experience, one continues to increase the challenge level (as from A1 to A4 and so on), and consequently continues to improve one's skills because doing so is necessary to stay in a flow state. A learning environment in which students are challenged at an appropriate level, which can produce flow, will be more productive.

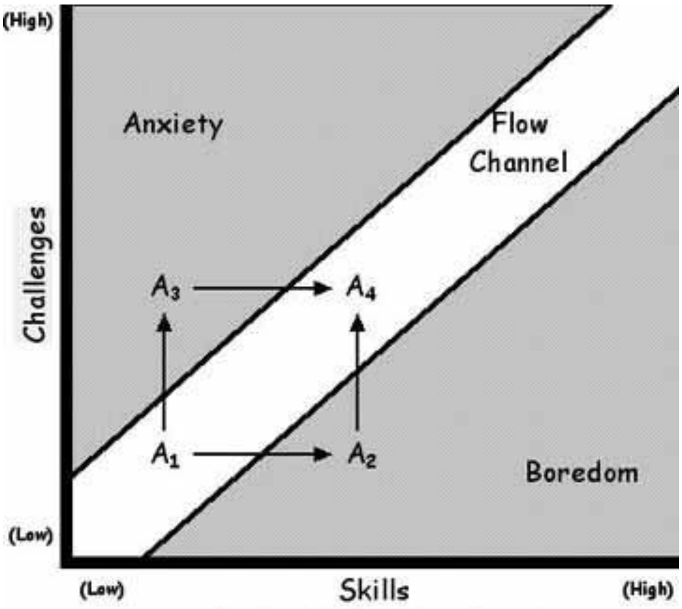


Figure 5. Flow: The Psychology of optimal experience (from secondlanguagewriting.com/explorations/Archives/2007/January.html)

2.5.5 The Learning Cycle

Kort and Reilly (2002; see Figure 6), after accomplishing preliminary pilot studies with elementary school children, offer a model of a learning cycle, which integrates affect, providing a framework for thinking and posing questions about the role of emotions in learning. They suggest six possible emotion axes that may arise in the course of learning together with a Four Quadrant model relating phases of learning to emotions.

Axis	-1.0	-0.5		0	+0.5	+1.0
Anxiety-Confidence	Anxiety	Worry	Discomfort	Comfort	Hopefulness	Confidence
Ennui-Fascination	Ennui	Boredom	Indifference	Interest	Curiosity	Fascination
Frustration-Euphoria	Frustration	Puzzlement	Confusion	Insight	Enlightenment	Euphoria
Dispirited-Enthusiasm	Dispirited	Disappointed	Dissatisfied	Satisfied	Thrilled	Enthusiasm
Terror-Excitement	Terror	Dread	Apprehension	Calm	Anticipatory	Excitement
Humiliated-Proud	Humiliated	Embarrassed	Self-conscious	Pleased	Satisfied	Proud

Figure 6. Emotion sets possibly relevant to learning (after Kort & Reilly, 2002)

A typical learning experience involves a range of emotions, cycling the student around the four quadrant cognitive-emotive space as they learn (see Figure 7). 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.We do not foresee trying to keep students in Quadrant I, but rather to help them see that the cyclic nature is natural in learning science, mathematics, engineering or technology (SMET), and that when he lands in the negative half, it is an inevitable part of the cycle. Our aim is to help students to keep orbiting the loop, teaching them to propel themselves, especially after a setback.

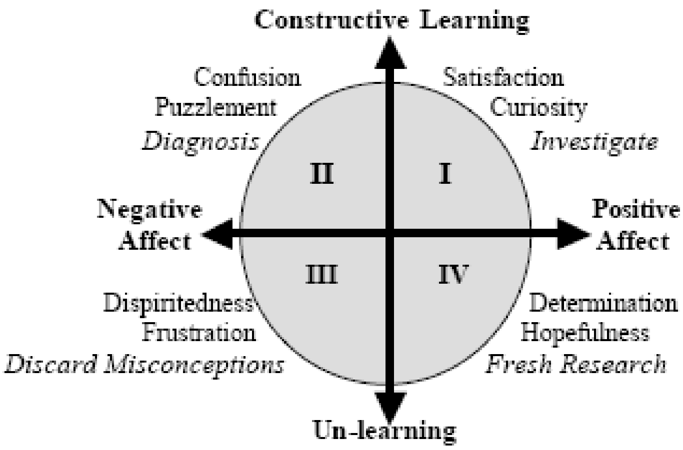


Figure 7. Four Quadrant model relating phases of learning to emotions (after Kort & Reilly, 2002)

2.5.6 Theories of emotion

Emotional scaffolding can improve the state-of-the-art, at least when provided by humans (Aist, Kort, Reilly, Mostow, & Picard, 2002), but still remains unexplored the way emotions can impact learning and vice versa. Which emotions are exhibited during learning? Is it possible to induce emotions to engage the student in a challenging and authentic learning experience and in which way? How does the learner manage frustration when learning is difficult? (D'Mello et al., 2005)

Pekrun et al. (2006) noticed that positive affect is positively related to mastery goals, and that negative affect (i.e., test anxiety) is positively related to performance-avoidance goals. However, there is some evidence that positive emotions do not necessarily have to foster learning whereas negative emotions lead to worse learning results (see Hascher, 2010). Therefore, Hascher argued that the valence of an emotion is only one aspect of its quality. She described eight factors that should be taken into account to analyze the quality of an emotion:

1. *Valence* (pleasant - unpleasant)
2. *Arousal level* (activating - deactivating)
3. *Intensity* (intense - low)
4. *Duration* (short - long)
5. *Frequency of occurrence* (seldom - frequent)
6. *Time dimension* (retrospective, actual, prospective)
7. *Point of reference* (self-related; related to others; referring to an activity)
8. *Context* (during learning, achievement etc.)

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 2 × 2 (or 3, respectively) taxonomy of achievement emotions, Pekrun et al. summarized their assumptions (see Table 3). 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. According to the taxonomy of Pekrun et al, Table 3 shows emotions with respect to achievement found for middle school, high school as well as university students.

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

Table 3. Taxonomy of achievement emotions proposed by Pekrun et al. (2006). See text for details.

2.5.2.1 Neuroanatomical Model of Emotion

LeDoux's (2000) systematic research underline the privileged position of Amygdala and Hippocampus; a point where everything converges (see Figure 8). Sensory signals go from the hypothalamus to the amygdala in 15 milliseconds and from the hypothalamus to the cortex in 25 milliseconds (LeDoux, 1999). As a result, the amygdala is creating emotional responses before the cortex has even received the signal to be processed.

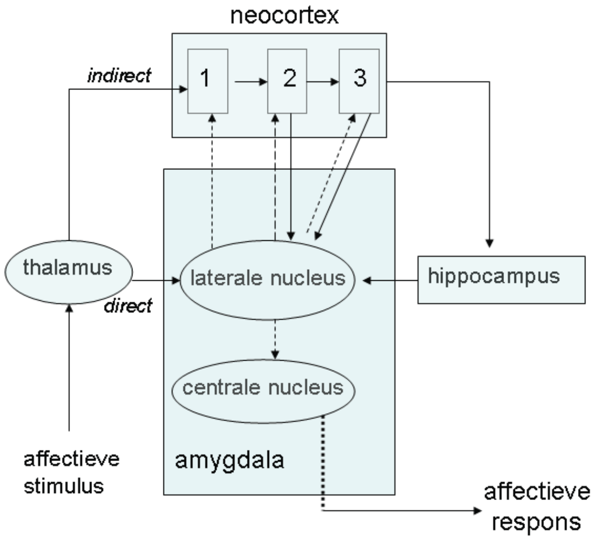


Figure 8 . Ledoux's Neuroanatomical Model of Emotion

However, the amygdala has limited pattern recognition capabilities compared to the cortex, and performs "a quick and dirty" pattern recognition and response. A stimulus is firstly, and above all, appraised if it is a threat. The amygdala has presumably been structured in answer to one critical question for survival: Do I eat it or does it eat me? Brain is able to sense fear before human can think of it *(D. Goleman)*.

Cytowic's studies (1994) point that all sensory inputs, external and visceral, must pass through the emotional limbic brain before being redistributed to the cortex for analysis, after which they return to the limbic system for a determination of whether the highly-transformed, multisensory input is salient or not. Damasio (1996) distinguishes between "primary" and "secondary" emotions. His idea is that there are certain features of stimuli in the world that we respond to emotionally first, and which activate a corresponding set of feelings (and cognitive state) secondarily. Such emotions are "primary" and reside in the limbic system. He defines "secondary" emotions as those that arise later in an individual's development when systematic connections are identified between primary emotions and categories of objects and situations. Paul Eckman (1997) has found that there are characteristic facial expressions to describe at least six basic emotions (i.e., fear, sorrow, anger, happiness, surprise and disgust), which have been found to be consistent in all cultures, including primitive ones with no access to the outside world.

2.5.6.1 The Circumplex Model of Affect

One further important model of emotions or affect, respectively, is the Circumplex Model of Affect (Larsen & Diener, 1992). The main idea of this model is that emotions are typically intercorrelated; i.e., when an individual is asked about his or her emotions he or she may have difficulties to describe them clearly. Rather, there might be some “mixture” of positive and negative emotions in a specific situation. As Posner, Russell, and Peterson, (2005) stated, those intercorrelations between emotions are rarely considered. In the model of Larsen and Diener, two dimensions are considered: *Activation* and *Pleasantness*. The affective states related to the two dimensions are ordered on the circumference of a circle (see Figure 9 for an example).

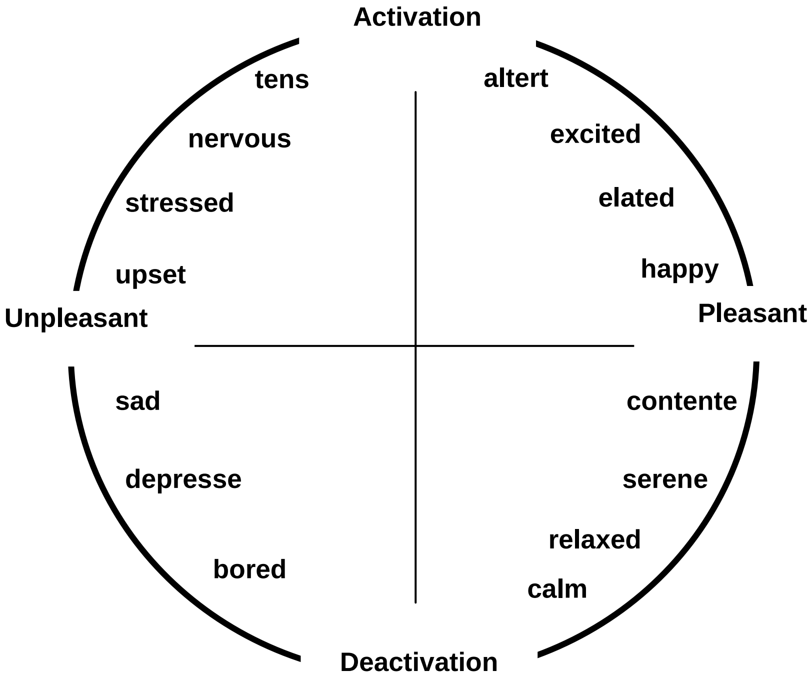


Figure 9. The circumplex model of affect. (adapted from Posner, Russell, & Peterson, 2005).

2.6 Learning styles

According to Wang, Wang, Wang, and Huang (2006), student learning is not only influenced by factors such as age, gender, and socioeconomic status but also by learners attributes such as affective expression (e.g., interest, motivation), learning experiences (e.g., mental models), and learning characteristics (e.g., cognitive or learning styles). From those factors, results of Wang et al. suggested that primarily learning styles have an important influence on e-learning.

One of the most commonly used instruments to measure different learning styles is Kolb’s Learning Style Inventory (LSI; see e.g., Kolb & Kolb, 2005). As stated before, Experiential Learning Theory

(ELT, Kolb, 1984) posits that the learning-process cycle can be divided into four learning modes: concrete experience, reflective observation, abstract conceptualization, and active experimentation. When the learning style of a person is specified with the LSI, a score for each of these modes is generated. With these scores, four different learning styles can be distinguished: *Divergers*, *Assimilators*, *Convergers*, and *Accommodators* (Figure 10).
Divergers prefer both the concrete experience and the reflective observations. Their strength lies in generating ideas and viewing concrete situations from different points of view. Assimilators prefer a combination of reflective observation and abstract conceptualization. Thus, they are more interested in abstract concepts than in people and are best at putting information into logical form. Convergers prefer abstract conceptualization and active experimentation. Their strength is to find practical uses for ideas and theories. Finally, accommodators prefer combinations of concrete experience and active experimentation. They often rely on information provided by others than by the result of their own analysis.

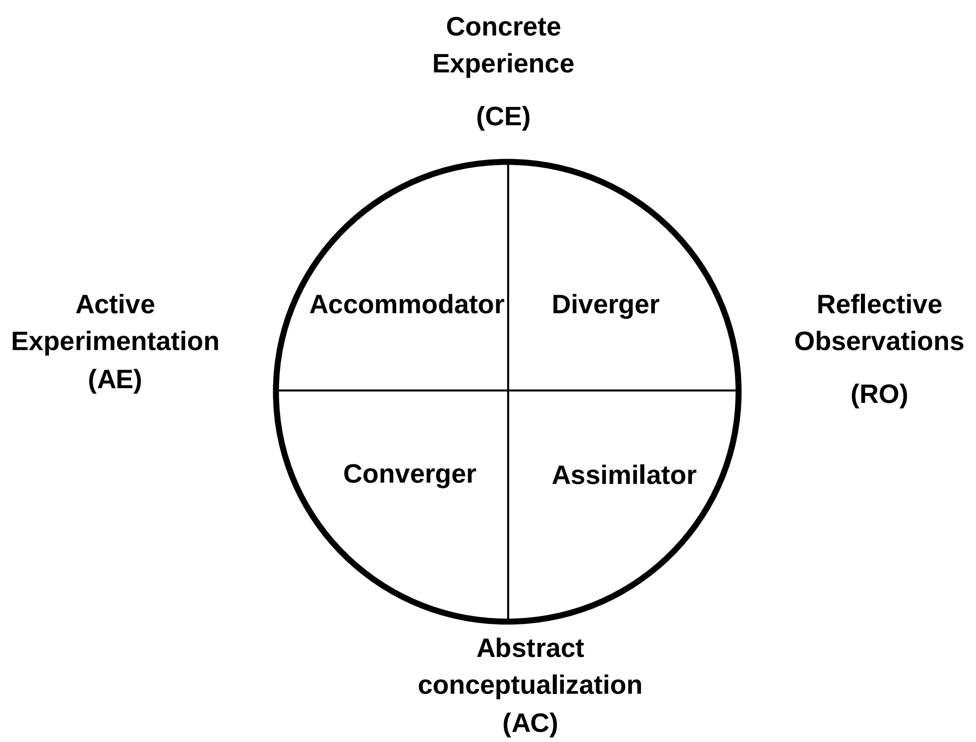


Figure 10. The relationship between experiential learning theory and learning styles (adapted from Kolb, 1984)

As proposed by the ELT, learning styles are rather a social-psychological construct than personality-based. Accordingly, learning styles are affected by educational specialization, career, job, and task skills (see Table 4 for a summarization how learning styles are determined at these levels).

Conole, de Laat, Dillon, and Darby (2008) were interested in how learners engage with e-learning and experience e-learning and furthermore, how e-learning does relate to learning experience. They suggested that there is a complex interrelationship between individuals and tools. They described eight factors that should be considered.

1. *Pervasive*: Students use technologies to support all aspects of their study; they share resources with peers.
2. *Personalized*: Students appropriate technologies to suit their needs; the learning is interactive, and students multitask.
3. *Niche, adaptive*: Students use technologies for particular purposes, not just for the sake of

using them.

4. *Organized*: Students search for and structure information.
5. *Transferable*: Students can apply skills gained through using technologies in other aspects to their learning context.
6. *Time and space boundaries*: Students can communicate with tutors and peers in a variety of ways which leads to a changing concept of "time" and "space" regarding learning.
7. *Changing working patterns*: new working practices are emerging that use a range of tools
8. *Integrated*: Students use tools in a combination of ways, they mix and match. This flexibility in using technologies enables them undertake learning anytime and anywhere.

Conole et al. (2008) concluded that students use technologies to support all aspects of their learning processes and that technology is central to how to they organize and orientate their learning. These findings have implications for how institutions provide technical infrastructure and the ways of how students should be supported during their learning process.

Behavior Level	Diverger	Assimilator	Converger	Accommodator
Personality types	Introverted Feeling	Introverted Intuition	Extraverted Thinking	Extraverted Sensation
Educational Specialization	Arts, History Psychology	Mathematics Physical Science	Engineering Medicine	Education Communication Nursing
Career	Social service Arts	Sciences Research, Information	Engineering Medicine Technology	Sales Social service Education
Current jobs	Personal jobs	Information jobs	Technical jobs	Executive jobs
Task skills	Valuing skills	Thinking skills	Decision skills	Action skills

Table 4. The relationship between learning styles and five levels of behavior (adapted from Kolb & Kolb, 2005)

2.7 Summary of the Chapter

The aim of Chapter 2 was to introduce and define general aspects of learning. In particular, we gave a definition of learning and collaborative learning and described recent models and frameworks. Moreover, we discussed basic theories like behaviorism, constructivism, and cognitive theories in order to cover the development in understanding the learner and the learning process. In two sections, we also considered motivational and emotional aspects of learning that are often neglected in research. Again, basic and important theories were presented regarding those aspects. Finally, we also addressed the question of different learning styles, considering the individual differences in learning.

3 Assessment

Assessment is an important factor in our everyday life. For instance, we are assessing the quality of the meal in the new restaurant around the corner, the performance of our favorite soccer team or the work of the mechanic. Likewise, our performance in our job is assessed by our colleagues and our supervisor. One field in our society in which the support and improvement of performance, knowledge, and skills, and hence, assessment plays a crucial role, is learning in educational setting. The assessment of students' knowledge, skills or performance is a crucial part in educational settings for more than 150 years (Buzzetto-More & Alade, 2006). Learners from pupils to graduates are regularly assessed to demonstrate and further develop their learned competences and abilities.

In this chapter we will give a broad overview and definition of assessment in general. The main focus lies on introducing the basic ideas of assessment that will be picked up again later on in this review. Furthermore, in order to form an extensive impression of assessment, we will also link aspects of assessment to the theoretical theories presented in Chapter 1. However, from the examples given before, it is clear that assessment is used in various settings and dimensions. To meet the needs of the ALICE project, we will only discuss assessment with respect to educational settings only although some of the definitions and descriptions could also be generalized in order to cover other research areas.

3.1 Definition and purpose of assessment

From a traditional viewpoint, the primer goal of assessment is to grade students' performance at the end of a course. However, from a more value viewpoint the goal of assessment is to support and improve one's performance, knowledge or skills; i.e., in this case, the progress on learning itself is rather taken into account. As Rovai (2000) stated, "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." (p.142). In their review, Black and Wiliam (1998) defined assessment "as encompassing all those activities undertaken by instructors and/or their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (p2). In this way, assessment cannot be seen as static but can rather be described as an ongoing process (Buzzetto-More & Alade, 2006).

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

Accordingly, assessment does not only affect students' learning processes but also instructors' behavior, teaching strategies or the effectiveness of the underlying curricular program (see also McAlpine, 2002). Furthermore, it has also to be taken into account that assessment is also affected by interactions between various aspects. For instance, it is not reasonable to assess the learning outcome when the used assessment itself is not effective. Furthermore, assessment might be also more valuable when the student gets valuable feedback afterwards (see Section 3.2).

In recent years, traditional face-to-face learning has been supplemented and extended by computer-based or e-learning activities. E-learning activities such as web-based learning, game based learning, or simulations, offer the possibility for the learners to learn what they want when they want. These changing educational goals and learning methods necessitate developing new forms of assessment. Note at this point that assessment in context of e-learning is called e-assessment.

However, instructors are often faced with the problem of how to measure learning processes and learning outcomes of their students in an adequate and valid way. Providing adequate types of assessment which meet all demands made by instructors, learners, and other involved persons is both a challenge and a chance for different research areas such as Computer science, Pedagogy, and Psychology. This challenge has been also increased due to the fact that the goals of education are changing (Dochy, 2001; see also Chapter 1). While in the past the focus primarily lay on the acquisition of basic knowledge within a domain, nowadays further aspects have become more and more important. Therefore, Dochy and McDowell (1997) suggested that assessment has to consider (a) *cognitive competencies* such as problem solving or critical thinking, (b) *meta-cognitive competencies* such as self-reflection or self-evaluation, (c) *social competencies* such as group working or communication skills, and (d) *affective dispositions* such as flexibility or independency. Note however that the aspect of motivation is not considered in their approach although motivation obviously affects learning and assessment. We will address this issue later on.

Dochy and McDowell (1997) stated that there is a change from 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. This agrees with Martell and Calderon (2005) who proposed that the assessment process already begins with the identification of learning goals and measurable objectives. Moreover, Sluijsmans, Prins, and Martens (2006) 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” (p.46; see also Dochy & McDowell, 1997; Dochy, 2001; Wang et al., 2006).

Furthermore, Wiggins (1990) suggested that assessment should be provided in a way 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. These methods will be discussed in Chapter 4 more detailed.

3.1.1 Referencing of assessment

One important issue when discussing assessment activities is the referencing of an assessment; i.e., the question regarding to which criterion the assessment takes place. McAlpine (2002) proposed three ways of referencing: (a) In *norm-related referencing* approaches the performance of a person is compared to the performance of his or her peers. Though frequently used, this approach gives little information about the actual abilities of the individual, however. (b) In *criterion-referencing* approaches, a comparison between the individual and a pre-defined domain takes place. This type of referencing allows disclosing lacks in knowledge and monitoring whether a specific level of performance has been achieved. (c) Finally, *ipsative referencing* can be defined as the comparison of a subject with himself or herself. Such approaches can be helpful when the main goal of assessment is to monitor the individual progress.

3.1.2 Types of assessment

In general, two types of assessment are distinguished: *summative* and *formative* assessment. Summative assessment occurs after the learning period has ended and summarizes the performance shown at the time of testing. This stands in contrast to formative assessment due to which the enhancement in performance can be diagnosed. Formative assessments can include short tests, quizzes, discussions, homework etc. As described by Boston (2002), formative assessment provides an appropriate opportunity for instructors and students to monitor and improve learning progress. Furthermore, it has also been shown that student anxiety is reduced when formative assessment rather than summative assessment is provided (Zakrzewski & Bull, 1998; cited after Wang et al., 2006). According to previous findings in context of assessment in general, results from Wang et al. suggested that the learning effect also depends on the variety of the used formative assessment strategies: The more diverse assessment strategies were used the better learning outcomes were found.

3.1.3 Strategies of assessment

Assessment is usually provided by the instructor and/or the tutor of the course. However, recent approaches of assessment increasingly tend to involve students in the process of assessment (e.g. Dochy & McDowell, 1997; Dancer & Kamvounias, 2005; Cano, 2011). As pointed out by McConnell (1995), learners will experience their studies qualitatively different when they are involved in decisions about how and what to learn compared to learners who are treated as recipients of teaching. This active involvement should also include the process of assessment: According to McConnell, learners should also be involved in decisions about criteria for assessment and the process of judging their own and others work. Usually, students can be involved in the assessment process in two ways, namely *self assessment* and *peer assessment*. There are two general benefits of such an involvement. From a students' viewpoint, important skills necessary for their learning progress can be developed. From an instructors' viewpoint the time investment in the assessment might decrease.

In this section we will provide a short introduction in the most important forms of assessment, namely *assessment by the instructor*, *self assessment*, and *peer assessment*. Furthermore, *group assessment* as a special case of peer assessment will also be discussed.

Assessment by instructors

Assessment by the instructor is the most traditional form of assessment. As "experts" in the learning domain, instructors usually define the objectives and standards of a course. In particular, they define the criteria to be fulfilled by the students to achieve the course objectives. However, assessment by instructors also has some disadvantages. In their review, Black and Wiliam (1998) summarized the weaknesses of instructor assessment as follows: First, instructors' evaluation practices usually concentrate on recall of isolated details rather than on the assessment of understanding and problem solving. Furthermore, instructors often do not critically review and reflect their assessment questions and over-emphasized the grading function of assessment. Finally, instructors tend to use normative rather than criterion referencing approaches (see Section 3.1.1). This leads to competition between students rather than to personal improvement of each student.

Dochy and McDowell (1997) suggested that an instructor should not simply transfer knowledge but should act as a mentor who supports students using their existing knowledge and skills in order to understand new topics. However, Black and Wiliam (2004) stated that it is very difficult to change inappropriate (teaching and) assessment practices because these practices are usually embedded within instructors' whole part of pedagogy. Thus, when instructors' assessment is supplemented or replaced by other forms of assessment, it has to be considered that such an intervention also changes the role of the instructor and the ways of teaching. Nevertheless, as it will be discussed in the next sessions, involving students in the process of assessment helps students in developing essential skills for their own learning and hence, future.

Self assessment

As the name suggests, self assessment allows students to evaluate their own learning and achievements. Boud (1995) stated that "self assessment occurs within a particular context, with respect to particular domains of knowledge and with particular goals in mind" (p. 15). Roberts (2006) defined self assessment as "the process of having learners critically reflect upon, record the progress of, and perhaps suggest grades for, their own learning" (p.3).

From these definitions it can be seen that self assessment is usually more than students' grading of their own work. It enables students to develop the ability to assess their own performance accurately. Such ability is a key foundation for lifelong learning. When students have to assess their own performance, they are also involved in the processes of finding criteria for the quality of a work. Furthermore, they have to develop meta-cognitive skills which support them in monitoring their work and in modifying their learning strategies depending on the task.

One important component of self assessment is reflection. Due to reflection, self awareness about the way of learning can be enhanced. However, Roberts (2006) noticed that students are usually not naturally skilled at self assessment but need guidance and practice to increase those skills.

Peer assessment

Besides self assessment, also peer assessment is an important form of assessment. In peer-assessment activities students are asked to value the performance of their peers although the learning may have occurred individually (Roberts, 2006). A general review of recent research on peer assessment can be found in Prins, Sluijsmans, Kirschner, and Strijbos (2005) or Wen and Tsai (2006). One advantage of peer assessment is that it allows students to develop learning at high cognitive levels (Bouzidi & Jaillet, 2009). Thus, peer assessment is linked with self assessment and therefore can enhance self assessment skills (Boud, 1995). When peer assessment is used, it is important to provide clear and concise guidelines (e.g., lists of points to be assessed; Roberts, 2006). Furthermore, the instructor should maintain the responsibility for the final grades. However, one weakness of peer assessment is that friends' work is often over-estimated when the assessment is not done anonymously. Because peer assessment especially with respect to e-assessment has raised increasing interest in the recent years, it will be covered more detailed in Chapter 4.

Furthermore, it is also possible that tutors and peers are involved collaboratively in the assessment process. This specific form is called *co-assessment* (Dochy & McDowell, 1997). For instance, Al-Smadi, Guetl, and Helic (2009a) showed that peer assessment and tutor assessment showed the highest agreement when the assessment tasks were relatively simple (e.g., definitions and enumeration answers).

With respect to peer assessment, Kollar and Fischer (2010) described four activities and their related cognitive and discursive processes (CaDP) that typically occur during peer assessment. First, as a prerequisite, a task has to be performed. This task can be done individually or collaboratively. CaDP that occur during task performance are for instance *planning*, *problem solving*, *explaining* or *questioning*. A second activity in peer assessment is then the provision of feedback. Here, the quality of the (performance) outcome or the process itself is assessed. In this activity, CaDP play a crucial role with respect to the question of how feedback should be formulated to be easily understood by the assessed person. The third activity in peer assessment is the reception of the feedback. Usually, the assessed person has no opportunity to reply to the feedback. However, for the purpose of a better understanding and hence, an enhancement in learning, it can be useful to allow such exchange after the feedback. In this stage of feedback reception, CaDP are necessary to take the feedback and to deal with it. Finally, the fourth activity in peer assessment is revision. Here, according to the feedback that was given before, the work will be reviewed by the assessed person alone – or more collaboratively – in interaction with the peer. From the view of the involved CaDP, this also enhances processes of critical thinking.

Group assessment

When group assessment is discussed it first has to be distinguished whether (a) the group as a whole or (b) an individual within a group is assessed, or whether (c) the group members assess the contribution of other group members to the group (Roberts, 2006). The assessment of group work as defined in the first case often raises the question of how to deal with free-raiders, i.e., with students who contribute only little to the work of the group but benefit from a shared group mark at the end. An appropriate way to deal with free raiders is to use a peer assessment strategy within the group, i.e., students are invited to assess the contribution of each group member. When the effects of collaborative learning are assessed, one has to consider that collaborative learning includes a variety of contexts. Thus, effects can only be investigated when the specific context has been clarified. One further issue regarding the assessment of collaborative learning is that often only individual performance - and not group's performance is measured.

3.2 Feedback

As noticed by Chickering and Gamson (1987), *"Students need appropriate feedback on performance to benefit from courses. In getting started, students need help in assessing existing knowledge and competence. In classes, students need frequent opportunities to perform and receive suggestions for improvement. At various points during college, and at the end, students need chances to reflect on what they have learnt, what they still have to learn, and how to assess themselves."*

Feedback provides a valuable tool to help learners become aware of gaps in their knowledge, skill or understanding of a topic (e.g., 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). Usually, the concepts of formative assessment (see Section 3.1.2) and feedback overlap (Black & Wiliam, 1998).

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
- A mechanism by which the information can be used to alter the gap

Kluger and DeNisi (1996) stated four potential actions when a gap between the actual and the reference level is recognized. Usually, there will be an attempt to reach the reference level. This will be the case when the individual's commitment is high. Another possibility is the abandonment of the reference level. This is often the case when the individual's belief in success is low. Furthermore, it is also possible to change the reference in order to reach this new reference level successfully. Finally, a person might also simply deny the existence of such a gap.

The quality of the feedback is a key feature in any procedure for assessment (Black & Wiliam, 1998). For instance, Kluger and DeNisi (1996) found out that feedback even reduced performance in one third of the cases. So, how should feedback be provided to help learners improving their learning progress? Considering that the goal of learning is to build up knowledge, to understand facts, and to apply knowledge, then assessment and feedback must also focus on understanding and not only on a reproduction of procedures or facts (e. g. Bransford et al., 2004). To this end, feedback should occur continuously, albeit not intrusively. Bransford et al. stated that effective instructors monitor both group and individual performances online, and that they assess the abilities of their students to link students' current activities to other parts of the curriculum and their life. Furthermore, they give feedback either formal or informal and help students to develop skills of self-assessment.

Usually, it is not recommendable to simply provide the right answer (e.g. Boston, 2002) but rather give comments about specific errors and specific suggestions for improvement. Gibbs and Simpson (2004) suggested that feedback should, for example, be sufficient in frequency and detail and that it should be less focused on the students' themselves but rather on their performance, learning, and actions. Finally, feedback is most beneficial for students when there is an opportunity to revise their work (Bransford et al., 2004).

3.3 Assessment from the viewpoint of cognitive and social theories

In Chapter 1, we draw an overview about the most important theories of learning, namely behaviorist, cognitive and social theories of learning. As these theories differ in their viewpoint of learning, also the conclusions for teaching assessment activities are different. We will briefly discuss this issue in this section.

For instance, assumptions from behaviorist theories lead to the suggestions that basic skills should be introduced before more complex skills and that positive feedback should be given to train the desired response to a stimulus (James, 2006). The implications for assessments are that learning progress should be monitored by unseen timed tests. The items tested should be formed from progressive levels using a skill hierarchy. Finally, when poor performance was shown in a test, it can be compensated by either more practice of the test items, by further deconstructing the skill or by relearning the more basic skills.

From the viewpoint of cognitive theories, Van Merriënboer and Ayres (2005) recommended that new forms of assessment are needed because

"… instructional methods that work well for novice learners may have no positive or even negative effects when learners acquire more expertise. … By adapting instructions according to levels of expertise, the difficult task of attempting to predict subsequent levels of expertise prior to the commencement of an instructional sequence is obviated."(p. 8).

Therefore, the aim of assessment is to close the gap between the learners' actual level of understanding and the level of the new understanding (James, 2006). The learners' models has to be updated by e.g. dialogues, concept-mappings or thinking aloud in order to scaffold their understanding and to support them to apply knowledge structures and strategies in new situations.

From a constructivist viewpoint, Brooks and Brooks (1999) suggest that assessment should occur in context of daily classroom investigations rather than as separate events because students usually demonstrate their knowledge and skills every day in a variety of ways.

With respect to the assessment of collaborative learning, it first has to be considered whether the individual outcome or the group outcome (or both) are going to be assessed. In general, however, findings from all learning theories already discussed before (behaviorism, constructivism etc.) can be helpful for assessment activities regarding collaborative learning. For instance, a constructivist approach in peer assessment leads to the assumption that an assessor can benefit when he or she are actively involved in assessment activities (Kim, 2009). Furthermore, an assessed person in peer assessment might profit from rewards and feedback as discussed with respect to behaviorist approaches.

3.4 Assessment from the viewpoint of motivational and emotional aspects

When assessment with respect to motivation and emotion is discussed, it first has to be clarified whether the task is (a) to assess motivation and/or emotion or (b) how motivation and emotion may affect the assessment. We will present a brief overview from literature for both viewpoints.

3.4.1 Assessment of motivation

With respect to how motivation can be assessed, Cocea (2006) described two directions, namely observation and explicit measurement. The latter one means that the learner is involved in the assessment. Typically, interviews, questionnaires and self-reports are used in this case. From the viewpoint of the instructor, Beghetto (2004) also suggested that pre-existing routines like observations, questionnaires, and class discussions can be used to assess motivation. Note that an important model in the assessment process is the ARCS-Model of Keller presented in Chapter 1.

One further important issue regarding assessment and motivation is that motivation is often affected by assessment (Harlen, 2006). Results of Maslovaty and Kuzi (2002) suggested that assessing through alternative methods (e.g., portfolio, experience enjoyment, challenges, and interest) supports adopting task goals and autonomy in learning. Hence, students may internalize the principles and activities of intrinsic motivation.

There are two main directions in assessment of motivation: one based on observation and one based on explicit measurement, involving the learner. From this perspective, of involving the learner in assessment, the main methods or instruments that can be used are: interviews, questionnaires, self-report and Wizard-of-Oz studies (Cocea & Weibelzhal, 2006).

There have been various ways that have been proposed to detect the motivational state of learners, including self-reports through sliders, behavioural cues in the interaction between learner and educational system, student's response times to tasks in combination with actual performance, as well as learner's attention, current task and expected time to perform (Kelly & Weibelzahl, 2006).

The way mostly used for grasping a learner's motivation is by answering questionnaires. There is a variety of studies that measure motivation with survey questions. In these studies, the researchers come up with a list of potential motivations and ask participants to respond, usually in the form of Likert-style[4] agreement statements. There are several examples of interesting papers that use variations of this method, for example:

1. The Motivation Assessment Scale (MAS) (Durrand & Crimmins, 1992)
2. The Motivated Strategies for Learning Questionnaire (MSLQ) (McKeachie, 2005) The MSLQ is a self-report measure that includes 81 items with both motivation and learning scales. The

[4] A Likert scale is a psychometric scale commonly used in questionnaires, and is the most widely used scale in survey research, such that the term is often used interchangeably with rating scale even though the two are not synonymous. When responding to a Likert questionnaire item, respondents specify their level of agreement to a statement. The scale is named after its inventor, psychologist Rensis Likert. [http://en.wikipedia.org/wiki/Likert_scale]

motivation scales include items related to intrinsic and extrinsic motivation (Shia, 1998) and are appropriate for a variety of students ranging in age and the academic area being studied.

3. The Academic Intrinsic Motivation Inventory uses a 7-point Likert scale with items on various factors relating to both intrinsic and extrinsic motivation (Shia, 1998).
4. The Instructional Materials Motivation Survey (IMMS) uses 36 Likert scale items to determine a variety of aspects related to motivation based on the ARCS model (Keller, 1987, Small, 1997).
5. Self-efficacy rating scales (Bandura,1997).

Other approaches try to infer automatically different motivational states by connecting the learner's actions (reading a page, solving a quiz, etc) and the time to perform them, with performance, which is typical information for educational systems. Activity tracking has been considered as a source of information for assessing users' motivation. Thus, a number of approaches have been presented trying to infer motivational states from the learners' interactions with the systems. For example:

1. A rule-based approach to infer relevance, confidence, satisfaction (from ARCS model, effort and sensory/ cognitive interest (de Vicente and Pain, 2003). De Vicente and Pain diagnose learner's motivation timely using a learner's actions based on a set of 85 rules. These rules are acquired by analyzing interactional actions between a learner and the learning system by different human tutors.
2. Inferring confidence, confusion and effort from: the learner's focus of attention, the current task and expected time to perform the task (Qu, Wang & Johnson, 2005). Qu et al. point out that by detecting the learner's motivational states, pedagogical agents will be able to promote the learner's motivation through interaction. They are suggesting a method for agents to assess learner's motivational states in an interactive learning environment. It takes into account the learner's attention, current task and expected time to perform the task.
3. Inferring attention and confidence from the learner's actions, using factor analysis to group the actions that indicate the two motivational states (Zhag,Cheng, He, & Huang, 2003). In a variety of research studies, the J. M. Keller ARCS Model is being employed to detect user's motivation rates. The ARCS Model uses four factors to describe motivation, namely *Attention*, *Relevance*, *Confidence*, *Satisfaction*. These four factors are thought to be the main points to maintain the learner's motivation. in a web-based learning environment.
4. Joseph E. Beck (August 2004) proposed a model based on the Item response theory, and used as input the difficulty of the question, how long the student took to respond, and whether the response was correct. He presented a means for analyzing the response times and correctness of the student's responses to model his/her overall level of engagement while using a computer tutor. From these data, the model determines the probability that a student was actively engaged in trying to answer the question. By analyzing 231 students' interactions, he showed that disengagement is better modeled by simultaneously estimating student proficiency and disengagement than just estimating disengagement alone.

3.4.2 Assessment of emotion

Assessing emotions is far more complex because of the difficulty to measure emotions objectively but also because of the lack of having a common definition of emotion (see also Chapter 1; c. f. Wong, 2006). As presented before regarding motivation, also emotion can be measured in various ways. Because individuals are often not aware about their emotions and/or simply because they cannot describe them adequately, therefore, self-reported emotional states should be taken and interpreted with caution. But also rather objective parameter like skin conductance or heart rate may not be able to exclude influences by other factors (e.g., a person might be angry not only through to the situation assessed but also because of a personal conflict before).

3.4.3 Emotion assessment systems

Humans use different sources of information to assess a person's emotions, including causal information context and individual traits, as well as information on the person's recognizable bodily reactions (Zimmermann, et al., 2003).

The most significant challenge in building emotion-oriented systems is the automatic recognition of affective states. Research literature has produced successful recognition techniques that classify physiological and neurophysiological signals, behavioural data and text/speech into different sets of emotions.

Affect measurement can be grouped into three areas (Zimmermann, et al., 2003, Leon & Nikov, 2010, Picard, 1997, Wong, 2006): Psychological, physicological, and behaviorial.

Psychological (Profiling tools)

Tools that fall into this category are recording a first-person report of subjective emotion experience (from moment to moment):

- self-reports,
- conductive chat
- rating scales
- standardized checklists
- questionnaires
- semantic and graphical differentials projective methods

Self-reporting allows users to express their emotions through verbal or non-verbal means. However, this technique suffers from problems of language, culture, and subjectivity. People can feel pressed to give wrong answers. They are considered retrospective and assess the conscious experience of emotion and mood, but much of the affective experience is non-conscious. Questions about affect are potentially influenced by when they are asked.

Physiological signals (use of sensors)

Another way is to use peripheral autonomic psycho-physiological measures:

- Skin conductance (SC)-Electrodermal Activity (EDA)
- Heart activity (electrocardiogram-ECG)
- Face muscle activity (electromyogram-EMG)
- Blood Volume Pulse (BVP)
- Respiration
- Pupillary dilation
- Muscle action potentials

Physiological sensors provide an objective measure of physiological signals. This measure is free from language and cultural problems. However, the instruments are often obtrusive or even invasive and thus hinder a user's experience with the interface. Furthermore, the reliability of these sensors is questionable since the mapping of sensor output onto emotion is not standardized. Moreover, they necessitate specialized and frequently expensive equipment and technical expertise to run the equipment.

Behavioral

This technique is inferring user's emotions from their observed behaviour.

- Face recognition/Facial expressions (face reader)
- Voice modulation/intonation (speech recognition)
- Gestures (hand tracking)
- Posture
- Motor behavior
  - Keyboard-Mouse (log files)
  - Hand muscles
  - Head movement
  - Corrugator's activity (eye-tracker)

This technique can pick up emotion cues that cannot be measured by self-reporting or physiological signals. However, it is a technique that requires experience and objectivity in the observer. These methods are tested almost exclusively on "produced" affect expressions. Recognition accuracy would drop heavily in natural situations. Furthermore, video cameras are considered obtrusive

Employing emotion assessment techniques to evaluate affective interfaces is open to two major problems: confounding errors and lack of a universal representation of emotions (Wong, 2006).

User's emotions can be easily affected by the external environment and experiment-related factors. Users who come into the experiment after a bad day react differently towards the interface compared to users in a good mood. Moreover, experiment-related factors, such as intrusive measuring apparatus and anxiety created by the experiment environment also contribute to measurement errors (Wong, 2006).

The second problem in using emotion assessment techniques for evaluation is the lack of a universal representation of emotions. Comparison and benchmarking of affective interfaces is very difficult without such a representation. The reliability of the evaluations is also weakened. Different emotion representations exist in the literature but no consensus exists. It will be a long road for researchers to achieve this common representation (Wong, 2006).

Research on emotions during assessment activities has not attracted much attention yet. For instance, from the instructor's viewpoint, Steinberg (2008) stated that [she] "did not find any studies that explored instructors' emotions towards marking, perhaps because the emotions are so uncomfortable that no instructor wants to dwell on them once the job is done" (p. 47). Besides the problems of assessing emotions in general discussed before and in Chapter 1, there is also some difficulty to define and measure emotions during assessment. For instance, using traditional questionnaires are again typically provided before or after the event but not during the activity. Rather physiological measurements might be again affected by other factors. However, there are some studies trying to investigate emotions during exams. The main focus here lies on test anxiety (e.g., Cassady & Johnson, 2002; Zeidner, 2007).

3.5 Adaptive Assessment

With the changing of the educational settings during the last decades as discussed before, also "individualization and personalization" in learning (and therefore also assessment) became more important. This changing is mostly caused by the fact that the latest generation of learners has grown up with digital media such as blogs, wikis, instant messaging etc. While the older generation had to become familiar with technologies like cell phone, computer, internet etc., members of the younger generation (i.e., born after 1980) can be seen as "digital natives" (McLester, 2007) or "New Millennium Learners" (NML; OECD, 2008). For instance, McLester described NML as social but also egocentric. Therefore, Baird and Fischer (2006; cited after McLester, 2007) suggested that NML have "personally tailored" learning paths; i.e., NML choose contents from different resources to meet their individual needs and learning styles.

Kareal and Klema (2006) stated that "different learners may have different characteristics, prior knowledge, motivation or needs" (p.260). Thus, the aim of personalization in context of (e-) learning is to identify the characteristics of the learner and to adapt the learning contents based on these individual characteristics. E-learning systems providing personalization typically have three components (e.g., Cocea, 2007): The *learning model* which stores the characteristics of the learner, the *domain model* which comprises the structure of the domain and the *adaptation model* which contains a set of rules. Due to this rules, the information from the learning model and the information from the domain model are combined. Furthermore, due to the adaptation model the information in the user model is changed and also a decision is made what content should be presented to the learner next.
When e-learning and e-assessment are used, such adaptivity, i.e., the adaptation of learning materials and assessment approaches to the actual state of the learner is an important factor that is assumed to enhance the learning outcome. Thus, it is important to develop and provide adaptive systems that consider various aspects of the individual learner. For instance, Kickmeyer-Rust (unpublished) stated that adaptivity in educational settings focuses either on *adaptive presentation* (i.e., providing individual additional information, or different variations of the same information or reordering information according to individual needs); *adaptive navigation support* (i.e., guiding a learner in the most suitable way through the learning material), and *problem solving support* (i.e., providing not only the final solution of a problem but by rather analyzing which knowledge might be missing).

Law and Kickmeyer Rust (2008) described the concepts of *macro adaptivity* and *micro adaptivity*. Therefore, macro-adaptivity refers to traditional techniques of adaption (e.g. adaptive presentation or navigation) whereas micro-adaptivity refers to interventions that are non-invasive and affect the presentation of a learning object. Such non-invasive, continuous interventions could be used for assessing the learning process and the motivational state of a person (e.g., Kickmeyer-Rust, Steiner, & Albert, 2009). However, providing intelligent and adaptive learning (and assessment) tools is often challenging and such systems did not open up the market yet (Kickmeyer-Rust & Albert, 2008). Accordingly to Kickmeyer-Rust and Albert, reasons for this are the difficulty of designing comprehensive data models, the interoperability and re-usability of learning media, and a lack of focus on the learner.

3.6 Quality of Assessment from a psychological viewpoint

When assessments are planned one typically has to consider what topic, issue or behavior should be assessed in particular and how this topic, issue or behavior can be assessed adequately. For instance, it is not feasible to assess the mathematical knowledge of a student by only measuring his or her intelligence. Furthermore, testing French vocabulary does not address every word a person might know but is rather a sample of his or her vocabulary.

Therefore, it is important that "every measuring instrument, if it is to be of any use, must demonstrate a number of important qualities." (McQueen and Knussen, 2006, p. 139). In this section we will give an overview about the most important quality standards, namely *objectivity*, *reliability*, and *validity*. Note that these standards are based on each other; i.e., reliability bases on the objectivity, and validity bases on reliability.

3.6.1 Objectivity

Objectivity means that a test results must be independently of the observer who performs the test. Hence, when different observers are assessing a person, they should come to the same conclusion.

3.6.2 Reliability

Reliability means that the test or tool must demonstrate consistency (McQueen & Knussen, 2006), i.e., when the same measuring instrument measures the same object for several times, the same results should be provided. This also includes the fact that a changing in the results of the measurements should be due to the changing in the object but not due to the measuring instrument. For instance, a scale should provide the same results when the same object is measured once again; a changing in the weight should be due to a changing in the object but not because of the scale. In context of learning this would mean, for instance, that when a mathematical test is presented before and after a course, the changing in the results should be due to the improvement of the learners but not because the difficulty of the test varied between the two times of measurement. There are several possibilities to measure the reliability of a test: First, to ensure that the test is consistent, the same thing can be measured twice (*test-retest reliability*). Note that in this case one has to consider practice effects; i.e., there might an improvement in the task anyway. Another form of reliability is *alternate form reliability* in which two forms of the tests are given and afterwards correlated. It is also possible to simply split the test in two halves and then to correlate them (*split-half reliability*). Finally, it is also possible to compute *reliability coeffients* to measure the internal consistence of a test.

3.6.3 Validity

Validity means that a tool or test is actually measuring what it is supposed to be to measure (McQueen & Knussen, 2006). Again, there are several ways to measure validity, though the methods are often not trivial. One form of validity is *content validity*. Content validity means that the content of the test must accurately and adequately reflect the content of the topic investigated. For instance, a knowledge test after a course should cover the main issues presented during the course. A further form of validity is *face validity* that is related to content validity. Due to face validity it is estimated whether the test appears to measure a criterion. Note, however, that it is not guaranteed that the criterion is indeed measured. Because of this consideration, face validity should not be used as a single form of quality criterion. A third form of validity is *criterion related validity*. This means that a test can be compared to a criterion that also measures the same construct. For instance, if a test should assess the ability of mathematical skills due to which a student is assigned to a specific course level, the later performance in the course could be checked against the test. Finally, *construct validity* indicates the extent to which a test measures a theoretical construct (e.g., intelligence). This form of validity is the most difficult to measure because the constructs themselves can only been measured by indirect measurements like observations. Thus, also construct validity can be measured only indirectly. One possibility here is to correlate a test with another test that measures the same construct.

3.7 Summary of the chapter

In this chapter we defined assessment and described forms and types of assessment. We furthermore discussed assessment in context of basic theories (behaviorism, constructivism, etc.) introduced in Chapter 1. In one section we introduced and outlined the importance of feedback. Finally, we shortly discussed adaptive assessment as a useful approach in supporting the individual learning and outlined the main quality criteria regarding assessment from a psychological viewpoint.

A·L·I·C·E

4 Assessment - Technology and Education

In general, assessment has different strategies according to its purposes. *Formative assessment* is part of the learning process; this assessment is used to give feedback to both students and instructors in order to guide their efforts toward achieving the goals of the learning process. *Summative assessment* is performed at the end of specific learning activity; and used to judge the students progression and also to discriminate between them (Bransford et al., 2004). According to Bennett (2002), technology is an essential component of modern learning system. As a result, technology is also increasingly needed for the assessment process to be authentic.

E-assessment can be distinguished as *Computer Based Assessment (CBA)* or *Computer Assisted Assessment (CAA)* which are often used interchangeably. CBA represents 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).

Types of useful assessment method highly depend on the learning objectives. These objectives have been classified in Bloom's Taxonomy into the following six levels: *knowledge*, *comprehension*, *application, analysis*, *synthesis*, and *evaluation* (Bloom, 1956; Krathwohl, Bloom & Bertram, 1973; see also Chapter 1). E-assessment systems can be classified according to the nature of the users' response to the test items into, *fixed response* or *free response* systems (Culwin, 1998). According to Culwin, fixed response systems -which also referred to as *objective* - force the user to have a fixed response by selecting an answer from a pre-prepared list of solution alternatives. Where, in the free response systems *non-objective,* unanticipated answers formulate the user's response. In free-response systems, skills like programming, essays writing, and meta-skills are assessed where knowledge assessment is mainly done using the fixed-response ones. Moreover, portfolios can also be used to assess learning outcomes. 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.

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).

4.1 Overview of projects and initiatives addressed to assessment

Because of the obvious relevance of e-assessments there are various numbers of projects funded under the Seventh Framework Programme (http://cordis.europa.eu/home_en.html) and other programs. In this section we provide a broad but certainly not complete overview about relevant projects regarding e-assessment (see Table 5). Note that further projects are also described more detailed within the next chapters.

Project /Initiative	(Main) Objectives/Contents	URL
AFEG Assessment framework for epistemic games August 2008 to July 2009	Assess innovative and creative thinking developed by computer games	http://cordis.europa.eu/fetch?CALLER=FP7_PROJ_DE&ACTION=D&DOC=24&CAT=PROJ&QUERY=012a9db42ff8:46d5:583697f9&RCN=88030
AfL Assessment for learning 8 schools project July 2005 to October 2006	Identify what helps pupils develop as motivated and effective learners	http://nationalstrategies.standards.dcsf.gov.uk/node/97897
EDNA Education Network Australia Running project; established in 1996	Network of the education and training community in Australia	http://www.edna.edu.au/edna/go
EERQI European Educational Research Quality Indicators April 2008 to March 2011	Build an advanced framework for relevance assessment of research documents in educational research	http://www.eerqi.eu/
FREMA Reference model for assessment April 2005 to March 2006	Develop a reference model for systems in the assessment domain of the JISC e-learning framework (Part of the JISC e-learning program)	http://www.jisc.ac.uk/whatwedo/programmes/elearningframework/reffrema.aspx
LTfLL Language technologies for lifelong learning March, 2008 to February, 2011	Develop services establishing the current position of the learner in a domain Develop support and feedback services based on analysis of the interactions of students Construe a knowledge sharing infrastructure that allows comparison and sharing of private knowledge	http://www.ltfll-project.org/index.php/index.html
METAFORA Learning learn together: A visual language for social orchestration of educational activities Start: July, 2010	Explore the potential of social learning for Science and Math Design new forms of assessment for individual and collaborative learning.	http://cordis.europa.eu/fetch?CALLER=FP7_PROJ_DE&ACTION=D&DOC=1&CAT=PROJ&QUERY=012a9d96257f:fa29:58b1a7da&RCN=95596
QUAL-PRAXIS Quality assurance and practice-oriented assessment in vocational education and training October 2003 to September 2006	Investigate innovative student assessment models from the perspective of different national vocational education and training traditions	http://www.peda.net/veraja/projekti/qualpraxis
PANdora; Sub-project7 E-assessment methods and models for student evualution in Asia 2005 to 2008	Study existing generic policies, practices, and methods of e-assessment, Identify human differences, institutional, technological, and operational issues,	http://www.pandora-asia.org/panprojects.php?main=panprojects_7.htm

	Develop and test an e-assessment generalized model Suggest future applications of the model.	
PISA Programme for International Student Assessment (Four assessments between 2000 and 2009)	Assess how students near the end of compulsory education have acquired knowledge and skills that are essential for full participation in society	http://www.oecd.org/pages/0,3417,en_32252351_32235731_1_1_1_1_1,00.html
TARGET Transformative, adaptive, responsive and engaging environment January 2009 to December 2012	Develop a new genre of technology-enhanced learning environment that supports competence develoment	http://www.reachyourtarget.org/moodle/
WebCEF Collaborative evaluation of oral language proficiency October 2006 to September 2009	Enable the collaborative assessment of oral language proficiency through a web-based environment	http://www.webcef.eu

Table 5. Overview of projects and initiatives addressed to assessment

4.2 Assessment Models

Do all assessment forms have the same framework or architecture? What are the common features between assessment forms? These and many other related questions have been discussed over the last years. Several frameworks, models, and design architectures have been provided either for assessment in general or for specific assessment forms or application domains.

Pellegrino, Chudowsky, & Glaser (2001) provided what they called assessment triangle that discusses three key elements for assessment in general. As depicted in Figure 11, 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 the set of beliefs about the kinds of observations that are constructed based on situations and tasks provided to the students so that 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.

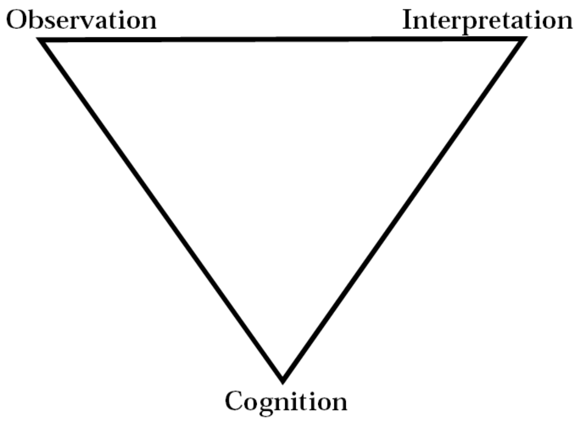


Figure 11. The assessment triangle (Pellegrino et al., 2001, p. 44)

Evidence-centered assessment design (ECD) (Almond, Steinberg, & Mislevy, 2002; Mislevy, Almond & Lukas, 2003) 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 12.

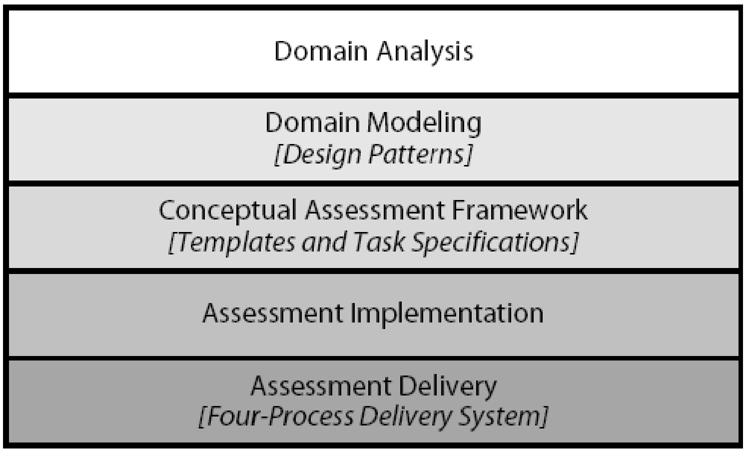


Figure 12. Graphical representation of ECD layers (Mislevy & Riconscente, 2005).

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 13, 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 modeling* 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 operational assessment 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, and*

*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.

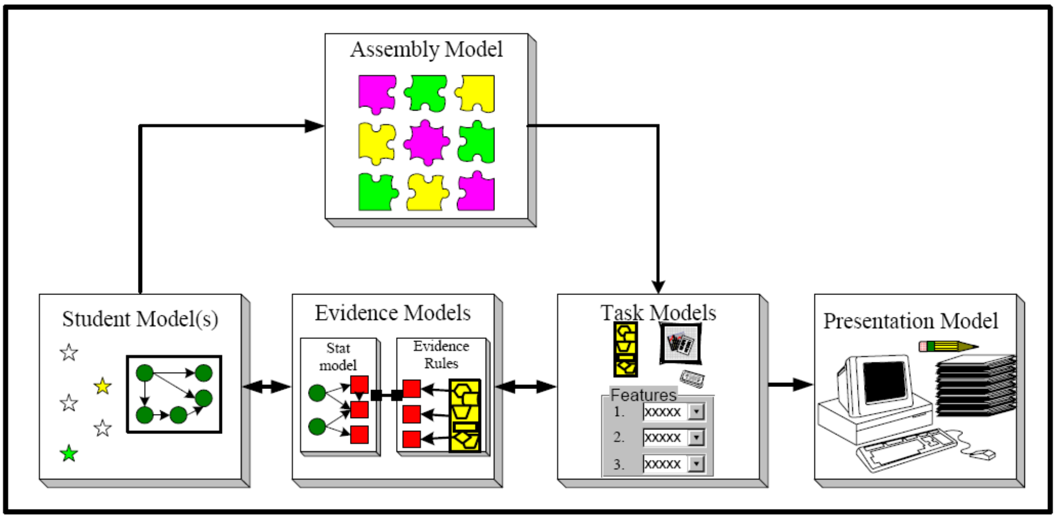


Figure 13. The Conceptual Assessment Framework (CAF) for the delivery system model (Mislevy et al, 2004)

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 14.

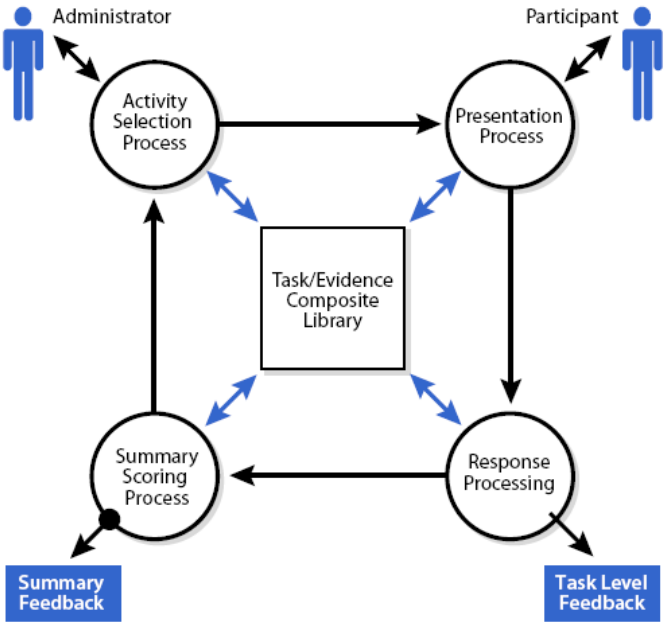


Figure 14. Four-process assessment architecture (Almond et al., 2002)

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 (point 3. *Personalization* is discussed in the original Koper (2001) reference but not cited in Brinke et al. (2007)):

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

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 15. 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. 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).

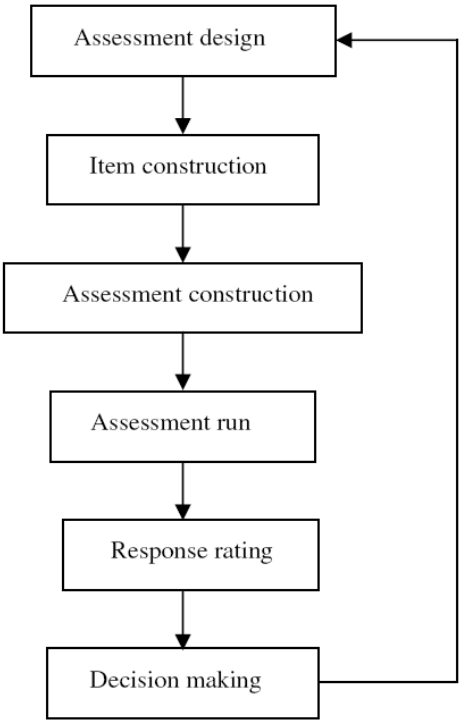


Figure 15. Main stages of the assessment process (Brinke et al., 2007)

Another useful framework is the Framework Reference Model for Assessment (FREMA, http://www.frema.ecs.soton.ac.uk). 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. 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 (Millard et al., 2005). The FREMA website provides interactive Flash® components to demonstrate the assessment domain as depicted in Figure 16.

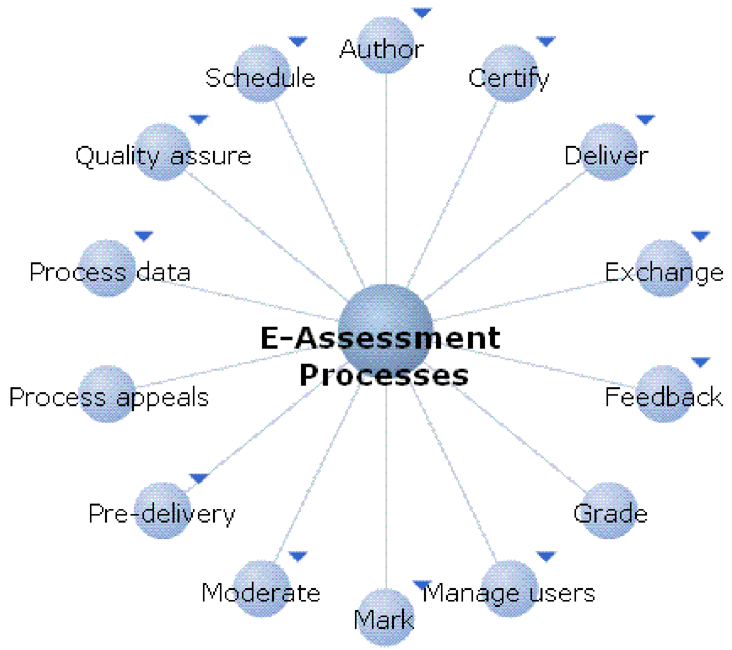


Figure 16. First level of FREMA e-assessment processes (http://www.frema.ecs.soton.ac.uk)

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.

Another example of assessment framework is the Service Oriented Framework for Assessment (SOFA). SOFA has been developed by AL-Smadi, Guetl, and Helic (2009b) as part of their research for flexible and standardized e-assessment systems. The authors suggested a flexible e-assessment system from the architectural point of view where the system can be used as a standalone e-assessment system or to be integrated with other systems such as LMS and authoring systems. Therefore, they distinguish between two levels of standardization in flexible e-assessment systems and their possible standards and specifications, as summarized in Figures 17, 18. The first level is an external level which represents possible services and standards that can be used to integrate the assessment system with external user agents such as LMSs. The second level is an internal level of standardization which discusses e-assessment related services and their possible standards. Based on that, they represented SOFA as a set of layers, as summarized in Figure 19.

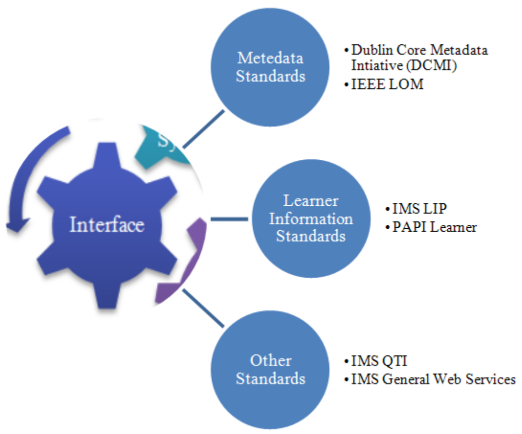


Figure 17. External level of standardization (AL-Smadi et al., 2009b)

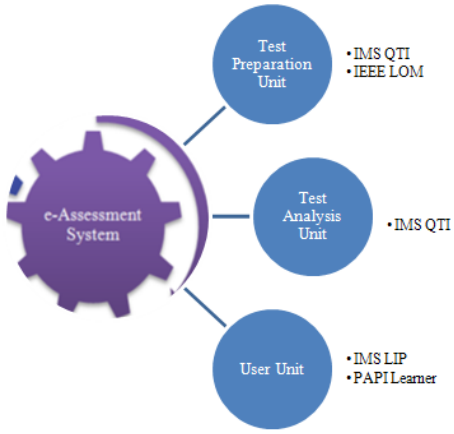


Figure 18. Internal level of standardization (AL-Smadi et al., 2009b)

SOFA concentrates on service-oriented architecture (SOA) where services should be designed to be standard-conform. SOA will foster e-assessment systems with flexible architecture so that they can be used as standalone systems or to be integrated with other systems and tools. SOFA abstraction layers discussed 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:* 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 here 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. The assessment services have been identified based on FREMA (Framework REference Model for Assessment) processes concept map.
- *Common Services:* a lower level of services those are not assessment-specific such as authorization and authentication.
- *Infrastructure:* represents the internal communications, storage and processing capabilities that the e-assessment system requires.

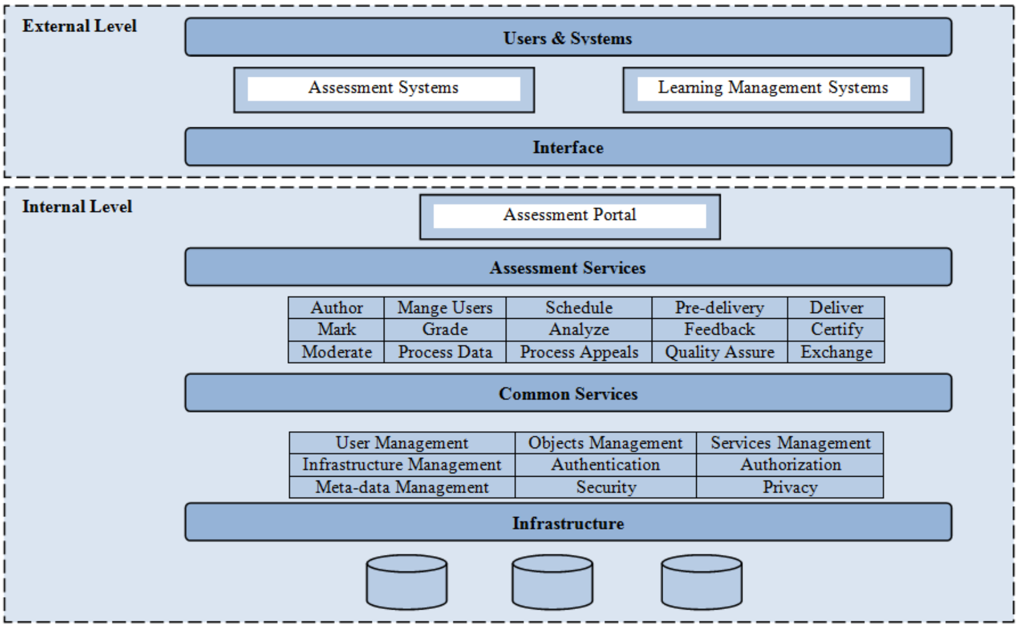


Figure 19. Service Oriented Framework for Assessment (SOFA) (AL-Smadi et al., 2009b)

4.3 E-Assessment Software

Computers have been used for decades to assist assessment. One of the earliest attempts of using computers to assist instructional and assessment process refers to the early 1960's when PLATO (Programmed Logic for Automatic Teaching Operations) project was started at the University of Illinois (Woolley, 1994). TICCIT (Time-Shared, Interactive, Computer-Controlled, Information Television), which has been started in 1967 is another example of a large-scale project for using computers in education (Hayes, 1999). The history of e-assessment can be referred to the use of computers to automatically assess the students' programming assignments (Douce, Livingston, & 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 instructor to supervise a larger number of students at the same course. Another application for the Automatic Grader was long distance teaching (Hollingsworth, 1960). Authors of (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 of data supplying, running time monitoring and keeping a "grad book" for recording problems.

Assessment played 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. Assessment as main part of the instructional design and media was affected by the revolution of micro-computers in the 1980's. According to Reiser, during the 1980's and afterwards an increasing interest of using computers in instruction had started and computers were used in automating some instructional design tasks. Examples of assessment systems from 1980s 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). Since then 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. Table 6 briefly describes the key features of some selected assessment tools and their web-links. The selected 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.

Assessment Management System (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 Langauge (CML) so that instructors can create algebraic questions as well as chemical formulas and exercises.

Blackboard™ and Questionmark Perception™ 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 gradebook and analyzed, as well as features such as item analysis and custom grading scales are available.

For free non-commercial examples, the Hot Potatoes 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.

AiM, CABLE, Maple T. A., and WaLLiS 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.

Assessment Tool	Description	Link
AiM	AiM is a system for intelligent computer-aided learning and assessment in mathematics and related disciplines, based on a symbolic mathematics program. It requires a commercial license for Maple to use it. Installation on own server is available, free for non-commercial use, source code available.	http://sourceforge.net/projects/aimmath/
ANGEL	ANGEL has two main products: ANGEL Learning Management System (LMS) and ANGEL ePortfolio. To enable client success, ANGEL	http://www.blackboard.com/Teaching-Learning/Learn-Resources/ANGEL-Edition.aspx

	provides hosting, consulting and training services. Recently, it becomes part of Blackboard products, commercial license is required.	
Assessment Management System	SARAS Assessment Management System is a tool designed to assess learners, evaluate performance and suggest further learning. SARAS AMS delivers tests to learners, diagnoses areas of weakness and provide remedial learning. AMS ensures secure delivery of tests and assignments to enhance participants' knowledge in a particular arena. Installation on own server is available.	http://www.excelindia.com/
ATutor	ATutor is an Open Source Web-based Learning Management System (LMS) used to develop and deliver online courses. Installation on own server is available, free for non-commercial use, source code is available under GNU General Public License (GPL).	http://www.atutor.ca/
BOSS	The BOSS Online Submission System is a course management tool, developed by the Department of Computer Science at the University of Warwick. BOSS allows students to submit assignments online securely, and contains a selection of tools to allow staff to mark assignments online and to manage their modules efficiently.	http://www.dcs.warwick.ac.uk/boss/index.php
Blackboard Academic Suite™	The Blackboard Learning System™ is a LMS that includes assessment. Additional features can be added using Building Blocks™. . Installation on own server is available, commercial license, annual renewal.	http://blackboard.com/
CABLE	CABLE (Computer Algebra Based Learning and Evaluation) is an online mathematical assessment tool for use with an open source computer algebra system such as Maxima or Axiom. Free for non-commercial use, source code is available.	http://sourceforge.net/projects/cable/ http://www.cable.bham.ac.uk/
Canvas Learning	Canvas learning Author and Course Builder constructs QTI-complaint items Flash™ Player is used to integrate with LMS sucha s Balckboard™ and Oracle iLearning®. Shockwave of Flash plugins are required on client side browsers. Windows authoring only, commercial license, perpetual.	http://www.the-can.com/
CASTLE	The Computer ASsisted Teaching & LEarning (CASTLE) toolkit is an on-line authoring tool that allows course tutors and managers to quickly create interactive quizzes for use on the Web. Stopped serving CASTLE requests on 30 September 2005 but the software is freely available for deployment and further development.	http://www.webarchive.org.uk/wayback/archive/20051116015929/http://www.le.ac.uk/castle/index.html
Claroline	Claroline is an Open Source eLearning and eWorking platform allowing teachers to build effective online courses and to manage learning and collaborative activities on the web. E-assessment is available. Translated into 35 languages. Installation on own server is available, free for non-commercial use, open source software.	http://www.claroline.net/
CQuest	CQuest Assessment Software is an e-learning tool that provides the means for workforce trainers or academic educators to create, generate, administer online testing , online training, and online assessments. Detailed response data is kept which allows one to report and analyze assessment results.	http://www.cquestsoftware.com
Criterion®	The *Criterion®* online evaluation service by Educational Testing Services (ETS). It provides instructors and students with reliable evaluations of	http://www.ets.org/criterion

	English-language essays. It delivers immediate score reporting and diagnostic feedback that students can use to revise and resubmit their essays. Instructors can use their own topics or select from the *Criterion* topic library of more than 400 essay assignments at various skill levels. Commercial web-based service.	
Desire2Learn	Desire2Learn Learning Platform is a LMS that includes e-assessment. Commercial license, perpetual.	http://www.desire2learn.com
Ensignia®	The Ensignia® is a Learning Management System (LMS) that includes assessment. Commercial license.	http://mcqi.com.au/lms/
ETH Lecture Communi-cator	ETH Lecture Communicator is a tool to improve the interaction in the classroom between instructor and students. The tool enables the instructor to create and carry out in-class online-assessments and facilitates organized instant communication for big classes. Installation on own server is available, free for non-commercial use, open source software.	http://lectcomm.sourceforge.net
ExamBuilder	ExamBuilder is an e-assessment service remotely hosted by ExamBuilder, commercial license.	http://www.exambuilder.com
Examine	Examine is a multiple-choice authoring and delivery system for use either as an adjunct to courseware or as a standalone means of on-line self-assessment A prototype of *Examine* (PC only) has been widely distributed, but a major new version (Macintosh and PC) has recently been developed.	http://ibis.nott.ac.uk/software/examine.html
FastTEST Pro	*FastTEST Web*, a comprehensive internet-based system for item banking, test assembly, and online test delivery. Macs or PCs, computerized adaptive tests (CAT) is possible based on item response theory. Commercial license.	http://www.assess.com/xcart/product.php?productid=273
Fronter	Fronter provides tools for your Learning Management System (LMS) assessment e.g. create tests and hand-in assignments. Commercial license.	http://com.fronter.info/
Hot Potatoes	The Hot Potatoes suite includes six applications, enabling you to create interactive multiple-choice, short-answer, jumbled-sentence, crossword, matching/ordering and gap-fill exercises for the world wide web. Hot Potatoes is freeware, and you may use it for any purpose or project you like. It is not open-source.	http://hotpot.uvic.ca/
i-Assess	i-Assess is an assessment system that can be used online, off-line, standalone or via LAN. Windows-based server, off-line versions are windows-based, commercial license.	http://www.iassess.com

IBM Lotus Learning Management System	IBM Lotus Learning Management System is a LMS that includes assessment. Installation on own server is available, commercial license.	http://www01.ibm.com/software/lotus/products/learning-management-system/
ILIAS	ILIAS is a LMS that includes assessment. Installation on own server is available, free for non-commercial use, available as open source software under the GNU General Public license.	http://www.ilias.de
IMS Assesst Designer	IMS Assesst Designer is an e-assessment system, installation on own server is available, windows platform, MySQL database, commercial license, perpetual.	http://www.xdlsoft.com/
INTERWRITEPRS®	INTERWRITEPRS® combines interaction and assessment to enhance classroom productivity. Using their radio-frequency or infrared wireless clickers, students can answer questions and record their responses with a simple click of a button. Commercial keyboards and license.	http://www.einstruction.com/products/assessment/prs/
JExam	JExam is a Java based testing package that enables question entry, test taking, test viewing, test creation, content management, hints, solutions, and statistical analysis. Installation on own server is available, free for non-commercial use.	http://exams.uga.edu/jexam/JExamStudent3/install/install.htm
Joomla LMS	JoomlaLMS is a learning management system, based on Joomla CMS - content management system platform. Installation on own server is available, commercial license.	http://www.joomlalms.com
Maple T.A.	Maple T.A. is an e-assessment system for mathematics. It supports complex, free-form entry of mathematical equations and intelligent evaluation of responses, making it ideal for mathematics, science, or any course that requires mathematics. Installation on own server is available, commercial license, annual renewal.	http://www.maplesoft.com/products/mapleta/
METRIC Maths	METRIC Maths consists of a series of self-tests in mathematics, java-based, client-side marking, free for non commercial use, access from remotely hosted site.	http://www.imperial.ac.uk/engineering/teaching/learningtechnology/metric
Moodle	Moodle is a Course Management System (CMS)that includes assessment. Installation on own server is available, free for non-commercial use, available as open source software under the GNU General Public license.	http://moodle.org/
OASYS GPL	OASYS GPL is an online peer assessment system. Using a web browser, students sit tests, and then mark the work of several of their peers. Installation on own server is available, free for non-commercial use, available as open source software under the	http://oasysgpl.sourceforge.net

	GNU General Public license.	
OLAT	OLAT (Online Learning And Training), is a Learning Management System (LMS) and includes assessment. Java-based, Installation on own server is available, free for non-commercial use, available as open source software under the GNU General Public license.	http://www.olat.org
Oracle iLearning	Oracle iLearning is an enterprise Learning Management System (LMS) and a core component of Oracle's E-Business Suite. Oracle iLearning provides e-assessment. Installation on own server is available, commercial license, annual renewal.	http://ilearning.oracle.com/ilearn/en/learner/jsp/login.jsp?site=TryMe
Pearson Learning-Studio	Pearson LearningStudio is a learning management system that includes e-assessment. Commercial license.	http://www.pearsoncustom.com
Question*mark*™ Perception™	The Question*mark*™ Perception™ assessment management system enables trainers, educators and testing professionals to author, schedule, deliver, and report on surveys, quizzes, tests and exams. Windows-based authoring. Can be used online, off-line, standalone or via a LAN. Installation on own server is available, commercial license, annual renewal.	http://www.questionmark.com
Question Tools	Question Tools is an integrated suite of products that allows anyone to create online lessons, exercises, surveys, tests & exams, and automatically collect & analyse results. Commercial license.	
Quia	Quia is an e-assessment service remotely hosted by Quia. Subscription license, annual renewal.	http://www.quia.com/
Quiz Factory 2™	Quiz Factory 2™ is e-assessment system. Can be used standalone or via LAN. Commercial license, perpetual.	http://www.learningware.com
QuizStar	QuizStar is an e-assessment system, hosted remotely by University of Kansas and free for non-commercial use.	http://quizstar.4teachers.org/
Respondus®	Respondus® is an e-assessment system that creats and manages exams that can be printed to paper or published directly to Blackboard, ANGEL, Desire2Learn, eCollege, Moodle, and other eLearning systems. Off-line authoring and assessments are uploaded to the LMS. Commercial license, annual renewal.	http://www.respondus.com/
RIVA e·test™	RIVA e·test™ is an e-assessment system. Installation on own server is available, or as a hosted service, commercial license, annual renewal. IMS QTI complaint.	http://rivatechnologies.com/etest/etest.htm

SAMigo	SAMigo also known as Quiz and Test is, an online assessment tool for teaching and learning that runs within the Sakai open-source LMS/CLE.	http://confluence.sakaiproject.org/display/SAM/Home
SharePointLMS	SharePointLMS is a Learning Management System and includes assessment. It is based on the Microsoft Office SharePoint Server 2007 platform. SharePointLMS is SCORM 1.2 and 2004 compliant. Commercial license.	http://www.sharepointlms.com/
SPIDER	The Strathclyde Personal Interactive Development & Educational Resource (SPIDER) virtual learning environment is an in-house development of a web portal that supports learning, teaching and administration. It uses the ShockWave IMS (SWIMS) question generator as a staff tool to create question banks, and the Shockwave InterNet Quiz viewer (SINQ) as a tool to deliver created assessments.	http://spider.science.strath.ac.uk
Test Generator	Test Generator (TG) is an e-assessment system that uses windows-based authoring of tests and platform-independent web-based delivery. Commercial license, perpetual.	http://www.testshop.com
Test Pilot	Test Pilot Online Assessment and Survey Engine, installation on own server is available, perpetual commercial license.	http://www.clearlearning.com/
TOIA	Technologies for Online Interoperable Assessment (TOIA) is an e-assessment system available free of charge to all UK further and higher education institutions.	http://www.toia.ac.uk/
TopClass e-Learning Suite™	TopClass e-Learning Suite™ is a learning and content management system LCMS that includes assessment. Commercial license, annual renewal.	http://www.wbtsystems.com
TRAIDS	The TRIpartite Assessment Delivery System (TRIADS) is an e-assessment system. TRAIDS is a multimedia interactive e-assessment system that requires Authorware™ for items authoring and the Shockwave plugin for web-based delivery. Can be used as standalone or via LAN.	http://pcwww.liv.ac.uk/apboyle/triads
WaLLiS	Web-based Assistant for Learning; on-Line Intelligent System (WaLLiS) is an e-assessment system for mathematics. Symbolic representations and mathematical equivalence are available. Contact website for availability.	http://www.maths.ed.ac.uk/wallis
WebAssign	WebAssign is an e-assessment system which provides homework assignment service, was initially developed at North Carolina State University. Some of the disciplines covered include mathematics, chemistry, statistics, physics, and biology. Subscription license, annual renewal.	http://www.webassign.net/

WebBoard	WebBoard is an online discussion board system. Windows server and database required. Perpetual commercial license.	
WebMCQ IQS	Interactive Question Server™ (formerly WebMCQ) is an assessment system. Commercial license, IMS QTI compliant.	http://mcqi.com.au
WebQuiz XP	WebQuiz XP is an e-assessment system that authors assessments in windows-based platform and deliver them over internet. Commercial license, perpetual.	http://eng.smartlite.it
WebTest	WebTest is a hosted, web-based system for the creation, delivery and administration of tests, tutorials and surveys.	http://www.chariot.com/webtest/index.asp

Table 6. List of e-assessment software (adapted from Crisp, 2007, P.69)

4.4 Other types of e-assessment

4.4.1 Peer-assessment

One of the first reported systems which has implemented peer assessment functionality was MUCH (Many Using and Creating Hypermedia), a tool for collaborative learning based on a multi-user database. In those days also a Macintosh application has emerged which has sent out assignment to two peers like in a peer review process. (Rada, Acquah, Baker, & Ramsey, 1993; Gehringer, 2000) At the end of the 1990s NetPeas (networked peer assessment system) has been introduced, and the Peer ISM system has been developed which combines assessment results of human peers with those of artificial peers (based on artificial intelligence) (Bull, Brna, Critchley, Davie, & Holzherr, 1999; Gehringer, 2000; Tsai, Lin & Yuan, 2002). Furthermore, first Web-based systems have emerged such as Web-based tool for collaborative hypertext authoring and assessment via email (Downing & Brown, 1997), Web-based assessment system for team member's contribution on engineering design projects (Eschenbach & Mesmer, 1998), the Calibrated Peer Review (CPR) system in 1999 which has already implemented a performance calibration task (Carlson & Berry, 2007) as well as the PG System, a portable Web-based peer-evolution system (Gehringer, 2000), and SPARK, a Web-based system for self and peer assessment of students' team work (Freeman & McKenzie, 2002). Further tools emerging at the beginning of the 21st century include the Computerized Assessment by Peers (CAP) system (Davies, 2003), and OASIS, a hybrid system handling multiple choice assessment automatically and free text assessment by peer grading (Trahasch, 2004). In 2004, OPAS (Online Peer Assessment System) included not only assignment and review functionality but also enabled involved persons to take part in discussion assignments and assessment results (Trahasch, 2004). The above mentioned systems do offer to some extend functionality for assignment uploading or online testing, scoring, commenting, presentation and discussion of results. However, there is a lack of sufficient functions for determined standards or reference answers, methods of scoring, forming of groups, and managing the workflow of the assessment in a flexible way. This situation has been improved by the Web-based self and peer assessment (Web-SPA) system (Sung, Chang, Chiou, & Hou, 2005). Examples of recent developments are the new enhanced open source implementation of the WebPA system originally developed in 1998 (WEBPA, 2008), and the Comprehensive Assessment of Team Member Effectiveness (CATME) to assess how effectively each team member contributes to the team (Ohland et al., 2009).

4.4.2 Computer Assisted/Based Assessment of Essays and Free Text Answers

Using Computers to assist the assessment process has been an important research for the last decades. Several domains have formulated the application domains of CAA/CBA, examples of such application domains are 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.

AEG and evaluating free text answers have been a field of interest during the last 40 years. Several applications based on different classifications have been emerged during this period. Page (1966), has distinguished between assessing content and style of the essay. Content refers to the body of the essay, what does the essay say, where style is related to the syntax and mechanics of writing. According to Christie (2003), in order to grade an essay both content and style are important. However, there are different techniques that only assess one of them during their grading process. AEG systems and the systems used to evaluate free text answers can be classified according to the technique background. 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. Another classification can be found in Chung and O'Neill (1997), where such systems are classified into systems which depend on documents classification, systems of this category are multilingual and do not perform any linguistics processes. The other category is the systems that understand the text meaning where a semantic, morphological and/or syntactic analysis is performed.

According to Gronlund (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, Pstoka, Laham,& MacCuish, 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, Bejar, & Hone, 1999). According to Valenti, Neri, & Cucchiarelli (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-Stephenson, 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 an interesting and valuable learning assessment activity, fast response marking and valuable feedback it will develop his writing skills as well as 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 and 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.

4.4.2.1 Historical Overview

The use of computers to assess free text answers goes back to the 1960s where a pioneer system Project Essay Grader (PEG) was developed by Page (1966). 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”. He 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”. According to Reed (1990), WH can be used to improve the students writing once it is utilized for revision.

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.

Webster (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. 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.

Going back to year 1998, another research was done by Larkey (1998) who 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 developed 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, Jerrams-

Smith, & Soh,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 came to view. Automark which 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). Rudner and Liang (2002), created another system called Bayesian Essay Test Scoring sYstem (BETSY), which based on statistical analyses. In the same year, the Paperless School free text Marking Engine (PS-ME) was developed by Manson and Grove-Stephenson (2002). Where Bloom's taxonomy (Bloom, 1956) and NLP are used to assess the answers.

In 2003, two systems were developed. The first one is Auto-marking (Sukkarieh, Pulmand, & Raikes, 2003) which is based on NLP and pattern matching methods. The other one is called CarmelTC and developed by Rose, Roque, and VanLehn (2003) to grade students' writing based on machine learning classification methods and a naïve Bayesian classification. 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 is a recent example for grading short-free answers, the system where developed at Graz University of Technology in the year 2007. E-Examiner is a web-based e-assessment system that serves as a complete assessment management system (Gütl, 2008).

4.4.3 Automatic Test Item Creation

Automatic test item creation from textual learning content has raised the interest of the community for quite a while, but research results and products in past were quite limited and basic. Please refer to Gütl et al. (2011) for an overview of developments in the past. The objective of this section is to give an overview of recent work on different approaches by exemplarily outlining one approach and tool each, which is finally used to compare system available with the Automatic Question Creator tool AQC (see also D5.2.1).

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 ontologies are resented 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 (see also Figure 20 and Figure 21). 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 found in the correct sentences (see also Figure 22).

$A(a)$: states that a is an individual of class A
$R(b,c)$: states that individuals b and c participate in binary role R.

Figure 20. Formal description of classes and roles (Papasalouros Kotis, & Kanaris, 2008).

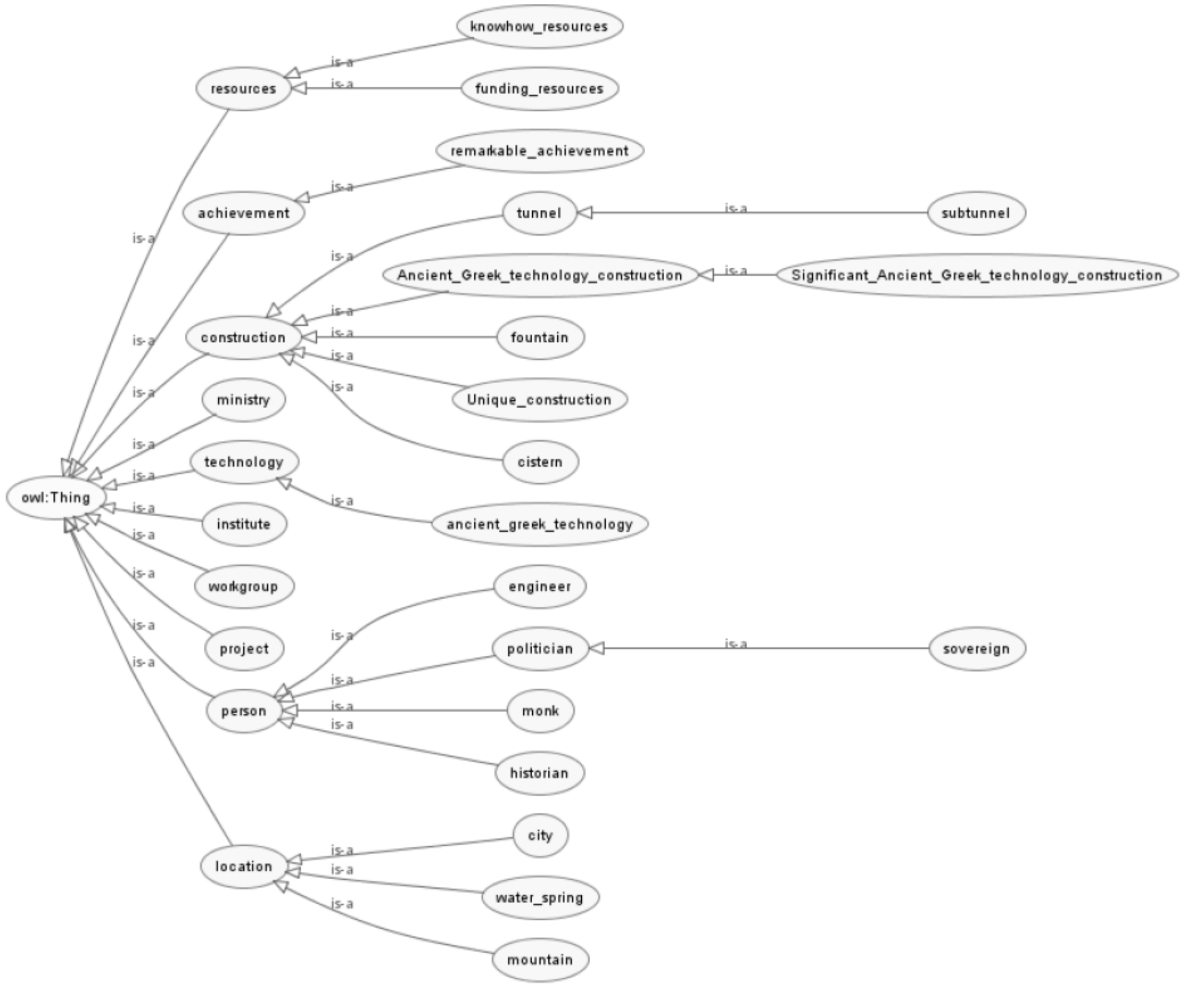


Figure 21. Eupalios Tunnel as an example of an ontology (Papasalouros Kotis, & Kanaris, 2008).

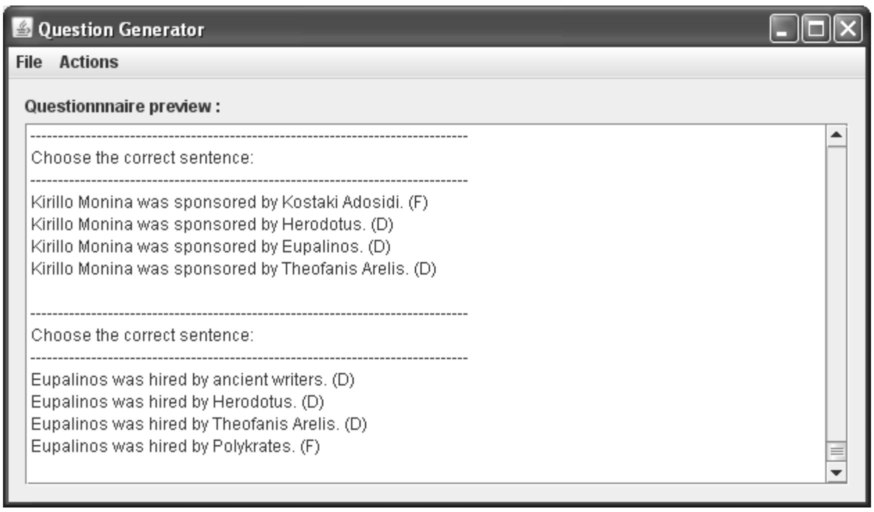


Figure 22. Multiple choice questions as an example of the tools output (Papasalouros Kotis, & Kanaris, 2008).

The authors in Sanz-Lobera, González Roig, and González Requena (2011) have proposed a parametric approach to create variants of exercises. In this parametric approach mathematic formula and models are the base for distractors of multiple choice test items; the application domain of such created test items are engineering and physics topics. Figure 23 outlines the applied methodology: (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 selected different variants which may include multiple correct and/or wrong answers; (d) questionnaires creation and maintenance created and managed the actual test items by combining text and computed values of variables; (e) results spreading and evaluation concerns the actual assessment activities. Figure 24 illustrates the way how the multiple choice questions are created by means of templates, Figure 25 shows concrete examples, and Figure 26 shows the result in an LMS manually copied by a tutor.

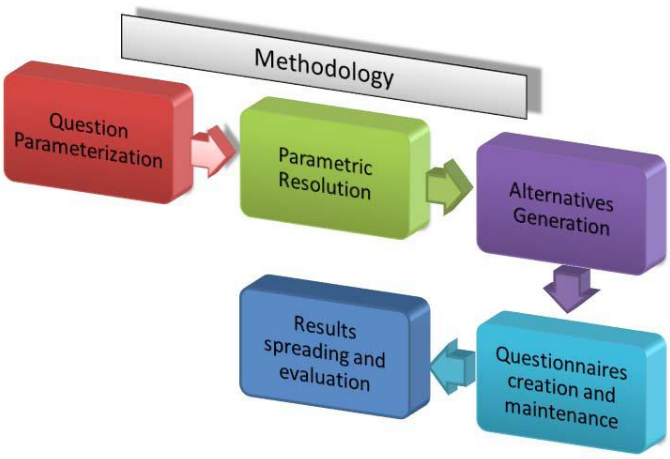


Figure 23. Methodology of parametrics-based question creation (Sanz-Lobera, González Roig, & González Requena, 2011)

→ A baseball player hits the ball with an initial velocity of **«v»** «**v_units**» and an elevation angle of «**alpha**» degrees. Neglecting the aerodynamic effects on the ball, the range in meters will be **{**
«**TFa**» → a) «**ra**»
«**TFb**» → b) «**rb**»
«**TFc**» → c) «**rc**»
«**TFd**» → d) «**rd**»**}**

Figure 24. Example of a template to create parametrics-based exercises (Sanz-Lobera, González Roig, & González Requena, 2011)

A baseball player hits the ball with an initial velocity of 150 km/h and an elevation angle of 25 degrees. Neglecting the aerodynamic effects on the ball, the range in meters will be {
=136
~152
~137
~149}
A baseball player hits the ball with an initial velocity of 60 km/h and an elevation angle of 30 degrees. Neglecting the aerodynamic effects on the ball, the range in meters will be {
~31,5
~25,5
=24,5
~18,5}

Figure 25. Example of two concrete test item in the GIFT format (Sanz-Lobera, González Roig, & González Requena, 2011)

Figure 26. Created example integrated (copied) into a LMS (Sanz-Lobera, González Roig, & González Requena, 2011)

The AEGIS system 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 (see also Figure 27 and Figure 28). 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 (see Figure 29).

⟨QUESTION SUBJECT="idioms"⟩
Data structures need to be studied ⟨DEL CAND="an,on,at,by"⟩ in ⟨/DEL⟩ order to understand the algorithms.
⟨/QUESTION⟩

Figure 27. Example of tagged data to create different types of test items (Mine, Suganuma, & Shoudai, 2000)

⟨QUESTION SUBJECT="W_S"⟩ *question region* ⟨/QUESTION⟩		
W_S	::=	word or symbol, where a backslash (\) must be added just before the symbol if it is a comma (,), double quotes ("), or a backslash (\).
⟨DEL CAND="CANDIDATE" LEVEL="PAIR" GROUP="ID" REF="ID"⟩ *hidden region* ⟨/DEL⟩		
CANDIDATE	::=	W_S | W_S,CANDIDATE
W_S	::=	word or symbol, where a backslash (\) must be added just before the symbol if it is a comma (,), double quotes ("), or a backslash (\).
PAIR	::=	LOW,HIGH
LOW	::=	an integer between 1 and 10
HIGH	::=	an integer between 1 and 10
ID	::=	keyword
⟨LABEL NAME="ID"⟩ sentences ⟨/LABEL⟩		
ID	::=	keyword

Figure 28. Overview on tags for test item creation (Mine, Suganuma, & Shoudai, 2000)

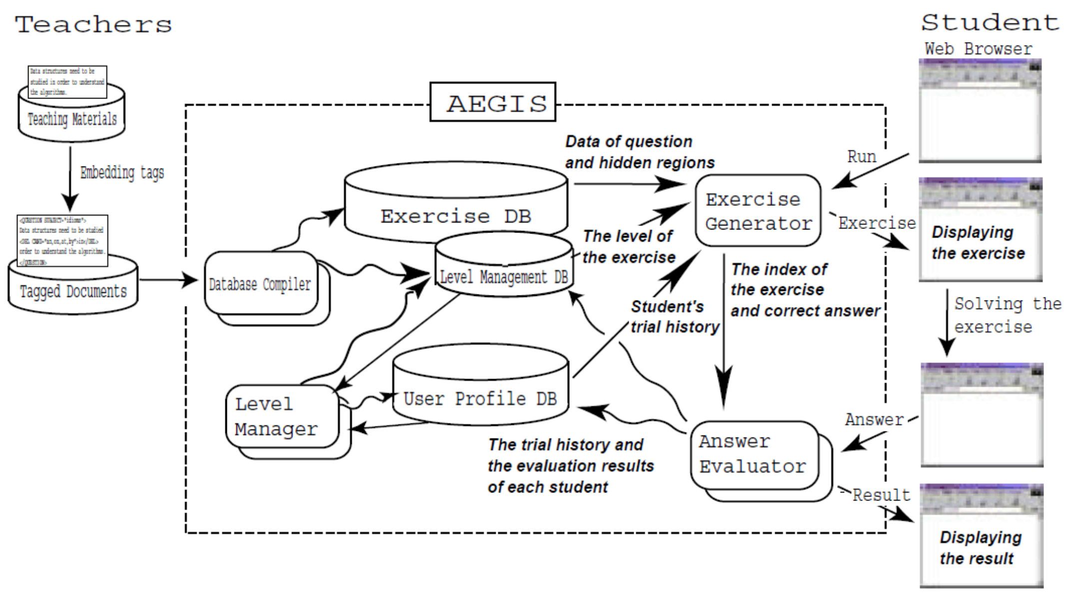


Figure 29. Overiew of the AEGIS system (Mine, Suganuma, & Shoudai, 2000)

The authors in (Goto et al., 2010) describe an approach and prototype to automatically create multiple choice questions from English texts for native or foreign language assessment. The learning/assessment environment is designed to receive texts from students and creates based on the text test items accordingly (see also Figure 30). This 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 (see also Figure 31). Figure 32 illustrated the flow into the system: textual input is tagged by a part of speech tagger, followed by the tree above mentioned process steps and finally a selected number of candidate distractors are selected and the test items are created.

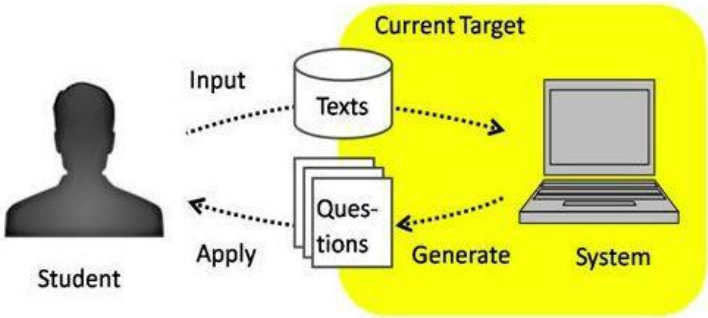


Figure 30. The learning and assessment environment at a glance (Goto et al., 2010)

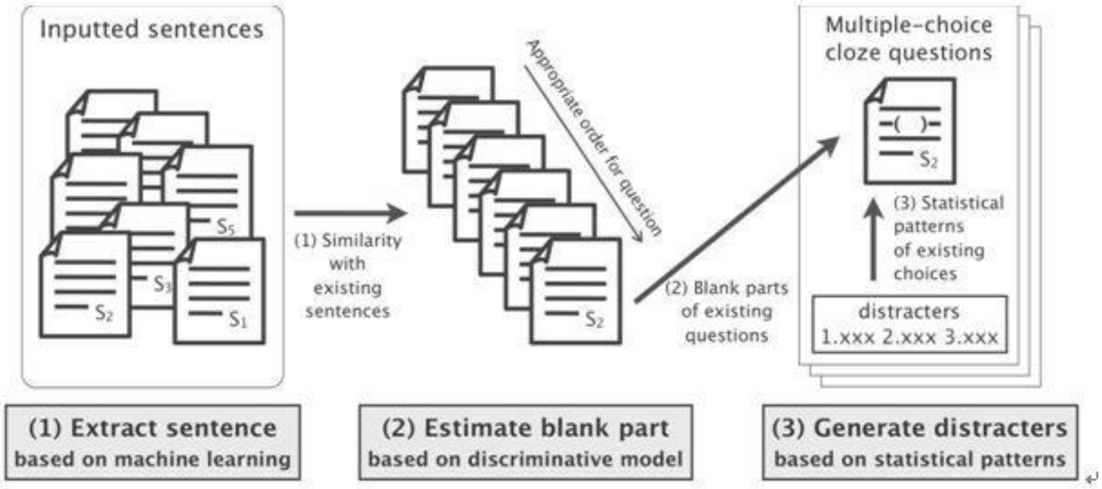


Figure 31. Overview of the proposed approach (Goto et al., 2010)

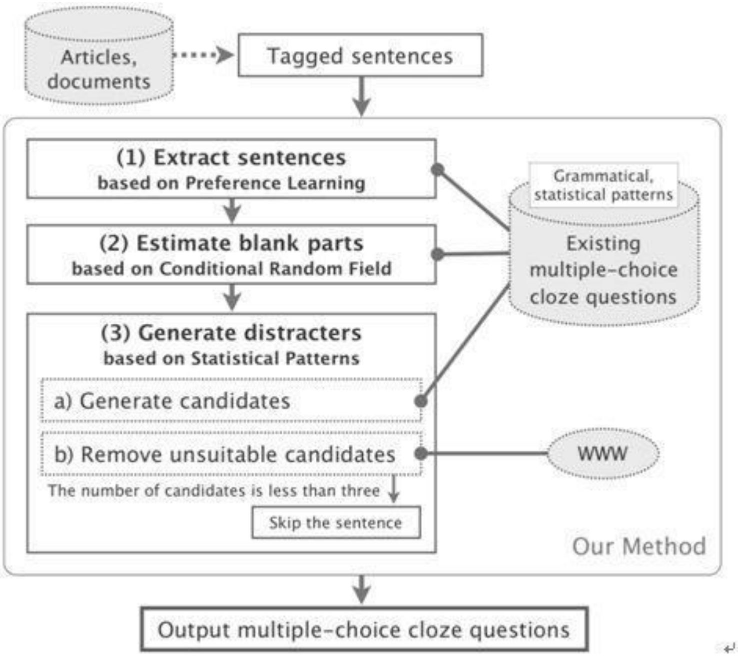


Figure 32. data and process flow (Goto et al., 2010)

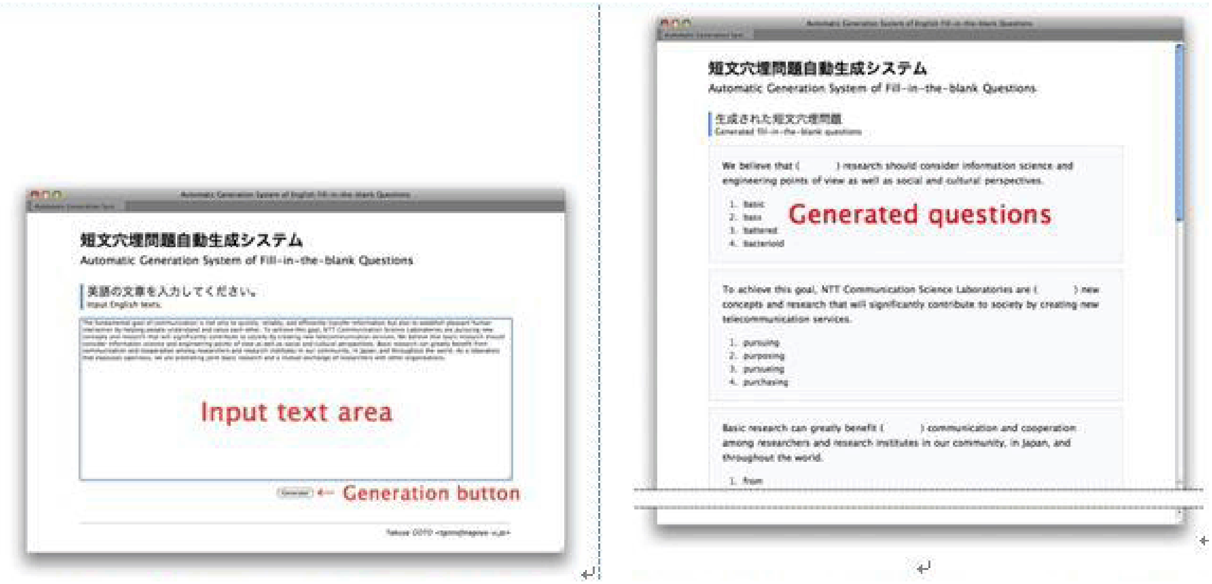


Figure 33a. Example of an input and created test items (Goto et al., 2010)

The authors in (Cubric & Tosic, 2010) extent the existing approach using ontologies to automatically create test items (such as described in (Papasalouros Kotis, & Kanaris, 2008; see also above) by the two following interesting aspects: (i) a meta ontology to model and create different question types, and (ii) a semantic interpretation on question types and respective levels based on Bloom's taxonomy. Finally, the approach applies question template for the test item creation process (see Figure 33a,b,c). The described approach make use of concepts and their "is-a" relationships only, a proof of concept is available as Protégé plugin.

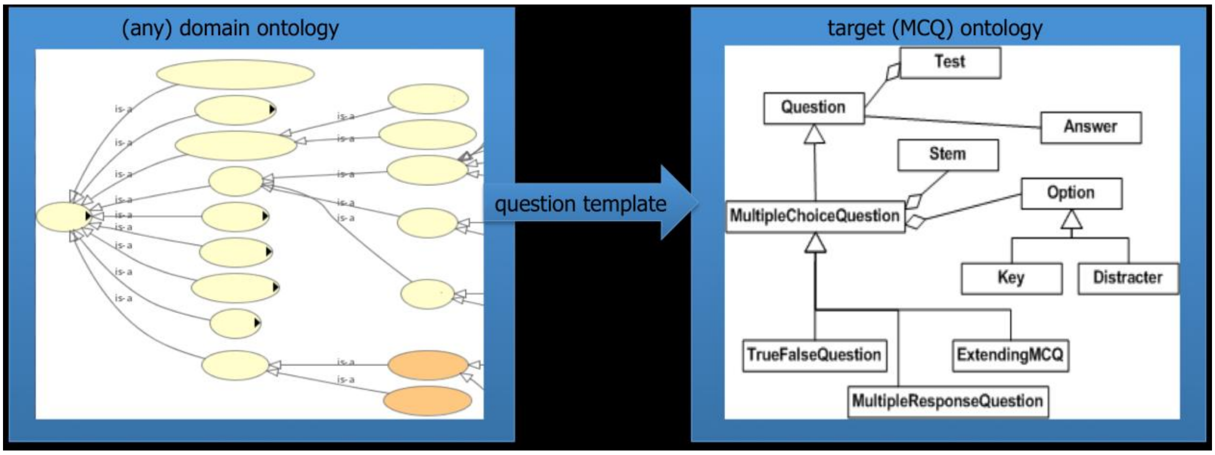


Figure 33b. mapping between domain ontology and meta ontology for different test item types (Cubric & Tosic, 2010)

Level	Question Stem	Correct answer
Knowledge	Which of the following definition describes the concept <A>?	any x where A:x
	Read the paragraph <x> and decide which one of the following concepts it defines	any A, where A:x
Comprehension	Which one of the following response pairs relates in the same way as <a> and <b> in the relation <R>?	(c.d) where R(a,b) and R(c,d)
Application	Which one of the following examples demonstrates the concept<A>?	any a, where A(a)
Analysis	Analyze the text <x> and decide which one of the following words is a correct replacement for the blank space in <x>	any A, such that x(A)
	Read the paragraph <x> and decide which one of the following concepts generalize the concept defined by<x>	any B, A<B and A:x

Figure 33c- Overview of question types and level (Cubric & Tosic, 2010)

The approach introduced by (Heilman, 2011) 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 (see also Figure 34 and Figure 35): (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.

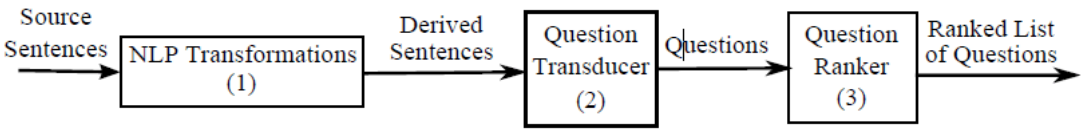


Figure 34. Three-stage conceptual architecture for automatic factual question creation (Heilman, 2011)

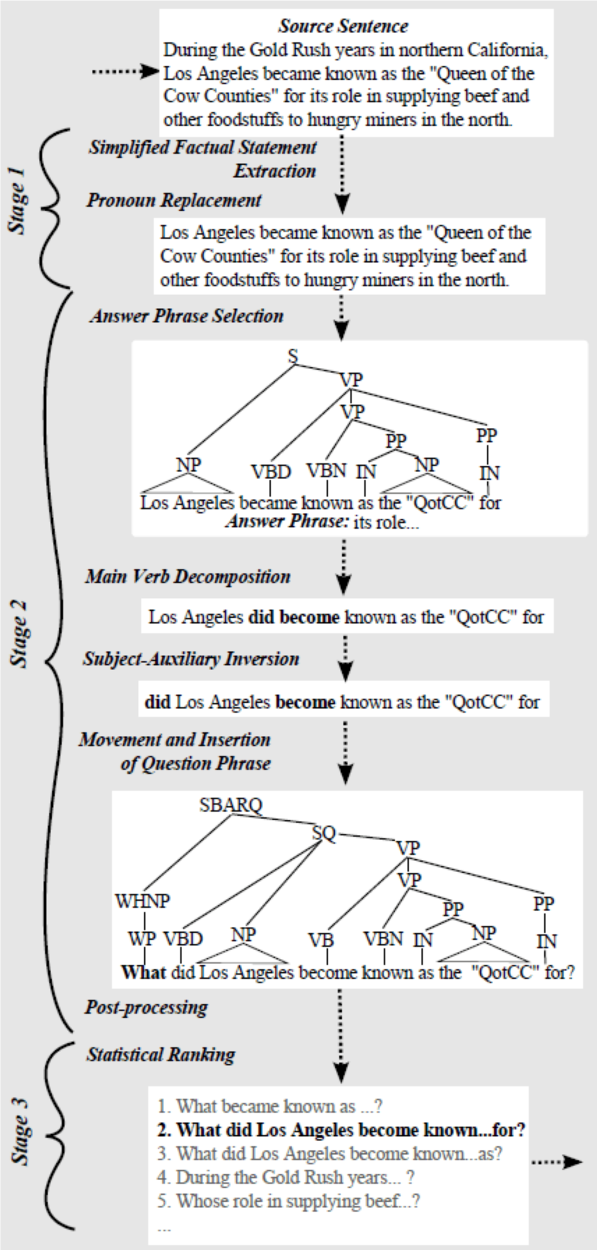


Figure 35. Simplified illustration of the process steps and data flow from input text to factual questions (Heilman, 2011)

System	Conten structure based	Seman-tic based	Ontology based	Required domain know-ledge	Interac-tive Mode	Automatic test item creation	Multi-lingual support	Web Service Integra-tion	Standard Compliant (QTI)
(Papasalouros et al., 2008)			✓	✓		✓			

(Sanz-Lobera et al., 2011)				✓		✓			
(Mine et al., 2000)	✓	✓		✓		✓			
(Goto et al., 2010)		✓				✓			
(Cubric & Tosic, 2010)			✓	✓		✓			
(Heilman, 2011)		✓				✓			
AQC	✓	✓			✓	✓	✓	✓	✓

Table 7. Features comparison among selected research of automatic test item creation

The automatic Question Creator (AQC) is a tool for semi-automated (interactive) and fully automated creation of various types of test items from learning content: this which 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, false/true, fill in the blank etc).

As depicted in Table 7 and referring to D5.2.1 and Gütl, Lankmayr, Weinhofer, & Höfler, M. (2011), competitive advantages include:

- Advanced tool supporting the **creation of four different test item types**: multiple choice questions, false-true exercises, fill-in-the-blank exercises, and open ended questions.
- **Learning setting dependent operating modes** supports fully-automatic test item creation and interactive process types taking into account student or teacher input.
- **Domain knowledge independent methods** allow test item creation of unseen textual content by applying statistical, semantic and structural analyses.
- 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.
- **Flexible design** supports an easy integration or exchange of modules in the system to offer improved processing tasks or even new features.
- **Easy integration into other systems and service provision** by a standard-conform web service interface.
- **Standard compliance** enables an easy export and reuse of test items created by the tool.

4.4.4 Assessment of 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, Koschmann, & Suthers, 2006). The emergence of Web 2.0 has supported CSCL with a variety of collaboration tools and software. Examples of such tools are e-mail, discussion forums, blogs, wikis, social networks, VOIP, and virtual words (Elliott, 2008; Crisp, 2007:181).

According to Elliott (2008), CSCL is pedagogically rooted to the social constructivist theories (Vygotsky, 1978) and the experiential learning theories (Kolb, 1984). 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 others 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).

However the pedagogical advantages of CSCL, integrating CSCL activities within a course must influence the assessment forms and procedures (Knight, 1995; Macdonald, 2003). Learning activities that encompass assessment tasks attract student's attention. 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 to record 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 has to be measured (Macdonald, 2003). Macdonald (2003) has distinguished between to 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.

According to Swan, Shen, and Hiltz (2006), the following issues have to be considered in designing assessment methods for CSCL activities:

- The variety of online collaboration domains and their learning goals: the authors provided some examples of application domains such as, collaborative construction of knowledge bases, group in game-based learning, peer-review of learning products, collaborative projects, and collaborative discussion groups. They also argued that according to the variance of the aforementioned domains assessment has to vary based of the learning goals.
- The complexity of both individuals and group assessment: the authors stressed on the importance of assessing both aspects of individuals and groups contribution. They depend on the research of Johnson & Johnson (1989, 1992) that recommended maintaining both *individual accountability*, in which students are responsible for their own learning, and *positive interdependence*, in which students satisfy learning goals if and only if the other students in the group satisfy them, in order to have a successful cooperative learning.
- Collaboration on assessment itself: as CSCL activities have to be linked with assessment forms to more engage and motivate students, it is also important to engage students in the design process of assessment and CSCL activities. The authors provided some examples of

collaborative assessment phases such as collaborative development of grading scheme, collaborative question composition, collaborative question answering, collaborative examinations, and peer, self-grading.

Learning in general is concerned to cognition and affect aspects where cognition is concerned with skills and processes such as thinking and problem solving and affect is concerning with emotional areas such as motivation, attitudes, and feelings. Affect is very important to the CSCL activities, aspects such as motivation and emotional state may influence the level of knowledge acquisition. 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 as a result may enhance their knowledge acquisition (Zumbach, Hillers, & Reimann, 2003; Daradoumis, Martínez Monés, & Xhafa 2006; Caballé, Daradoumis, & Xhafa, 2008). Issrof and del Soldato (1996) provide the following features of collaborative learning settings that are crucial for motivation (Jones & Issrof, 2005):

- *Social affinity between partners*: the partner level of respect and willingness to work together. Social affinity has a significant effect on the nature and effectiveness of collaborative interactions.
- *Cognitive ability:* both actual and perceived cognitive ability have an effect on the students' motivation in collaborative interactions.
- *Feedback:* feedback is crucial to motivation and directly affects the student's level of knowledge acquisition.
- *Distribution of control*: the collaborative learning environment has to be carefully designed to guarantee the balance of control amongst students. There are two aspects of control in relation to CSCL: control of student own learning, and control of the tool. Both of these aspects are crucial to motivation and may affect the collaborative interactions and product.
- *Nature of task*: how flexible are the tasks to be sub-divided and distributed over individuals.
- *Time:* students lose their motivation over time and as a result the nature of their collaborative interactions changes significantly.

Peer-assessment is often used as a component to assess CSCL activities (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 assessment software that can be used for group work assessment are: web-based self- and peer-assessment system for group work Web-SPA (Sung et al., 2005). Web-SPA is designed to allow individuals or groups to upload their assignments for self and/or peer-assessment and second phase of mark modifications based on group analysis. Another example is Self and Peer Assessment Resource Kit (SPARK). SPARK is an open-source assessment software designed to facilitate group work assessment (Freeman et al., 2002). Using SPARK instructors can set assessment criterion, individuals rate their peers' contributions against the assessment criterion, tutors grade individuals' contributions as well, and feedback is provided based on that. Assessment of CSCL activities is more concerned in assessing the activities that utilize the aforementioned CSCL tools and software; for instance, the assessment of discussion forums or assessing contributions and versions within wikis.

4.4.4.1 Assessment of Online Discussion

Discussion forums (also referred to an online discussion) are form of computer mediated communication (CMC) where students and teachers can interact in asynchronous manner. The assessment of online discussion posts can be either formative of summative; the assessment criteria should reflect the task goals and consider both qualitative and quantitative participations (Caballé et.

al., 2008; 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 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).

In order to have a successful online discussion, the online discussion activities have to be linked to assessment (Swan et al., 2006). But how to assess online discussion activity? Possible solutions can be either by online discussion content analysis 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), content analysis by categorizing the discussion post (Garrison, Anderson, & Archer, 2001; Hara, Bonk, & Angeli 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 in assessing online discussion are the work of Baron and Keller (2003), Pelz (2004), Swan et al. (2006), and Ho, (2004).

4.4.4.2 Assessment of Wiki-based CSCL

The term wikis was first introduced in the year 1995 by Bo Leuf and Ward Cunningham (Ebner, Kickmeier-Rust, & Holzinger, 2008). The first implementation of the wiki principle was in wikiwikiweb (http://www.c2.com/cgi/wiki) in 1995 by Ward Cunningham. According to Ebner et al. (2008) the original aim of the wiki principle is to have an easy-to-use knowledge management system that enables users to collaborate online efficiently and effectively. Wikis are websites that can be authored in a mass collaboration of user, where they are capable to add, edit, delete, and rollback to previous versions of the wiki-page. Moreover, email and RSS (Really Simple Syndication) notifications of page edits as well as pre-, post-comments of page content (Judd et al., 2010). Wikis has pedagogical advantages as well, wikis provide students with the so called “structured bulletin board” (Leuf & Cunningham, 2001) where they can reflect and receive feedback with easy-of-use, wikis support different learning styles as they form as “inherently democratic medium” (Leuf & Cunningham, 2001), moreover, they provide students with 24h/day 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).

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; Cole, 2009) none of the students enrolled in the course contributed to the wiki neither by editing nor by creating new pages, when the students are given the freedom to contribute or not. 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 (Judd et al., 2010).

However, 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). 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, 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 visualise 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.

4.4.4.3 *Collaborative Learning Systems (Sharda et al., 2004)*

Over the past four decades researchers, educators, and corporate trainers from many varied disciplines have explored using computer systems in teaching and learning and several areas of research and practice have emerged.

Computer-Supported Learning Systems have traditionally been labelled Computer-Aided/Assisted Instruction (CAI) systems. These systems contributed significantly to the use of computers in education. However, they traditionally focused on individual learners working on a local computer to accomplish cognitive learning objectives. Distance Learning, at its most basic level, is an extension of CAI to enable remote students to access course content.

Collaborative Systems are often referred to by the all-encompassing term “GroupWare” that was coined by MIS researchers Paul and Trudy Johnson-Lenz Circa 1980. Collaborative systems can range from email to online discussion groups and Internet chat rooms to sophisticated **Group Decision Support Systems**. Most Group Support Systems (GSS) research for education has involved same-time, same-place classroom situations.

The intersection of computer-supported learning systems and collaborative systems includes systems that extend Distance Learning by integrating collaborative learning and information technology, which is commonly referred to **Computer-Supported Collaborative Learning (CSCL)**. Many MIS researchers have used Group Support Systems (GSS) in the classroom to enhance learning, while others in IS and related fields have developed Asynchronous Learning Networks (ALNs). Combinations of these two system types have enabled affective learning objectives related to interactive communication and teamwork to be achieved, in addition to more traditional cognitive learning objectives.

The field of CSCL can be contrasted with earlier approaches to using computers in education. Koschmann (1996) identified the following historical sequence of approaches:

(a) Computer Assisted Instruction (CAIs)

(b) Intelligent Tutoring Systems (ITSs)

(c) Logo as Latin (LOGO)

(d) Computer-Supported Collaborative Learning (CSCL).

Collaborative Virtual Design Environments (CVDEs) use Virtual Reality to view and review complete systems, assembly processes, and individual parts. CVDEs provide realistic 3D displays and enable rotational capability for complete 360-degree visualization as well as views from top, bottom, inside, and underneath objects.

5 Standards in Modern Learning Settings

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. Specialists argue that conforming to standards during the design and development of our e-learning tools in general and e-assessment in particular may foster them with the following abilities (Shepherd, 2006).

- *Durability:* no need for further redesigns or redevelopments 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?
- *Accessibility*: are the contents accessible and deliverable from anywhere and anytime?

5.1 National organization standards and guidelines

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. Examples of those organizations are (Crisp, 2007), The Quality Assurance Agency (QAA) for Higher Education in UK (http://www.qaa.ac.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), an independent, not-for-profit national agency that promotes, audits, and reports on quality assurance in Australian higher education (http://www.auqa.edu.au). AUQA has published a database for good practices to quality assure Australian higher education. The Association of Universities and Colleges of Canada (AUCC) has published principles for institutional quality assurance in Canadian higher education (http://www.aucc.ca). The Higher Education Quality Committee (HEQC) is responsible to promote the quality of higher education in South Africa (http://www.che.ac.za/about/heqc).

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, The British Standards Institute (BSI) which published in April 2002 a standard named "BS 7988: a code practice for the use of information technology in the delivery of assessment" (www.bsi-global.com). 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 Electrotechnical Commission), and has been adopted by the Joint Technical Committee ISO/IEC JTC 1, *Information technology*. The Scottish Qualification Authority (SQA) has published guidelines for adopting computer-based assessment in further education (http://www.sqa.org.uk). SQA has collaborated with other organizations which regulate qualifications in the UK such as the Qualifications and Curriculum Development Agency (QCDA) (http://www.qcda.gov.uk), 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.

5.2 Professional association standards and 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, 2005). The following aspects have been discussed in the document:

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

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

3. Provide appropriate levels of control over CBT and Internet testing:
   - Detail the level of control over the test conditions
   - Detail the appropriate control over the supervision of the testing
   - Give due consideration to controlling prior practice and item exposure
   - Give consideration to control over test-takers authenticity and cheating

4. Make appropriate provision for security and safeguarding privacy in CBT and Internet testing:
   - Take account of the security of test materials
   - Consider the security of test-takers data transferred over the Internet
   - 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) principles of good assessment and feedback, developed as a result of the REAP project funded by the Scottish Funding Council during 2005–2007 (http://www.reap.ac.uk). 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?
- *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) benchmarking club (http://www.heacademy.ac.uk/resources/detail/ourwork/assessment/MESATool_Resource_Form). 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 (http://www.assessment-reform-group.org)
- Assessment Standards Knowledge Exchange (ASKe) http://www.brookes.ac.uk/aske)
- Centre for Excellence in Teaching and Learning in Assessment for Learning (http://www.northumbria.ac.uk/sd/central/ar/academy/cetl_afl)
- Institute of Education University of London (http://www.ioe.ac.uk)
- JISC TechDis (for guidance on inclusivity) (http://www.techdis.ac.uk)
- The Higher Education Academy (http://www.heacademy.ac.uk/ourwork/teachingandlearning/assessment)
- The Higher Education Academy Subject Centres (http://www.heacademy.ac.uk/ourwork/networks/subjectcentres).

5.3 Content standards, specifications and guidelines

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, 2009), 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).

Specifications and standards can be classified according to their level of approval into the following (Devedžic, 2006):

- 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).*
- De facto standards: *the same as the official one, but accepted only by the community and industry.*
- 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).*
- 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).*

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.*
- 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" (www.bsi-global.com). Moreover the Scottish Qualification Authority (SQA) has published guidelines for adopting computer-based assessment in further education (http://www.sqa.org.uk).

- Semantic Standards: *specify how we can organize content and refer to it in the semantic web. Such as Resource Description Framework (RDF) (http://www.w3.org/RDF), W3C Semantic Web Activity (http://www.w3.org/2001/sw/), Web Ontology Language (OWL) (http://www.w3.org/2001/sw/wiki/OWL).*

According to (Shepherd, 2006, p. 80) e-learning standards and specifications can be grouped into the following categories:

- Authentication: *specifications or standards on how systems can provide single-sign-on access to individuals and tools.*
- 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.*
- 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.*
- Data Transport: *specifications or standards that explains and describes how data can be transferred among systems.*
- Launch and Track: *specifications or standards that explains how content (courses, assessments, etc) can be launched and tracked by LMSs.*
- Metadata: *specification or standards that describes data-about-data which mainly used by LMS for content tagging so to facilitate content search and retrieval.*
- Philosophical: *specification or standards that represents a framework for describing the overall learning process, materials, services and tools.*

5.3.1 E-Assessment Content standards, specifications and guidelines

Despite the variance in e-learning content specifications and standards, e-assessment content has a limited number. 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 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.

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 36 summarizes the role of assessmentItem *and* assessmentTest *within authoring tools and learning management systems.*

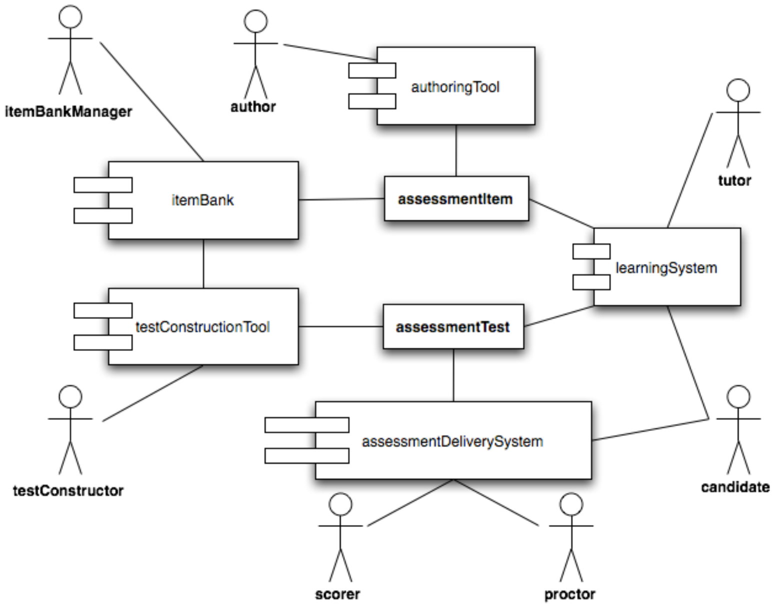


Figure 36. Assessment (item, test) role in assessment system (IMS QTI v2.1, http://www.imsglobal.org).

The IMS QTI information model as depicted in Figure 37 consists of two main data structures:

1. ASI (Assessment, Section, and Item) data structure: for assessment content representation.
   - *Assessment:* represents the test unit.
   - *Section:* is a group representation of sub-sections and assessment items that may share common learning objectives.
   - *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.
2. Results Reporting: represents the results from the candidate interactions.
   - *Context:* holds information session variables such as participant username, ID, and institution.
   - *Assessment Results:* used to report the results of candidate's interaction on both levels test (testResult) and item (itemResult).

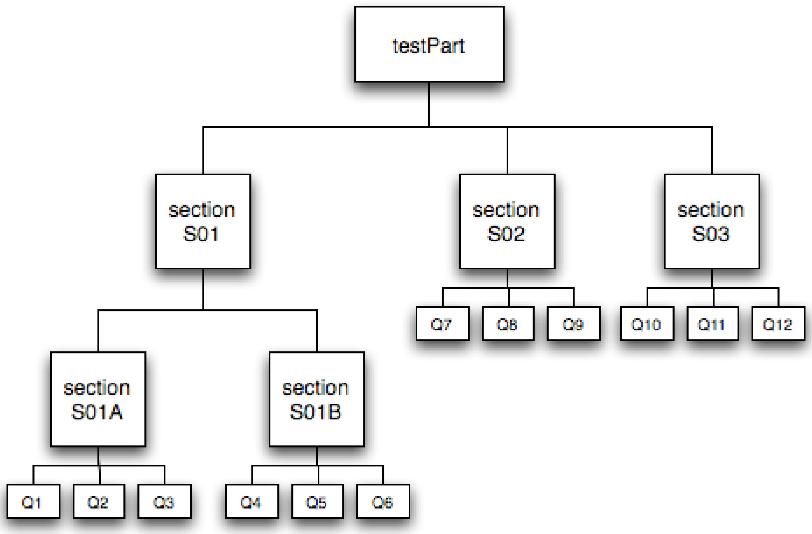


Figure 37. IMS QTI Test structure (IMS QTI v2.1, http://www.imsglobal.org).

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

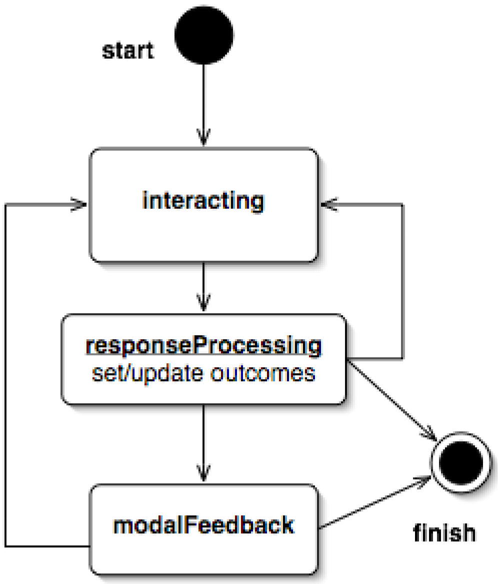


Figure 38. Response Processing based on candidate interaction(s) (IMS QTI v2.1, http://www.imsglobal.org).

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 38. 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.*

Based on the discussed interactions and response processing templates, QTI provides a set of different question types as summarized in Table 8. 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).

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 piece of text	textEntryInteraction	Map Response
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	

Table 8. QTI question types and their corresponding interaction type and response processing template.

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 39 explains the two types of feedback using a single-response question.

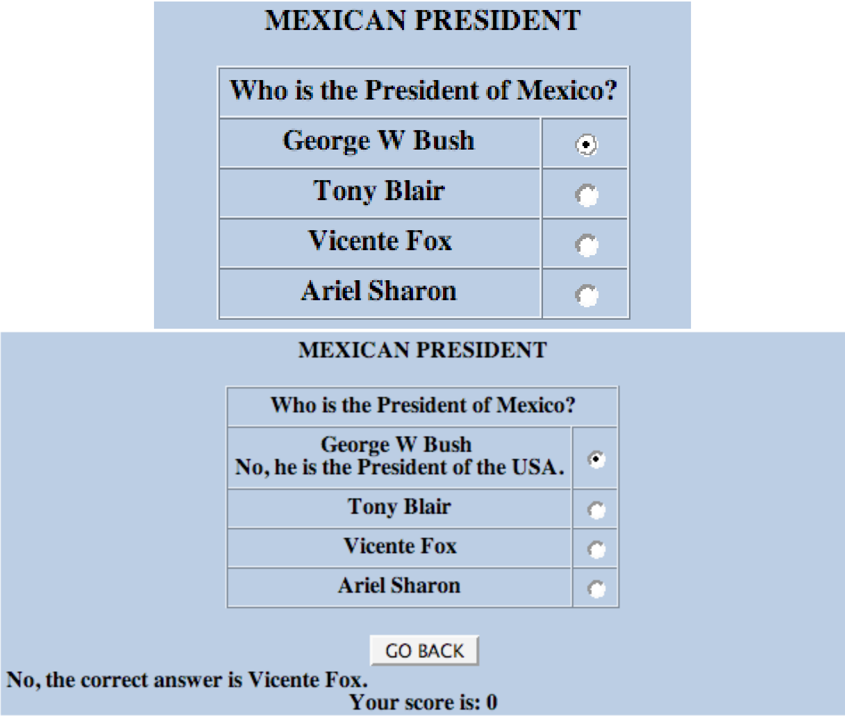


Figure 39. Both types of feedback with QTI-based item.

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, see Figure 40. Therefore, the design of items user interfaces is left to the software designer/developer. Some common practices is to 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 hotspot and graphical order items (require drag and drop objects), Java applets are attached to the XTML files in order to visualize and handle those interactions. The emergence web 2.0 holds a great promise of visualizing and delivering QTI items with more platform independent manner. QTI XML items can be mapped into HTML5 based files and visualized independently of the platform.

```xml
- <assessmentItem xsi:schemaLocation="http://www.imsglobal.org/xsd/imsqti_v2p1 http://www.imsglobal.org/xsd/imsqti_v2p1.xsd" identifier="choice" title="Unattended Luggage" adaptive="false" timeDependent="false">
  - <responseDeclaration identifier="RESPONSE" cardinality="single" baseType="identifier">
    - <correctResponse>
        <value>ChoiceA</value>
      </correctResponse>
    </responseDeclaration>
  - <outcomeDeclaration identifier="SCORE" cardinality="single" baseType="integer">
    - <defaultValue>
        <value>0</value>
      </defaultValue>
    </outcomeDeclaration>
  - <itemBody>
      <p>Look at the text in the picture.</p>
    - <p>
        <img src="images/sign.png" alt="NEVER LEAVE LUGGAGE UNATTENDED"/>
      </p>
    - <choiceInteraction responseIdentifier="RESPONSE" shuffle="false" maxChoices="1">
        <prompt>What does it say?</prompt>
        <simpleChoice identifier="ChoiceA">You must stay with your luggage at all times.</simpleChoice>
        <simpleChoice identifier="ChoiceB">Do not let someone else look after your luggage.</simpleChoice>
        <simpleChoice identifier="ChoiceC">Remember your luggage when you leave.</simpleChoice>
      </choiceInteraction>
    </itemBody>
    <responseProcessing template="http://www.imsglobal.org/question/qti_v2p1/rptemplates/match_correct"/>
  </assessmentItem>
```

Figure 40. QTI XML representation for Single Choice item

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 (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 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 cognition aspects, 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). Chang et al. (2004) have proposed 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 41 depicts the proposed MINE SCORM as an assessment metadata model.

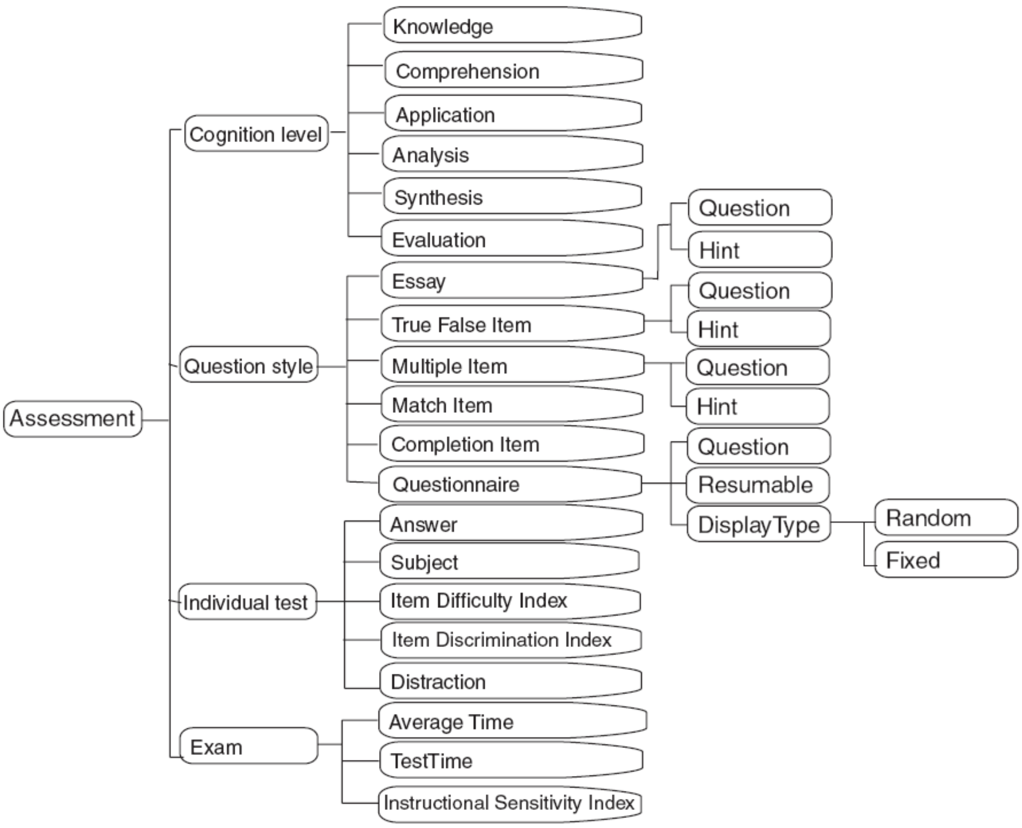


Figure 41. MINE SCORM e-assessment metadata model (Chang et al., 2004)

5.4 From Specifications to Standards

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 by that 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.

5.5 Benefits and added values of standards

In addition to the benefits we have discussed earlier of having a interoperable, durable, manageable, scalable, affordable, reusable, and accessible e-assessment system, specifications and standards can promote the process of e-assessment systems development with the following (Shepherd, 2006):

- Simplifying the overall process of e-assessment systems design and development.

- Supporting market vendors to more quickly provide interoperable products that are desired by customers.
- Maintaining the growth of the market by providing more robust and matured products.
- Facilitating the collaboration between to interoperable pieces of software.
- Maintaining longevity of products as more independent and reusable.
- Improving the available contents by using contents from standards-compliant systems.

5.6 Challenges of Releasing Standards

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; the first category is about specifications and standards in general which has two main factors:

- *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. (Shepherd, 2006).

The second category addresses the 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, 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 modeling. (AL-Smadi et al., 2009b)

5.7 Interoperability

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 arisen the limitation of these LMSs to cover 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 away form 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 LMS.

Several definitions have been provided to the term interoperability. The Oxford English Dictionary (http://oxforddictionaries.com) 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". He discussed the aforementioned two definitions and argued that both of them don not take integration into consideration. Moreover he stressed 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) 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: information (content, user data) level and on tools level (tools interoperability). Information interoperability has been a major research area for years. Several specifications and standards have been published. For content examples are IEEE 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 and emerging research where limited examples of specifications are available. Among these specifications we can mention the Open Knowledge Initiatives (OKI; http://www.okiproject.org) and its Open Service Interface Definition (OSID) (http://www.okiproject.org), 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.

5.7.1 OKI Open Service Interface Definitions (OSID)

The Open Knowledge Initiative (O.K.I.; http://www.okiproject.org) 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 42 the core of the architecture is the "institutional infrastructure layer" which includes file systems, databases, authentication servers etc.. 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.

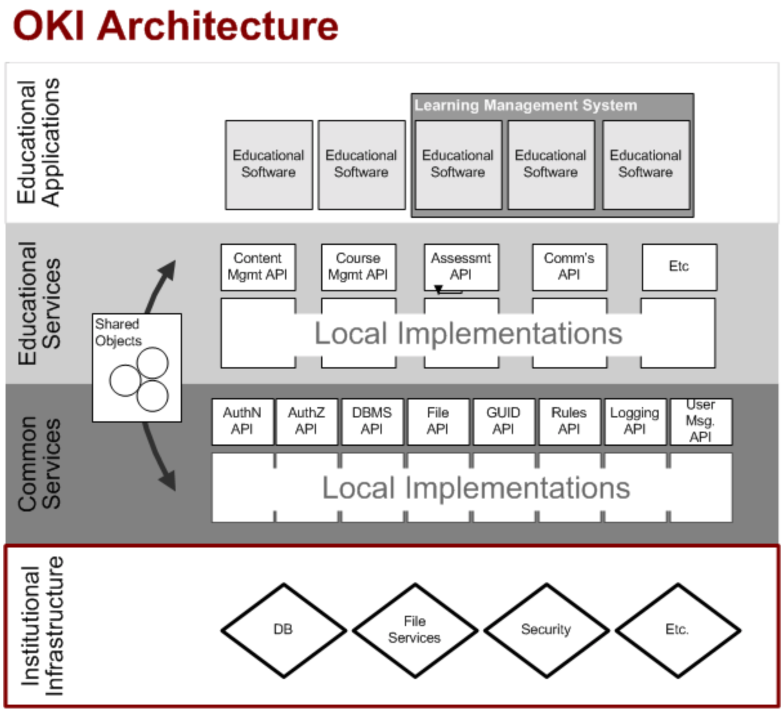


Figure 42. OKI Architecture

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 as follows:

- Flexible integration: simple integration with existing infrastructure.
- Tools and services interoperability: tools and services can be easily shared among different campuses and universities.
-

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# (http://sourceforge.net/projects/okiproject). The OSIDs has a variety of classes (http://www.grids.ac.uk/ETF/public/WebServices/classes.html) following some of them are mentioned:

- Common services:
    - Agent: a representation of a principal or actor known to a system.
    - Authentication: The Authentication OSID gathers required credentials from an agent, proves for their authenticity and introduces the agent to the system.
    - Authorization: The Authorization OSID allows an application to establish and check a user's privileges to view, create, or modify application data, or use application functionality.

- Dictionary: The Dictionary OSID provides a means to support multiple languages, domain-specific nomenclature and culture-specific conventions through interchangeable property files.
- Filing: The Filing OSID provides platform-independent means to handle files arranged in simple hierarchical containers.
- Hierarchy: The Hierarchy OSID manages parent-child relationships among elements. In addition to simple tree structures, the OSID supports hierarchy that is recursive and have nodes with multiple parents.
- ID: The OSIDs supports creating, storing, and retrieving unique identifiers (Ids)
- Logging: The Logging OSID records and retrieves a variety of application activity history
- Scheduling: The Scheduling OSID manages events in shared calendars.
- Shared: The Shared OSID contains fundamental objects used in the other OSIDs to provide their functionality.
- SQL: The SQL OSID provides relational database access functionality at a higher level of abstraction.
- User Messaging: The User messaging OSID supports communication and notification among users.
- Workflow: The Workflow OSID provides a way to manage an interdependent succession of activities each of which has completion constraints.

- Educational services:
  - Assessment: OKI Assessment OSID provides APIs for managing banks of items, sections, and assessments, and for publishing assessments.
  - Course Management: A Course management service allows applications or services to access and manage courses, modules and other units of learning.
  - Grading: Part of the Assessment process. A Grading service supports submitting grades against courses, modules, and other units of learning.
  - Repository: A Resource management service supports the management of finite physical resources, such as equipment and rooms.

An Example of a project that utilizes the O.K.I. OSIDs is the Campus project (http://www.campusproject.org). 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 43 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. It is worth mentioning that both Moodle and Sakai gateways and the OKI Bus are available as open source from (http://www.campusproject.org/en/download.php).

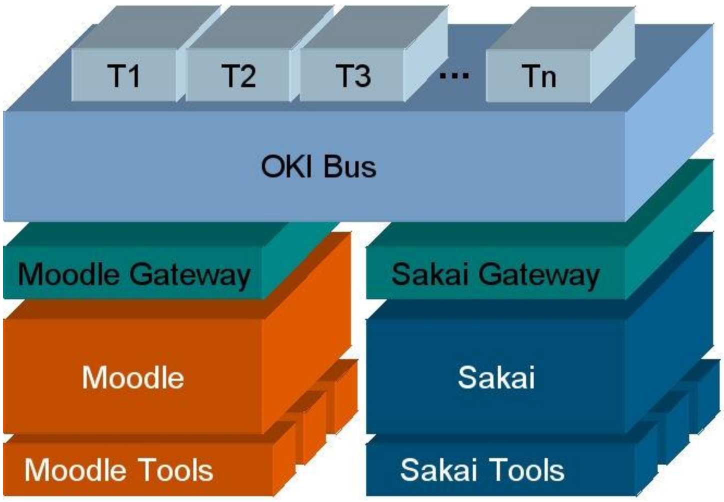


Figure 43. Campus architecture of layers and components

5.7.2 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) (http://sled.open.ac.uk/). 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 44, 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). 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 (http://tencompetence-project.bolton.ac.uk/).

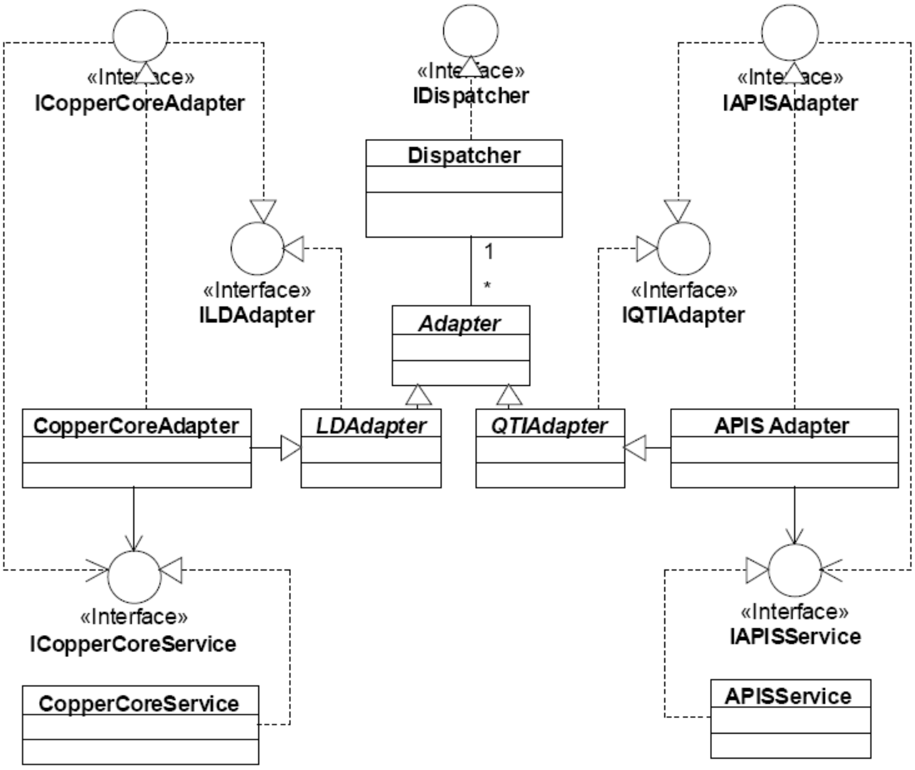


Figure 44. Integrating APIS with IMS LD using CopperCore Service Integration architecture (Vogten et al., 2006)

5.7.3 IMS Tools Interoperability

Tools interoperability concerns the ability of aggregating third-party tools to cooperate with a LMS platform. Third-party tools 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 Tools Interoperability Guidelines v1.0 has described this architecture as well as its main components (IMS TI, 2009). As summarized in Figure 45, the suggested architecture has introduced two main concepts:

- Proxy Tool: from its name this tool will be used by the LMS to communicate with the external tools. 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.
- Tools Interoperability Runtime (TIR): is a set of services that have to be implemented to the hosting environment (e.g. LMS). The TIR facilitates the deployment, configuration, and launching of the proxy tool.

TIR has the following services:

- *Deployment Service*: The main function of this service is to interpret and load the Proxy Tool definition into the host TIR via its deployment descriptor. Thus this service is also expected to perform validation of the Proxy Tool settings in order to ensure correctness of and compatibility with the hosting environment TIR.

- *Configuration Service*: The main function of this service is to manage the runtime settings of the Proxy Tool in order to provide the proper set of the same response for/during any given launch context.
- *Launch Service*: This service provides two main services, depending upon the context:
    - *Proxy Tool Host*: Performs all the functions related to launch of a Proxy Tool, including generating the relevant Proxy Tool launch message, utilizing the appropriate security profile, etc.
    - *Tool*: Exposed as a web service that accepts launch messages from the hosting environment TIR, understands the security profile used therein and responds back to the hosting environment TIR using the Proxy Tool core protocol as to the status of the launch.
- *Outcome Service*: This service provides two main services, depending upon the context:
    - *Tool*: A web service client that generates outcome messages from the Tool's TIR conforming to a specific outcome profile type, for a given interaction of a user with the Proxy Tool/Tool, including utilizing the appropriate security profile.
    - *Proxy Tool Host*: Exposed as a web service that accepts outcome messages from the Tool's TIR, understands the security profile used therein and responds back to the Tool TIR using the Proxy Tool core protocol as to the status of the outcome processing.

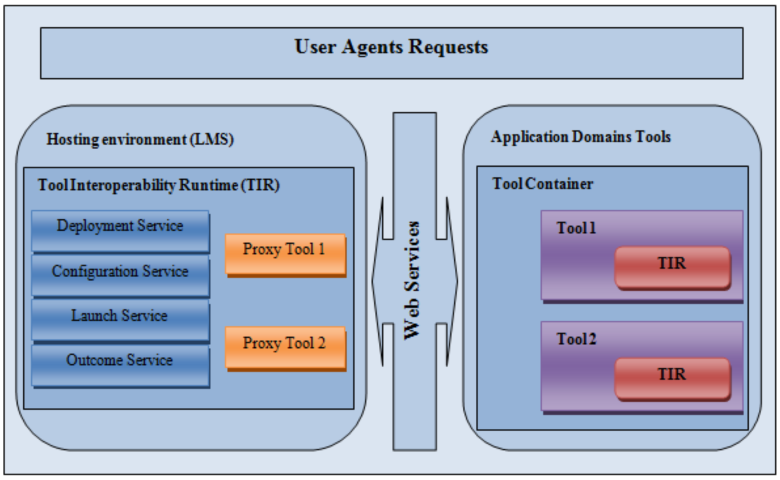


Figure 45. Tools interoperability architecture (AL-Smadi & Gütl, 2010)

The communication between the TIR/Proxy Tools is handled by a core protocol defined for this purpose. The core protocol is based upon the IMS General Web Services (GWS) specifications v1.0 (IMS GWS, 2009) which utilize XML with WSDL for defining services and 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).

5.8 Flexible Integration

In the world of Technology Enhanced Learning (TEL) extending the LMS services via 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 an interoperable tools and services:

- *Data & 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 & 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 represents 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 people in the domain 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).

Similar to that AL-Smadi and Gütl (2010) suggested the following requirements for a flexible e-assessment system:

- Clear guidance represented by a well-formed *framework.*
- *Standards and specifications* that represents the whole process of assessment as well as the communication between the services and components.
- *Cross-domain requirements* analysis in order to define the specific requirements for each application domain (such as, educational editor in the mathematic domain).
- *Web services* that provide the cross-domain requirements and interact through well-defined interfaces.

Dagger et al. (2007) discussed the flexibility and interoperability challenges for what he called "next-generation" LMSs. Dagger stressed on the importance of that LMS should exchange both information's syntax and semantics which goes in line with IEEE definition as systems have to be able to share information and to use them as well. Moreover, he argued 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. He also recommends a shared semantic view about services (such as Semantic Web) instead of APIs as in OKI OSIDs (Open Knowledge Initiatives Open Service Interface Definition) so that services can be easily selected, orchestrated, and consumed based on a common understanding of the learning process.

6 Game-based & Simulations

Modern electronic games have evolved to incorporate a wealth of components of interest to educators. Taking serious games to represent the state-of-the-art in the convergence of electronic gaming technologies with instructional design principles and pedagogies, in this section, we explore a range of assessment traits specific to electronic games and simulations. The focus of this consideration is to define how they may be capitalized upon both in the serious games developed within ALICE, and with reference to supporting background material, in a broader context. We do so by first describing the intrinsic relationship between assessment and feedback, and its particularly vital role in game-based learning. By reflecting on current pedagogic thinking with respect to game-based learning, we reach a range of conclusions regarding how assessment and feedback mechanisms should be designed and implemented in serious games. Subsequently, we reflect on a range of game elements from this pedagogic perspective, including leaderboards, competition, achievement systems, scoring, simulation, and mission structuring. Hence, this section links the broader discussion of assessment methods and mechanisms in previous sections to the specific case of game-based learning, and identifies the key opportunities presented by their integration. The work will build upon previous work undertaken by members of the Research Group (de Freitas & Oliver, 2005), where the four dimensional framework (4DF) has been developed to aid evaluation, validation and development of serious games. The framework outlines four dimensions: the learner dimension (user modeling and profiling, specified needs and requirements of learners), the pedagogic dimension (using associative, cognitive and situative models), the representation of the game (differing levels of interactivity, fidelity and immersion required to support the learning objectives) and the context within which learning takes place (including disciplinary context, place of learning and resources available). This mapping allows for end-to-end development, validation and evaluation of the game, using a participatory design approach model. The group has also developed an exploratory learning model (de Freitas and Neumann, 2009), the ELM model utilises and expands on the scaffolded learning approaches, extending the experiential learning model (Kolb and Fry, 1975).

6.1 Assessment and Feedback in Game-based Learning

Literature reviews of feedback (Shute, 2008, Mory, 2004) suggest the importance of formative models yet provide little conclusive evidence on how these models are best implemented. As described in previous sections, assessment as a whole must consider not only how to evaluate learners, but also how this evaluation may form a basis for feedback which induces a positive increase in learning transfer. With specific regard to serious games, we have explored in previous research the link between frequency and format of feedback and performance - for example in a control study of Triage Trainer (Jarvis and de Freitas, 2009, Knight et al., 2010), the authors identified a significant positive impact on performance when feedback was simplified and made more frequent. Though the sensitivity of training efficacy to context and learners precludes general conclusions, this work along with that of others (Mautone et al., 2008), supports the concept that a well-designed feedback mechanism is an integral part of effective game-based learning. In turn, this suggests that accurate assessment, as a core component of effective feedback, is equally vital.

Such a mechanism may take many forms; game-based feedback may be either conveyed directly to users as the outcome of a periodic summative assessment, or as immediate responses to errors, or in an evolutionary form which allows users to see a simulation evolve and interpret the consequences of their actions. We have previously explored game-based feedback in terms of Rogers' (1951) classification into evaluative, interpretive, supportive, probing, and understanding forms (Dunwell et al., 2010a), as shown in Table 9.

Feedback Type	Example for a score-based element	Technical demands (cumulative)
Evaluative	You got a score of 120/200	Measure variables
Interpretive	You got a score of 120/200 because you failed to respond quickly enough	Measure variables and model their relationships
Supportive	You got a score of 120/200, and need to improve your response times to challenges	Present and format measured data in a form relevant to the learner
Probing	You got a score of 120/200, because your response times were too low, was this because the user interface was too complex, or due to the game being too hard, or was it something else?	User interaction model and support for responsive dynamicism and adaptivity through (for example) intelligent agency
Understanding	You got a score of 120/200, because you found the user interface too complex, and as a result you responded too slowly to the challenges, you should complete the tutorial on the user interface	Link expert system to intelligent agency to determine root causes of failure

Table 9. Types of Game-based Feedback

As Table 9 illustrates, the simplicity of assessment variable itself, in this case a score, belies the technological complexity required in order to autonomously interpret it and feed it back to the user in a useful form. It is for this reason that we frequently see serious games more successfully deployed in blended contexts (Kirkley & Kirkley, 2007), wherein a human instructor takes responsibility for the higher levels of feedback, since generic techniques for autonomous feedback which satisfies the requirements of higher level responses remain beyond the state-of-the-art. However, as the techniques defined in previous sections demonstrate, methods such as the integration of semantic web services into learning environments can offer a basis for autonomising expert knowledge generation and application (Dunwell et al., 2010b). Furthermore, adaptive and integrative assessment techniques couple well with the use of the game engine as a tool for data collection – a key advantage of the use of an electronic game as a training medium is the ease with which interactions can be tracked, monitored, and logged in a holistic fashion.

Following Csikszentmihalyi's (1991) principles of flow, a further role of feedback is to support a balance between perceived difficulty of task and learner ability (Figure 46, taken from Dunwell et al., 2010a, originally adapted from Csikszentmihalyi, 1991). In this case, flow is defined as an optimal mental state for learners, in which they exhibit a maximum degree of focus on-task and learning transfer. Achieving this flow state requires that the level of perceived challenge must be closely matched to learners' self-perceptions of skill level – too simple and learners become apathetic or bored, too complex and they become worried or anxious. Flow is frequently considered with respect to serious games due to clear parallels between the behavioral traits of individuals undergoing a 'flow experience' described by Csikszentmihalyi (1991), and those of engaged gamers. Feedback is intrinsically related to learners' perception of challenge level, and therefore may be manipulated to induce flow. Parallels in leisure games are immediately observable: short-term attainable goals to engage and retain the attention of players, and adjustable or dynamic difficulty levels are frequently employed.

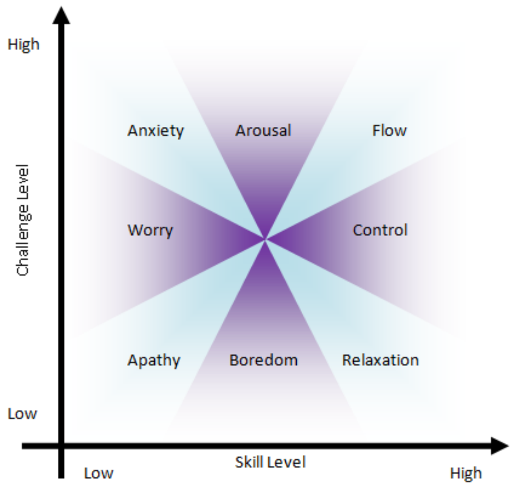


Figure 46. The Affective Relationship between Challenge and Skill

To summarize, although we can define with relative ease the *requirements* of any given game-based assessment and feedback system, as one which considers assessment and feedback in terms of depth, frequency, format, and challenge level, defining the mechanisms through which these requirements can fully be analyzed is a more complex task. Participatory design approaches are often advocated for serious games (de Freitas & Oliver, 2005), and certainly, given the lack of evidence on which to base prescriptive designs, end-user involvement throughout the development cycle is a key consideration in the development of any serious game. The well-established tension between 'fun' game and effective instructional tool (Zyda, 2005), requires that users not only be asked the simple question – "is this game fun?", but also be carefully assessed to establish the quality of the game as an intervention. This is compounded by the diverse range of desired learning outcomes encompassed by serious games, which range beyond Bloom's (1956) taxonomy from fact-based knowledge transfer, to problem-solving skills, to interventions with behavioural outcomes at their core such as *Re-Mission* (Kato et al., 2005). Hence, accurate assessment of learners is not only an integral part of a serious game, but also a common element of the development process behind it.

The concept of a 'pedagogic element', a composable, self-contained pedagogy purposed to a specific learning objective, is desirable in serious game design, since it can be more readily equated to technical features and functionality than higher level conceptual models. Moreover, it can be reused on both a conceptual and technical premise; integrated succinctly into a design addressing a clearly specified learning objective. ALICE notably considers the notion of 'complex' learning objects (CLOs), and a view we put forward in this section is that a CLO is effectively an instance of such an element. However, methods to achieve learning outcomes through serious games remain broadly untested to the extent that a learning outcome cannot typically be satisfactorily decomposed into such pedagogic, and hence technical, elements. What little empirical evidence exists is limited to the learners, context and technological medium that define the measured intervention. In particular the potential for game-based learning to provide feedback indirectly, or even tangentially, to user interactions is of conceptual note but extremely difficult to assess in practice. Bearman (1997), for example, demonstrates abstraction in feedback to be beneficial for learning of certain concepts, and a broader Vygotsky (1978) view of the learning process emphasizes its important role in early-stage development. Such a view plays well to the affordances of game based learning: many games are inherently abstract, and concepts central to Vygotskan theory such as more-able partners can be reborn as virtual characters, capable of synthetically and scalably providing the instructor dimension that technology-led training interventions have often notably lacked.

As a result of this aforementioned difficulty in showing conclusively the efficacy of any given approach, coupled with the wide range of potential game designs that can arise from a single learning objective, the state-of-the-art in assessment and feedback in serious games is predictably diverse. Thus whilst a single, notable, best-practice approach remains elusive, in the next section, we outline some common considerations in the development and implementation of serious games.

6.2 Key considerations in game-based learning

Educational environments have in the past been evaluated around variety of characteristics to determine the most desirable properties of a pedagogic element or learning object. However, the main purpose of an educational environment must always remain the *facilitation* of learning, rather than rote transfer of knowledge. Instilling intrinsic motivation to learn about a concept is generally more effective than extrinsically 'forcing' the knowledge transfer across (e.g. Kozma, 1991), and more likely to stimulate broader learning and greater retention. Following this logic, and excluding specific cases where intrinsic motivation can be fully assumed as independent of the game, or where learning objectives are narrowly defined and an extrinsic approach is valid, serious games should exhibit the following features:

- **Challenging tasks:** tasks should continually and appropriately challenge students to stimulate engaged learning. If a task is too easy or too difficult it could de-motivate the students. Task must be increasing in difficulty as student progress on the game and thus ensuring the students are engaged to the task (Wang & Kang 2006). This relates to the concept of flow mentioned previously; a learner left unchallenged will become bored, whereas one faced with an impossible task will become anxious. Providing such flow relates to learning profiling: content must be tailored to their ability level to provide constant challenge, and activities designed such that more and less able students do not hit difficulty thresholds below or above which the game is ineffective.

- **Interaction with others**: Collaboration is frequently shown to promote engaged learning experiences (Allen, 2003). Similarly, competition, though frequently warned against in classroom teaching, is an everyday element of players lives, and in particular common to leisure games. A key area, therefore, in which game-based learning differs from other methods is the need to consider this interaction in a competitive context, and capitalize on gaming elements which promote this behaviour.

- **Focused objectives:** Students should be aware of or discover which outcomes they need to achieve. This can be accomplished by defining a clear set of compelling goals that relate to the learning objectives (Quinn, 2005). Again, the nature of games lends itself to such objectives being clearly defined as 'quests' or 'missions' within the structure of the game.

- **Identity and role:** being a member of a community has a strong motivational influence on an individual's identity with their peers. It is shown that research emphasises the importance of the social factors of learning engagement (Rosenberg, 2001). Games often adopt a roleplaying approach, where the player is placed into another identity - in fact, and often as a technical than narrative decision, very few leisure games have the player playing themselves. This is of interest when considering feedback and assessment in serious games: is it the player being assessed, or their avatar? It has been hypothesised that presence, in this case used to describe the transfer of sense of self into the virtual space (Pausch, Proffitt, & Williams, 1997), is beneficial in scaffolding learning transfer from the virtual back to the real space, proven in the case of simple cognitive tasks, less conclusively so for behavioural outcomes.

- **Immediacy of feedback**: Students should receive feedback on their work throughout the process, which informs them of their progress towards the learning goals. As in education, feedback in games is important in providing players with timely and relevant information on their progress towards goals and identifying their level of achievement so far (Mitchell & Savill-Smith, 2004). Progress within the game will often be summarised in a map, and achievement indicated though ongoing game statistics, measuring attributes such as player skill, strength and health

- **Standard of Asssements**: the standards for assessing performance should be clear and are important to students (Schlechty, 1997). This may be a consideration in game based learning where related to assessment, since students enjoyment of the game may be overshadowed by concerns over assessment and monitoring of play. In the authors' experience, this has been the case in a game developed to address infection control in hospital wards, since nurses were concerned that poor game performance and survey response would have a negative effect on the perceived competency of the ward in controlling healthcare-related infection.

Given these considerations, we go on to view them in terms of resulting pedagogic elements which can be readily composed between technical implementations and provide a framework in which learning objects may be deployed. These serve as a high-level indicator of elements which may be relevant to ALICE, as it seeks to implement a game for civil defence training within an intelligent and dynamic tutoring environment:

6.2.1 Leaderboards

Gaming consoles and desktop computers offer access to online networks such as PSN (http://uk.playstation.com/psn), Steam (http://www.steampowered.com/), or Xbox Live (http://www.xbox.com/en-US/live), and both single and multiplayer titles typically offer online leaderboards on which users can track their progress and compare their performance to other players. The basic paradigm at the core of this concept is behaviourist: rewarding desirable behaviour and encourage engagement through regular feedback on relative performance measured against those on the same course. As such it is subject to the same criticisms leveled at serious games based in behaviourism, that learners may learn to defeat the game and best other players without addressing the learning requirements. This has been observable in a range of serious games, for example MathBlaster was shown to result in children with significantly improved ability at an irrelevant sub-task (shooting balloons) but less significance in improved numeracy, the target of the intervention (BinSubaih, Maddock, S., & Romano, 2006). We therefore consider the next point as a potential mechanism for aligning the learning outcomes with the game objectives, without compromising either the entertaining and intrinsically-motivating nature of the game, or its viability for use in situated, exploratory, or cognitivist learning.

6.2.2 Achievements

Leaderboards are frequently supplemented in leisure games by a 'trophy' or 'achievement' system which defines specific achievements in-game that a player needs to accomplish in order to add them to their collection. Such systems have been explored in a range of contexts, for example the application to photo sharing studied by Montola et al. (2009), who conclude that immediate and explicit feedback is a critical component of such a reward system. Their psychology in a pure gaming context has been broadly analyzed as a topic of interest for leisure game publishers; for example Weber and Shaw (2010) provide a qualitative analysis, suggesting that players can be grouped similar to Bartle's *achievers, explorers, killers* and *socialisers* (Bartle, 2003), examining the application of the various traits ascribed to these users in the case of achievements. For the achiever, this is relatively straightforward – they provide a structure for what must be achieved, define the means of doing so, and hence support the playstyle directly. For an explorer, the completionist element provided by a need to explore every niche of the game to gather all achievements is a suggested to be the driving force. For killers (defined by Bartle as players whose sole purpose is to kill or otherwise defeat all other players), then the competitive element – having more trophies than other players – comes to the forefront.

6.2.3 Competition

One of the most striking dissimilarities between games and traditional education is the broad inclusion of competitive elements. The competition itself takes many forms, including competition between players or computer-driven AI, and ranging from individual to team challenges, yet it remains, as Zyda (2005) argues, a fundamental component of any game. Competition is an intrinsic element in game systems and occurs routinely as a player pursues goals. Some goals may be achieved with ease while others require a higher level of skill that has to be developed (Salen & Zimmerman, 2003). The struggle of players to complete these goals, either in opposition or collaboratively, encourages engagement by building on their natural competitive drive.

It can be immediately seen that serious games offer a multitude of avenues for assessment, though a paucity of empirical research into their various efficacies means a highly predictive, prescriptive model for developers remains elusive. Furthermore, the efficacy of any given approach is tied closely to the target demographic, usage context, choice of technology, and underlying pedagogy (de Freitas and Oliver, 2007), and hence an attempt to evaluate any given model typically results in evidence which lacks composability when transferred to other groups of learners, subject matter, and situations.

6.3 Summary and conclusions

In this chapter, we have briefly presented the state-of-the-art in assessment and feedback in game based learning for review within the ALICE consortium as well as for public dissemination. We have noted that ubiquitous and conclusive models for best-practice in serious game design and development remain elusive, a consequence of a lack of generalisable evidence. Certainly best-practice methods for software engineering are applicable to the development of a serious game, though experience has shown that this needs to be delivered concurrent to a rigorous programme of participatory and user-centric design if success is to be guaranteed. It is the costs associated with such an approach, which must by nature deliver frequent prototypes to users and make substantial revisions as a result, that is a major barrier to the more widespread deployment of serious games to address the challenges faced by education as learners become increasingly disengaged with traditional methods of teaching and learning.

7 An Integrated Model for E-assessment

7.1 Purpose

The purpose of this chapter is to define an integrated model for e-assessment. We will first describe the most important objectives of ALICE (Section 7.2) in order to motivate the need for such a model. Based on this, we will describe the model and its components in Sections 7.3 to 7.4 as well as a framework for the use of the model in Section 7.5. The results of a first application of the model in a CLR are outlined in Section 7.6.

7.2 Objectives of ALICE

As described in the proposal of the ALICE project, e-learning systems can support traditional (face-to-face) teaching and be used when no such teaching is available or appropriate. E-learning also gives greater freedom to students in relation to when and where they study, and there are significant advantages for teaching administrators. Moreover, it is important to note that the profile of the learner has changed in recent years. With the advent of the so-called "information age" there is an expectation that the workforce will adapt their skills or even change careers to keep in step with technological advancements. Hence, learning is a lifelong process with many returning to education to retrain. Also, since the early 1990's, the proliferation of technology means that students have grown up with computers, MP3 players, mobile phones and digital games. This new generation is variously described as "Digital Natives", "Net-geners" and the "Nintendo Generation" (OECD, 2008; see also Section 3.5). The new learner has so different needs that have to be addressed if e-learning is to be successful. First, the new learner expects to be in *control of his or her learning experience* while in a supportive, collaborative and simulative environment. Thus e-learning systems should promote self-directed learning that is also more motivating. Unfortunately, many e-learning systems have a linear structure with a single path through the learning. While this design is cost-effective, the lack of choice reduces control of the learning experience. Second, although current e-learning systems allow learners to move at their own pace it isolates them from their peers participating in the same learning process. This inhibits the learning achieved through *social interaction and collaboration*, with some learners feeling "lost". Research indicates that a sense of belonging to a social group improves motivation and effective learning overall. Third, an *authentic learning experience* is necessary. Learners expect the material to be linked to prior knowledge and be relevant to their everyday lives and careers. Generally, learners are more engaged when they are participating in activities that they can relate directly to prior knowledge and make connections between what they are learning and the real world. If such links are missing, learners are less inclined to participate in the learning process and may see it as pointless.

ALICE aims at building an innovative adaptive environment for e-learning combining personalization, collaboration and simulation aspects within an affective/emotional based approach able to contribute to the overcoming of the quoted limitations of current e-learning systems and content. In other words the proposed environment will be interactive, challenging and context aware while enabling learners' demand of empowerment, social identity, and authentic learning experience. The defined system will be able to effectively involve learners in educational, cultural and informative activities in two specific contexts: university instruction (with particular emphasis on scientific topics) and training about emergency and civil defence (as for example the behaviour to take at a personal and collective level when the treat of a big risk shows up e.g. a natural event like earthquake, or a fraudulent one like terrorist attack).

The ALICE project intents to answer the following questions:

- How is it possible to create collaboration conditions and therefore to encourage the learner to choose a collaborative-type education also when collaboration is actually difficult?
- How can the effectiveness of learning actions be supported by interactive simulations and serious games that may be created with low costs thanks to techniques of reusability?
- In what way can the storytelling be integrated with Learning Experiences having contents of different types?
- Eventually, how to create a learning additivity related to the earlier themes, being not the simple sum of various aspects, but a real integration and subsequent super-additivity with respect to single components?

Nowadays no implementation, except for some experiments and attempts limited to single aspects among those indicated, is able to offer a complete methodological-technological-industrial solution covering all these aspects. This include contents and tools of simulative, collaborative, storytelling and assessment type, as well as the ability to manage the emotional-affective responsiveness, creating personalised learning paths that take into account the learner profile from the points of view of both learner preferences and the emotional reaction to given stimuli.

Consequently, the motivation of ALICE is to build a learning environment enabling new forms of learning that are dynamically adaptive and enhance the following basic aspects (see Figure 47).

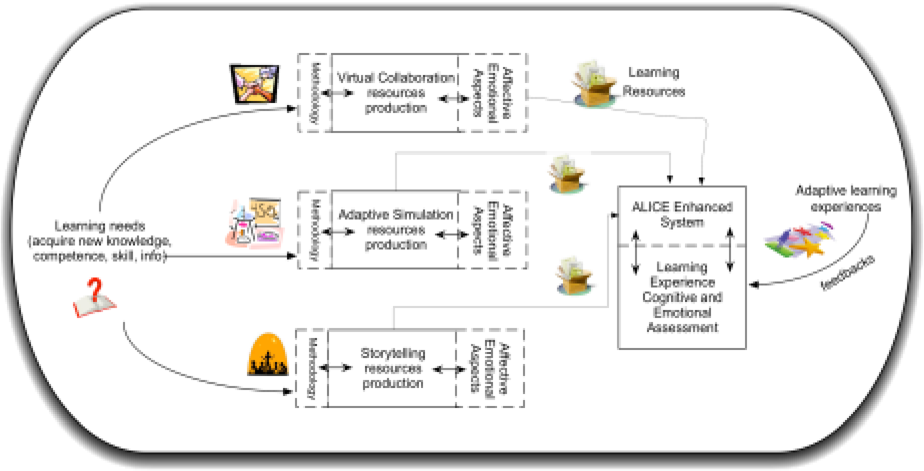


Figure 47. Providing an adaptive learning experience in ALICE

- *Collaborative learning* (live and virtualized collaboration);
- *Simulation and serious games* (experience, simulation and intuition driven learning);
- *Storytelling* (exploitation of storytelling role in learning experiences);
- *Affective and emotional approaches* (capability of intercepting emotional and affective aspects and to manage and adapt them with respect to the educational offer)
- *New form of assessment* (development of an advanced assessment system that is not only accurate with respect to content, but also looking at the educational process meaning as a whole).

7.3 An integrated model for enriched learning experiences

Assessing students' learning by using information technologies has become very popular within the recent years. E-assessment can be *computer based* (i.e., test delivery and feedback provision is done by the computer) or *computer assisted* (i.e., the whole process of assessment involving test marking, analysis and reporting is covered), although both terms are often used interchangeably (Charman & Elms, 1998; see Chapter 4).

In Chapter 4 we presented a range of frameworks and models which either discuss key elements for assessment in general (e.g. Chudowsky, & Glaser, 2001) or emphasize specific aspects of the assessment process (e.g., Almond, Steinberg, & Mislevy, 2002). 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 (see Section 4.4). Because the aim of Work package 5 of ALICE is to develop new forms of assessment, an integrated model for e-assessment (IMA) is needed that addresses the requirements arising from such new forms. According to the goals of ALICE (see Section 7.2) the model is expected to renew the approaches of e-assessment in a structural (i.e., testing approaches), adaptive (backward feedback guided), and distributive level. Moreover it should represent complex learning resources like simulations, collaborative experiences, virtual experiences, and storytelling as well as emotional elements and it should be able to evaluate the results of those learning experiences. In the next sections we will define such a model and discuss its components[5].

In the following an updated version of the model is presented (see Deliverable 5.1.1 for the original model). After a first evaluation by an expert from the field of cognitive sciences, as well as a first round of experimentation several small changes within the overall model have been necessary. Furthermore, we added a chapter on assessment (Section 7.4) in order to present this central part of the model in more detail as well as a framework of how to use the model in real applications (Section 7.5). The already conducted studies also allowed us to give an example of how the model can be used in a real learning scenario. Section 7.6 shows how the single components of the model can be applied in a self-directed learning course with a co-writing Wiki assignment. The revised model together with the added assessment part was then given to five more experts who evaluated the model as well as the tools (Automatic Question Creator AQC and co-writing WIKI) developed in WP5. The detailed results of this validation by experts are reported in Section 7.7. Finally, Section 7.8 through 7.10 deal with the special requirements arising in the context of the three main learning forms we are dealing with in the ALICE project, namely collaborative and social learnin, serious games, and storytelling.

7.3.1 General Model

Figure 48 shows the abstract level of a model that addresses the requirements of the ALICE project. In general, the model represents an enriched learning experience that consists of four main components: This core methodology includes the *didactical objectives*, *complex learning resources*, *assessment activities* (including feedback), and indicators for its *evaluation and validation*, respectively. Results from the validation and evaluation processes can again influence the first three components. Thus the development of efficient learning environments should be seen as cyclic process, which is open to improvements. The enriched learning experience is influenced by several components like *educational and psychological aspects*, *technical issues* and *existing standards and best practices*, respectively. 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 addresses all components of the enriched learning experience is also considered in the model. The quality assurance is also relevant with respect to indicators that are expected to result from the enriched learning experience: indicators for its *educational efficiency and effectiveness*. Finally, in order to ensure that the learning experience allows *adaptivity*, the model also interacts with three other important models: the learner model, the knowledge model, and the didactic model, respectively. In the next sections, components of the model are discussed more detailed.

[5] Note that according to the time schedule of ALICE, this model will be refined later on during the project.

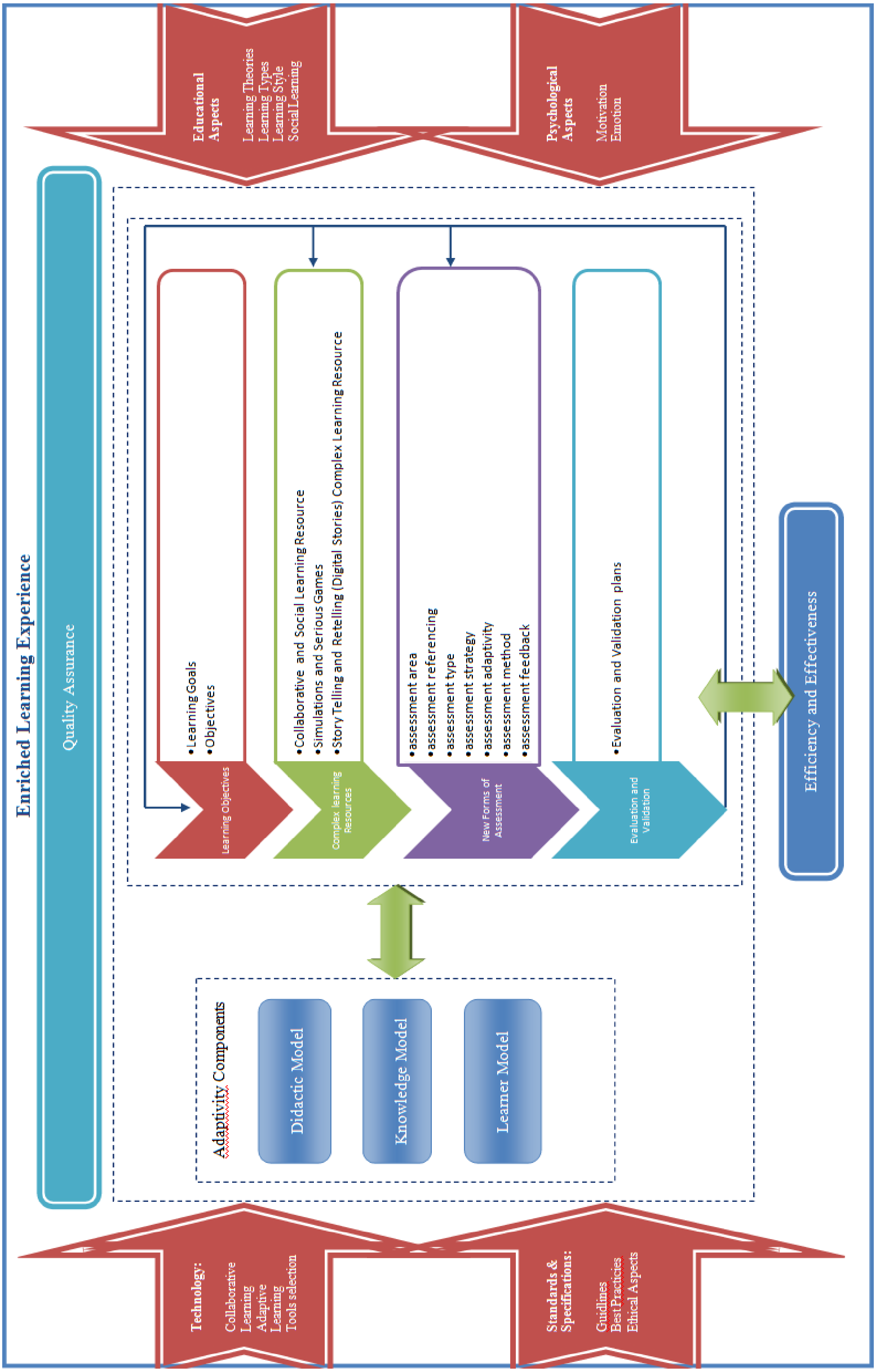


Figure 48. Integrated model of e-assessment (abstract level).

7.3.2 IMA-Core Methodology

7.3.2.1 Learning objectives

The first component of the model 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. However, there might be some relating didactical objectives during the learning process that are not immediately linked to the learning goals in a narrower sense. Such further objectives could be e.g., gaining social competences (due to collaborative work) or meta-cognitive skills (due to self-regulated learning activities). Because those skills might also be very important it is necessary to consider them already at the beginning of a learning experience. Moreover, it has to be mentioned that didactical objectives (consisting of immediate learning goals and related learning objectives) affect the type of learning resources and assessment activities that are chosen in an enriched learning experience.

For instance, according to Bloom's Taxonomy, learning goals can be divided into the following six levels: *knowledge*, *comprehension*, *application, analysis*, *synthesis*, and *evaluation* (Bloom, 1956; Krathwohl, Bloom & Bertram, 1973; see also Chapter 1). To define the criteria when a learning goal has actually been reached is not easy. For instance, one learning goal might be that a student is able to apply the knowledge he or she acquired during the learning activities. How does the instructor know that this specific learning goal has indeed been reached? Thus, the successful achievement of a learning goal has to be operationalized, i.e., a criterion has to be found that ensures that it is measurable. Obviously, operationalizations of a learning goal are the underlying assessment activities. For instance, if the learning goal is the appliance of knowledge, a simple knowledge test will not provide any information whether the learning goal was reached successfully or not. Rather, assessment activities should include a task where the student has to apply the knowledge in this case (for instance, to apply a previously learned formula). Furthermore, these assessment activities may also depend on/affect the second component of the enriched learning experience: the complex learning resources.

7.3.2.2 Complex learning resources

To address the needs of an "active learner" who is actively involved in the learning process (see Chapter 1), an enriched learning experience is generated made up of complex learning resources (CLR). According to the aims of the ALICE project, those CLR are expected to add moments of *collaboration*, *simulation*, and *storytelling* in order to support the learners in achieving the learning objectives. Of course there are other complex learning resources such as problem-based and project-based learning. Nevertheless, in the following we focus on the three mentioned aspects, which are relevant in our case.

- *Collaboration.* Collaboration can enhance the learning efficacy because people learn from one another (e.g., due to observational learning, imitation, and modelling; see also Chapter 2). Moreover, due to collaboration, learners can be supported in the achievement of specific skills (e.g. communication, problem solving, decision making, etc.)

- *Simulation and Serious Games.* Serious games are intuitive learning systems used to train (individual but also groups of) learners while achieving their optimum potential and include all aspects of education (teaching, training and informing). We have already discussed serious games and simulation as a useful method of learning in Chapter 5.

- *Story Telling.* 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-centered teaching approach able to involve learners emotionally, provide guidance and support reflection.

7.3.2.3 New forms of assessment

Assessment activities represent the third component of the enriched learning experience. There is a need for new forms of assessment which meet the high demands arising from the CLR. Therefore, innovative assessment activities as outlined in the ALICE proposal are considered to be based on Bloom's taxonomy of educational objectives (knowledge, comprehension, application, analysis, synthesis, and evaluation) and on effective kinds of learning such as reflective learning, experiential learning, and socio-cognitive learning (see also Section 7.3.3). As our focus is on the assessment activities that are embedded within the enriched learning experience, Section 7.4 gives an overview about the aspects innovative forms and types of assessment activities should cover.

7.3.2.4 Evaluation and Validation

To ensure that learning and assessment activities have a high quality standard, these activities should regularly be evaluated and validated. *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 and hence, all students passed. Therefore, evaluation criteria should consider best practices and standards as described in 7.3.3.4 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 also be validated regularly in order to ensure their validness. In Section 3.6 and Chapter 5 we presented an overview for quality criteria that may be considered. These include quality criteria from a psychological point of view (objectivity, reliability, and validity), as well as technical standards and guidelines.

Results from those evaluation and validation processes form valid indicators for the quality of the enriched learning experience in order to adapt to/enhance the learning experience. Adaptations might concern the chosen assessment forms, the complex learning scenario, or the learning objectives themselves.

7.3.3 Inputs to the enriched learning experience

An enriched learning experience is affected by several components such as *educational* and/or *technical aspects*. Also *psychological aspects* such as motivation and emotion are expected to influence learning experiences. Here we present influences arising from these educational, psychological, and technical aspects and discuss them with respect to the CLR and assessment activities. It has to be mentioned that there is some reciprocal relationship between these educational and psychological components and used technologies (although not added to the graphic): Educational aspects should be considered in the development of technologies and the development of new technologies certainly affects educational, i.e. psychological and pedagogical theories of learning.

7.3.3.1 Educational Aspects

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 (see Chapter 2). For instance, Blooms' Taxonomy (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 learning. In this context, also learning processes of social entities like classes or communities play an essential role. Thus, social learning theories and frameworks for collaborative and social assessments have to be considered.

7.3.3.2 Psychological Aspects

Other important issues are motivational and emotional aspects during learning and assessment. We have presented relevant theories and models in Chapter 2. Due to the measurement of the motivational and/or emotional status of a learner he or she can be supported in a suitable and personalized manner in order to enhance his or her affective inclination and hence, to stimulate the learners' attention and learning.

7.3.3.3 Technological aspects

From a technological viewpoint, learning and assessment activities can be supported in many ways. In Chapter 4, we have already discussed some examples for e-assessment software. Therefore, enhanced technologies are necessary that not only allow generating CLR but also flexibly adapting the learning path with respect to the individuals needs and learning progress. One aim of ALICE is to develop innovative e-assessment tools that support assessment activities in the enriched learning experiences. Those tools will not only consider the assessment of individual 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.

7.3.3.4 Standards and Specifications

Standards for e.g. learning content reusability and interoperability, learner's information accessibility and share ability are essential for any learning management system (including e-assessment) and therefore, also for quality assurance (see Section 7.3.3.6). Different standards and best practices have been developed to design and develop e-learning content and components. In Chapter 5 (see also Section 3.6) we have already presented state-of-the art guidelines and best practices when developing and conducting assessment activities.

7.3.3.5 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 between what teachers want to teach, how they teach, and how they assess students' performance. Kellough and Kellough (1999, quoted after Buzzetto-More & Alade, 2006) posited that one aim of assessment is to improve teaching effectiveness. Hence, during the evaluation process, effectiveness and efficiency should be considered. Checklist for educational efficiency and effectiveness can be found for instance at https://www-internal.jcu.edu.au/internal/groups/public/documents/strategic_plan/jcuprd_055358.pdf .

One 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 CLR and assessment might be meaningful. When selecting an assessment tool, both CLR and didactical objectives have to be considered. For instance, did learning occur during a collaborative activity or not? Should there be an individual assessment, or a group assessment, or a 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 appliances based on the knowledge they acquired?

7.3.3.6 Quality assurance

Learners profit from an enriched learning experience mostly 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 meets the requirements. The quality can be assured when several aspects are considered.

Learning and assessment activities should consider the state-of-the art of best practices and standards in the field. As discussed in Chapter 5 (see also Section 3.6), such guidelines should be consulted when assessment is generated but also when it is delivered, scored, and interpreted (e.g., 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, behavior 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 (See Section 7.3.5). 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).

7.3.4 Interaction with other models

In order to provide adaptive and personalized learning, the IMA is interacting with three other models, namely the *Learner model*, the *Knowledge model*, and the *Didactic mode* (Capuano et al., 2009). In co-operation with the learner model, the cognitive status of the learner in terms of knowledge and skills is updated, in co-operation with the knowledge model the ontology of learning and in co-operation with the didactic model, eventual alternative models are recovered.

- The *Knowledge Mode*l 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, etc.
- 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.

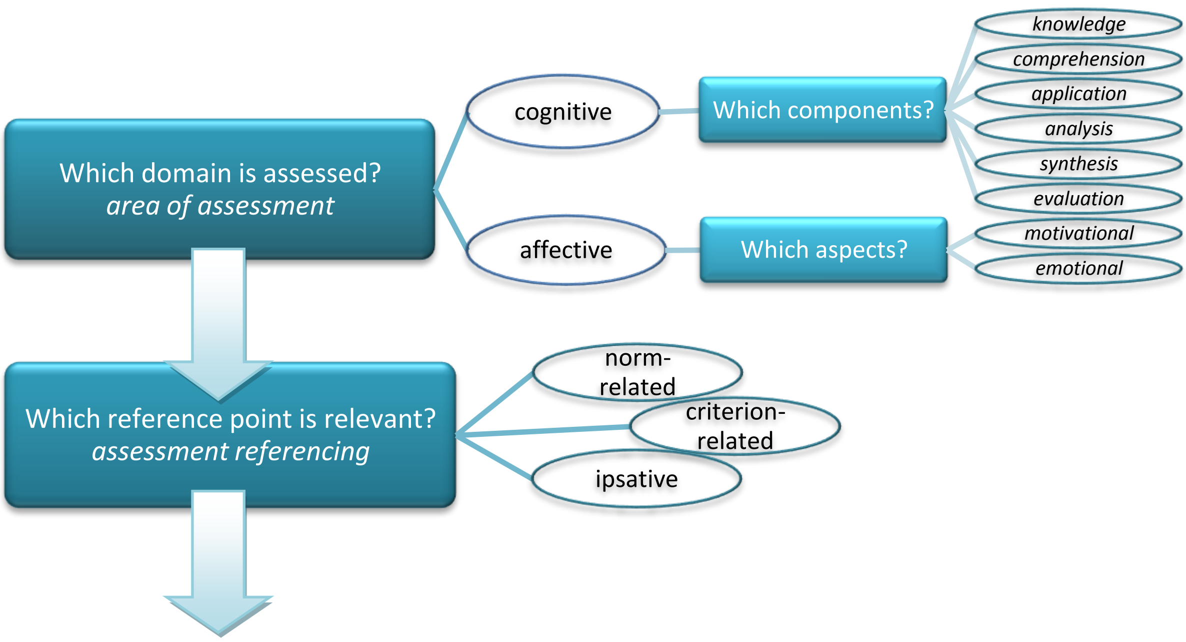
7.4 New forms of e-assessment

In this section we want to focus on the assessment activities that are part of the enriched learning experience and explain it in more detail. As outlined in Chapters 3 and 4, modern forms of assessment have to cover several aspects based on cognitive and educational findings, as well as technological standards. Thus, based on already existing ways of assessment, the Alice assessment model combines these assessment forms in order to provide a comprehensive assessment of knowledge, skills, behavioural, motivational, and emotional aspects for complex learning resources. In the following we summarize the most important features to be considered when developing e-assessments. Figure 49 depicts the different assessment forms (as they are listed in Figure 48) as eight questions that should be answered when planning an assessment. For each question the respective specifications are listed. Depending on the learning objectives and the respective learning scenario, adequate assessment forms can be found by going through the specified aspects of assessment and selecting all the relevant ones. Thus, by answering each of the seven questions, a full assessment plan can be developed. Thereby, it has to be considered that the different forms cannot be seen as independent aspects, but influence each other. Hence, the representation does not mean a linear order of the relevant assessment forms. Nevertheless, it can be seen as a suggested way of proceeding. The listed options are a summary of the most relevant assessment forms, but the selection is of course open to change and/or extensions. However, before starting the assessment, the learning objectives should be mapped into a set or dictionary of competencies, which are then used to build assessment rubrics that give a detailed overview of the learning goals. Furthermore, each goal should be connected to a criterion that specifies how and when a goal is achieved.

In the following the considered assessment forms are explained in more detail:

- Assessment area/domain: cognitive competencies (knowledge and skills) vs. affective dispositions (motivation and emotion).

  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., Blooms 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 Blooms Taxonomy can be used, namely knowledge, comprehension, application, analysis, synthesis, and evaluation.



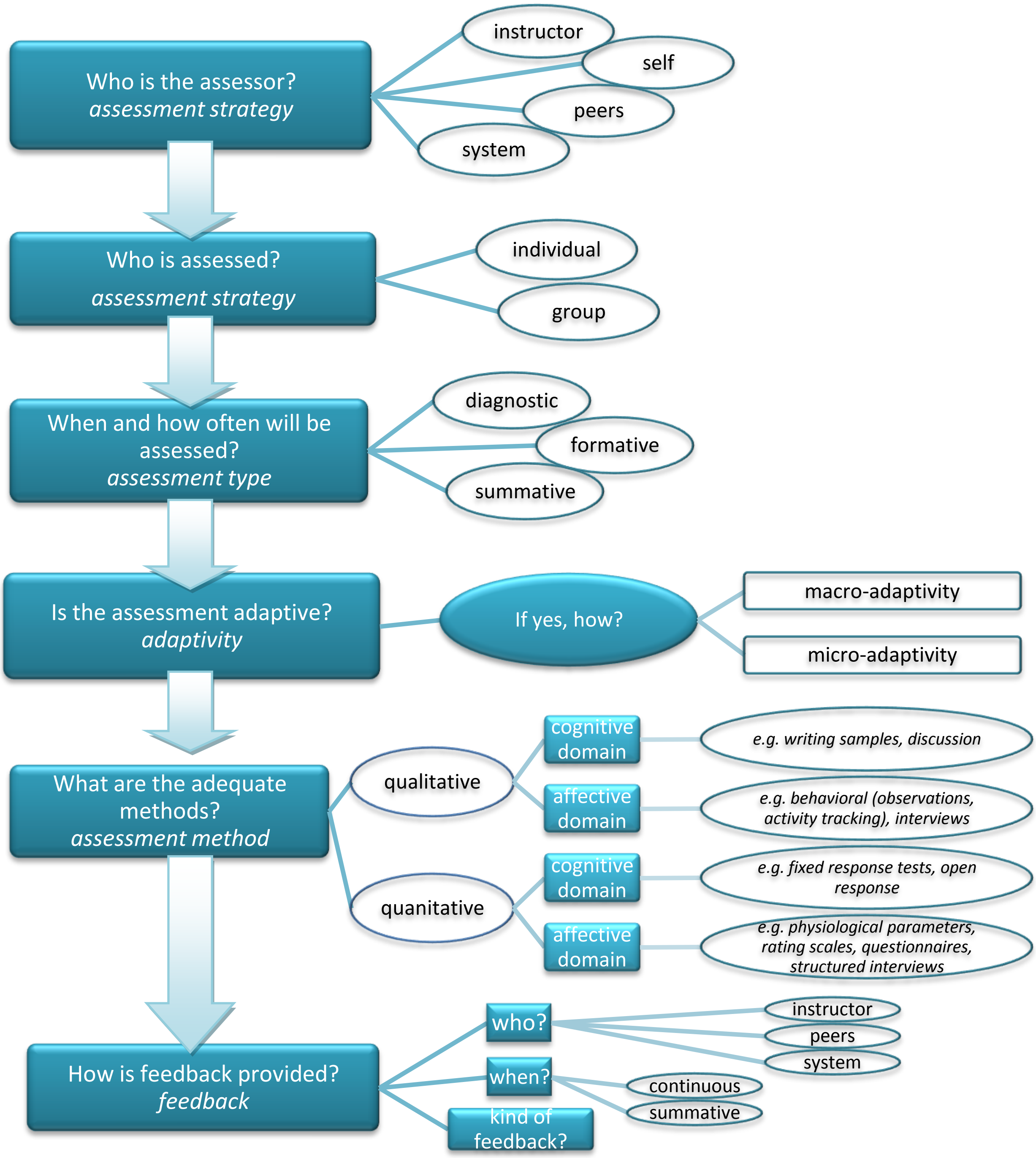


Figure 49. New forms of assessment

Additionally, innovative forms of assessment should always cover the learner's affective disposition (see Section 7.3.3) 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. Futhermore, 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 differenent 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:* norm-related, criterion-related, vs. Ipsative 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, we speak of norm-related referencing, if the comparison concerns the individual's actual status with a pre-defined domain, we are dealing with criterion-referencing. Finally, ipsative referencing means a comparison of 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. When assessing motivation or emotion, the reference can be used to e.g. 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 I:* 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, new forms of assessment that are based on CLR also involve students in the process of assessment. Particularly, students are asked to assess their own work (self assessment) or the work of their peers (peer assessment). Also the performance of a whole group can be assessed. 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. 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 e.g. serious games, system based assessment can be used. In this case the system or 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, see section 3.5). 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.

  Assessment strategy II: individual vs. group assessment. As already mentioned, it is important to focus on the role of the involved persons. In a first step 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 her self 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*: diagnostic vs. formative vs. summative
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. To assess at the end of a learning activity whether a learner has reached the learning goal (= summative assessment) is certainly the most common form of assessment. However, formative assessments might provide a more valuable outcome for the learning process (see Chapter 3). Due to formative assessment, learners are assessed more or less regularly during the learning activity. Such assessment does not only provide the possibility of giving feedback to the learner but is also helpful to meet the needs of a learner: Due to formative assessment, learners are supported in reflecting their learning performance. Typical methods (see below) used for cognitive formative assessments are e.g. quizzes, discussions, homework, or short tests. In social settings (such as in collaborative learning contexts), the assessment of knowledge can be divided into deferred and immediate assessment, where the latter one basically corresponds to formative assessment. In deferred assessment the collaborative process is replayed 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. For a more detailed discussion of these approaches, see Section 7.8.

- *Adaptive assessment*: E-assessment has the great advantage that it allows personalized testing. Thereby, it can be differentiated between 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). For a more details description of the item see Section 3.5. The adaptive assessment constitutes an important aspect within the whole assessment process, because it is directly related to the three adaptivity components (see Section 7.3.4). The outcome of each (invasive or non-invasive) assessment leads on the one hand to an adaptation of learning content, navigation support, and/or presentation form of the learning objects 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, reaching from simple tests, instructor observations, or writing samples to discussions or the analysis of student work. Generally, we can differentiate between quantitative (e.g. points or percentage achieved in a test, ratings, physiological parameters) and qualitative assessment (e.g. open ended questions in interviews, behavioural observations) methods. For e-assessment, computer-assisted (CAA) fixed (multiple choice; for assessing knowledge) or free response formats are common. Fixed formats are usually multiple choice items to test the state of knowledge, while free response formats are used to assess competencies in programming, essay writing, or meta-skills. The chosen assessment method depends on the assessment area (e.g. multiple choice items for knowledge tests vs. rating scales for motivational assessments), the assessment type (instructor, self, or peer as well as individual vs. group assessment), the assessment form (formative or summative) and last but not least the learning objective (as described in Bloom's taxonomy, bloom, 1956). For more examples of assessment methods please check Sections 7.8 through 7.10 which discuss special features of assessment in collaborative and social settings, serious games, as well as storytelling.

- *Feedback:* Giving feedback is an important issue in the context of assessment (see Sections 3.2 and Chapter 6) because 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 change their learning behaviour. 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). Generally, feedback overlaps with formative assessment.

7.5 Framework for the use of the IMA-Model

The complexity of IMA is driven from its coverage of enriched learning resources of 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 variety of external aspects such as educational settings, technology and standards, and affective aspects. Therefore, AL-Smadi et al (AL-Smadi, Hoefler, Guetl, 2011a) have designed and implemented a framework by which using IMA in modern learning settings can be facilitated. The authors have discussed a methodology of how to use IMA in developing educational tools that represents the CLR with integrated forms of e-assessment such as self-, and/or peer-assessment through a bottom-up layered framework.

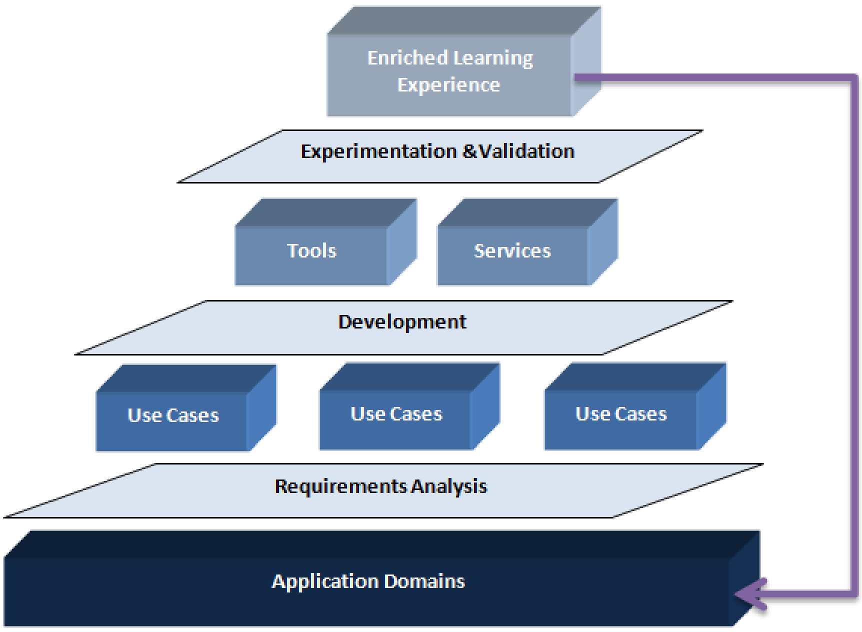


Figure 50. Bottom-up framework to use IMA.

The provided framework pyramid is built up of layers consisting of components and a continuous process. The top of it represents the main outcome of an enriched learning experience of the assessment of CLR in modern learning settings. 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 Fig. 50 the framework provides a methodological approach where the following steps are followed:

- *Define application domain:* as this framework has been designed to build on IMA in this step the components and related external aspects from the model are explored. This step tries to answer questions about the learning scenario to be developed. More precisely 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 didactical objectives and how they can be achieved with the learning scenario and what are suitable assessment forms are 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.
- *Identify requirements:* 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 input from technology aspects in IMA is used to investigate available tools and services, standards and specifications, and the suitability of these tools and standards for the designed learning scenario. 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.

7.5.1 Show Case

This section explains a real case of how we have used IMA and the bottom-up framework to develop and evaluate a tool entitled "Co-Writing Wiki" for collaborative writing and peer-review as part of our contribution to ALICE project. As discussed in (AL-Smadi, Hoefler, Guetl, 2011a) the following procedure has been followed:

1. Domain Definition
   As discussed before this step tries to answer questions about the learning scenario to be developed. The problem statement has been formatted based on the following Personas:
   - 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.
   - 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.
   - 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.
2. Requirements Identification
   Based on the problem statement from the step before the following scenario has been followed to formalize the requirements:
   - An instructional designer has been asked to recommend possible tools to support Eric in his course. The instructional designer is following the bottom-up framework discussed before as a methodology to identify available tools, to define aspects related to learning and teaching styles, to define learning objectives and goals, possible strategies, and methods for assessment . The instructional designer started with the first step to define the application domain and reported the following aspects:
     "…the course should be enriched with collaborative learning activity by which students can be grouped into small groups of (3-5) students. The learning activity should be applicable for both blended and distance learning where variety of students can participate and learn. Moreover, the learning activity should integrate self, peer-assessment activities by which students can reflect on themselves as well as evaluate the progress of their peers and provide feedback. The learning activity should reflect a continuous assessment where teachers/tutors can assess the progress of the collaboration, and the performance of individuals and groups. Therefore, software should be developed to deliver collaborative learning and provide a variety of learning styles (i.e. visual, verbal and non verbal) via a flexible way by which both students and

teacher can receive valuable feedback regarding the group work and the learning progress represented by contribution and assessment. Moreover, I recommend a collaborative writing assignment by which students performs in groups and collaborate to provide solutions based on problems. Nevertheless, the outcome can be also a scientific report for specific topic in Software Engineering.

The tutors dispose of a variety of collaborative strategies, methods and tools to support and enhance collaboration, debate and information exchange among peers so as to lead them to complete the required project successfully. Each group should be able to choose an adequate subset of the given collaborative strategies and to build their own collaboration strategy that best suits the group's dynamics, interests, and goals. The tutor should provide a well-structured project with suitable learning activities, well defined tasks, as well as rules and procedures that group members have to follow in order to accomplish the project."

3. Use Cases

   From the discussed Personas and the instructional designer recommendations, the software architect or system analyst has identified functional requirements and generalized the following use cases: *assignment author*, *contributor*, *reviewer*, and *evaluator*. The same user can take one or more use cases. As assignment author, user is allowed to author and schedule assignments, assign topics, configure groups, and create assessment rubric. The author has administrator role, and is responsible for organizational aspects. A contributor is allowed to create pages, to create links among pages, to add comments. A reviewer is allowed to view contributions and reviews of each participant and to grade a page by using the assessment rubric provided by the author. An evaluator is allowed to grade contributions and reviews.

   Moreover, the developed tool should be enhanced to provide the recommendations from the instructional designer as well as to consider the use cases. Therefore the developed tool should achieve the following requirements: assignment management, group management, multiple roles, tools for integrated self, peer-assessment, assessment rubrics management, collaborative writing and contribution, web-based to support distance and blended learning, enhanced visualization tools to support different learning styles (i.e. visual, verbal and non verbal) and feedback provision, and interaction logging to evaluate performance.

4. Tools and Services

   The generalized use cases and the functional requirements have been given to the developer in order to develop the tool. In this step services and tools that are suitable for the identified requirements and use cases are identified. An enhanced wiki system with the required functionalities has been developed as we could not find one tool to cover all the functionalities. Moreover, wikis have capabilities to manage groups and they are web-based applications. Nevertheless, they are suitable for collaborative writing assignments. Therefore, an enhanced wiki for collaborative writing and peer-review has been developed. The tool is entitled Co-writing Wiki and it provides integrated assessment of self, peer assessment as well as assessment rubrics for assignment grading and feedback (AL-Smadi, Hoefler, Guetl, 2011b; D5.2.1).

5. Experimentation and Validation

   In order to evaluate Co-writing Wiki tool and to validate the integrated assessment forms against the requirements identified in step 2, the tool has been experimented in learning activities as part of courses in the University. For more details about the experiments you can refer to D8.1.1, R8. The next section will explain in more detail the impact of this tool on students learning in complex learning scenarios.

7.6 First application of the model in selected complex learning scenarios

In this Section, we give a first example of how the components of the IMA model are reflected in a real learning scenario, namely in a self-directed learning course with a collaborative writing assignment using co-writing Wiki. Applications for scenarios connected to work-packages WP3 (live and virtualized collaboration), WP4 (simulation and serious games), and WP6 (storytelling) will be outlined after completion of the respective experimentation.

7.6.1 Enriched learning experience for a self-directed learning course with collaborative writing assignment

In this Section, we want to demonstrate how the different components of the enriched learning experience can be applied within a real learning setting. Therefore, we will explain how each of the features specified within the abstract model (see Figure 48, p.110) was used for the development and application of a self-directed learning course with a collaborative writing assignment using co-writing Wiki. Automatic assessment and feedback were embedded in the course. For the former, students could use the AQC. For a detailed description of the study, refer to D8.1.1, R9 (ALICE, 2011).

7.6.1.1 Learning Objectives

The main objective of this study was to create a complex learning environment in order to support students in self-regulated learning and working collaboratively. We also considered that these goals are related to further objectives such as gaining social competences (due to collaborative work) or meta-cognitive skills (due to self-regulated learning activities). In order to reach the main objective of the study, we provided a complex learning environment as follows:

The students participated in an online course about “Scientific Working”. First, they were asked to study two articles from a provided course material. During reading the articles the students could test themselves with automatically generated questions provided by the AQC. They had the opportunity to take a pre- intermediate and/or a post-test during reading content. Testing themselves with questions should stimulate their learning process and support them in self-regulated learning.

Then the students were asked to collaboratively write essays about these articles by using the Co-writing wiki. After that they received automatically created questions as part of a stage test and got a grade on it. Finally they collaboratively planned a study within the co-writing Wiki. This task was accompanied by self- and peer-assessments.

In order to investigate whether the students could really benefit from the provided learning environment, the students were asked to fill in a Pre-, Intermediate and Post-Questionnaire. Besides demographic data and general questions about the pre-knowledge of students, the questionnaires covered usability, task-awareness, motivational, and emotional aspects. Additionally we could observe their activities, learning processes and outcomes within the Co-writing wiki. Besides, we had a look on their frequency of testing themselves with automatically created questions and the grade they got in the stage test.

7.6.1.2 Complex Learning Resources

In order to ensure an enriched learning experience for active learners, the provided complex learning resource is a self-directed learning course integrated with a collaborative writing assignment using co-writing Wiki. For the development of the wiki tool, see Section 7.5.1. Besides the self-assessment function within the wiki, the AQC (automatic question creator) was provided as tool for self-assessing learners' knowledge in a given field (in this scientific working). For a more detailed description, refer to Deliverable D5.2.1 (ALICE, 2010).

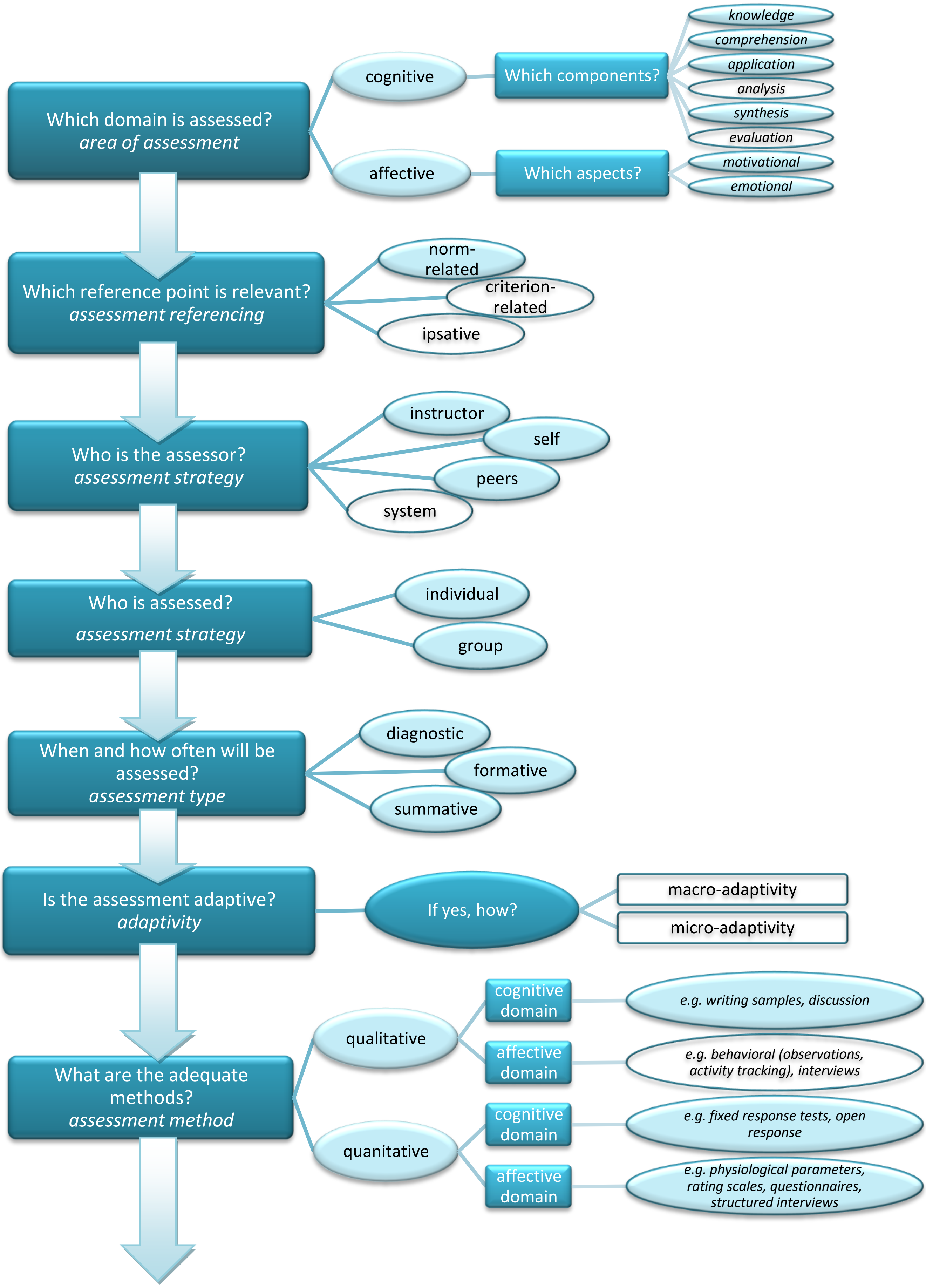
Co-writing WIKI: As discussed above, collaborative work supports students' learning efficacy and achievement of specific skills. The Co-writing wiki ensures students to work collaboratively, because it is an enhanced wiki for collaborative writing assignments. The developed tool supports task and social awareness as well as group well-being and group production function during collaborative work. Additionally it provides self-, peer-, and group-assessment functions to give valuable feedback to individual learners and learning groups.

In the presented study, the students collaboratively planned a study and wrote their essays within the Co-writing wiki. By using the Co-writing wiki the students also assessed their own work according to its importance (self-assessment), reviewed their peer's work and got feedback for their own contribution (peer-assessment). In a final step, the students were asked to review the product of two other groups (group-assessment). For this, assessment rubrics with 5-star rating scales covering the three categories content, style, and literature were provided to ensure a fair and consistent assessment.

Automatic question creator (AQC): The AQC uses an automated process to create different types of test items from a given textual learning content. More specifically, it generates four different types of questions, namely open-ended, fill in the blank, single choice, and multiple choice items. It is also capable to process learning content stored in various file formats and to extract the most important concepts, which are then used as basis for the generation of test items and reference answers. Furthermore, the creation of questions from concepts previously extracted by humans is possible. Pre-studies showed that the quality of automatically created questions is comparable with questions provided by humans.

7.6.1.3 Forms of e-assessment

As described above, there are a lot of different forms, types, strategies, etc. of assessment. In this section, we want to give an overview about the assessments, which were provided in the presented study in order to reach the stated learning objectives. In Figure 51 the forms of assessment that are applicable to this study are emphasized by means of a green background. As far as the criteria for mastering a learning objective are concerned, we used assessment rubrics for the group- and instructor assessment to insure a fair and consistent assessment over all learning groups. The rubrics for assessing the quality of the collaborative writing task concerned the three categories literature, content, and style with several subcategories each (see D8.1.1 for a more detailed description; ALICE, 2011).



Which domain is assessed?
area of assessment
cognitive
Which components?
knowledge
comprehension
application
analysis
synthesis
evaluation
affective
Which aspects?
motivational
emotional
Which reference point is relevant?
assessment referencing
norm-related
criterion-related
ipsative
Who is the assessor?
assessment strategy
instructor
self
peers
system
Who is assessed?
assessment strategy
individual
group
When and how often will be assessed?
assessment type
diagnostic
formative
summative
Is the assessment adaptive?
adaptivity
If yes, how?
macro-adaptivity
micro-adaptivity
What are the adequate methods?
assessment method
qualitative
cognitive domain
e.g. writing samples, discussion
affective domain
e.g. behavioral (observations, activity tracking), interviews
quanitative
cognitive domain
e.g. fixed response tests, open response
affective domain
e.g. physiological parameters, rating scales, questionnaires, structured interviews

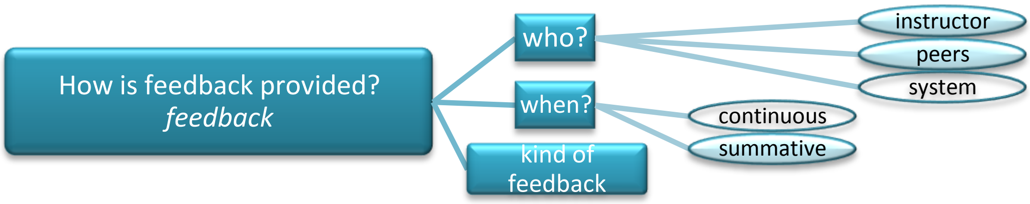


Figure 51: Assessment forms used in the application study (green background)

With respect to the *assessment area*, we investigated both, cognitive competencies by testing the students with automatically created questions and affective dispositions by collecting data on students' motivation and emotional status. More detailed information concerning the affective assessment is described below (see 7.6.2.2 Psychological Aspects). In the cognitive domain, the questions provided by the AQC tested on Bloom's level of remembering, whereas the collaborative assignment covers the levels of comprehension (e.g. identify important steps for planning a study), application (e.g. apply steps to own research questions), and synthesis (e.g. plan and formulate research design for given research question).

Concerning the *assessment referencing*, in our case the students compared their product with the work of other peers (norm-related).

The *assessment strategies* provided in our study included self- and peer-assessments as well as assessment by an instructor. Additionally individual contributions and group products have been assessed (group-assessment). Hence, both individuals and groups were assessed by their peers or by an instructor.

Regarding the *assessment type*, we provided formative assessment in order to monitor and improve students' learning process (the used strategies were self- and peer-assessments and a knowledge test provided by the AQC). Additionally, after the students had finished their papers they were asked to evaluate the work of other groups, which would count to summative assessments. For the group-assessment, we provided assessment rubrics with the following three categories: literature, content and style. As a rating scale we used 5 stars, in which 1 star is the minimum and means the worst evaluation and 5 stars are the maximum and the best possible evaluation. We also integrated a diagnostic assessment in order to investigate the students' learning progress. At the beginning the students were asked about their previous knowledge concerning scientific working and after the course they got the same questions again. This allowed us to compare their knowledge status before and after the course, in order to evaluate their learning progress.

Adaptivity was only achieved on a very low level, namely regarding the process of collaboratively creating a document, because each review given by a peer influences the next steps taken within the learning and working process. Personalized adaptation of learning content or test-items (e.g. based on students' current knowledge, motivational, or emotional status) was not embedded yet. However, these adaptive components should be provided after the full integration of the co-writing Wiki and the AQC into the IWT (for detailed information on the integration, please refer to D1.2.2)

The *assessment methods* included quantitative as well as qualitative methods and fixed as well as open response formats. The three questionnaires consisted mainly of rating-scales, which are quantitative assessments, just as the number of correct questions achieved in the computer-generated tests (AQC), and the ratings given in self-, peer-, and group reviews. Regarding the test-items themselves, ratings, single- and multiple choice questions and fill in the blank items are fixed formats, the open response items by the AQC as well as the essays were open response formats. Qualitative assessments included the open answers in the questionnaires (e.g. regarding improvements of the tool), comments in the self-, peer-, and group reviews, and the review of the essays by the tutors.

Finally, the tutors gave a detailed individual *feedback* at the end of the course regarding students' contributions concerning planning a study.

7.6.1.4 Evaluation and Validation

For the evaluation of the tools, we analysed the quality of the automatically created questions provided by the AQC. Besides, the students were asked whether the questions they received supported them in self-directed learning.

For validation purposes, we had a look on the impact of the whole tool. Hence we provided rating scales in order to measure for instance students' intrinsic and extrinsic motivation. Besides we also considered emotional aspects and preferred learning styles and whether these components had an influence on students' learning process.

7.6.2 Inputs to the enriched learning experience

7.6.2.1 Educational Aspects

In order to investigate students' learning styles, we consulted the learning styles of Wild (2000). In our case we concentrated especially on the repeating and the elaborating learning style to find out whether the students' learning process is rather superficial or aims at a deeper understanding. Besides, it should be mentioned that an important condition for the elaborating learning style is that people are intrinsically motivated. So the educational and motivational aspects are related and have an influence on each other.

Therefore, due to their preferred learning style, we can assume whether they answered the automatically created questions out of pleasure with the aim to deepen their knowledge or whether answering the questions was rather a superficial revision of the content.

7.6.2.2 Psychological Aspects

We also measured motivational and emotional aspects during learning and assessment. The assessment was performed by means of rating scales, thus on a psychological level. According to students' motivation, we had a closer look on students' intrinsic and extrinsic motivation regarding the peer-assessment. Additionally we used an intrinsic orientation scale, an extrinsic orientation scale and a task value scale in order to investigate their motivation concerning the whole course and its task. The students stated their level of agreement on a 5-point Likert scale. The rating scale ranged from "I strongly disagree" (1), "I disagree" (2), "neither/nor" (3) to "I agree" (4), "I strongly agree" (5).

Finally the students were asked about their level of motivation according to the different learning activities such as planning a study, working with the Co-writing wiki or doing the assessment. The students stated their level of motivation with the following answer categories: "absolutely unmotivated" (1), "unmotivated" (2), "motivated" (3), "very motivated" (4).

Moreover, we investigated in which emotional mood the students were during using the AQC-tool and working within the Co-writing wiki. Therefore we consulted a measure scale regarding emotional aspects developed by Kay and Loverock (2008). This scale includes 12 items and is used especially to measure emotions related to learning new computer software. Research showed that these 12 items are describing four emotions, namely happiness, sadness, anxiety and anger. The answer categories in this case are “None of the time”, “Some of the time”, “Most of the time” or “All of the time”.

7.6.2.3 Technological aspects

As ALICE aims to develop innovative e-assessment tools that support assessment activities in the enriched learning experiences, a set of educational tools integrated with assessment forms have been developed. More precisely for this section we will consider the prototypes named “Co-Writing Wiki: Enhanced Wiki for Collaborative Writing and Peer-review” and “Assessment in self-directed learning” as reported in D5.2.1.

Co-writing wiki is an enhanced wiki for collaborative writing assignments. Co-writing wiki is enhanced with some tools to maintain task, social-awareness and group well-being. ScrewTurn wiki 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. 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 Manager utilize the group management and document management provided by the wiki system to author and deliver Co-writing assignment with peer-review. For the sake of integration, the Co-writing wiki will be fully integrated with IWT where a single sign-on (SSO) mechanism will be applied. Moreover, Co-writing wiki should interact with the learner model and knowledge model services from IWT to maintain personalisation and adaptation of the learning activities, for more information about integration please refer to D1.2.2 (ALICE, 2011).

For the sake of “Assessment in self-directed learning” three main components have to be utilized or implemented in order to realize the procedure: an e-assessment system capable to interact with LMS and selects learning material (represented by IWT), an Automatic Question Creator tool (AQC), and Question and Test Interoperability (QTI) web service in order to manage, interpret, validate, and create QTI questions which adhere to the IMS Questions and Test Interoperability specifications. For more information about AQC and automatic item creation form textual material please refer to section 4.4.3.

7.6.2.4 Standards and Specifications

As this section is concerned with standards and best practices that have been developed to design and develop e-learning content and components, we will mention related standards and specifications or guidelines to “Co-Writing Wiki: Enhanced Wiki for Collaborative Writing and Peer-review” and “Assessment in self-directed learning”.

For Co-writing wiki, the literature shows that learning activities linked to assessment attract students more and increase their motivation (Macdonald, 2003; Reimann & Kay, 2010). According to Reimann and 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). Macdonald (2003) provides guidelines for the assessment of CSCL by which he highlights the importance of linking collaborative learning activities to assessment procedures, for more information please refer to section 4.3.4. Although 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.

According to (Reimann & Kay, 2010), "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-centred Assessment Design (Mislevy, Steinberg, & Almond, 1999), where this relationship is maintained by an evidence model that determines which of the students interactions should be logged and how to use the tracked interactions to update the student model. In order to make this feasible a detailed understanding and representation of the task model should be available. This reference understanding is provided by maintaining the interaction with the learning model. For more information about Evidence-centred Assessment Design please refer to assessment models in section 4.2.

In case of "Assessment in self-directed learning", AQC is automatically creating assessment items based on textual material. IMS-QTI Assessment content specifications have been used to represent the created items, for more information please read section 5.3.1. Moreover, a web service with clear interfaces has been designed for flexible integration between AQC and IWT. The development of these interfaces is following the guidelines discussed in the interoperability section 5.7.

7.6.2.5 Efficiency and Effectiveness

To evaluate the assessment tool (Co-writing wiki) and the AQC-tool, on the one hand we provided a usability scale in order to compare our tool with others, and on the other hand the students were asked to answer some open questions concerning suggestions for improvements.

So after the students had worked with the Co-writing wiki and the AQC-tool, they were asked to assess the usability/functionality of these tools. In order to compare the usability of our tools with others, we consulted the SUS (System Usability Scale) developed by Brooke (1996). Furthermore, we could put the SUS scores we received in relation to the average SUS score from 500 studies. Additionally, we provided open questions such as: "Do you have any suggestions for improvements?" or "Please describe what you did not like regarding the tool."

7.6.2.6 Quality Assurance

From the psychological point of view, the quality criteria, which are discussed in 3.6 were considered in the planning of the study.

7.6.2.7 Interaction with other models

We conducted this study in order to improve the tools and thus used the tools stand-alone. So instead of focusing learners' needs, we concentrated more on the tools themselves, that they are able to support the learners in their learning process. Hence, we just investigated some aspects such as learning styles, but a whole interaction with the adaptive components was not considered in this study yet. However, after full integration with the IWT (see above), interaction with other models will be provided.

7.7 Validation of the IMA and assessment model by experts

In order to improve the IMA and assessment model, nine e-learning experts from different European universities were asked to validate the model concerning the importance of its components, their accurate relations between each other and its application and relevance in the field of e-assessment. Additionally they were asked to test and evaluate two tools, which were developed in WP 5.

Five experts, two men and three women participated in the study. One expert is habilitated in computer science, three experts have a PhD and one a Magister degree (MA). Their working fields vary from psychological research, information technology, computer science to teaching at university and on high school level. A look on their research interests also shows a variety of expertise. Their research interests range from media psychology, evaluation, usability, e-learning, utilization of information technologies in an industrial context (area of knowledge management), adaptive learning systems, workplace learning, collaborative learning, social software, community information systems, mobile social software, mobile multimedia services to digital libraries, social network analysis, technology enhanced learning and new product development.

For the validation, we sent them Chapter 7 “An integrated model for enriched learning experiences” of this Deliverable 2.0 (current sections 7.1 to 7.6), guidelines of the tools and a questionnaire (see Annex A.3). In addition, we provided access to the tools, so that the experts could go through the functionalities and check whether the tools are user-friendly. The experts reported that they spent on average 4 hours on reading the Chapter, testing the tools and answering the questionnaire.

7.7.1 Validation of the model

Five experts were asked to answer eleven questions regarding different aspects of the model. They stated their level of agreement or disagreement on a rating scale which ranged from “I strongly disagree”(1), “I disagree”(2), “neither/nor”(3) to “I agree”(4), “I strongly agree”(5). Expect for one question we used a 7 point rating scale wich ranged from “not relevant” (1) to “very relevant” (7). Additionally they were asked to justify their rating values by giving comments and suggestions for improvements. Table 10 gives an overview of the ratings and comments.

Question	Mean value (SD)	Comments
The model provides an accurate representation of the real world	2.80 (0.84)	- too abstract - need of including mobile technologies or multimedia - no linear order in reality - model focuses on learning of individuals, learning processes of social entities are missing - lots of important elements are considered
The model provides a substantially complete representation of the real world.	2.20 (1.10)	- missing aspects: social context, group dynamics, working/learning context, problem based or project based learning as complex learning

		resources - assessment of relational factors
There is an obvious error in the model.	2.20 (1.14)	- learner/user/student model instead of learning model - 4 experts found no error
The components of the model are easy to comprehend.	2.80 (1.30)	- interplay of components - illustration by a concrete example - adaption part is not clear - some components require reading the details
All of the included components are relevant and priorities are set appropriately.	3.80 (0.45)	
The relations between the components make sense.	2.80 (1.30)	- add relation between educational/psychological aspects and learning goals and technology - inside (single learning episode) vs. outside (whole educational design) the box
The flows are correct.	2.80 (0.84)	- no linear order - different order (text vs. model)
The model fits the requirements/objectives to “specify and design a functional innovative framework to evaluate didactic experiences in adaptive learning systems”.	3.75 (0.96)	- clearer guidelines on how to evaluate didactic experiences
All in all, how would you rate the integrated model regarding to its relevance in the field of e-assessment?	4.69* (0.89)	- emphasize benefit/advantage of this model - add more components - adaptive to underlying system - elaborate and well justified assessment part
What would you especially improve regarding the model?		- priorities of the model more visible - Skip red arrows → background/context - Integration of relational factors
Do you have any further comments?		- Focus on individual learning experience, although talking about social interaction and collaboration

Table 10. Mean values (standard deviation) and comments from the validation of the model by experts

(Note. Mean values are based on 5-pt. rating scales from (1) strongly disagree to (5) strongly agree. *for this question a 7-pt. rating scale from (1) not relevant to (7) very relevant was used.)

First, the experts were asked whether the model provides an *accurate representation of the real world*. They neither agreed nor disagreed on that (M = 2.80, SD = 0.84). According to their comments, the description of the model is too abstract and they also criticized the linear order of the steps in the assessment model. Furthermore an expert stated that mobile technologies or multimedia should be included. Another expert mentioned that the focus of the model is on learning of individuals and that learning processes of social entities like classes or communities are missing. However, one expert also thought that the model has a lot of the important elements. In order to give concrete examples for the model, we added Sections 7.5 through 7.10. Regarding the missing aspect of social learning, Section 7.8 deals with this issue. Furthermore, we added social learning as educational input and extended the collaborative complex learning resource to a collaborative and social learning resource.

Then, the experts were asked whether the model provides a *substantially complete representation of the real world*. Almost all experts disagreed on that (M = 2.20, SD = 1.10). From their point of views, the social context such as the socioeconomically situation of the learner and his/her family is lacking and group dynamics within the collaborative learning setting should be considered. One expert also mentioned that the working/learning context is missing. Another expert suggested adding problem-based or project-based learning as a complex learning resource. One expert stated that he misses assessment of relational factors such as availability of the tools, centralities and contribution structures in groups and in the web. Regarding the scope of the ALICE project, we especially considered the consistent criticism of missing social context by adding a Section on assessment for collaborative and social learning (Section 7.8).

Four experts stated that there is no *obvious error in the model* (M = 2.40, SD = 1.14). Just one expert found an obvious error regarding the term “learning model”. This was corrected throughout the deliverable.

Furthermore the experts were asked whether *the components of the model are easy to comprehend*. For two experts the components were easy to comprehend. However, two of the experts also disagreed on that and one of them neither agreed nor disagreed on it (M = 2.80, SD = 1.30). A closer look on their comments showed that their disagreement does not refer to a single component they did not understand, it was rather the interplay of the components which was hard to understand. An expert mentioned that the whole description is very abstract and that an illustration by a concrete example would be helpful. Another expert stated that especially the adaptation part was not clear for him. Finally one expert explained that the components were easy to comprehend, just some require reading the details such as “new forms of assessment” or “complex learning resources”. Since the explaining text is meant to be an accompanying part of the model, focused on giving more concrete examples in which also the other mentioned points are dealt with.

Regarding the components’ relevance and priorities, the experts agreed that all of the included *components are relevant and that priorities are set appropriately* (M = 3.80, SD = 0.45).

Furthermore the experts were asked whether the *relations between the components make sense*. For two of them the relations make sense. However, two of the experts disagreed on that and one of them neither agreed nor disagreed on it (M = 2.80, SD = 1.30). One expert stated that there should also be a relation between educational/psychological aspects and learning goals and technology. As this relation is explained in the text, we assume that this expert failed to read this term in detail. Another expert suggested to focus on single learning episodes (inside the box) and on the whole educational design (outside the box). As our model provides a framework to evaluate didactic experiences in adaptive learning systems, it doesn’t make sense in our case to focus single learning episodes. Moreover, in addition to educational aspects, there are other components such as technological and psychological ones which are also relevant in an enriched learning experience.

Then, experts were asked whether *the flows are correct*. One of them stated that the flows are correct. However, two of the experts disagreed on that and two neither agreed nor disagreed on it (M = 2.80, SD = 0.84). In addition, the experts were asked about what they think is incorrect concerning the flows and should be improved. One expert mentioned that the flows should not be assumed to be linear. Besides, one expert noted that in the textual description the order is different to the model. We matched the order in the figure and text. Regarding the linearity, it is explained in Section 7.4 that the various assessment areas are interrelated and influence each other in a non-linear way.

The experts agreed that the *model fits the requirements/objectives* specified in the ALICE proposal (WP5, D5.1.1), namely *to "specify and design a functional innovative framework to evaluate didactic experiences in adaptive learning systems"* (M = 3.75, SD = 0.96). Just one expert suggested to provide clearer guidelines on how to evaluate didactic experiences and one of them did not give an answer. As the model represents an overall framework to describe assessments in very rich adaptive learning systems, a general guideline for all different kinds of didactic experiences cannot be provided. Thus, the description of the assessment part of the model points at the fact that the order of questions is just a suggested way of proceeding. Additionally the final forms of assessment are highly dependent on the learning resource at hand and also influence/restrict each other. The three added chapters on assessments in different learning environments should improve this issue by providing examples for assessment procedures in different scenarios.

Moreover, the experts were asked to rate the integrated model regarding to its *relevance in the field of e-assessment* on a seven-point relevance scale, which ranged from "not relevant"(1) to "very relevant"(7). The mean value of 4.69 (SD = 0.89) shows that the experts rated the integrated model as relevant in the field of e-assessment. One expert suggested emphasizing the benefit or rather the advantage of this model compared to already existing models and to other e-learning approaches. Another expert stated that the e-assessment components are a relevant contribution, but he would add more components. However, he didn't mention which ones. One expert was in favour of the assessment part and described it as elaborate and well justified. As the experts only read a chapter or rather an excerpt of our contribution which described the model in detail, they missed that in the whole theoretical background before, the benefit or rather the advantage of the model was emphasized. We are also aware of the fact that the e- assessment components we used are just a selected sample of a variety of components. Nevertheless the model is not meant to be exhaustive but to cover the most important components which are relevant in the field of e-assessment.

Finally, the experts were asked what they would especially improve regarding the model. Additionally to suggestions for improvements, which are mentioned above, one expert suggested skipping the red arrows and defining them as background or rather context. After rethinking the idea of skipping the red arrows, we decided to keep these components in the model, because of their importance and influence on the enriched learning experience. Another one stated that the priorities of the model should be made more visible. By explaining "new forms of e-assessment" in detail (Section 7.4), we tried to focus on assessment activities as part of the enriched learning experience. One expert also mentioned the integration of relational factors such as the underlying social network of the learning peers or the tool learner interaction patterns. As already mentioned above, we considered the social context by adding a Section on assessment for collaborative and social learning (Section 7.8)

Further comments were only stated by one expert, namely that the social interaction and collaboration stressed in the objectives of Alice (Section 7.2) are not elaborated enough in the subsequent sections. Here we also refer to the Section on assessment for collaborative and social learning (Section 7.8).

Summarized, the validation by experts lead to the following improvements of the model. First, the social aspect of learning was considered more thoroughly by adding it to the model itself (educational inputs in the IMA model and differentiation between assessing individuals and groups in the assessment part). Secondly, for each kind of complex learning resource (collaborative and social learning, serious games, and storytelling), we added a separate section discussing the special aspects to be considered in these scenarios. This input is also meant to meet the request for more concrete examples. Finally, minor changes were done throughout this Chapter to correct errors and especially clarify the description of the model were necessary.

7.7.2 Validation of the tools

Besides the development of the IMA model, two tools were developed in WP 5, which are integrated into the overall learning platform IWT and thus support the e-learning experience of students. The same experts who evaluated the model itself, were also asked to evaluate the tools. However, only three experts could finish all three tasks (Model and AQC and co-writing WIKI). Thus the following values and comments are from only four experts for the AQC and three experts for the WIKI.

7.7.2.1 Automatic Question Creator (AQC)

The Automatic Question Creator (AQC) was evaluated by four experts. Table 11 gives an overview of the mean ratings given by the experts as well as a summary of their comments.

Question	Mean value (SD)	Comments
The Automatic Question Creator (AQC) adequately supports self-regulated learning environments.	3.25 (0.96)	- Need of question evaluation - Concept and implementation problems - Support focused learning - AQC assesses low levels of rote learning/fact learning
The AQC can also be used in other learning scenarios such as Story Telling and game-based learning	2.75 (0.96)	- Storytelling and game-based learning require learning by understanding and not learning by remembering facts - depends on the algorithms - Test questions disturb game experience and storytelling flow
The AQC is also useful to support the instructor of a course.	4.00 (0.82)	- Saves time and re-analysis of learning material - Instructor has to check whether the questions make sense and the learning content is entirely covered - Using AQC for low level questions instructor can focus on more high level questions - A question creation wizard would be helpful for the instructor - Depends on instructor and course
The four different question types (open ended, fill in the blank, multiple and single choice) generated by the AQC are	3.00 (0.82)	- Generally suitable, but also depends on the content and the instructor - Need of questions that aim to test the students' understanding of interrelations

suitable to test students' knowledge.		and connections between topics
Would you change any question types or add new ones?		- question types are enough for getting a first impression about a student's knowledge - No suggestions for changes or improvements
Could you imagine using AQC in your working field?		- As a help for generating questions for a test in a school/university setting - To give learners a tool for self-reflection - Not when understanding is an issue. - Not in a mathematical or engineering working field

Table 11. Mean values (standard deviation) and comments from the validation of the AQC by experts

(Note. Mean values are based on 5-pt, rating scales from (1) strongly disagree to (5) strongly agree.)

First, the experts were asked whether the Automatic Question Creator (AQC) *adequately supports self-regulated learning environments*. Two of them agreed, one expert disagreed and another expert neither agreed nor disagreed on that (M = 3.25, SD = 0.96). According to their comments, one expert pointed out that the AQC could support students in reflecting their learning progress so that students can consolidate knowledge and know what they have not learned yet. Two experts criticised that the AQC only assesses very low levels of rote learning or rather fact knowledge. The experts also discovered that the quality of the questions depends on the quality of the source material. For instance: if the source material is factually incorrect or includes references at the bottom (like a Wikipedia page), the AQC will create questions based on incorrect or nonsensical information. From the point of view of one expert, it is necessary to evaluate the questions carefully in order improve their quality. Here it needs to be pointed out, that the quality of generated questions and concepts is evaluated by means of several studies, parts of which are still in progress. Experts did not have access to the study results.

Furthermore the experts were asked whether the *AQC can also be used in other learning scenarios such as storytelling and game-based learning*. Two experts disagreed, one expert agreed and another expert neither agreed nor disagreed on that (M = 2.75, SD = 0.96). One expert reported that he/she already has experiences with manual multiple choice tests from repositories in game-based learning environments. According to students test questions may disturb game experience and storytelling flow. As storytelling and game-based learning environments require learning by understanding instead of learning by rehearsing and remembering facts, one expert doesn't think that AQC can be used in these learning scenarios. Thus, when using AQC in connection with this learning resources, one has to carefully choose the point in time it can be applied. For example, a diagnostic assessment using the AQC at the beginning of a story or serious game could help to choose the right scene according to the user's previous knowledge.

Nevertheless, the experts are convinced that *AQC can be useful to support the instructor of a course* (M = 4.00, SD = 0.82). The experts stated that using the AQC saves time and re-analysis of learning material. However, the instructor has to check whether the questions make sense and the learning content is entirely covered. One expert suggested using AQC in order to ask more low level questions, e.g. check whether the students have actually read the book, so that the instructor can focus on more high-level questions.

Regarding the different question types generated by the AQC, the experts were asked whether the *four different question types are suitable to test students' knowledge*. One expert agreed, two experts neither agreed nor disagreed and another expert disagreed on that (M = 3.00, SD = 0.82). According to their comments, the experts stated that generally the questions are suitable, but that their quality also depends on the content and the instructor. One expert would add more open questions that aim to assess students' understanding, especially regarding their understanding of interrelations and connections between topics.

An open question followed, investigating whether the experts would *change any question types or add new ones.* None of the experts would change any question types or add new ones. One expert pointed out that the question types are enough in order to get a first impression if a student has learned anything about a knowledge field.

Finally the experts were asked whether they could *imagine using the AQC in their respective working fields*. One expert disagreed and one of them agreed on that without giving a reason. Another expert disagreed on that and explained that most of the knowledge he/she assesses in his/her working field is of mathematical or engineering nature. Therefore, students have to solve mathematical or programming exercises. Two experts stated that they would not use it in a setting where understanding is an issue. One expert mentioned that he/she could imagine using it as a help for generating questions for a test in a school/university setting or to give learners a tool for self-reflection.

Summarized the expert validation of the AQC confirmed the intended use of the tool. According to the experts it is a valuable tool to test knowledge (on a lower level) and to get a first impression of what the students have learned. However, it is no suitable to test students' deeper understanding of a subject.

7.7.2.2 *Co-writing WIKI*

Co-writing WIKI was evaluated by three experts. Table 12 summarized their mean ratings and comments.

Question	Mean value (SD)	Comments
Do you think the following components of the wiki are helpful for the instructor?		
– Actions feed in assignment homepage	4.33 (0.58)	
– Contribution graphs in assignment homepage	3.00 (1.73)	
– Revision player in the contribution tool	4.00 (1.00)	
– Charts in the contribution tool	2.67 (1.53)	
– The rate control (stars)	4.33 (0.58)	
– Rubrics for assessment	4.00 (1.00)	
If you found one or more components not helpful for the instructor, please state your reasons		– Contribution graphs/charts: number of signs is not an informative aspect – Rubrics for assessment: helpful, but too many sub-rubrics to consider

why.		
Do the components support the students in their learning process?		
– Actions feed in assignment homepage	4.33 (0.58)	
– Contribution graphs in assignment homepage	3.00 (1.73)	
– Revisions player in the contribution tool	3.67 (1.15)	
– Charts in the contribution tool	3.00 (1.73)	
– The rate control (stars)	4.33 (0.58)	
– Rubrics for assessment	4.00 (1.00)	
If you found one or more of the components not helpful for the students, please state your reasons why.		– Referring to the feedback above
Can you think of any further components you would integrate into the tool (to support students and/or teachers).		– Discussion option – Information about contributors – Search function – awareness and visual feedback
Imagine you have to evaluate students' contributions. Which assessment forms would be helpful?		
– Self-assessment	3.33 (2.08)	
– Peer-assessment	4.67 (0.58)	
– Group-assessment	3.67 (0.58)	
In your opinion, in which fields could the Co-writing wiki be used?		– in any field – Collaborative essay writing, collaborative note taking, collaborative reflection – University level writing courses
Could you imagine using the Co-writing wiki in your working field?		– Co-writing wiki combines advantages of ether pad and internal wiki – Only for more advanced pupils
Regarding the components and the visual presentation of the Co-writing wiki, do you have suggestions or comments for improvements?		– Design of co-writing wiki is not motivating – More web 2.0 looks

Table 12. Mean values (standard deviation) and comments from the validation of the WIKI by experts

(Note. Mean values are based on 5-pt, rating scales from (1) strongly disagree to (5) strongly agree.)

With respect to the *components of the wiki*, the experts agreed that the actions feed in the assignment homepage (M = 4.33, SD = 0.58), the revision player in the contribution tool (M = 4.00, SD = 1.00), the rate control (stars) (M = 4.33, SD = 0.58), and the rubrics for the assessment (M = 4.00, SD = 1.00) *are helpful for the instructor*. Just one expert disagreed that the contribution graphs in the assignment homepage (M = 3.00, SD = 1.73) and the charts in the contribution tool (M = 2.67, SD = 1.53) can support the instructor. For him the number of signs is not an informative aspect since it's about quantity alone. He further criticised that the rubrics for the assessment consider too many sub-rubrics.

Next, the experts were asked whether these *components support students in their learning process*. According to the experts, the actions feed in the assignment homepage (M = 4.33, SD = 0.58), the rate control (stars) (M = 4.33, SD = 0.58), the revisions player (M = 3.67, SD = 1.15), and the rubrics for the assessment (M = 4.00, SD = 1.00) support students in their learning process. Regarding the contribution graphs (M = 3.00, SD = 1.73) and the charts in the contribution tool (M = 3.00, SD = 1.73), the same expert which criticized these components before was not convinced that these components would support students.

When asked *which further components they would like to see integrated into the tool* to support students and/or teachers, the experts suggested a discussion option, a search function and information about contributors (such as attended courses, fields of interest, etc.). One expert mentioned that everything that creates awareness and visual feedback is good as long as it does not prevent students from learning and collaborating.

For the purpose of *evaluating students' contributions*, the peer-assessment form received very positive ratings (M = 4.67, SD = 0.58), while the self- (M = 3.33, SD = 2.08) and group-assessment (M = 3.67, SD = 0.58) forms received still positive results, however one expert disagreed that self-assessment would be helpful in order to evaluate students' contributions.

Furthermore, as open comments, the experts were asked *in which fields the Co-writing wiki could be used*. The experts listed collaborative essay-writing, note taking and reflection, as well as university level writing in general and even any field at all. The next question asked, whether the experts could *imagine using the Co-writing wiki in their respective working fields*. One expert agreed and stated that the Co-writing wiki combines the advantages of two tools he/she previously used in his/her research group. Another one disagreed and gave no reason. The third expert would use it but only for more advanced pupils.

Finally the experts were asked to give suggestions or comments for *improvements regarding the components and the visual presentation of the Co-writing wiki*. One expert criticized the simple looks of the wiki and suggested to integrate more web 2.0-esque features (such as colors, charts, icons, menu-bars) in order to make its looks more appealing.

In summary, the experts considered the co-writing WIKI for the most part as supportive for students as well as instructors, especially the actions feed in the assignment homepage, the revision player in the contribution tool, the rate control and the assessment rubrics were found to be very helpful components. In general, the experts saw the fields of application very broad, but would improve its design and add some components, such as a search function and more information about the contributors.

Whereas the validation of the model had a direct influence on the work of this deliverable, the results of the tool validation are subject to the technical part of this workpackage.

7.8 Assessment for collaborative and social learning

One of the objectives of the ALICE project is the provision of a learning environment that is not only suitable for individual learners, but meets the needs of learning groups. Thus, this chapter deals with the special aspects that have to be considered when developing assessment procedures for collaborative environments.

Collaborative learning assessment requires a broad perspective about learning and the involved processes. It is necessary to encompass the asynchronous and synchronous interactions produced between group members. This assessment method has a significant effect on CSCL (computer-supported collaborative learning) because it motivates learners through accountability and constructive feedback. The bases for an enriched learning experience are familiarization with the formative activities, contents through discourse and encouraging students' participation (Caballé et al., 2008).

Assessment is a systematic process for making inferences about the learning and development of students. Also, assessment is a process of defining, selecting, analyzing, interpreting, and using information to increase students' learning through motivation, engaging, awareness and real experiences (Daradoumis et al., 2006).

Collaborative learning has an important social foundation. Collaborative and social assessment involves making expectations explicit and public; setting appropriate criteria and high standards for learning quality; systematically gathering, analyzing, and interpreting evidence to determine how well performance matches those expectations and standards; and using the resulting information to document, explain, and improve performance (Kreijns et al. 2003).

However, Kreijns, Kirschner, and Jochems (2003) identified as main pitfall in collaborative learning environments, the tendency to restrict social interaction to educational interventions aimed at cognitive processes, while social (psychological) interventions aimed at socio-emotional processes are ignored, neglected or forgotten. Students need to trust each other, feel a sense of warmth and belonging, and feel close to each other before they will engage wilfully in collaboration and recognize the collaboration as a valuable experience (Rourke, 2000, as cited in Kreijns, Kirschner, & Jochems).

Incorporating emotional assessment (sensing and responding to user's emotions) into collaborative learning can offer more interactive and challenging learning collaborations, enabling learners' social identity and authentic learning experience. Additionally, the thorough testing of user's emotion's transitions over time, can lead to more precise results.

Next both knowledge and emotional processes involved in collaborative and social assessment are presented.

7.8.1 Knowledge assessment

7.8.1.1 Deferred and immediate assessment

It is possible to describe knowledge assessment in collaborative learning by two assessment models or approaches: deferred and immediate assessment.

Deferred assessment uses learning scenes to show how the learning process has evolved after it took place. These scenes can be generated in many formats, for example: storyboards, forum dialogs, collaborative material creation, etc. In a deferred assessment, students observe the collaborative scene generated during his/her collaborative work. At the time the scene is played, it is also assessed. While students observe how the collaborative process has been developed they are fostered to get a better individual learning understanding and improve the social experience. The deferred assessment can be realized as individual or group activity.

Immediate assessment is focused on real time actions, understood as interaction moves between learning partners (students, tutors, environment, learning materials and resources, etc). Immediate assessment lets the partners increase the social awareness and shows how any cause produces effects in the collaborative learning process. Most importantly, social interactions are based on the empathy produced by a set of interaction moves between collaborative learning partners. Empathy is not only a feeling between human partners but it is possible to create empathic ties with non-human elements, such as material, activities, resources, etc. A quick feedback from social interaction can be essential to minimize the lack of awareness and improve the group engagement.

Both assessment types can be joined in a unique learning action, in which each evaluates a special aspect of learning produced by collaborative interactions. Using the immediate assessment it is possible to evaluate the students' learning problems and redirect the scene discussion to a specific scene point or resource of interest that helps getting a better understanding of the task or problem in hand.

Collaborative and social assessment has the mission of detecting problems in the interaction moves produced during the collaborative work sessions. Within the context of social and instructional interaction, Northrup (2001) proposes a framework of interaction attributes; each embedding possible strategies and tactics that can be used to facilitate instructional and social interactivity. This framework includes interaction with content, collaboration, conversation, interpersonal interaction, and performance support. All these attributes are interesting to evaluate the group social performance.

Collaborative and social aspects of learning are developed in a sequential process that can be evaluated step by step to give a useful feedback to partners. This feedback should provide the purpose of producing an enriched experience of the collaborative learning process.

7.8.1.2 Scene reference and description

The context where the collaborative and social learning is developed is the learning scene. A scene represents a set of environmental and social circumstances where the collaborative learning takes place through partners' interactions and interaction with environment and resources. Collaborative learning uses the scene to develop partners' abilities and competences through a sequential and integrated process where the interaction moves are evaluated to determine the scene sequence and the effective use of resources. Immediate assessments evaluate data input by students, such as options, actions, words used in dialog, test or questions proposed as well as time elapsed in every action or response. These informative data is to be processed and optionally enriched with personal and contextual information about users, groups, environment, etc.

The most effective way to process this information in order to obtain quick and safe results is to develop a set of rubrics that take diagnose inputs and return a diagnosis response. The diagnose inputs are the interaction moves data and some information related to a personal user, group, a resource and also environmental data. It is possible to detect problems with the following interactions: human-human, human-resource and human-environment. These responses can be processed as human feedback or as changes in the interaction response of resources and environment.

As a result, group partners are able to understand and manage the feedback supplied by the assessment system, in order to have an enrichment experience and deep control of his/her learning process. The assessment system can propose changes in the interaction way used by a partner if the interaction produces negative effects in social empathy or collaborative learning. Also, it is possible to adjust the interaction way with the resources or environment.

Despite real time (immediate) learning assessment can be vivid and rich there is an additional assessment level to be considered: deferred assessment. This assessment takes place once the scene step has passed and lets the group or partner review the scene developed. While the scene is reviewed, the assessment system evaluates global factors not easily evaluable in immediate mode.

For example the contribution proportion of every partner, the relevance of the contributions, the empathy generated from every partner, the time continuity effort, etc. It is possible to discern global or personal problems related with the group scene performance.

7.8.1.3 Inputs to the enriched learning experience

The incorporation of social aspects is important to create an enriched learning experience. However, just placing students in groups does neither guarantee collaboration nor social benefits. The incentive to collaborate has to be found within the groups. A complex set of educational approaches simultaneously applied, each reinforcing and/or complementing the other, can enhance collaborative learning and social interaction amongst group members. These instructional approaches result in group members socially interacting in ways that encourage elaboration, questioning, rehearsal, and elicitation. Three approaches are to be considered (Kreijns et al. 2003):

a. the cognitive approach of promoting 'epistemic fluency',

b. the direct approach of structuring task-specific learning activities, and

c. the conceptual approach of applying a set of conditions to stimulate/stress collaboration

The inputs to enrich the learning experience must promote cognitive approaches, a careful design of learning tasks, resources and scenes and endow the learning environment with conditions to stimulate/stress collaboration.

The conceptual approach usually makes use of:

- Positive interdependence: team members are linked to each other in such a way that each team member cannot succeed unless the others succeed and/or that each member's work benefits the others (and vice versa).
- Promotive interaction: individuals encourage and help each other's efforts so as to in order to reach the group's goals.
- Individual accountability: all group members are held accountable for doing their share of the work and for mastery of all of the material to be learned.
- Interpersonal and small-group skills: specific skills are needed when learners are learning within a group; students who have not been taught how to work effectively with others cannot be expected to do so; thus these skills must be developed (Sharan & Sharan, 1992).
- Group processing: the group determines which behaviors should continue or change for maximizing success based upon reflection of how the group has performed so far.

The knowledge assessment system must detect problems regarding all of these social aspects and develop an immediate or deferred feedback to make possible the feeling of control in the learning experience. The tasks and resources must create an empowerment and learning control in the learner.

7.8.2 Emotional assessment

7.8.2.1 Affect detection

Based on the models, methods and techniques for detection of Affectivity/Emotivity that are analysed in D2.1.2 and D2.2.2, the assessment of the student's emotion will be accomplished into three time shares:

(i.) **Before the collaborative task:**

- *Profile creation:* Vital information about user's interests and inclinations can be collected through guided questions, based on Multiple Intelligence tests. This can be done at any time, independently from the task that the user is asked to accomplish. Additional information can be added from the student in the form of Wikis of personal blog.
- *Mood inspection*: There is no use to bomb a student with questions when he/she is simply in bad mood. Mood is directing the affective state of a person as tendency in positive or negative direction. Before proceeding into user's emotion assessment, it is crucial to simply ask them to indicate their general mood. Photo-realistic images can be exploited for more realistic mood induction.

(ii.) **In real-time**: While the user is taking part in collaboration tasks, he/she will be able to report upon his/her affective state following three approaches:

- *Labelling*: Students will be able to describe their feeling from a set list of emotions or by picking expressive images or colours from available pools.
- *Dimensional view of emotions*: Users are asked to place their feeling in a 2- or 3-dimensional grid based on how aroused or how positive or negative, for example, they feel.
- *Use of Emoticons & Avatar*. Emoticons, smileys and avatars still remain an instant way of communicating emotions. In general, the more complex the emoticon, the less emotional content it carries. The key is to use emoticons sparingly and accurately in context to have maximum potential impact. The selection of emoticons will be in two layers: (a) *Standard* for instant emotion expression and *(b) Cue + Ornamental* for text enrichment

(iii.) **Retrospective**: Evaluation of the user's affective state after the accomplishment of the task. Identify the respondent's affective state by his/her content only through sentiment analysis techniques. For sentiment analysis the GATE system can be used that is quite common, easy to use and install and free of charge. The system can combine corpora based approaches ie. the WordNet-Affect corpus and the MPQA Opinion Corpus. The Subjectivity Sense Annotations that rely on WordNet 2.0 and the sense inventory are available for download at http://www.cs.pitt.edu/mpqa.

7.8.2.2 Affective Feedback

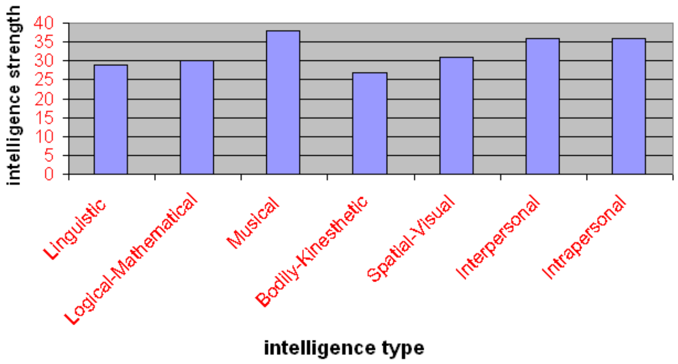
In a similar way to the means to capture emotion info in real-time, an important aspect to consider is transparency and that the feedback provided is as little intrusive as possible. Interventions include emotional scaffolds that encourage student's positive attitude towards learning and empathetic strategies that assure student's emotional safety and foster their meta-cognitive and meta-affective skills. Our focus is on developing effective production rules, straightening students' engagement in the learning experience. Feedback will be enriched by practices that have been tested and evaluated for years in Social and Emotional Learning applications.

Reactive- empathetic feedback can be enriched by practices that have been tested and evaluated for years in Social and Emotional Learning (SEL) applications (http://casel.org/), especially when trait emotions are identified. We will try to exploit the theory of Multiple Intelligence (Gardner, 2006) by providing feedback adapted to the profile of the user.

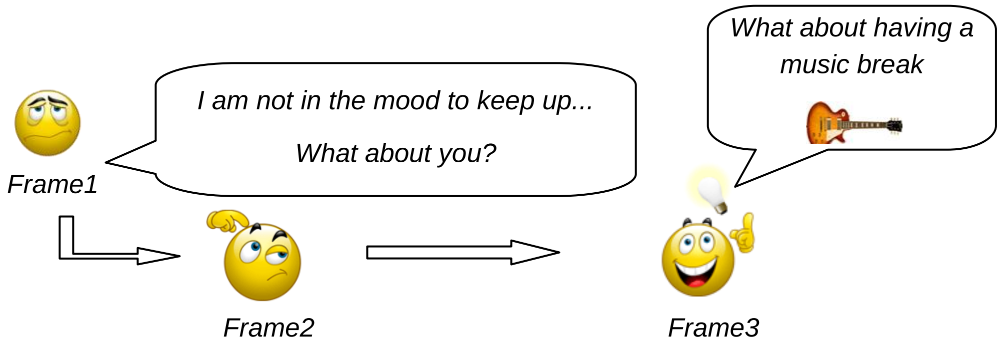
7.8.2.3 Hypothetical Scenario

Below we provide an example for a hypothetical student named Jelka:

1. When Jelka logged in the platform, the system motivates the student to update her profile, to write about interests, hobbies as well as to take a MI (Multiple Intelligence) test to figure out about her specialties and inclinations.
2. At another point in time, while Jelka is taking part in a collaborative task, the state of disappointment is detected (trait emotion that often leads to abandonment, difficult to change across short time), through her report on the emot-meter tool (she picked to use the Geneva Emotion Wheel to report on her affective state). This information was also combined with her mood assessment (rank 2 out of 5 in the mood-meter) that took place in the beginning of the collaboration task.
3. The system advised by the respective production rule, selects the "*Reactive- empathetic feedback strategy for disappointment state"*
4. The system is looking at Jelka's profile information: *"…I like books, movies, skiing and all kinds of music, especially classical music and rock. My favorite pianist is Maksim Mrvica, my favorite violinist is David Garret and my favorite band is Guns N´Roses. I come from Ptuj, the oldest town in Slovenija.."*
5. According to MI Model, Jelka's inclinations point to the musical, interpersonal and intrapersonal intelligence types of intelligence.



6. The system sends the respective information to the instructor, while it responds to the user:



7. The users click on the "guitar icon" and the system fetchs the http://www.last.fm/music/Guns+N%27+Roses/+similar web page.

In the other end, the instructor is able to combine Jelka's emotion report sent from the system with the result from the sentiment analysis of the collaboration content that took part.

7.9 Assessment for experiential learning and serious games

The three main complex learning resources dealt with in ALICE are collaborative and social learning, simulations and serious games, and storytelling (see Figure 48). Thus this chapter will discuss the special features that have to be considered when serious games are used as learning resource.

Experiential learning (Kolb, 1984) implies a method of experience, reflection, concept formation, and concept application in a continuous cycle. In a game-based learning environment, a separation is induced between experience and reflect, and between concept formation and application as a result of the use of a simulative or abstract environment for learning. In the case of learners whose capacity for proximal development (Vygotsky, 1978) does not include the ability to reflect and interpret fully the distinction between action in-game and outcome in the real world, scaffolding is required in some form to address this rift. As detailed elsewhere (Dunwell et al., 2011), if left unsupported, this rift can lead to games which are unable to satisfactorily or comprehensively convey sufficient feedback to the learner to achieve the desired outcome. Consequently, assessment in serious games must seek to understand individual learner needs, and in particular identify where a blended approach or intervention from a tutor or more able partner is required to ensure continued learning transfer. In many cases games are deployed as motivational agents within the learning process rather than standalone solutions, and it is no coincidence that games deployed in such contexts tend to demonstrate greater efficacy (Egenfeldt-Nielsen, 2005). To achieve this, it is important to consider the role of both the tutor and the learner when receiving and interpreting feedback from a game.

As an example, consider the specific case of the civil defence game constructed within WP4 of the ALICE project. On an individual level, players can be instantly chastised for incorrect actions such as collecting their possessions before evacuating. But does this necessarily indicate a knowledge deficit, or an intuitive learner seeking to explore by deliberately performing incorrect actions? If it does indeed show a knowledge deficit, can we say with certainty that by conditioning the learner to perform a certain action in a specific manner in-game, they will be able to independently reflect upon the underlying principle, and develop independently the understanding required to invoke the same knowledge in a real-world evacuation? Guideline principles for addressing these issues would be firstly to make in-game feedback limited in depth and complexity, and as unintrusive as possible (Jarvis and de Freitas, 2009), to avoid hindering the exploratory and intuitive learner with feedback explaining principles they already understand, and are simply performing incorrectly to explore the game. A macroscopic understanding, for example at the class level, might provide a tutor with a better understanding of how well core concepts are understood and applied - for example if all learners are religiously collecting their possessions before evacuating, it would certaintly suggest that intervention from the tutor at a group-level to explain correct and safe procedure. Hence in this case the output from the game engine at a meta-level to the tutor may well prove more valuable than its individual output to the learner, who is still hindered by their own zone of proximal development.

Further consideration of feedback on experiential learning in serious games might consider established models such as that of Rogers (Rogers, 1951, Dunwell et al., 2011). In Table 9 (see Section 6.1) we have previously shown the various levels of this model and how they might be supported in a serious game. The core consideration put by this model is a need for the technologist to consider carefully the levels required, and whether technology is capable of facilitating them alone.

If not, then a facilitator must be introduced through a blended approach. At higher levels the required technology becomes complex, requiring a degree of intelligent agency and expert systems knowledge to implement; even then more subtle nuances of these interactions require an understanding of affective and emotional state, itself an objective within other work packages within the ALICE project. Whilst the example in Table 1 (see Section 2.1) applies well to any activity that can be scored, more complex interactions such as collaborative learning induce further complications in how feedback is structured and conveyed (Dillenbourg, 1999).

In ALICE, we posit that for the case study game in civil defence, at least the third level of supportive feedback within this model is achievable through technology, and move into the fourth level through dialogues with virtual characters. To do so we apply knowledge of metrics such as evaluation time, and binary responses to actions and opportunities, together with multiple-choice responses from dialogues with virtual characters. This requires an understanding of the maxima and minima for these performance metrics, which can be implemented at the design stage. Knowing that under the constraints of the simulation, for example, a fast evacuation might take 60 seconds, we can associate a score with this metric and feed back to the learner based on their own evacuation scores in comparison to this measure. Yet we must be careful to avoid a behaviourist approach (Binsubaih et al., 2008); assuming this metric to be an ideal proxy for learner ability risks encouraging learners to defeat the game, rather than demonstrate a full knowledge of safety process and procedures, and due to the inevitable and aforementioned rift between simulation and reality, imposes the danger that dangerous practices might be misconstrued as correct ones. Adding further score-based metrics might ameliorate this risk somewhat, for example in the ALICE prototype running is swiftly penalised as an unsafe action, and the high scores associated with fast and unsafe evacuation are offset by penalities for dangerous actions. Despite this, it would be naive to suggest, as educator or designer, that this score alone is an adequate feedback mechanism.

It is therefore important to look beyond this score into a wider pedagogic design. Metrics of learner performance can often be of more use to the educator or facilitator than the learner themselves, and assuming their presence, educators and facilitators can be better positioned to supply feedback than a technology-driven environment (McGreal, 2006). Game environments provide a rich environment for data capture, and a broad stroke approach at the capture phase can generate large data sets suitable for quantitative analyses (Calvillo Gamez et al., 2010). Such analyses might offer a basis for identifying relationships in terms of correlations without requiring understanding of the underlying processes, in turn feeding this back into models of assessment which can act as predictors of learner performance. In the integration between WP4 and WP5 in ALICE we have explored how data might be exported from a game in a meaningful and ubiquitous formal (XML), opening access to this data to any system suitable for processing it. The simple model in Table 9 may then be refined to look beyond arbitrary scoring mechanisms, and instead at models of learner performance supported by empirical research into their relationship to a wide range of low-level metrics.

Social learning is one aspect that might benefit from such an approach. Similarly, more adaptive games, which can customise themselves to learners' needs, are another avenue by which issues could be addressed. In the following two subsections, we outline how these methods might be considered under an experiential model of learning, and applied to support game-based learning outcomes.

7.9.1 Assessing social learning in serious games

In applying established social theories of learning (Bandura, 1977) in game-based contexts it is important to note the discrepancies between social technologies, which provide a means for communication and collaboration, and social learning, which implies a more holistic outlook addressing not only how socialisation is facilitated, but also how learning is enabled in a meaningful fashion. Simply allowing communication does not guarantee socialisation, though the use of the game as a data capture mechanism as previously outlined (Calvillo Gamez et al., 2010) can allow these communications to be analyzed, provided adequate ethical and consensual infrastructures are in-place. As well as qualitative or quantitative analysis of text or transcribed audio logs, common elements to social infrastructures such as “friends” lists can be constructed into networks suitable for analysis via techniques such as clique percolation (Palla et al., 2005), and in turn compared to real-world network structures shown to hold the most effective social bonds. Translating this to evidence of learning and pedagogic outcome again requires some mapping of network structure or collaborative effort to assessment outcomes. It should be noted, however, that as the benefits of social aspects may lie principally in areas such as the stimulation of intrinsic motivation (Luckin et al., 2010), as well as continued collaboration beyond the confines of the classroom, immediate assessment may not reflect its impact as effectively as a longer-term approach.

Noting that the ultimate role of any assessment of a learning experience or object should be to feed in directly or indirectly to its improvement, a question of particular relevance when creating game-based social learning environments is how these environments might effectively be adapted or respond to assessment outcomes. If, for example, a social network analysis suggests learners are becoming segregated into small cliques of 2-3, when and how should action be taken to integrate these cliques into larger social groups? Iteration is an obvious though seldom pragmatic solution, requiring extensive investment in repeatedly adapting then rigorously re-evaluating an intervention. Empowering educators here with the capability to specify and monitor clique sizes and formations, reflect upon, and integrate this knowledge into their direct and didactic interactions with learners could prove a useful tool (Dillenbourg, 1999). Similarly, the outcomes of textual analysis could be fed back to the tutor, though particular care should be taken in performing this in a consensual form. Even with learners’ explicit consent for their actions to be monitored and reported on, knowledge this is occurring could impact peer interactions strongly as well as the emotional state of learners (Schultz et al., 2010).

For serious games facilitating experiential or exploratory learning, the benefits of such social mechanisms readily become apparent. A key technological question is whether collaboration in-world has substantive training benefits when compared to a course in which simulative assets are used in a standalone fashion, and collaboration is enabled through wider integration of game or simulation into the learning process. Both methods have their merits; socialisation in-world can offer a means for more realistic human interactions to be facilitated without complex artificial intelligence being required (Riedl et al., 2008), yet this requires other learners to be able to act and react to each others behaviour in a plausible fashion. Taking the context of an emergency evacuation, whilst collaboration would enable more effective crowd behaviour, the realism and actions of this crowd would remain dependent on the behaviour of individual learners. Returning to the rift that can emerge in experiential learning scenarios which rely on reflection on virtual action then real-world application (Dunwell et al., 2011), this could also result in a self-propagating situation in which learners are individually unable to exhibit correct – or even realistic – behaviour, hence adversely rather than positively impacting the experience of other learners. Avoiding this requires either a highly iterative and participatory approach to design, restrictive rules in-game which enforce plausible behaviour, or scaffolding the role of more-able partners within the simulation.

The alternative approach is to utilise virtual characters in-game as a means of social learning. Whilst such characters lack the adaptivity for deeper, more affective and meaningful social interaction, they can provide a useful means for creating a more plausible backdrop or conveying information in a more immersive form. A dialogue engine can here play a pivotal role in allowing learners a degree of interactivity, though this must be adequate supplied with information on learner behaviour in order to create an adaptive experience in-line with the flow approach described earlier in this deliverable (Cziksentmihalyi, 1997). In particular narrative and dialogue have been shown as an effective means for scaffolding learning interaction and experience within a serious game (Mott et al., 2006), drawing on parallels in entertainment gaming which seek to induce immersion through narrative rather than visualisation components of the game. In ALICE, therefore, we seek to apply such adaptive characters as a means for avoiding the complexities that might arise with a peer-interaction based model for social learning, whilst integrating technologies from beyond game based contexts, such as an assessment engine, and repurpose it for use within the game (see D5.1.2 and D4.2.2). Through interaction and assessment the virtual characters role can transition from a peer, needing help from the learner and thus testing their own knowledge and understanding, to one of a more able partner identifying and correcting incorrect actions.

7.9.2 Adaptivity in serious games

Adaptivity is a particular challenge in high-fidelity virtual worlds, given the technical challenges in adjusting complex content dynamically and in response to learner action. Customisability might describe a less substantive approach to adaptivity, with the game recognising the learner's name and gender and hence addressing them correctly. More sophisticated adaptivity requires an underlying architecture defined around a notion of composability, whereby game elements can be extracted, adapted, and reinserted in response to either user feedback or explicit design changes. This is demonstrated within ALICE through the infrastructure underlying the developed game (D.4.3.2), which allows content to be extracted and repurposed. In an example, text in extracted from the game and formatted into plain text, which is then used along with the Google Translate service to localise the game into different languages. Moreover, as learner interactions and actions are monitored and exported as XML, they can be used to create an adaptive learner experience by adjusting the game in response to these actions. The virtual agent described in the previous section is one example of this, and, indeed, social elements in serious games imply some degree of adaptivity. However, other aspects of the game, such as challenges and performance metrics might also by dynamically adjusted to sustain flow. Adaptivity can also bring with it increased or divergent learning potential. Consider for example the linguistic example; in allowing the game to seamlessly change between languages at run-time, we create, albeit indirectly, a tool with potential use as a resource for learning and teaching languages. Effectively, by inducing a degree of adaptivity, we have introduced increased repurposability for the game as a learning object. Given the value of cultural or linguistic learning resources, it could be suggested these areas are of particular interest when creating an adaptive game-based learning environment.

Game-based learning is also unique in that difficulty is often a central gameplay component and requires particular balance to avoid instilling disengagement or apprehension amongst learners (Cziksentmihalyi, 1997). This can complicate the introduction of instructional or educational material as this difficulty balance and flow of the game must be sustained, else the risk exists of creating a less-effective reincarnation of existing materials (Zyda, 2005). Adapting the level of challenge in response to user performance or selection is a technique common to entertainment games, but if used too freely in an educational context might allow learners to circumvent learning requirements. Avoiding this behaviourist approach is a key concern when developing serious games, as learners have been shown to approach serious games as entertainment games rather than as an educational resource, and in this case will be seeking to win by all means necessary, particularly if this can be achieved by avoiding the more challenging elements (Binsubaih et al., 2008). Returning again to the levels of feedback presented in Table 9, a solution here is to blend and scaffold learner interaction, and adjust the tone rather than content of educational material. In particular individual elements such as scores may have little pedagogic value or relevance, yet play a central role in keeping learners engaged in the educational process as they seek to improve. Dynamicism in these areas has potential to scaffold engagement whilst remaining detached from pedagogic implications (Warburton, 2008). In Figure 52 we demonstrate how this is afforded by the adaptive game architecture within ALICE.

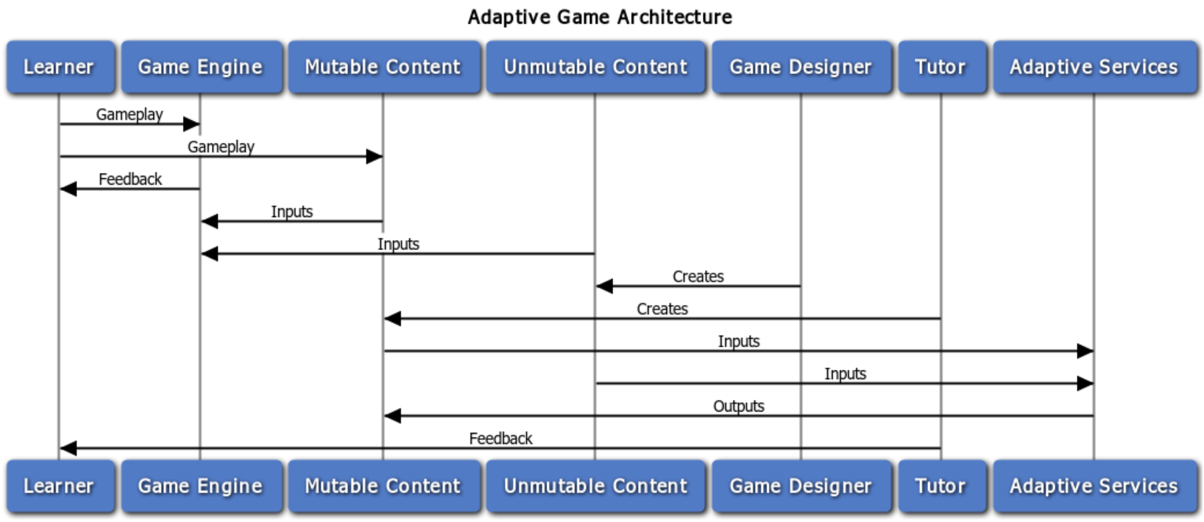


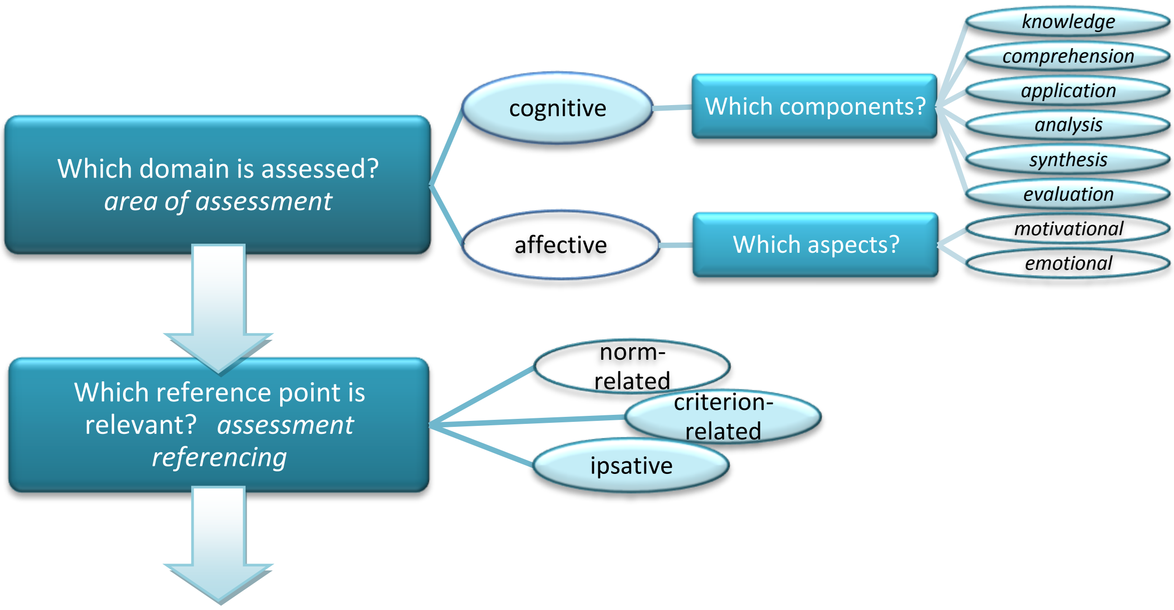
Figure 52: Adaptive Game Architecture

Critical to this input-process-output driven approach are both the learner and tutor, and the definition of content which is mutable (can be changed) post-development, and content which is unmutable (can only be changed by the game designer himself). This demarcation assumes a level of technical skills and time investment required to modify 'unmutable' content; here we use mutability as a measure of intrinsic capacity to be changed by the tutor. Adaptive services external to both learner and tutor provide a further level of adaptivity; in ALICE we demonstrate this through the integration with Google Translate to provide a localised version of the game automatically without requiring either the learner or tutor to speak the same language as the game's developer. Expanding this further to the semantic web services described in D4.2.2 offers the potential to replace other elements of content, such as poster images, with material derived similarly from web-based resources: ultimately the game developer is tasked not with sourcing and implementing content, but with identifying the required content and relying on automated semantic search processes to source and localise this content into an adaptive game. This also allows for greater adaptiveness to learner need. Embedding sourced information within the environment allows learners to progress and consume educational content whilst the adaptive environment responds, allowing learners to modify content and experience dynamically-sourced material as they progress through the game. This supports well the notion of an intuitive learning experience, wherein an expansive environment provides a backdrop against which learning can occur.

7.10 Assessment for storytelling learning resources

The objectives of ALICE aim at an innovative adaptive learning environment, where personalization, collaboration, and emotional aspects are combined to an interactive, challenging, context aware, and authentic learning experience. In this, learning materials and learning paths should be related to prior knowledge and connections between the learning material and the real world should be provided.

Besides collaborative learning and serious games, storytelling constitutes another form of complex learning resource. In this Section, we want to show how the developed assessment model can be applied within the context of narrative learning and storytelling. In Figure 53 the assessment options relevant for complex storytelling learning objects (SCLOs) are emphasized by means of a green background. Regarding the area of assessment, the affective domain is not highlighted in this context, because it is part of WP2. Generally, emotion and motivation are of course applicable to storytelling learning objects and an emotional test is also integrated in the IWT. Please refer to D6.1.2 (ALICE, 2011) for a detailed description of the storytelling design model developed within the ALICE project. D6.1.2 also contains a detailed description of how the IMA and assessment model can be applied in the context of storytelling. As personalization and adaptivity are core elements of ALICE, we focused on the different kinds of adaptivity, namely micro- and macro adaptivity as they are described in Section 3.5 of Deliverable 6.1.1. Furthermore, it is outlined, how the tools developed in WP5 (see D5.2.1, ALICE 2011), namely the Automatic Question Creator (AQC) and the Co-writing wiki, can be used within SCLOs and how they can be utilized for a micro- and macro-adaptive personalization of learning paths. Besides, other forms of assessment relevant to the field of storytelling (see D6.1.1, Section 3.8) are discussed in the context of IMA. Thus, this section only holds the overall Figure of the main assessment forms for storytelling, whereas the explaining text is to be found in D6.1.2 (Section 4.1).



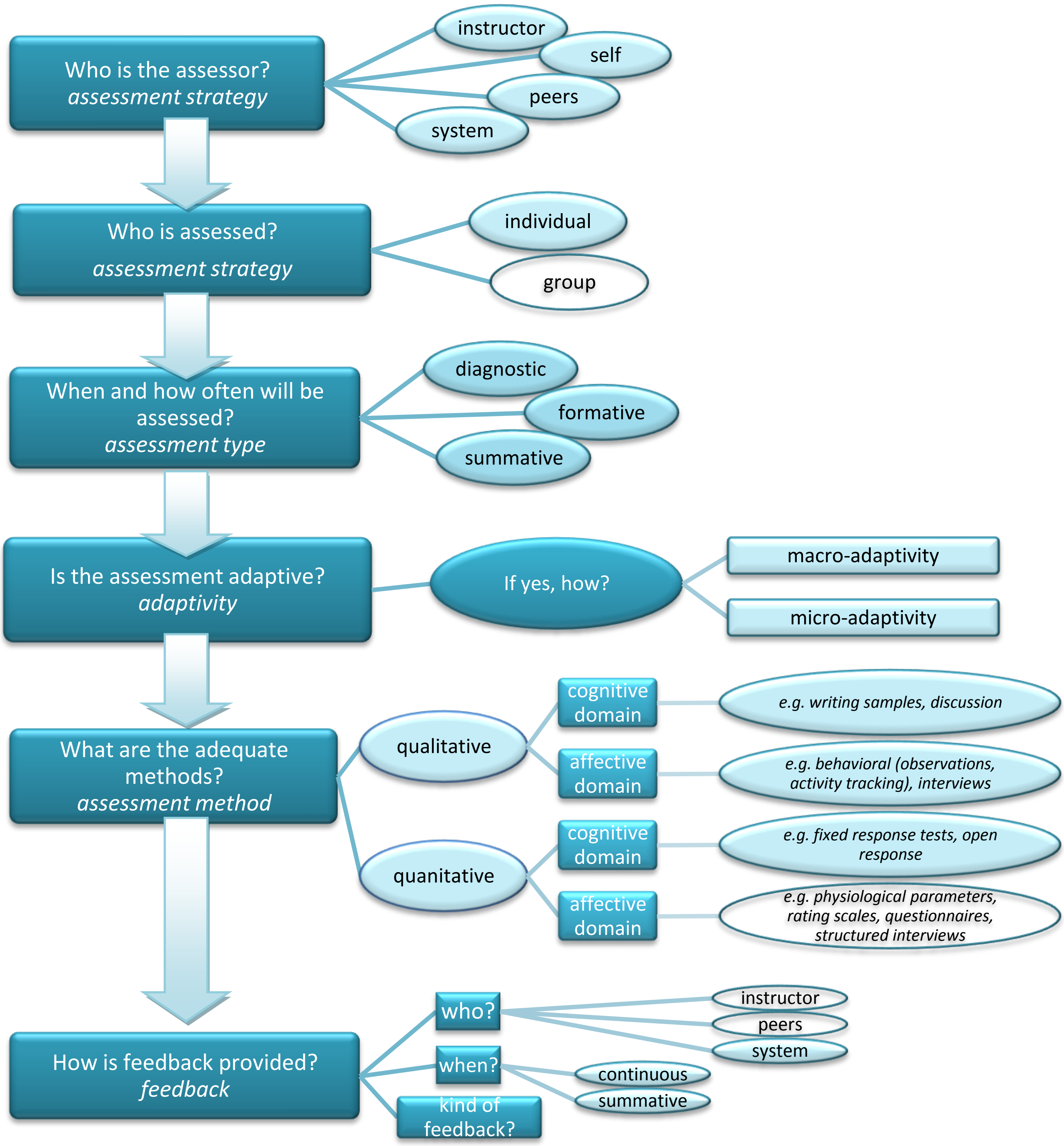


Figure 53: Forms of Assessment regarding Storytelling

7.11 Summary of the chapter

The aim of this chapter was to define an integrated model for e-assessment (IMA). We first described the objectives of ALICE regarding the expectations about the model. Then we presented a model of e-assessment from a generic viewpoint. This model describes an enriched learning experience that is made up of the didactical objectives, different learning resources, and assessment activities. It also considers influences arising from the viewpoints of pedagogy and psychology as well as from the viewpoint of technology. Furthermore, the relationship to other models (didactic model, knowledge model and learner model) is emphasized. Finally, to assure a high quality standard of the model, efficiency and effectiveness as well as evaluation and validation processes are mentioned as indicators coming up from the model.

In order to use IMA in real learning scenarios a bottom-up framework has been proposed as a reference to facilitate IMA deployment. Moreover, a case study from ALICE project has been presented to show how the framework can be used step-by-step in IMA deployment. The application of the model a real learning setting was demonstrated by showing how each component of the model was considered in a self-directed learning course (using two tools developed within this context). These first version and application of the model was presented to a sample of five experts who were asked to validate the IMA model with its special assessment part as well as the tools, which were applied for the model's first application. This expert validation resulted in a few changes of the model and a detailed description of assessment in the three main complex leanring ressources considered within the ALICE project. Thus assessment in collaborative and social learning, in experiential learning and social games, as well as in storytelling is discussed in the last three Sections of this Chapter.

References

ADL (2008). The Advanced Distributed Learning. Last retrieved October 7th 2008 from http://www.adlnet.gov/about/index.aspx.

AICC (2008). The Aviation Industry CBT (Computer Based Learning) Committee. Last retrieved October 7th 2008 from http://www.aicc.org/index.html.

ALICE (Adaptive Learning Via Intuitive/Interactive, Collaborative And Emotional Systems) project (2011), Deliverable D1.2.3 *"Integrated Prototype v2*" (revision), project co-funded by the European Commission within the 7th Framework Programme (2007-2013), n. 257639 (2010).

ALICE (Adaptive Learning Via Intuitive/Interactive, Collaborative And Emotional Systems) project (2011), Deliverable D2.1.2 "Learning Experience and Advanced Learning Resources v2", project co-funded by the European Commission within the 7th Framework Programme (2007-2013), n. 257639 (2010).

ALICE (Adaptive Learning Via Intuitive/Interactive, Collaborative And Emotional Systems) project (2011), Deliverable D2.2.2 "Management of Emotional-Affective feedback v2, project co-funded by the European Commission within the 7th Framework Programme (2007-2013), n. 257639 (2010).

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).

ALICE (Adaptive Learning Via Intuitive/Interactive, Collaborative And Emotional Systems) project (2011), Deliverable D5.2.1 "*Enriched learning experience V1"* (revision), project co-funded by the European Commission within the 7th Framework Programme (2007-2013), n. 257639 (2010).

ALICE (Adaptive Learning Via Intuitive/Interactive, Collaborative And Emotional Systems) project (2011), Deliverable D8.1.1 *"Inititial Experimentation and Evaluation Results"* (revision), project co-funded by the European Commission within the 7th Framework Programme (2007-2013), n. 257639 (2010).

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.

Allen, M. (2003). *Guide to E-learning*. John Wiley & Sons Inc, United States of America.

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.

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., 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 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.

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

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.

ARIADNE (2008). *ARIADNE Foundation*. Last retrieved October 7th 2008 from http://www.ariadne-eu.org/.

Artino, A. R. (2008). Cognitive load theory and the role of learner experience: an abbreviated review for educational practitioners. *AACE Journal, 16*, 425-439.

Atherton, J. S. (2010). *Learning and Teaching; Assimilation and Accommodation* [On-line]. Last retrieved 12nd October 2010 from http://www.learningandteaching.info/learning/assimacc.htm.

ATP Association of Test Publishers (2002). *Guidelines for computer-based testing.* ATP.

Baddeley, A. (2005). *Human Memory*. Hove: Psychology Press.

Bandura, A. (1977). *Social learning theory*. New York: General Learning Press.

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, 46*, 157-162.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman: New York.

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/.

Bartle, R. (2003). *Designing Virtual Worlds*. New Riders.

Bearman, M. (1997). *Narratives and cases: Implications for Computer-based Education*. In proceedings of ASCILITE '97 Conference, Perth, Australia.

Beck, J. E. (2004). *Using response times to model student disengagement*. In Proceedings of the ITS2004 Workshop on Social and Emotional Intelligence in Learning Environments.

Beghetto, R. A. (2004). Toward a more complete picture of student learning: assessing students' motivational beliefs. *Practical Assessment, Research & Evaluation*, *9*(15). Last retrieved October 19th 2010 from http://PAREonline.net/getvn.asp?v=9&n=15.

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.

BinSubaih A., Maddock S., Romano, D.M.(2008). A Practical Example of the Development of a Serious Game for Police Training. In Richard E. Ferdig (Ed.). Handbook of Research on Effective Electronic Gaming in Education.

BinSubaih, A., Maddock, S. & Romano, D. (2006). A serious game for traffic accident investigators. *Special Issue of International Journal of Interactive Technology and Smart Education on Computer Game based Learning, 3*(4), 329-346.

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*.

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.

Boekaerts, M. (1993). Being concerned with well-being and with learning. *Educational Psychologist.*

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.

Boud, D. (1995). *Enhancing learning through self assessment*. New York: Routledge/Falmer. Retrieved August 18, 2010 from http://books.google.at/books ?hl=de&lr=&id=mJ9qQ2IyuykC&oi=fnd&pg=PR7&ots=Lzf-TuYIMS&sig= IZjeIg3fjRmmBZhfxIhkEgb1juU#v= onepage&q&f=false

Bouzidi, L. & Jaillet, A. (2009). Can online peer assessment be trusted? *Educational Technology & Society, 12*, 257-268.

BPS (2002). *Guidelines for the Development and Use of Computer-based Assessments.* Leicester: British Psychological Society.

Bransford, J. D., Brown, A. L. & Cocking, R. R. (2004). *How people learn.* Washington D.C.: National Academy Press.

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.

Brooks, M. G. & Brooks, J. G. (1999). The courage to be constructivist. *Educational Leadership 57*. Last retrieved October 12nd 2010 from http://www.ascd.org/publications/educational-leadership/nov99/vol57/num03/The-Courage-to-Be-Constructivist.aspxl.

Brown, J. S., Collins, A. & Duguid, S. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Bruner, J. (1990). *Acts of meaning.* Cambridge, MA: Harvard University Press.

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.

Bull, J. & McKenna, C. (2001). *Blueprint for Computer Assisted Assessment.* London: RoutledgeFalmer.

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.

Cabanac, M. (2002). What is emotion? *Behavioural Processes, 60*, 69-83.

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.

Callear, D., Jerrams-Smith, J. & Soh, V. (2001). Caa of short non-mcq answers. *In Proccedings of the 5th International Computer Assissted Assessment conference*.

Calvillo Gamez, E. H., Cairns, P., Gow, J., Back, J. & Capstick, E. Video games as research instruments. *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems,* 2010 New York, NY, USA. ACM Press, 4493–4496.

Cano, M. D. (2011). *Students' involvement in continuous assessment methodologies: A case study for a distributed information systems course*. Last retrieved 26th January 2011 from http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=05590268.

Capuano, N., Gaeta, M., Marengo, A., Miranda, S., Orciuoli, F., & Ritrovato, P. (2009). LIA:

an Intelligent Advisor for e-Learning. In Taylor & Francis (Eds.), *Interactive Learning Environments, vol. 17, no. 3, pp. 221-239.*

Cassady, J. C. & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology, 27*, 270-295.

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, K. C. & Jang, S. J. (2010). Motivation in online learning: Testing a model of self-determination theory. *Computers in Human Behavior, 26*, 741-752.

Chickering, A. W. & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. American Association of Higher Education Bulletin, *39*, 3-7. Last retrieved August 23rd 2010 from http://www.aahea.org/bulletins/articles/sevenprinciples1987.htm.

Chun, M. (2002). Looking where the light is better: A review of the literature on assessing higher education quality. *Peer Review. Winter/ Spring*.

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.

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.

Clarke, T., Ayres, P. & Sweller, J. (2005). The impact of sequencing and prior knowledge on learning mathematics through spreadsheet applications. *Educational Technology Research & Development, 53*(3), 15-24.

Cocea, M. (2006). *Assessment of motivation in online learning environments.* [online]. Last retrieved October 19th 2010 from http://www.ncirl.ie/dynamic/File/Research/Technical Report Aug 2006.Mihaela Cocea.doc.

Cocea, M. (2007). *Assessment of motivation in online learning environments*. School of Informatics, National College of Ireland: Unpublished diploma thesis. Last retrieved October 19th 2010 from http://www.dcs.bbk.ac.uk/~mihaela/publications/MSc%20NCI.pdf.

Cocea, M. & Weibelzhal, S. (2006). *Can log files analysis estimate learners' level of motivation?* Proceedings of 14th Workshop on Adaptivity and User Modeling in Interactive Systems. Hildesheim, Germany.

Cole, M. (2009). Using wiki technology to support student engagement: Lessons from the trenches. *Computers & Education, 52*(1), 141-146.

Conlon, G. (1986), Objective measures of writing ability, New York: Longman, chapter Writing assessment: issues and strategies, 109-125.

Conole, G., de Laat, M., Dillon, T. & Darby, J. (2008). 'Disruptive technologies', 'pedagogical innovation': What's new? Findings from an in-depth study of students' use and perception of technology. *Computers & Education, 50*, 511-524.

Costall, A. & Still, A. W. (1987). *Cognitive psychology in question.* New York: St. Martin's Press, Brighton: Harvester Press.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral & Brain Science, 24*, 87-114.

Crisp, G. (2007). "*The e-Assessment Handbook"*. Continuum International Publishing Group, New York.

Crook, C. (1994). *Computers and the Collaborative Experience of Learning*, London: Routledge.

Csikszentmihalyi, M. (1990). *Flow:* The *Psychology of Optimal Experience.* New York: Harper and Row.

Csikszentmihalyi, M. (1991). *Flow: the Psychology of Optimal Experience*. London: Routledge.

Csikszentmihalyi, M. (1988). *Optimal Experience*. New York: Cambridge University Press.

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.

Culver, D. A. (1998). *Review of Emotional Intelligence by Daniel Goleman: Implications for Technical Education*. Watson School of Engineering and Applied Science, SUNY-Binghamton.

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.

Cytowic, R. E. (1994). Personal Communication.

Cziksentmihalyi, M. (1997). *Finding Flow: The Psychology of Engagement with Everyday Life,* London: Routledge.

Damasio, A. (1996). *Descartes's Error: Emotion, Reason, and the Human Brain.* London: Papermac.

Dancer, D. & Kamvounias, P. (2005). Student involvement in assessment: A project designed to assess class participation fairly and reliably. *Assessment & Evaluation in Higher Education, 30*, 445-454.

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.

Daradoumis, T., Xhafa, F. & Marques, J. M. (2002). A methodological framework for project-based collaborative learning in a networked environment. *International Journal Continuing Engineering Education and Lifelong Learning, 12*, 389-402.

Daradoumis, T., Xhafa, F. & Pérez, A. A. J. (2006). A framework for assessing self-, peer, and group performance in e-learning. In T. S. Roberts (Eds.), *Self, Peer and Group Assessment in e-Learning* (pp. 279-294). Hershey, PA, USA: Idea Group Press.

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.

Davou, B. (2000). *Thought Processes in the Era of Information: Issues on Cognitive Psychology and Communication*. Athens: Papazissis Publishers.

Davou, B. (2007). Interaction of Emotion and Cognition in the Processing of Textual Material', *Meta : journal des traducteurs / Meta: Translators' Journal*, *52*, 37-47.

DC (2008). *The Dublin Core Metadata Initiative*, Last retrieved October 7th 2008 from http://dublincore.org/.

Deci, E. L. (1975). *Intrinsic Motivation*. New York: Plenum Press.

De Freitas, S. & Neumann, T. (2009). The use of 'exploratory learning' for supporting immersive learning in virtual environments. *Computers and Education, 52*, 343-352.

De Freitas, S., Oliver, M. (2005). *A four dimensional framework for the evaluation and assessment of educational games.* Paper accepted for Computer Assisted Learning Conference 2005.

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science, 38*, 105-134.

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.

De Vicente, A. & Pain, H. (2003). *Validating the Detection of a student's Motivational State.* In S. A. Cerri, G. Gouarderes, F. Paraguacu (Eds.), *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems* (pp. 933-943).

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.

Dillenbourg, P. (1999). *Collaborative Learning: Cognitive and Computational Approaches. Advances in Learning and Instruction Series.,* New York, NY, Elsevier Science Inc.

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.

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.

Dunwell, I., Jarvis, S., and de Freitas, S., (2011) Four-dimensional consideration of feedback in serious games. In *Digital Games and Learning*, P. Maharg and S. de Freitas Eds. Continuum Publishing.

Dunwell, I., Jarvis, S. & de Freitas, S. (2010a). *Four-dimensional consideration of feedback in serious games.* In Digital Games and Learning, P. Maharg & S. de Freitas (Eds.) Continuum Publishing.

Dunwell, I., Petridis, P., Protopsaltis, A., de Freitas, S., Panzoli, D. & Samuels, P. (2010b). *Automating Content Generation for Large-Scale Virtual Learning Environments using Semantic Web Services.* In proceedings of SemWiki2010, Heraklion, Crete, May 2010.

Durand, V. M. & Crimmins, D. B. (1992). *The Motivation Assessment Scale (MAS) administration guide*. Monaco & Associates (Topeka, Kan.).

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.

Egenfeldt-Nielsen, S. (2005). *Beyond Edutainment: Exploring the Educational Potential of Computer Games.* IT-University.

Ekman, P. (1997). *Facial Action Coding System',Consulting*. Psychologists Press.

Elliott, B. (2008). *Online Collaborative Assessment*, Scottish Qualifications Authority.

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.

Foltz, P., Laham, D. & Landauer, T. (1999), 'The intelligent essay assessor: Applications to educational technology', *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*.

Ford, M. E. (1992). *Motivating Humans: Goals, emotions, and personal agency beliefs*. Newbury Park: Sage. Last retrieved October 18th 2010 from http://books.google.at/books?id=_bRmaOI8bQC&printsec=frontcover&dq=ford+1992&hl=de&ei=dvu7TKbbGdHPjAfSqfTRDg&sa=X&oi=book_result&ct=result&resnum=9&ved=0CFQQ6AEwCA#v=onepage&q=ford%201992&f=false.

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, J. (2006). *Assessment and Learning*. London: SAGE.

Gardner, H. (2006). *Multiple Intelligences: New Horizons in Theory and Practice*. Paperback.

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.

Gibbs, G. & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education, 1,* 3-31.

Glaser, B. (2005). The Grounded Theory Perspective III: Theoretical Coding. Mill Valley, CA :Sociology Press.

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.

Gronlund, N. E. (1985). Measurement and evaluation in teaching (5th ed.). New York: Macmillan.

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.

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.

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.

Harter, S. (1978). Effectance motivation reconsidered: Toward a developmental model*. Human Development, 1*, 661-669.

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]. Last retrieved on 7th July 2008 from http://jtc.colstate.edu/vol1_1/multimedia.htm.

Hearst, M. (2000). The debate on automated essay grading. *IEEE Intelligent Systems, 15*(5), 22-37, IEEE CS Press.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: John Wiley & Sons.

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.

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.

Hoyles, C., Healy, L. & Pozzi, S. (1992). Interdependence and autonomy: Aspects of groupwork with computers. *Learning & Instruction, 2*, 239-257.

Hillman, D. C., Willis, D. J. & Gunawardena (1994). Learner interface interaction in distance education. An extension of contemporary models and strategies for practioners. *The American Journal of Distance Education, 8*(2), 30-42.

Hollingsworth, J. (1960). Automatic graders for programming classes. *Commun. ACM*, *3*(10), 528-529.

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 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 LD (2008). *IMS Learning Design Specification*. Last retrieved October 7th 2008 from http://www.imsglobal.org/learningdesign/.

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 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.

James, M. (2006). Assessment, Teaching and Theories of Learning. In J. Gardner (Ed.). *Assessment and Learning*. London: Sage Publications, 47-60.

Janssen, J. H. & van den Broek, E. L. (2009). *Guidelines for mobile emotion measurement*. Paper presented at MobileHCI09, Bonn, Germany.

Jarvis, S. & de Freitas, S. (2009). *Evaluation of an Immersive Learning Programme to Support Triage Training: In-game Feedback and its effect on Learning Transfer.* In proceedings of IEEE Games and Virtual Worlds for Serious Applications (VS-GAMES '09), 117-122.

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.

Johnson, D. W. & Johnson, R. (1989). Computer-assisted cooperative learning. *Educational Technology, 26*(1), 12-18.

Johnson, D. W. & Johnson, R. (1992). Positive interdependence: Key to effective cooperation. In R. Hertz-Lazarowitz and N. Miller (Eds.), *Interaction in Cooperative Groups: The Theoretical Anatomy of Group Learning.* Cambridge: Cambridge University Press, 174-199.

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.

Kalyuga, S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction, 19*, 402-410.

Kareal, F. & Klema, J. (2006). Adaptivity in e-learning. *Current Developments in Technology-Assisted Education*, 260-264.

Kato, P., Cole, S. W., Bradlyn, S. & Pollock, B. H. (2008) A video game improves behavioural outcomes in adolescents and young adults with cancer: A randomised trial. *Pediatrics, 122*(2), 305-317.

Kay, R. H. & Loverock, S. (2008). Assessing emotions related to learning new software: The computer emotion scale. *Computers in Human Behavior, 24*, 1605-1623.

Keller, J. M. (1983). Motivational design of instruction. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: An overview of their current status*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Keller, J. M. (1987). Development and use of the ARCS model of motivational design. *Journal of Instructional Development, 10*, 2-10.

Keller, J. M. & Suzuki, K. (2004). Learner motivation and e-learning design: A multinationally validated process. *Journal of Educational Media, 29*(3). Last retrieved August 12nd 2010 from http://www.gsis.kumamoto-u.ac.jp/ksuzuki/resume/journals/2004a.pdf.

Kelly, D. & Weibelzahl, S. (2006). *Raising confidence levels using motivational contingency design techniques.* In Proceedings of International Conference on Intelligent Tutoring Systems (ITS06), 26-30 June 2006, Jhongli, Taiwan.

Kester, et al. (2006). Enhancing social interaction and spreading tutor responsibilities in bottom-up organized learning networks. *IADIS International Conference Web Based Communities*, 80-87.

Kickmeyer-Rust, M. D. (unpublished). *Adaptive Educational Technology*. Unpublished Technical Report. Last retrieved October 6th 2010 from http://wundt.uni-graz.at/staff/kickmeier/03.pdf.

Kickmeyer-Rust, M. D. & Albert, D. (2008). *The ELEKTRA ontology model: A learner-centered approach to resource description.* Advances in Web Based Learning- ICWL 2007, Lecture Notes in Computer Science, 4823/2008. Berlin: Springer. 78-89.

Kickmeyer-Rust, M. D., Steiner C. M. & Albert, D. (2009*). Non-invasive assessment and adaptive interventions in learning games.* [online]. Last retrieved October 20th 2010 from http://www.eightydays.eu/uploads/tx_takoscientificpublications/IWASCL2009_Kickmeier-Rust_et_al_Final_.pdf.

Kilpatrick, S., Barrett, M. & Jones, T. (2010). *Defining learning communities*. Last retrieved October 8th 2010 from http://www.aare.edu.au/03pap/jon03441.pdf.

Kim, M. (2009). *Peer assessment as a learning method*. Saarbrücken: Müller.

Kirkley, S. E. & Kirkley, J. R. (2007). Creating next generation blended learning environments using mixed reality, Video Games and Simulations. *Humanities, Social Sciences and Law, 49*, 3, 42-53.

Kluger, A. N. & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254-284.

Knight, P. (1995). Assessment for learning in Higher Education. London: Kogan Page.

Knight, J., Carly, S., Tregunna, B., Jarvis, S., Smithies, R., de Freitas, S., Dunwell, I. & Mackway-Jones, K. (2010). Serious gaming technology in major incident triage training: A pragmatic controlled trial. *Resuscitation 81*(9), 1174-9.

Kolb, A. Y. (1984). *Experiential Learning: Experience as the source of learning and development*. New Jersey: Prentice Hall.

Kolb, D. A. & Fry, R. (1975) Toward an applied theory of experiential learning. In C. Cooper (Ed.) *Theories of Group Process*. London: John Wiley.

Kolb, A. Y. & Kolb, D. A. (2005). *The Kolb learning style inventory – Version 3.1 Technical Specifications*. Technical Report.

Kollar, I. & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*, 344-348.

Koper, E. J. R. (2001*). Modelling units of study from a pedagogical perspective: The pedagogical metamodel behind EML*. Open University of the Netherlands.

Kort, B. & Reilly, R. (2002). *Analytical Models of Emotions, Learning and Relationships: Towards an Affect-sensitive Cognitive Machine*, M.I.T. Media Laboratory.

Koschmann, T. (1996). *CSCL: Theory and Practice of an Emerging Paradigm.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Kozma, R. (1991). *"Learning with media"*, Review of Educational Research, *61*(2), 179-211.

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.

Kukich, K. (2000). Beyond Automated Essay Scoring. *IEEE Intelligent Systems, 15*(5), 22-27.

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.

Larsen, R. J. & Diener, E. (1992). Promises and problems with the Circurnplex model of emotion. Last retrieved October 12nd 2010 from http://www.csun.edu/ ~gk45683/ Larsen%20and%20Diener%20% 281992%29.pdf.

Lave, J. & Wenger, E. (1990). *Situated Learning: Legitimate Periperal Participation*. Cambridge, UK: Cambridge University Press.

Law, E. L. C. & Kickmeier-Rust, M. (2008). 80 Days: Immersive Digital Educational Games with Adaptive Storytelling. STEG 08.

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.

LeDeux, J. E. (1999*). The Emotional Brain: The Mysterious Underpinnings of Emotional Life.* London: Phoenix.

LeDoux, J. E. (2000). Emotion Circuits in the Brain. *Annual Review of Neuroscience, 23*, 155-184.

Leon, S. & Nikov, A. (2010). *Intelligent Emotion-Oriented eCommerce Systems*. In Proc. 9th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, pp. 202-207.

Leuf, B. & Cunningham, W. (2001). *The Wiki Way Quick Collaboration on the Web.* Addison Wesley.

Lin, A. C. H. & Gregor, S. (2006). Designing websites for learning and enjoyment: A study of museum experiences. *The International Review of Research in Open and Distance Learning, 7*. Last retrieved August 24th 2010 from http://www.irrodl.org/index.php/irrodl/article/view/364/735.

Luckin, R., Shurville, S. & Browne, T. (2010). Initiating e-learning by stealth, participation and consultation in a late majority institution. *In Shurville, S. and Browne, T. (eds) ICT-driven organisational transformation in HE. A special issue of: Journal of Organisational Transformation and Social Change* 3**,** 317-332

Macdonald, J. (2003). Assessing Online Collaborative Learning: Process and Product. *Computers & Education, 40*(4), 377-391.

Maclean, D. P. (1990). *The Triune Brain In Evolution: Role In Paleocerebral Functions.* New York: Plenum Press.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 5*, 333-369.

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.

Maslovaty, N. & Kuzi, E. (2002). Promoting motivational goals through alternative or traditional assessment. *Studies in Educational Evaluation, 28*, 199-222.

Maslow, A. H. (1970). *Motivation and Personality*. New York: Harper & Row.

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.

Mautone, T., Spiker, V. A. & Karp, M. R. (2008). Using serious game technology to improve aircrew training. In *proceedings of the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*.

Mayer, R. E. & Moreno, R. (2000). 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. (1995). *A methodology for designing post graduate professional development distant learning CSCL programmes. CSCL Proceedings* [on-line]. Last retrieved October 18th 2010 from http://delivery.acm.org/10.1145/230000/222192/p234-mcconnell.pdf?key1=

222192&key2=6119837821&coll=GUIDE&dl=GUIDE&CFID=106231274&CFTOKEN=44180163.

McGreal, R. (2006). *Online education using learning objects.* London, New York: Routledge.

McKeachie, W. J. (2005). *Learning styles can become learning strategies.* Last retrieved February 8th 2005 from http://www.ntlf.com/html/pi/9511/article1.htm.

McLester, S. (2007). Technology literacy and the myspace generation: They're not asking permission. *Technology & Learning, 27*, 16-22.

McQueen, R. A. & Knussen, C. (2006). *Introduction to research methods and statistics in psychology.* Edingburgh: Pearson.

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.

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.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

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. 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. & 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.

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.

Mitchell, A. & Savill-Smith, C. (2004). T*he use of computer and video games for learning*. Learning and Skills Development Agency, London.

Montola, M. (2010). A Ludological View on the Pervasive Mixed-Reality Game Research Paradigm. Personal and Ubiquitous Computing journal. Springer.

Moreno, R. & Mayer, R. E. (2000). A learner-centered approach to multimedia explanations: Deriving instructional design principles from cognitive theory. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning, 2*. Last retrieved August 19th 2010 from http://imej.wfu.edu/articles/2000/2/05/index.asp.

Mory, E. H. (2004). *Feedback research revisited.* Handbook of Research on Educational Communications and Technology, 2nd Edition, D. H. Jonassen Eds. Lawrence Erlbaum Associates, 745-783.

Mott, B. W., McQuiggan, S. W., Lee, S., Lee, S. Y. & Lester, J. C. Narrative-Centered Environments for Guided Exploratory Learning. *5th International Joint Conference on Autonomous Agents and Multiagent Systems*. 2006, Hakodate, Japan.

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.

Nagi, K. (2006). *Solving ethical issues in elearning.* Third International Conference on eLearning for knowledge-based Society, Bankok, Thailand.

Naidu, S. (2006). *E-learning. A guidebook of principles, procedures and practices.* [online]. Last retrieved October 21st 2010 from http://www.cemca.org/e-learning_guidebook.pdf

Neisser, U. (1967). *Cognitive psychology*. Appleton-Century-Crofts New York.

Norman, D. A. (1980). Twelve issues for cognitive science. *Cognitive Science, 4*, 1-32.

Ochs, M. & Frasson, C. (2004). *Emotionally Intelligent Tutoring Systems*. FLAIRS Conference.

OECD (2008). *New Millenium Learners*. Last retrieved October 4th 2010 from https://www.oecd.org/dataoecd/39/51/40554230.pdf.

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.

O' Regan (2003). Emotion and e-learning. *Journal of Asynchronous Learning Networks, 7*(3), 78-92.

Paas, F., Tuovinen, J. E., van Merriënboer, J. J. G. & Darabi, A. A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research & Development, 53*(3), 25

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.

Palla, G., Derényi, I., Farkas, I. & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature,* 435**,** 814–818.

Palmer, J., Williams, R. & Dreher H. (2002). Automated essay grading system applied to a first year university subject – How can we do it better. *Proceedings of the Informing Science and IT Education (InSITE) Conference, Cork, Ireland*, 1221-1229.

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.

Papasalouros A., Kotis K., Kanaris K. (2008). Automatic generation of multiple-choice questions from domain ontologies. IADIS e-Learning 2008 conference, Amsterdam.

Paulsen, M. F. (2007). *Cogs, clips e outros instrumentos de apoio à aprendizagem cooperativa realizada em ambientes virtuais*. In B. Almeida et al. (Eds), *Livro de actas encontro de educacao em fisca do ensino basico ao supirior no sec*, XXI.

Pausch, R., Proffitt, D. & Williams, G. (1997). *Quantifying immersion in virtual reality.* In Proceedings of the 24th Annual Conference on Computer Graphics and interactive Techniques International Conference on Computer Graphics and Interactive Techniques. ACM Press/Addison-Wesley Publishing Co., New York, NY, pp. 13-18.

Pekrun, R., Elliot, A. J. & Maier, M. A. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology, 98*, 583-597.

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.

Perry, W. (1968). *Intellectual and Ethical Development in the College Years: A Scheme', Holt*, New York: Rinehart & Winston.

Picard, R. (1997). *Affective Computing*. Cambridge: MIT Press.

Posner, J., Russell, J. A. & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology, *Development and Psychopathology, 17*, 715-734.

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.

Pugh, E. V. (2008). *Recognising emotional intelligence in professional standards for teaching*.

Qu, L., Wang, N. L. & Johnson, L. (2005). Detecting the Learner's Motivational States in an Interactive Learning Environment. *Artificial Intelligence in Education*.

Quinn, C. (2005). *Engaging Learning,* Pfeiffer, San Francisco, CA.

Rada, R., Acquah, S., Baker, B. & Ramsey, P. (1993). Collaborative Learning and the MUCH System. *Computers and Education, 20*(3), 225-233.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28*, 4-13.

Redecker, C. (2009). Review of learning 2.0 practices: Study on the impact of web 2.0 innovations on education and training in Europe. JRC Scientific and Technical Reports. Last retrieved 5th October 2010 from ftp://ftp.jrc.es/pub/EURdoc/JRC49108.pdf.

Reed, W. (1990, April), The effect of composing process software on the revision and quality of persuasive essays. Paper presented at the Eastern Educational Research Association, Clearwater, FL.

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. & 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.

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.

Riedl, M. O., Stern, A., Dini, D. & Alderman, J. (2008). Dynamic Experience Management in Virtual Worlds for Entertainment, Education, and Training. *International Transactions on Systems Science and Applications, Special Issue on Agent Based Systems for Human Learning,* 4.

Roberts, T. S. (2006). *Self, peer and group assessment in e-learning*. Hershey: Information Science Publishing.

Rogers, Carl. (1951). *Client-centered Therapy: Its Current Practice, Implications and Theory.* London: Constable.

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.

Rosenberg, M. (2001*). E-learning Strategies for Delivering Knowledge in the Digital Age*. New York: McGraw-Hill.

Rottmann, R. M. & Hudson, H. T. (1983). Computer Grading As an Instructional Tool. *Journal of College Science Teaching, 12*, 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.

Salen, K. & Zimmerman, E. (2003). *Rules of Play-Game Design Fundamentals*, Cambridge: MIT Press.

Salmon, G. (2000). *E-moderating: The Key to Teaching and Learning Online*. London: Kogan-Page.

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.

Scardamalia, M. & Bereiter, C. (1996). Computer support for knowledge-building communities. In T. Koschmann (Ed.) *CSCL: Theory and Practice of an Emerging Paradigm* (pp. 249-268). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schlechty, P. C. (1997). *Inventing better schools: An action plan for educational reform*. San Francisco, CA: Jossey-Boss.

Schultz, D., Ambike, A., Logie, S. K., Bohner, K. E., Stapleton, L. M., Vanderwalde, H., Min, C. B. & Betkowski, J. A. (2010). Assessment of Social Information Processing in early childhood: development and initial validation of the Schultz Test of Emotion Processing-Preliminary Version. *Journal of abnormal child psychology,* 38**,** 601-13.

SCORM (2008). *ADL Sharable Courseware Object Reference Model. Last r*etrieved October 7th 2008 from http://www.adlnet.gov/scorm/index.aspx.

Seligman, M. (1975). *Helplessness: On Depression, Development, and Death*. New York: Freeman.

Sharan S. & Sharan Y. (1992). Expanding Cooperative Learning through Group Investigation. New York: Teachers College Press.

Sharda, R., Romano, N. C. Jr., Lucca, J. A., Weiser, M., Scheets, G., Chung, J.-M. & Sleezer, C. M. (2004). A Foundation for the study of computer-supported collaborative learning requiring immersive presence. *Journal of Management Information Systems, 20*, 31-63.

Shellnut, B. J. (1998). *John Keller: A motivation influence in the field of instructional systems design*. Last retrieved August 18th 2010 from http://www.arcsmodel.com /pdf/Biographical%20Information.pdf.

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.

Shia, R. M. (1998). *Assessing academic intrinsic motivation: A look at student goals and personal strategy.* Last retrieved February 8th 2005 from http://www.cet.edu/research/student.html

Shute, V. J. (2008). Focus on formative feedback*. Review of educational research*, *78*(1), 153-189.

Siang, A. C. & Rao, R. K. (2003). Theories of learning: A computer game perspective. Proceedings of the IEEE Fifth International Symposium on Multimedia Software Engineering (ISMSE'03).

Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review, 24*(2), 86-97.

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.

Small, R. (1997). *Motivation in instructional design.* ERIC Digest. Last retrieved March 1st 2005 from http://www.eric.ed.gov.

Smith, M. & Berge, Z. L. (2009). Social learning theory in Second Life. *Journal of Online Learning and Teaching, 5*. Last retrieved August 27th 2010 from http://jolt.merlot.org/vol5no2/berge_0609.htm.

Smythe, C. & Roberts, P. (2000). *An Overview of the IMS Question & Test Interoperability Specification"*, Computer Aided Assessment (CAA'2000), Leicestershire, UK.

Stahl, G. (2000) book review: Professional Development for Cooperative Learning: Issues and Approaches, Teaching and Learning in Medicine: An International Journal. Available at: http://www.cs.colorado.edu/~gerry/publications/journals/ medicine/coop_learn.html.

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).

Städler, T. (2003). *Lexikon der Psychologie* [Encyclopaedia of Psychology]. Stuttgart: Kröner.

Steinberg, C. (2008). Assessment as an "emotional practice". *English teaching: Practice and Critique, 7*, 42-64.

Stemler, S. (2001). An overview of content analysis. *Practical Assessment, Research and Evaluation, 7*(17).

Streeter, L., Pstoka, J., Laham, D. & MacCuish, D. (2003). The credible grading machine: Automated essay scoring in the dod.

Sukkarieh, J., Pulmand, S. & Raikes, N. (2003). Auto-marking: Using computational linguistics to score short, free text responses.

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.

Svinicki, M. D. (1999). New directions in learning and motivation. *New Directions for Teaching and Learning, 80,* 5-27.

Swan, K., Shen, J. & Hiltz, R. (2006). Assessment and collaboration in online learning. *Journal of Asynchronous Learning Networks, 10*(1), 45-62.

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.

Valenti, S., Neri, F. & Cucchiarelli, A. (2003). An overview of current research on automated essay grading, *Journal of Information Technology Education 2*.

Van Merriënboer, J. J. G. & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research & Development, 53*(3). 5-13.

Verhoeven, L. Schnotz, W. & Paas, F. (2009). Cognitive load in interactive knowledge construction. *Learning and Instruction*, 19, 369-375.

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.

Vygotsky, L. S. (1996) *Thought and Language*. Cambridge, MA: MIT Press.

Wang, M. & Kang, M. (2006). *Cybergogy for Engaged Learning: A framework for creating learner engagement through information and communication technology. In Engaged Learning with Emerging Technologies* Springer, Netherlands, pp. 225-253.

Wang, K. H., Wang, T. H., Wang, W. & Huang, S. C. (2006). Learning styles and formative assessment strategy: enhancing student achievement in web-based learning. *Journal of Computer Assisted Learning, 22,* 207-217.

Warburton, S. Defining a Framework for Teaching Practices Inside Virtual Immersive Environments: the Tension Between Control and Pedagogical Approach. In *proceedings of RELive '08 Conference,* 2008.

Weber, R. & Shaw, F. (2009). Player types and quality perceptions: A social cognitive theory based model to predict video game playing. *International Journal of Gaming and Computer-Mediated Simulations, 1*, 66-89.

WebPA (2009). http://webpaproject.lboro.ac.uk. Last retrieved 5th February 2009.

Webster, W. (1990). *'Computers and writing assessment: A preliminary view'.*

Weiner, B. (1986). An *attributional theory of motivation and emotion*. New York: Springer.

Wen, M. L. & Tsai, C.-C. (2006). University Students' Perceptions of and Attitudes Toward (Online) Peer Assessment. *Higher Education, 51*, 27-44.

Wertsch, J. V. (1985). *Cultural, Communication, and Cognition: Vygotskian Perspectives.* Cambridge University Press.

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review, 66*, 297-333.

Whitelock, D. & Brasher, A. (2006). *Roadmap for e-Assessment* Report for JISC.

Wiggins, G. (1990). The case for authentic assessment. [online]. Last retrieved August 16th 2010 from http://eric.ed.gov/PDFS/ED328611.pdf.

Williams, R. & Dreher, H. (2004). Automatically Grading Essays with Markit, In *proceedings of Informing Science*, Rockhampton, Australia.

Williamson, D., Bejar, I. & Hone, A. (1999), Mental model comparison of automated and human scoring, *Journal of Educational Measurement, 36*, 158-184.

Wong, M. (2006*). Emotion assessment in evaluation of affective interfaces.* [online]. Last retrieved October 19th 2010 from http://www.cgl.uwaterloo.ca/~wmcowan/research/essays/maria.pdf.

Woo, Y. & Reeves, T. R. (2007). Meaningful interaction in web-based learning: A social constructivist interpretation. *Internet and Higher Education, 10*, 15-25.

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.

Zeidner, M. (2007). Test anxiety in educational contexts: Concepts, findings and future directions. In P. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 165-184). Amsterdam: Elsevier.

Zhang, G., Cheng, Z., He, A. & Huang, T. (2003). *A WWW-based Learner's Learning Motivation Detecting System.* In Proceedings of International Workshop on "Research Directions and Challenge Problems in Advanced Information Systems Engineering, Honjo City, Japan.

Zimmermann, P., Guttormsen, S., Danuser, B. & Gomez, P. (2003). *Affective Computing – A Rationale for Measuring Mood with Mouse and Keyboard*. Swiss Federal Institute of Technology, Zürich, Switzerland.

Zumbach, J., Hillers, A. & Reimann, P. (2003). Supporting distributed problem-based learning: the use of feedback in online learning. In T. Roberts (Ed.), *Online Collaborative Learning: Theory and Practice* (pp. 86-103). Hershey, PA: Idea Group.

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer, 28*(9), 25-32.

Annex A

A.1 First validation of model by querying experts

A.1.1 Questionnaire (Validation of the Model)

Demographic data

Gender

- Female x
- Male

Highest level of education

University

Working field …

Research, Teaching

Research interests (main topics)

Knowledge assessment, inductive reasoning, cognition

Questions concerning the model

1. Do you think that the model provides an accurate representation of the real world?

I strongly disagree	I disagree	neither/nor	I agree x	I strongly agree

Why / Why not?

Yes, but title and model emphasize enriched learning experience not assessment (assessment is part of model)

2. Do you think that the model provides a substantial complete representation of the real world?

I strongly disagree	I disagree	neither/nor	I agree x	I strongly agree

Why / Why not?

IMA is a very comprehensive model including the main components involved in e-learning and e-assessment and therefore provides a good representation of a high-quality e-learning environment. Comparing the model, for example, to the 10 Pedagogic principles of E-learning (Anderson & McCormick, 2005), seven of the principles can easily be retrieved within the model (match to the curriculum, learner engagement, innovative approaches, effective learning, formative, assessment, summative assessment), the remaining three (inclusion, e.g. support of physical disabilities, coherence/consistency/transparency, and cost-effectivness) are not directly pointed out, but presumably covered by the inputs "Standards & Specifications" as well as "Technology". (see also next question)

3. Is there something missing?/ Would you add something?

   The model seems to be sufficient regarding its main components, inputs, and interactions with other models, however, the specifications for the main components appear to be selective at some places. E.g. it is not clear, whether the listed CLRs are exhaustive or only examples of resources to be used in the environment (in the latter case, the question arises why other learning objects like text elements, videos, etc. are excluded).

4. Is there any obvious error in the model?

   No

5. Are the components of the model comprehensible?

- Regarding the assessment, it should be clarified, if the different forms of assessment are optional or if at least one type per item needs to be fulfilled in each scenario. Meaning, if you can choose between e.g. assessment of knowledge and skills vs. behavioral assessment or if you always have assessment of skills, behaviors, emotions, etc. ….
- Generally the difference between evaluation/validation and efficiency/effectiveness should be elaborated, e.g. by defining sets of indicators/criteria and/or by specifying what is to be evaluated/validated and which efficiency/effectiveness is to be tested. Also the relationship between quality assurance and educational efficiency/effectiveness is not totally clear (which indicators –> same/different; what about technical quality assurance?)
-

6. Would you skip components or set other priorities?

   I wouldn't skip anythink, but try to emphasize the assessment-component (see also question 9)

7. Do the relations between the components make sense?

   Evaluation and validation results refer to and might therefore influence the chosen assessment forms and learning activities (without changing the didactic objectives). Thus the arrow back should probably branch off to the CLR and the assessment forms (question: is linearity from didactical objective to evaluation correct or does objective influence learning resources, assessment form and eval/val equally, while the latter three show reciprocal influences???).

8. Are the flows correct?

   Regarding the three adaptivity components, it should be indicated within the abstract model, that the three mentioned models are interacting in order to provide adaptivity. Furthermore, it might make sense to show where exactly and at what point in time the adaptivity components take effect/interact? E.g. ontologies from the knowledge model are used with metadata associated with learning resources; the learning model is also associated to CLR (preferred approach); the didactic model defines the learning paths, thus it also seems to effect CLR, but is itself influenced by results of assessment.

9. Do you think that the model fits the requirements/objectives?

… the requirements/objectives of WP 5, D5.1.1, I suppose: The Name of the Model (IM**A**) and the WP refer to a model which specifies and designs a framework to **evaluate** didactic experiences in adaptive learning systems. The model itself is in my opinion a model for an “Enriched Learning Experience” (as it is labeled in the graph itself), which includes the topic of evaluation, but does not emphasize it. A framework to “evaluate didactic experiences…” should at least give guidelines or criteria for a qualitatively sound evaluation.

WP5: D5.1.1: SOA about models and assessment methods; specify and design a functional innovative framework to evaluate didactic experiences in adaptive learning systems.

D5.2.1: Enriched learning experience: prototype to verify how the modalities and the formats stated in the model are functional to measurement of the acquired processes of the learners. It will introduce the defined new forms of assessment in complex learning experiences

10. All in all, how would you rate the integrated model regarding to its relevance in the field of e-assessment?

1	2	3	4	5	6 x	7

not relevant very relevant

Comments

What would you especially improve regarding the model?

Generally, one main problem, which is however common in this area, is that the number of possible combinations is too high for an empirical validation. E.g. learning type x learning theory x learning style x learning resource x type of assessment, etc. Thus the model cannot be validated empirically, only by face-validity. Also, to judge the efficiency and effectiveness of the model (question: of the assessment or the entire learning environment?), indicators should be specified (e,g. usability, functionality, satisfaction, motivation, …..).

Do you have any further comments?

To the bottom up framework (Fig. 2 in Al.Smadi et al, 2011): In the first step, namely the application domain, all questions regarding the IMA main components and all IMA –Inputs are supposed to be answered. Is this really necessary before the requirements are identified? And where are the adaptivity components in the bottom-up framework as well as in the Case-Study?

A.2 Second validation of the model and the tools by experts

A Demographic data

Gender

- Female
- Male

Highest level of education

Working field …

Research interests (main topics)

B Integrated Model for e-Assessment

11. The model provides an accurate representation of the real world.

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

Why / Why not?

12. The model provides a substantially complete representation of the real world.

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

If not, what is missing?

13. There is an obvious error in the model.

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

If you agree, please explain the error you observed.

14. The components of the model are easy to comprehend.

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

If not, which components do you think are difficult to understand?

15. All of the included components are relevant and priorities are set appropriately.

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

If you disagree, would you skip components or set other priorities?

16. The relations between the components make sense.

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

If not, how would you change the relations between the components?

17. The flows are correct.

I strongly	I disagree	neither/nor	I agree	I strongly agree

disagree				

If not, what do you think is incorrect concerning the flows and should be improved?

18. The model fits the requirements/objectives to “specify and design a functional innovative framework to evaluate didactic experiences in adaptive learning systems”.

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

Why not?

19. All in all, how would you rate the integrated model regarding to its relevance in the field of e-assessment?

1	2	3	4	5	6	7

not relevant very relevant

Why / Why not?

20. What would you especially improve regarding the model?

21. Do you have any further comments?

C Automatic Question Creator (AQC)

1. The Automatic Question Creator (AQC) adequately supports self-regulated learning environments.

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

Reasons for agreement or disagreement:

2. The AQC can also be used in other learning scenarios such as Story Telling and game-based learning.

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

Reasons for agreement or disagreement:

3. The AQC is also useful to support the instructor of a course.

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

Reasons for agreement or disagreement:

4. The four different question types (open ended, fill in the blank, multiple and single choice) generated by the AQC are suitable to test students' knowledge.

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

Reasons for agreement or disagreement:

5. Would you change any question types or add new ones?

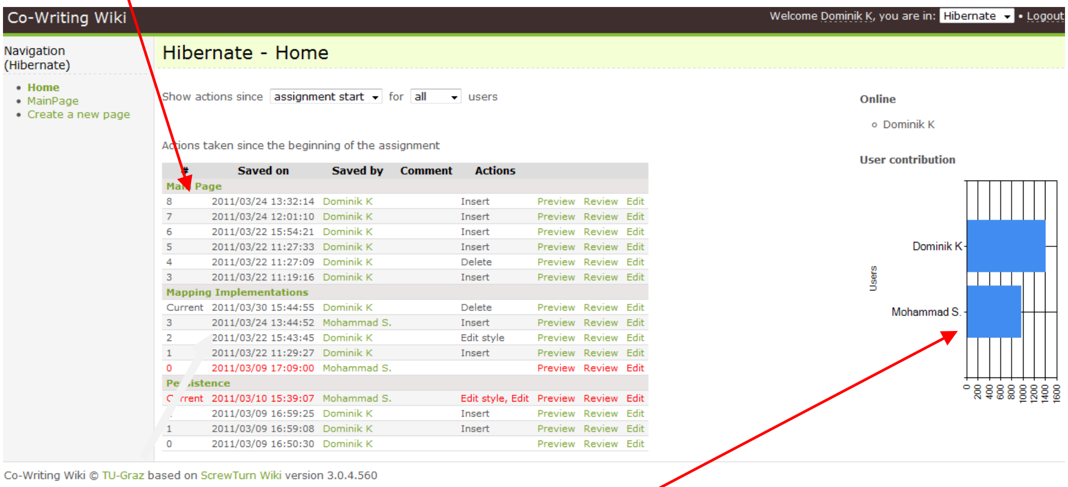
6. Could you imagine using the AQC in your working field?

D Co-writing WIKI

7. Do you think the following components of the wiki are helpful for the instructor?

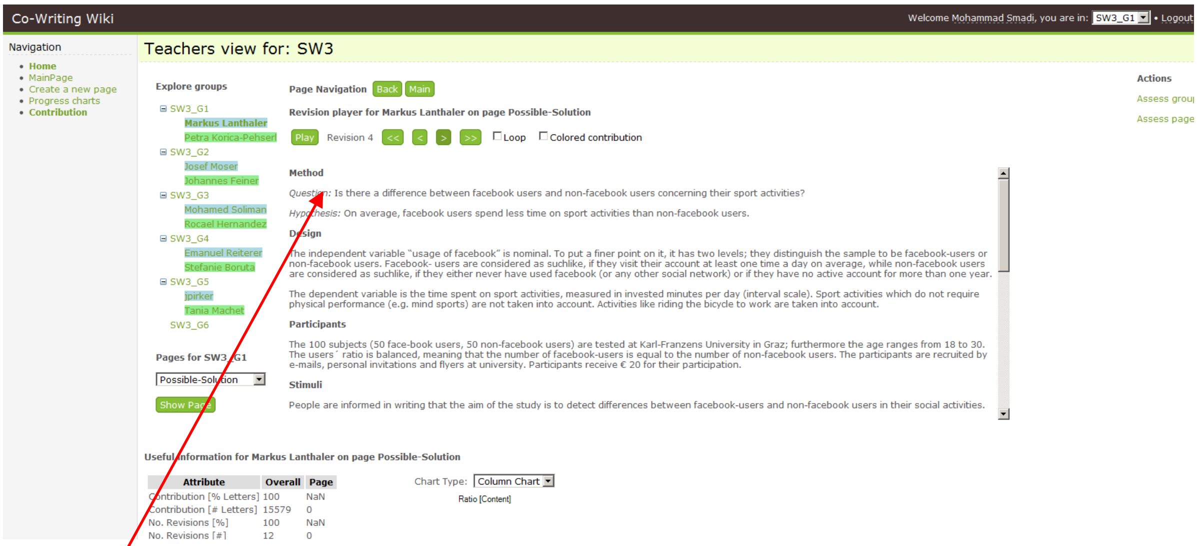
Actions feed in assignment homepage

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree



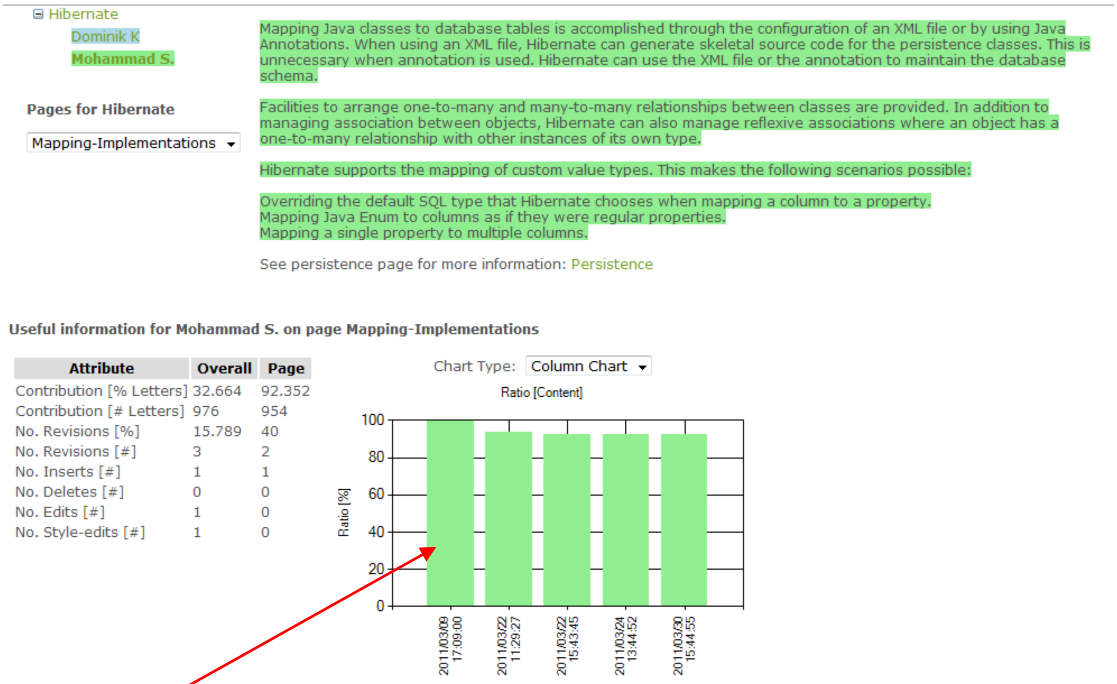
Contribution graphs in assignment homepage

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree



Revisions player in the contribution tool

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

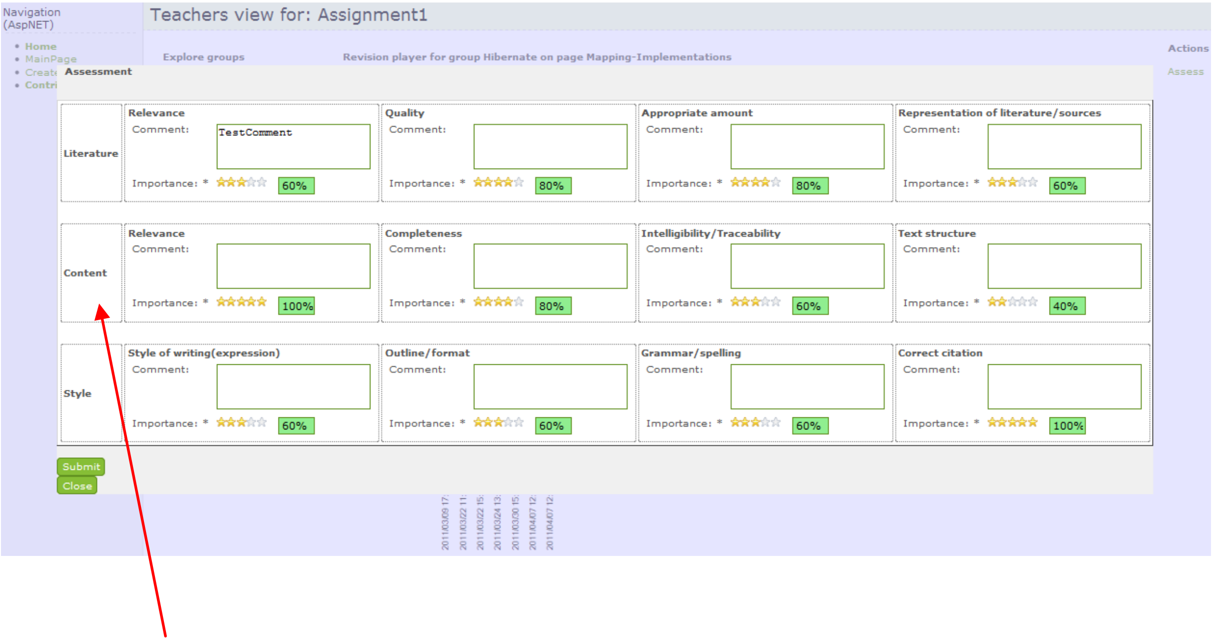


Charts in the contribution tool

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

The rate control (stars)

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree



Rubrics for assessment

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

8. If you found one or more components not helpful for the instructor, please state your reasons why.

9. Do the components support the students in their learning process? Please indicate your level of agreement.

Actions feed in assignment homepage

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

Contribution graphs in assignment homepage

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

Revisions player in the contribution tool

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

Charts in the contribution tool

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

The rate control (stars)

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

Rubrics for assessment

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

10. If you found one or more of the components not helpful for the students, please state your reasons why.

11. Can you think of any further components you would integrate into the tool (to support students and/or teachers).

12. Imagine you have to evaluate students' contributions. Which assessment forms would be helpful?

Self-assessment

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

Peer-assessment

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

Group-assessment

I strongly disagree	I disagree	neither/nor	I agree	I strongly agree

13. In your opinion, in which fields could the Co-writing wiki be used?

14. Could you imagine using the Co-writing wiki in your working field?

15. Regarding the components and the visual presentation of the Co-writing wiki, do you have suggestions or comments for improvements?

User Guide for Co-writing wiki

You can enter the Co-writing wiki with your test accounts, which we provided for you.

You received three accounts:

- With the teacher account you can have a look on the wiki from the viewpoint of an instructor
- Two student accounts will show you the wiki from the viewpoint of the students

So on the one hand, you will find all the features which are available for the teacher, on the other hand you get an insight into actions and collaborative functions, which are provided for students.

This User Guide should help you to find the most relevant components in the system.

Student View

Step 1

Please enter Co-writing wiki with your student account. First you can see the Main Page, go to Home.

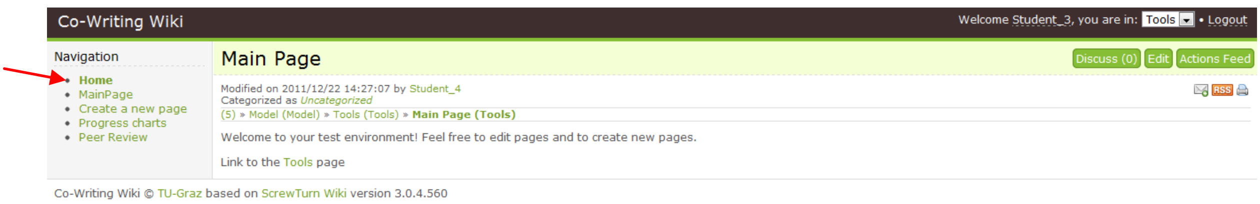


Figure 1: Main Page

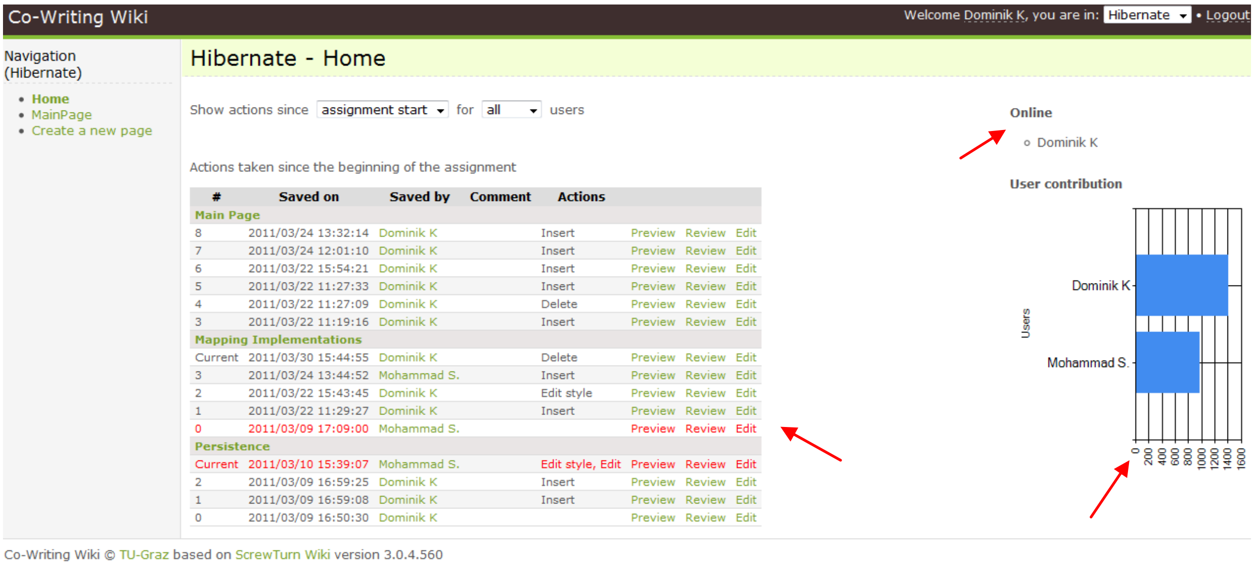


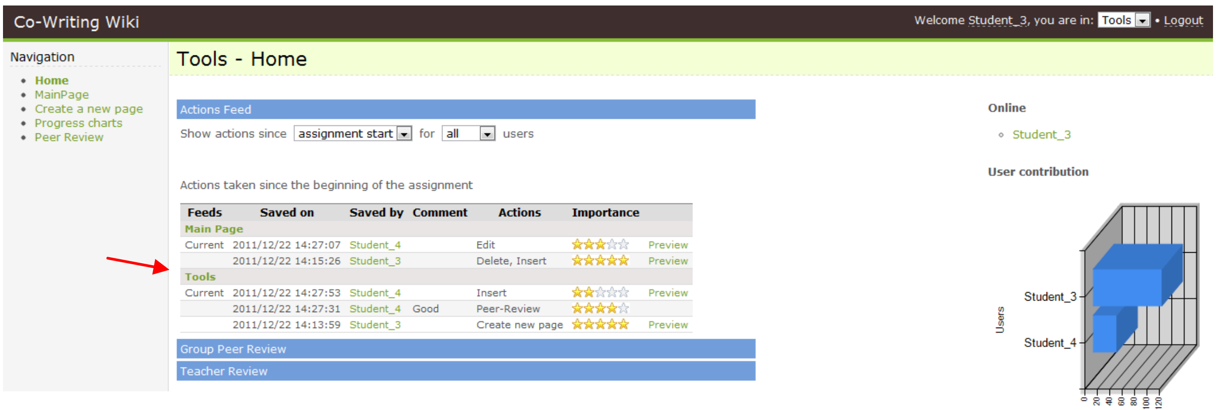
Figure 2: Assignment Homepage

The **assignment homepage** has been designed to maintain social and task awareness as well as to support group function production and well-being. As depicted in Figure 2, the assignment homepage consists of:

- **Actions feed**: the group members' actions on the assignment pages are fed back to the assignment homepage. 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.
- Online peers: shows the currently online group members.
- **Contribution charts**: this graph represents the amount of letters each group member has contributed to the assignment wiki.

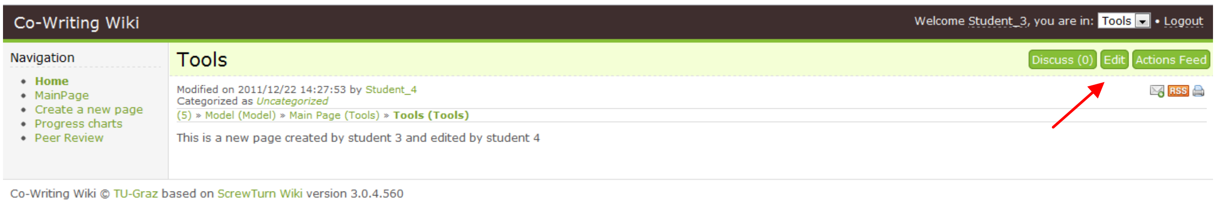
Step 2

Click on the page called Tools/Model to see the contribution of the students.



Step 3

This is the contribution of the students. To edit the page, click on edit.



Step 4

The difference page shows the actions on changes on the last version. Now click on edit current version to change or add something.

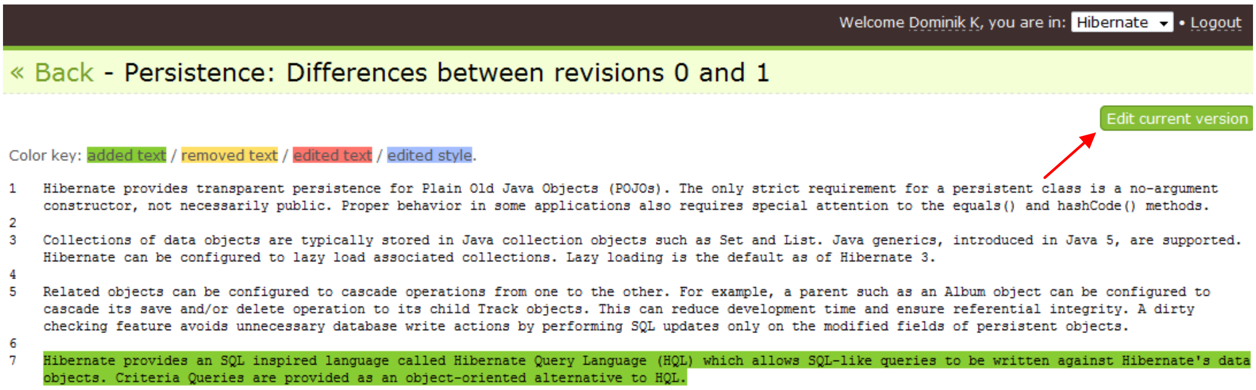
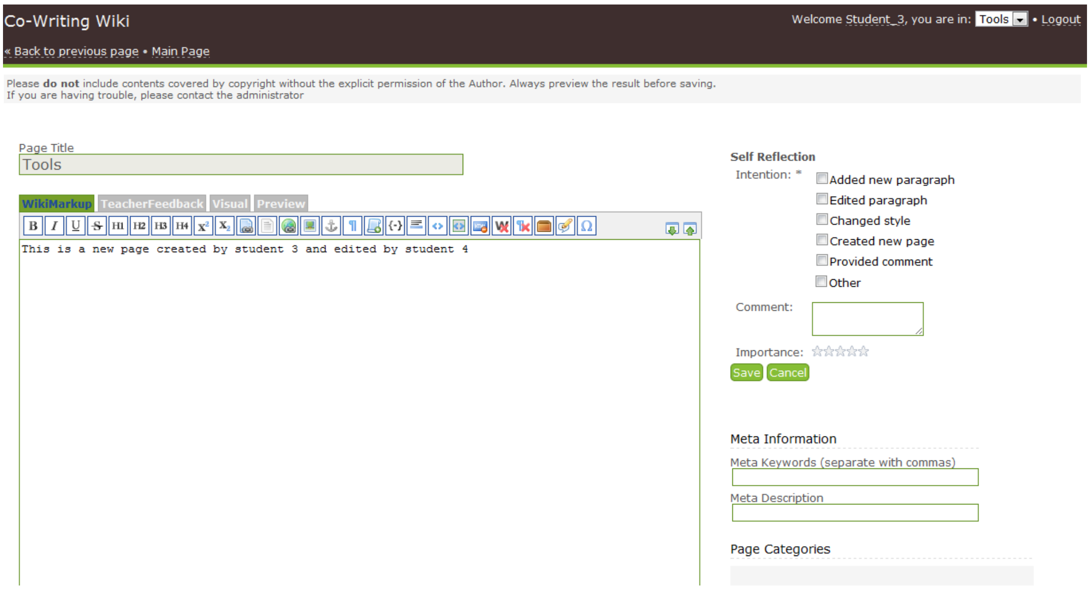


Figure 3: Difference page

Step 5

Now you can add your changes and do the self-reflection before you save your changes.



Group-, Self- and Peer-assessment

Co-Writing wiki is enhanced with tools to conduct both individual's self-assessment and group's peer-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 4.
- **Peer-assessment**: an internal peer-review follows the "Edit" action by which 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 "Peer Review" right upper-corner of Figure 5.
- **Group-assessment**: Co-writing wiki provides a tool for group-assessment. By using this tool students and teachers can assess other group's final product of the assignment and provide feedback. The tool is enhanced with a rubric to facilitate the assessment process, provide feedback, and to maintain persistence and reliable assessment. Moreover, preliminary information about group's product quality can be provided to the teacher in order to support him/her to grade the group product, see Figure 6.

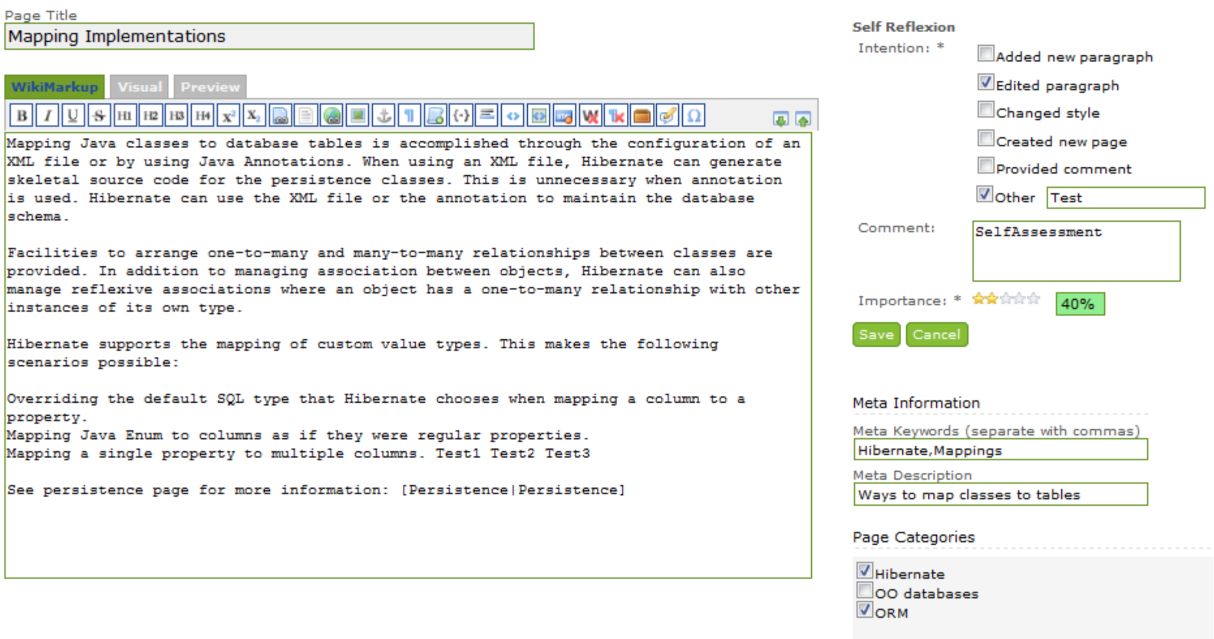


Figure 4: Self-assessment as part of the Edit page

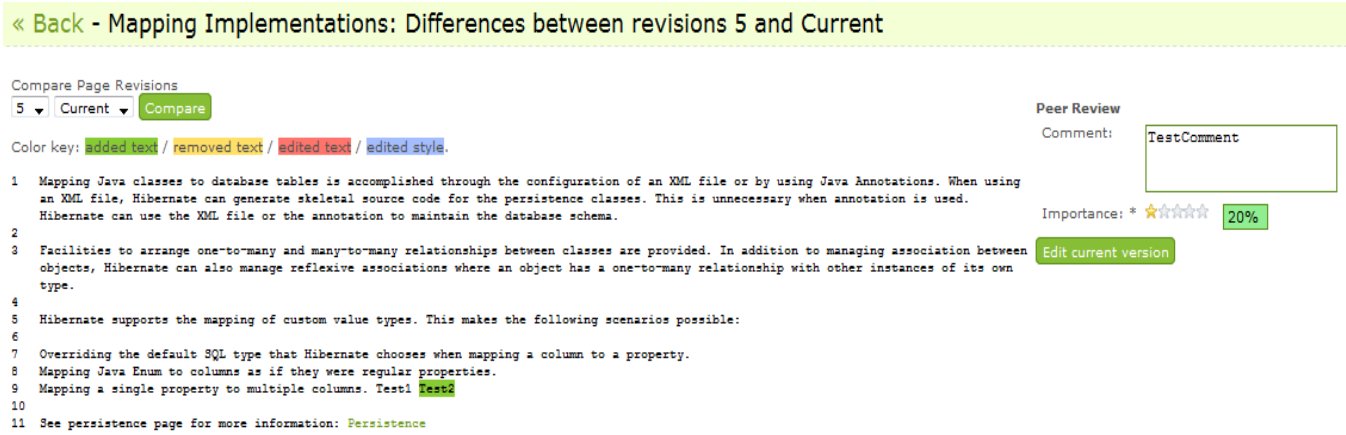


Figure 5: Latest Action Peer-assessment (Right-upper Corner)

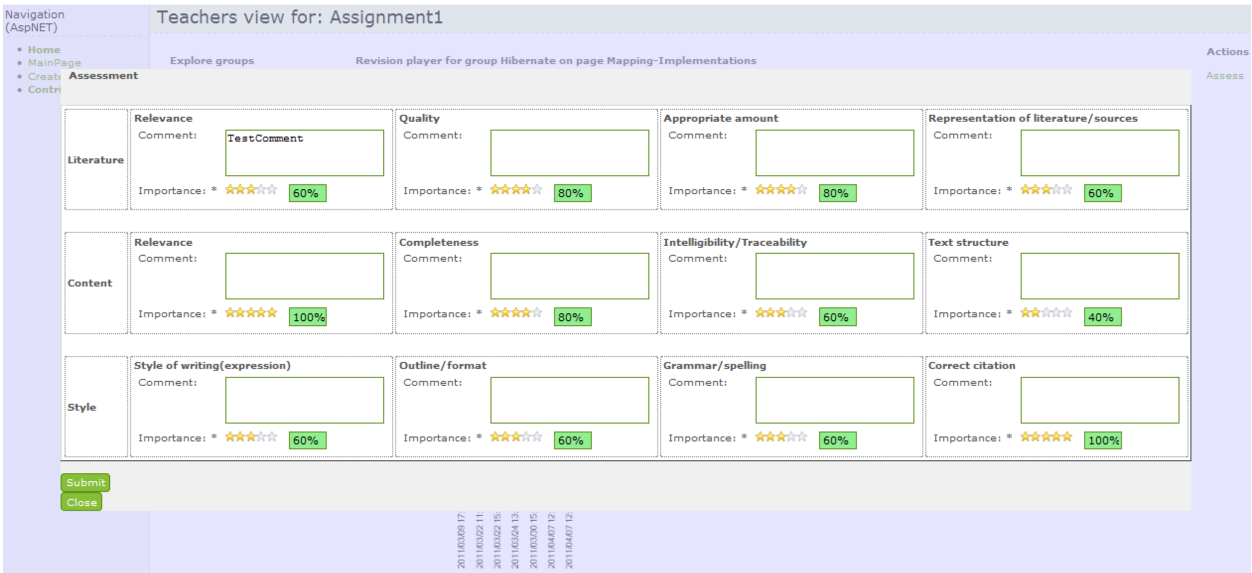
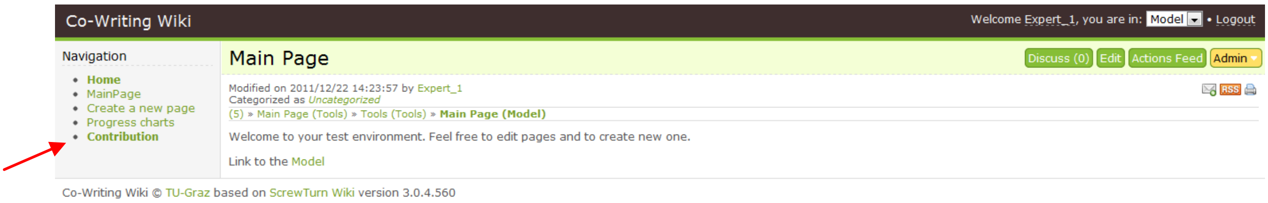


Figure 6: A Rubric used to peer-assess group's final product

Teacher View

Step 1

Enter with your teacher account. Click on contribution.



Step 2

Choose the contribution you want to have a look on (Tools/Model) and click on show page.

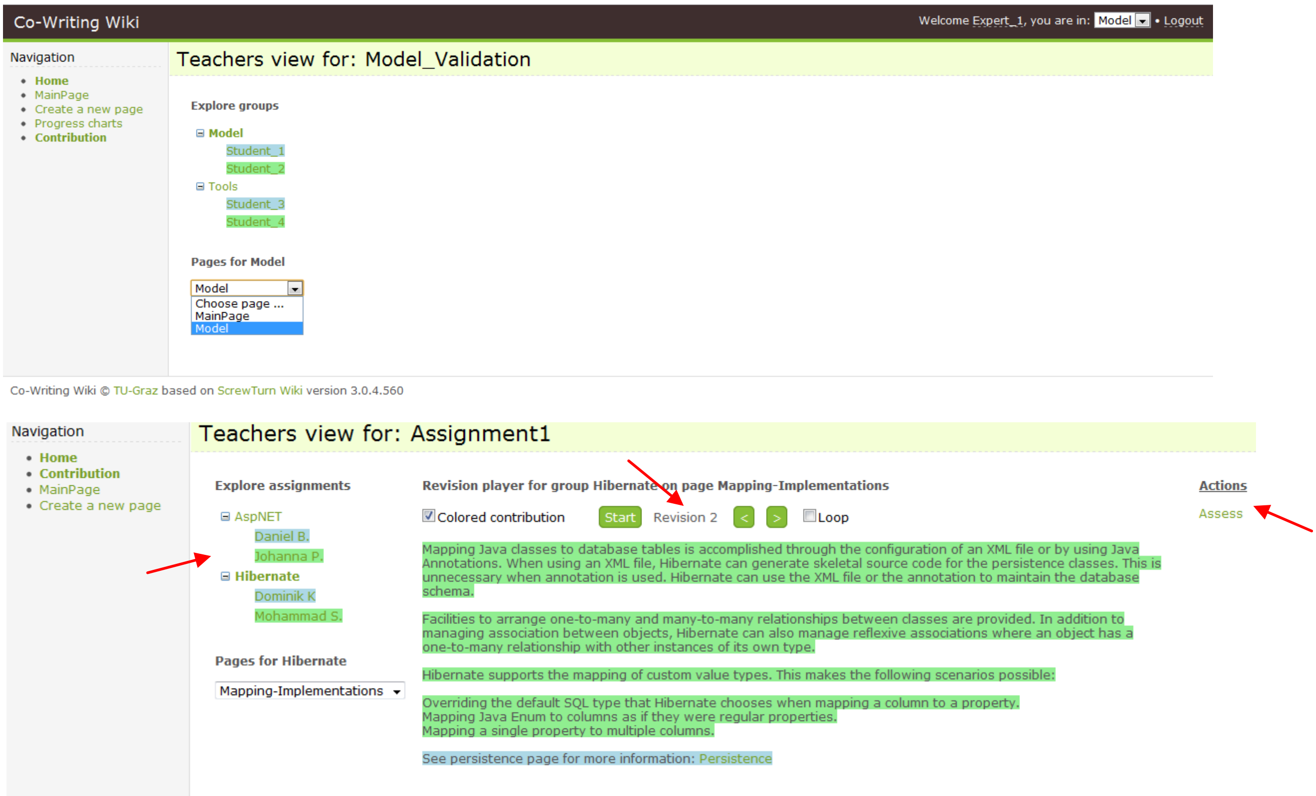


Figure 7. Group's explorer and the History player as part of Teacher View page

The teacher view page has been designed to support teachers with valuable information about the assignment. Quantitative and qualitative textual and visual information representing groups' member's collaboration and contribution to the assignment wiki document are provided within this page.

The teacher view page consists of the following:

- Group navigation: a tree-view has been provided to explore the assignment related groups of students, see Figure 3.
- **Revision player**: the revision player is a tool that demonstrates the color-based wiki document as a slide show. As decited in Figure 3, the player has a start button the plays the page revision from the first revision until the final one. Moreover, it has navigation buttons to navigate forward and backward in the page revisions with color-based contribution of the user.
- Action list: this list contains the possible actions that the teacher may take to evaluate individuals and groups contributions, see the right side of Figure 3.
- Useful information: in this part of the page the teacher gets some useful information about the collaboration process and student's contribution, see Figure 4.
- Chart panel: in this panel, the information is visualized in different charts by which useful information is provided to the teacher, see Figure 4.

Hibernate supports the mapping of custom value types. This makes the following scenarios possible:

Overriding the default SQL type that Hibernate chooses when mapping a column to a property.
Mapping Java Enum to columns as if they were regular properties.
Mapping a single property to multiple columns.

See persistence page for more information: Persistence

Useful information for Mohammad S. on page Mapping-Implementations

Attribute	Overall	Page
Contribution [% Letters]	32.664	92.352
Contribution [# Letters]	976	954
No. Revisions [%]	15.789	40
No. Revisions [#]	3	2
No. Inserts [#]	1	1
No. Deletes [#]	0	0
No. Edits [#]	1	0
No. Style-edits [#]	1	0

Chart Type: Column Chart

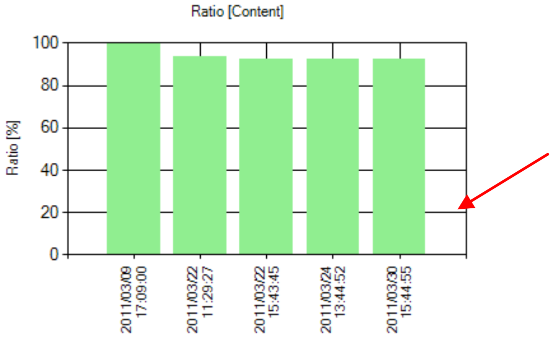


Figure 8. Useful information as part of Teacher View page

A.3 Example of Use regarding the AQC

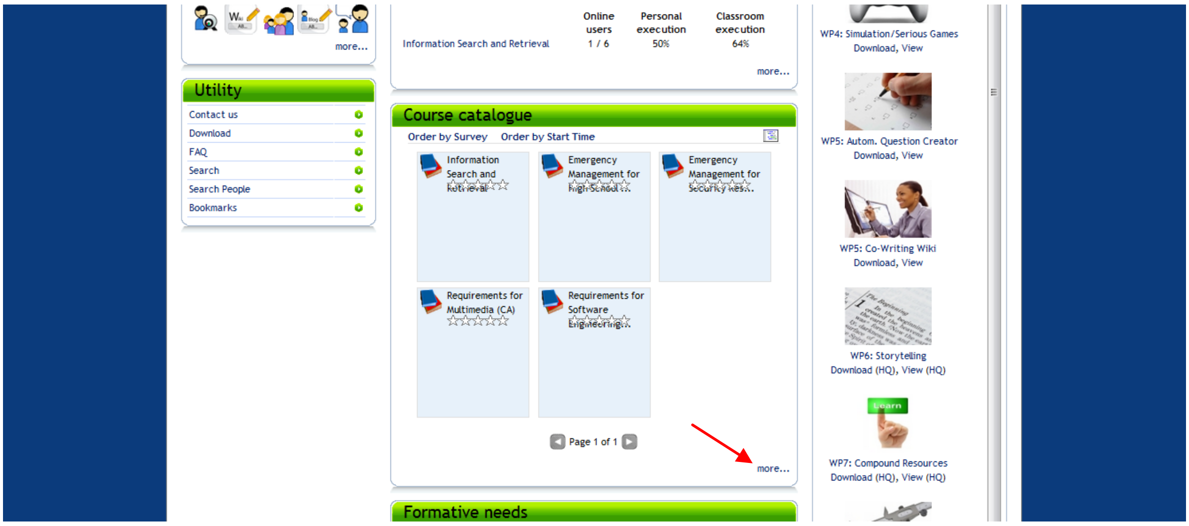
The AQC is integrated in the Intelligent Web Teacher (IWT). To enter the IWT, please register yourself at first. The following example gives you an overview of the AQC used for self regulated learning.

AQC for Self Regulated Learning

Suppose that a student wants to access a self-regulated course on „Information Search and Retrieval“, for short ISR. The student wants to read through the course material on his own and also wants to take automatically created formative tests based on the material.

Step 1

To do this a student logs into IWT (Intelligent Web Teacher) and subscribes to the ISR course in the course catalogue.



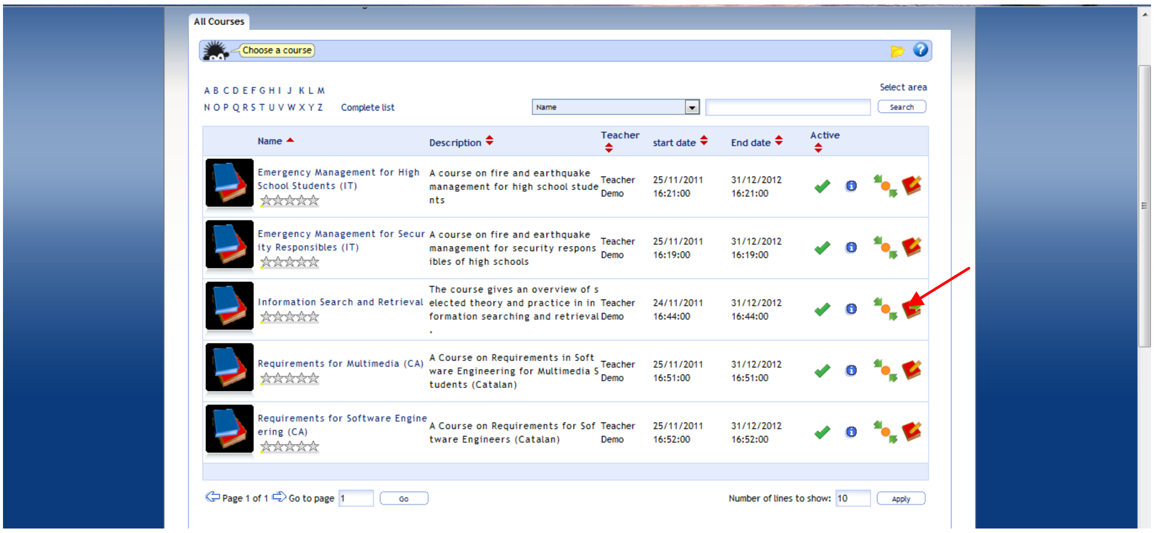


Figure 9. Course selection in IWT

Step 2

After that the student enters the course and reads through a detailed description of the topic.

Click on introduction to isr.

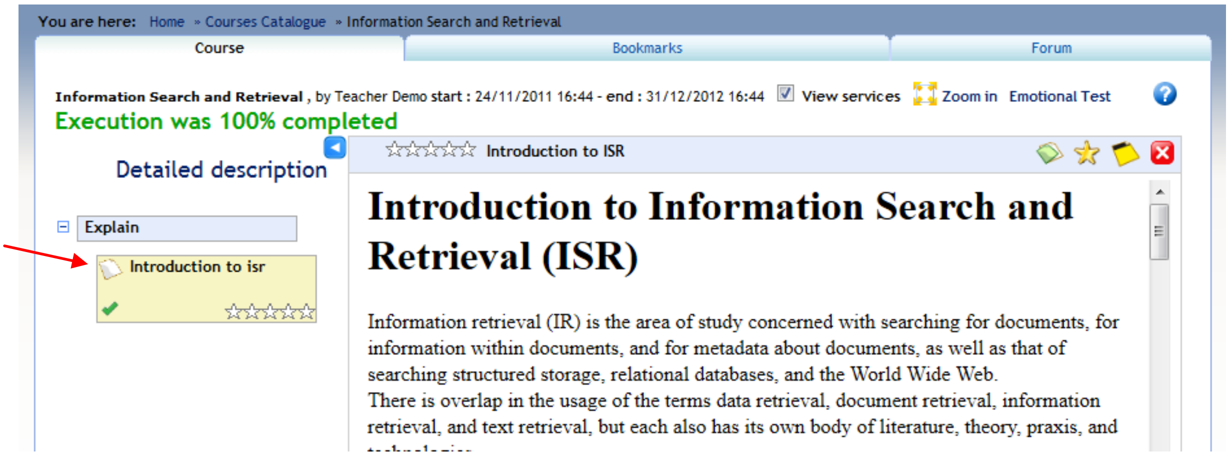


Figure 10. Course description and text material

Step 3

The student can use an integral search engine to find additional material in IWT or on Wikipedia. Click on detailed description, enter a topic, you are interested in and click Wikipedia and search.

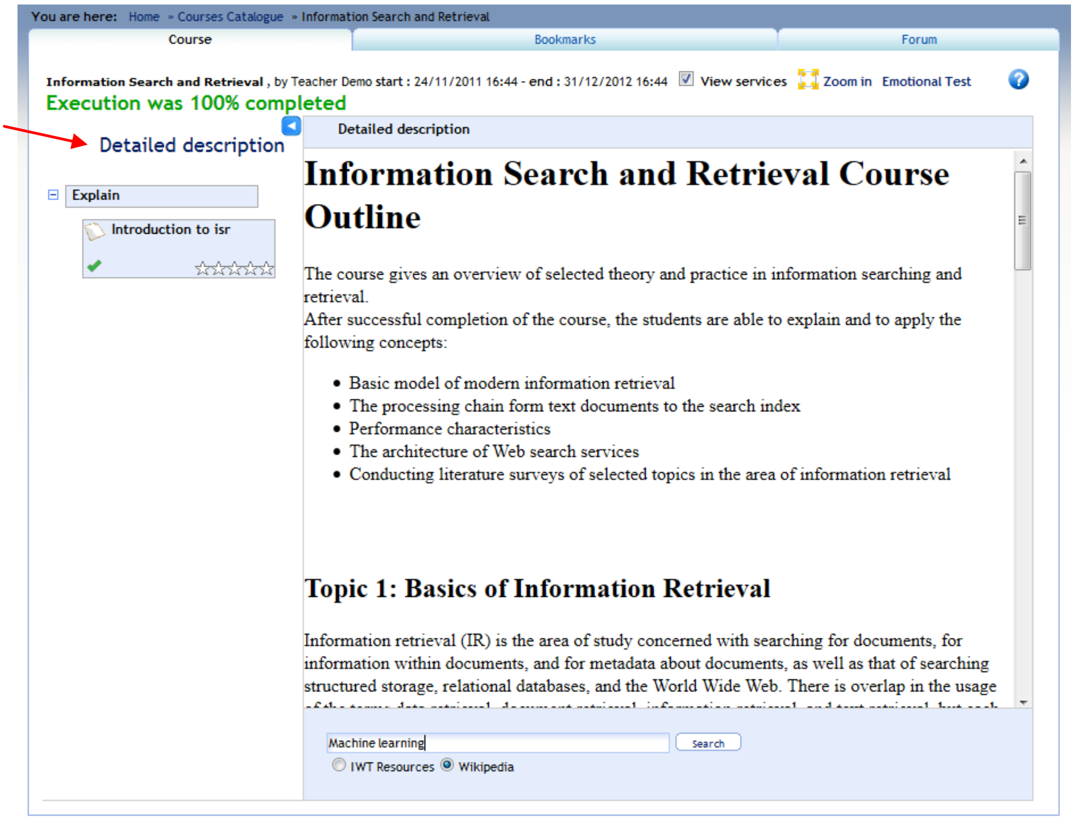


Figure 11. Example for searching additional material in IWT

Step 4

The system shows you the results regarding your topic. Choose one and click on add.

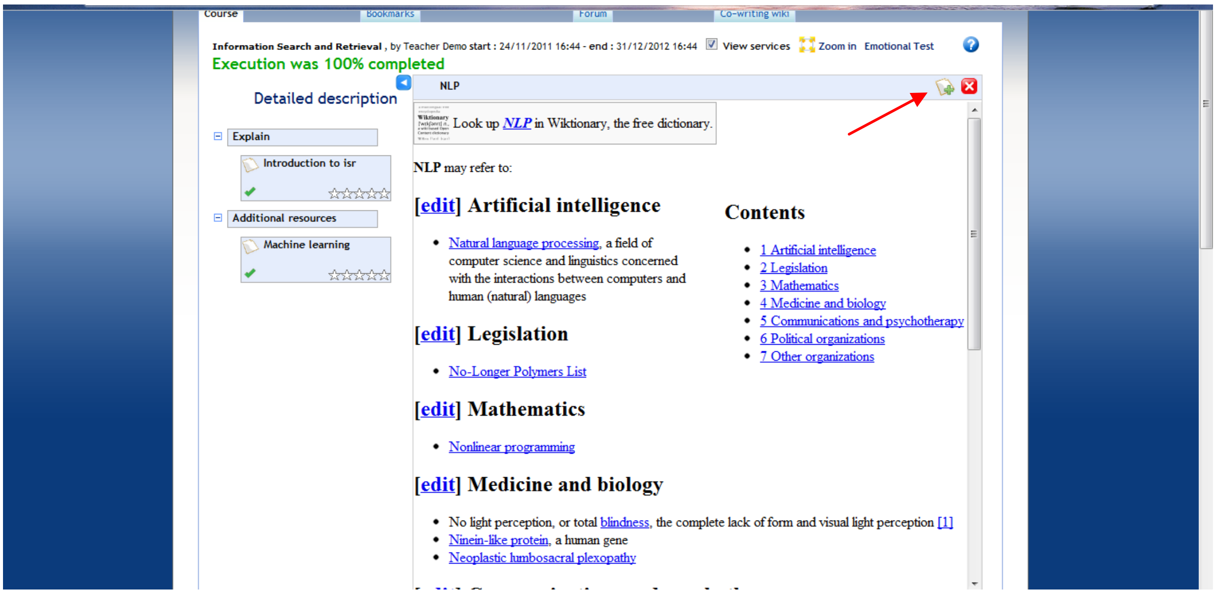


Figure 12. Results of a search for additional material in IWT

Step 5

Once a useful learning resource is found, the learner reads it. The reading panel has two buttons, the first is used to add the resource to a list of interesting resources, the second is used to generate on-the-fly a test on the resource (through a call to the AQC). Click on generate Test.

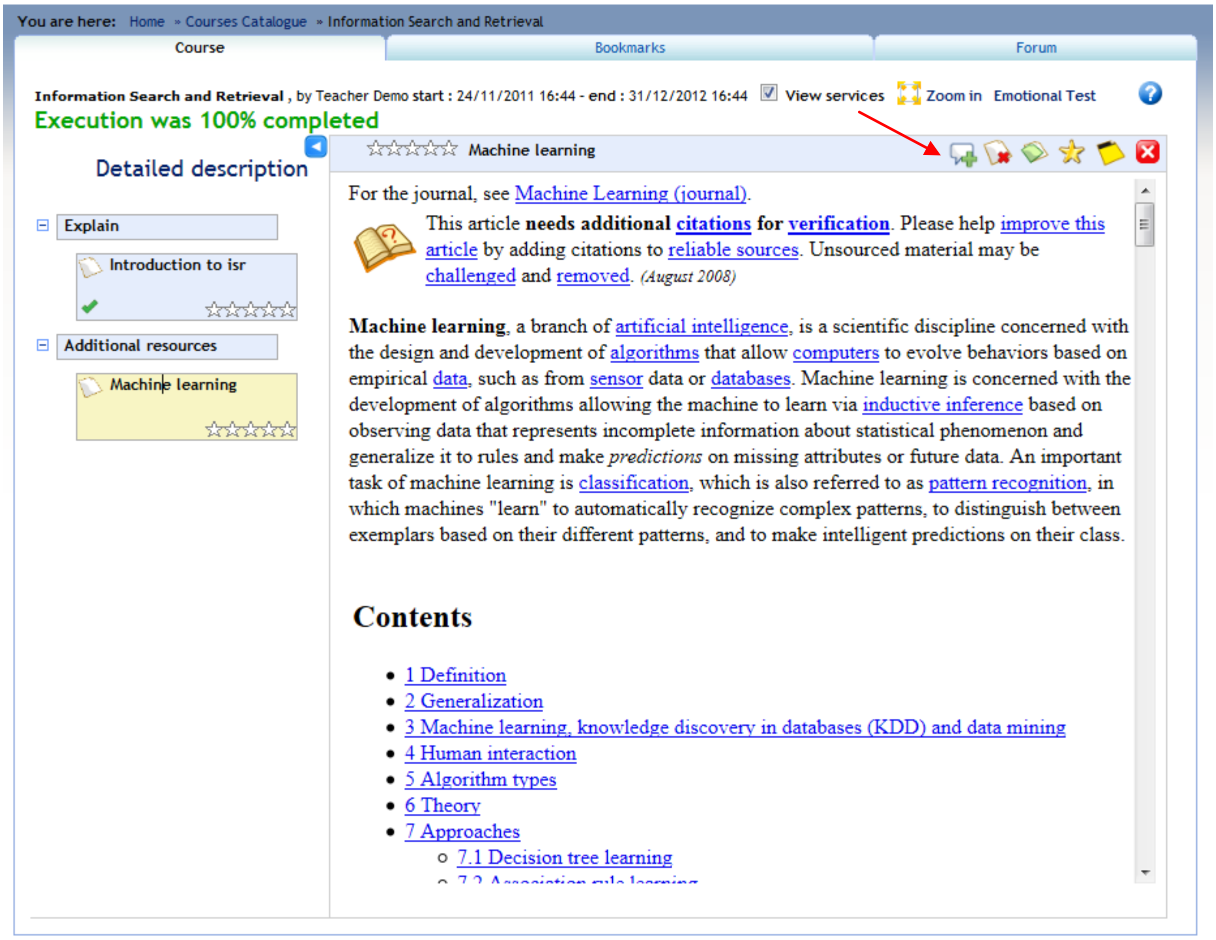


Figure 13. Button for generation tests

Step 6

The system can generate automatically a test made of fill-in blank, multiple choice and true/false questions from the text composing selected resources. Here you can set the numbers of questions you want to create for each question type. Please be patient, the item generation can take a few minutes.

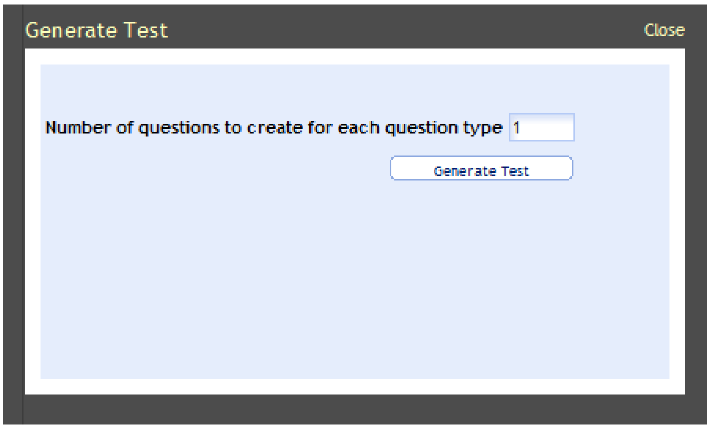


Figure 14. Example for the generation of test questions

Step 7

The learner executes the test and receives a feedback about his performances.

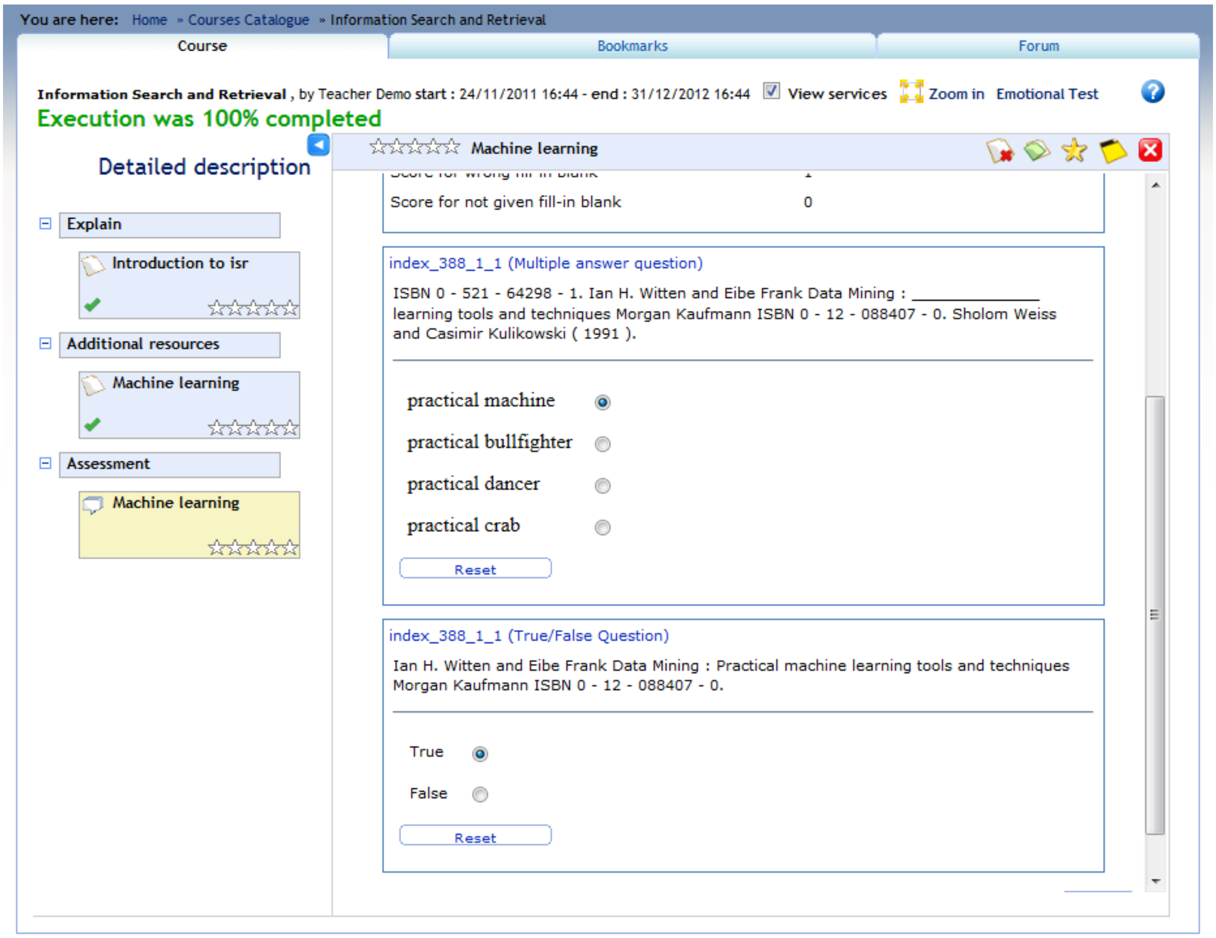


Figure 15. Examples for automatic test questions and feedback