Assessment for Complex Learning Resources

Development and Validation of an Integrated Model

http://dx.doi.org/10.3991/ijet.v8iS1.2354

G. Wesiak[1], M. AL-Smadi[1], M. Höfler[1,2] and C.Gütl,[1,3]
[1] Graz University of Technology, Graz, Austria
[2] University of Graz, Graz, Austria
[3] Curtin University of Technology, Perth, WA.

***Abstract*—Today's e-learning systems meet the challenge to provide interactive, personalized environments that support self-regulated learning as well as social collaboration and simulation. At the same time assessment procedures have to be adapted to the new learning environments by moving from isolated summative assessments to integrated assessment forms. Therefore, learning experiences enriched with complex didactic resources - such as virtualized collaborations and serious games - have emerged. In this extension of [1] an integrated model for e-assessment (IMA) is outlined, which incorporates complex learning resources and assessment forms as main components for the development of an enriched learning experience. For a validation the IMA was presented to a group of experts from the fields of cognitive science, pedagogy, and e-learning. The findings from the validation lead to several refinements of the model, which mainly concern the component forms of assessment and the integration of social aspects. Both aspects are accounted for in the revised model, the former by providing a detailed sub-model for assessment forms.**

***Index Terms*—e-Assessment, Assesesment Model, Expert Validation, Complex Learning Ressouces**

I. INTRODUCTION

With the continuous development of information and communication technology (ICT) in the context of learning, the adjustment of educational goals, settings, and assessment methods become a major challenge. Today's e-learning activities are expected to be interactive, challenging, and personalized. Learners should be in control of their learning experience, but simultaneously experience a supportive, collaborative, and simulative learning environment. Thus, self-regulating learning combined with social aspects and high levels of motivation are asked for. These changing e-learning activities also entail the need of changing assessment activities [2]. E-assessment, i.e., assessment in context of e-learning activities is a challenging field of research for Computational Science, Pedagogy, and Psychology. Within the EC-funded project "Adaptive Learning via Intuitive/Interactive Collaborative and Emotional System" (ALICE)[1] our research group at Graz University of Technology designed an integrated framework for e-assessment that is based on the requirements of different complex learning resources, such as collaborative learning, storytelling, and serious games [3]. The resulting integrated Model for e-Assessment (short IMA) describes the components involved in an enriched learning experience, including not only the learning objectives, resources, and assessment methods, but also inputs to the learning experience and interactions with other models [4]. In order to meet the needs of different learning environments and resources, the proposed model was evaluated and improved in two steps. After a first round of experimentation and a model-validation by an expert from the field of cognitive science, the model was extended by means of a sub-model dealing with the different forms of assessment. Then, the extended IMA was presented to a round of experts from the fields of cognitive science, e-learning, and pedagogy, who evaluated the model with regard to its relevance and applicability in the field of e-assessment.

In Chapter II of this paper, which is an extended version of [1], the IMA and its sub-model on assessment forms is presented in detail. Chapter III outlines an example application in a collaborative learning environment. Chapter IV gives an overview on the methodology used for the expert validation and the derived results. Finally, in Chapter V, we discuss our findings and give a short outlook for future research.

II. INTEGRATED MODEL FOR E-ASSESSMENT (IMA)

The paradigm shift for online learning and assessment has caused researchers to rethink assessment practices. Traditional assessment practices – often based on objective testing – are neither adequate for testing meta-cognitive skills such as critical thinking, creativity, and self-reflection nor for the assessment of authentic learning or for supporting life-long learning [5]. Thus, rethinking e-assessment practices towards advocating alternative assessment has emerged. Alternative assessment practices - including self- and peer-assessment, portfolio-assessment, behavioral assessment, and performance assessment [6] – point to the need of considering learning and pedagogy theories by advocating constructive, authentic, contextualized, and deep learning assessment. Consequently, educators are faced with the challenge of developing authentic, reliable, and ethical e-assessment methods that are integrated with the learning process, evaluate learning, engage students, appraise students' learning process, and promote further learning [7].

In order to provide quality assessment a set of assessment models has been designed, which are discussed in [8]. Based on this review, assessment models are either generic and discuss key elements for assessment in general [9] or they are specialized and emphasize specific aspects of the assessment process, e.g., [10]. However, the discussed assessment models lack to some extent aspects

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

such as: (a) pedagogical flexibility and an alignment with theories of learning, (b) design of the suitable assessment form for the learning activity or task, (c) available technology - in terms of systems, tools, and services, (d) standards, specifications, and guidelines of how to design, and develop assessment for the target learning practice, (e) feedback as a crucial component for quality assessment practice, (f) guidelines or frameworks of how to use these models to support developing learning tools with integrated assessment.

The remainder of this section discusses an integrated model for e-assessment which addresses the aforementioned limitations in assessment models.

A. Integrated Model for Enriched Learning experiences

The general IMA addresses the requirements of an enriched learning experience as it is defined in the ALICE project [3], namely as an experience that is based on complex learning resources (e.g. collaborative and social learning, storytelling, simulation and serious games) and integrated assessment methods (e.g. cognitive and affective assessment). This combination is expected to yield effective learning processes such as reflective and experiential learning [11] as well as socio-cognitive learning [12]. Fig. 1 depicts the abstract level of IMA with its core-methodology, inputs to the learning environment and adaptivity components interacting with the learning resources and assessment. IMA's core methodology consists of the following four main components: (1) the learning objectives, which usually refer to the goals defined by the instructor of a course but also to related didactical objectives such as gaining social competence or meta-cognitive skills due to collaborative work or self-regulated learning. Learning objectives influence the type of learning resource as well as the assessment forms appropriate in a given learning experience. For instance, if the learning objective is to apply knowledge (see [13] and [14] for a taxonomy of educational objectives), the provision of text material and a simple knowledge test will not suffice. In this case a more complex learning resource and an assessment including the application of knowledge are required (a very simple example would be the application of a previously learned formula). (2) Complex learning resources (CLR) should be provided to support learners in achieving the learning objectives by means of an active involvement in the learning process. According to constructivist theories (see e.g. [15] for a review) we build explanations of ourselves and our environment to actively create knowledge. To meet the needs of an active learner, enriched learning experiences are made up of CLR including collaboration, simulation and serious games, as well as storytelling. (3) New forms of assessment should meet the high demands arising from the CLR by considering different levels of educational objectives and effective kinds of learning (see Section B of this Chapter for more details). (4) Evaluation and validation processes should be included on a regular basis to ensure a high quality learning experience. Evaluation refers to the assessment of the used methods and procedures, whereas validation means that the measures provide a valid conclusion about the status of a learner. Results from the evaluation and validation process can again influence the first three components. Thus the development of efficient learning environments should be seen as cyclic process open to improvements.

Besides the core methodology, several components influencing the learning experience have to be considered (big red arrows on the left and right in Fig.1). These include educational aspects (e.g. different learning styles or social learning), psychological aspects (emotion or motivation), technical issues (e.g. adaptive learning or tool selection) and existing standards and specifications (e.g. best practices or ethical aspects).

To ensure a high quality standard of all activities in this complex learning environment quality criteria should be defined. Therefore, *quality assurance* which addresses all components of the enriched learning experience is also part of the model. Aspects to be considered include best practices and standards in the field in general, guidelines for delivering assessment, scoring and interpreting, e.g. [16], or ethical aspects (plagiarism, cheating, but also data protection, voluntariness, and transparency of assessment activities). A comprehensive framework for e-learning quality, which includes criteria for infrastructure, technical standards, content development, pedagogic practices, and institutional development is given by [17].

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*. For instance, the theory of constructive alignment [18] 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. Thus, 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, a group, or a peer assessment? Should the assessment activity be formative or summative? What exactly should be assessed? The knowledge of learners or whether they can apply the knowledge or even create new appliances based on the knowledge they acquired?

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. In co-operation with the learner model, the cognitive status of the learner in terms of knowledge and skills is updated, with the knowledge model the ontology of learning is recovered and with the didactic model, individual sequences of learning activities are build and eventual alternative models are recovered.

B. New forms of assessment

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, IMA combines these assessment forms in order to provide a comprehensive assessment of knowledge and skills, behavioral, motivational, and emotional aspects for complex learning resources. Fig. 2 depicts the different assessment forms as eight questions methodology 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 eight

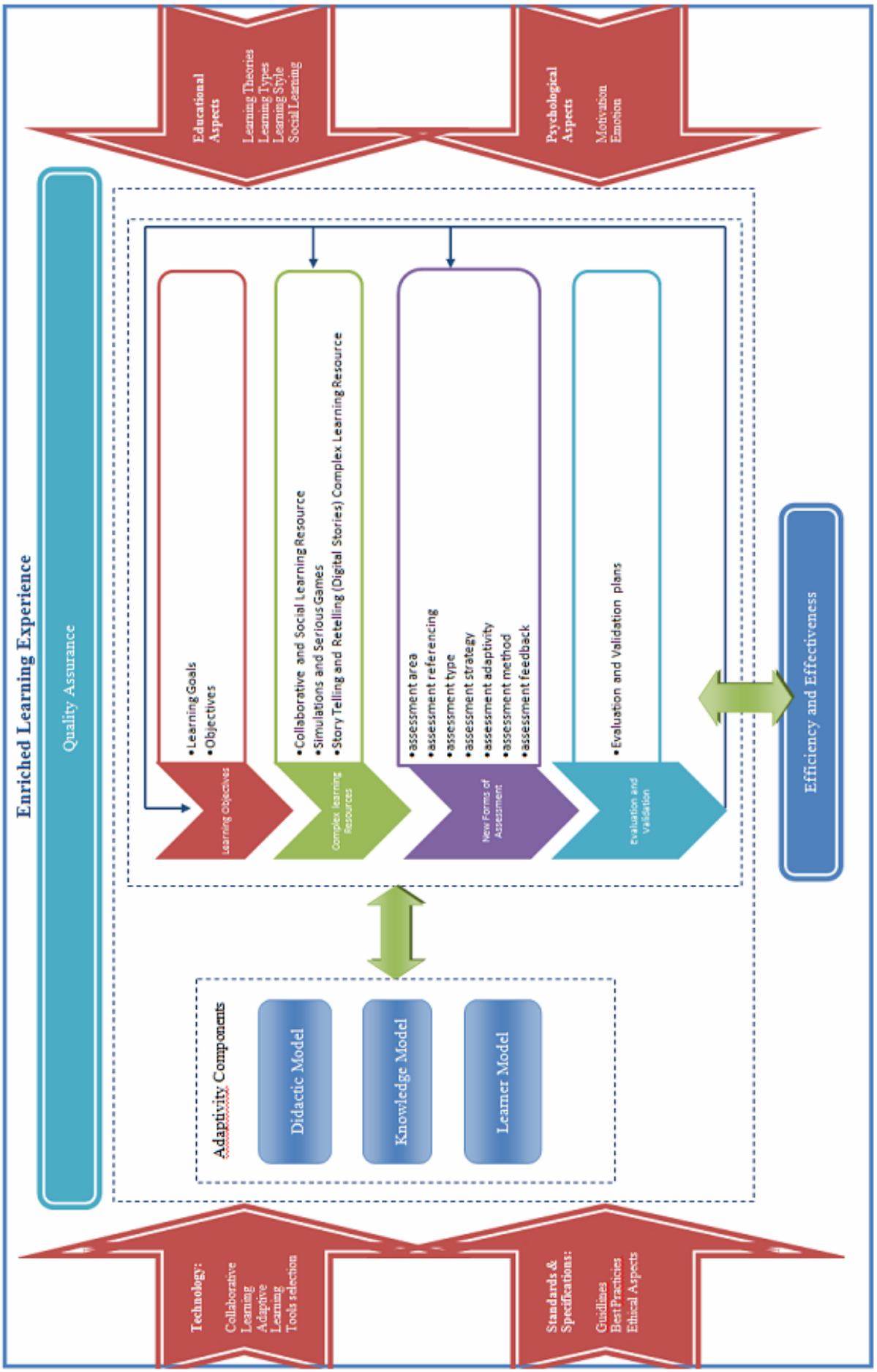

Figure 1. Integrated Model for e-Assessment

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 imply 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. In practice, 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 eight considered assessment forms are explained in more detail:

- **Assessment area**: In line with the learning theories that built the background of the enriched learning experience, for the *cognitive* domain 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 necessary to specify the level of difficulty, i.e. which competence should be assessed. For this, the six levels according to Bloom's Taxonomy [13] can be used (knowledge, comprehension, application, analysis, synthesis, and evaluation). Additionally, innovative forms of assessment should always cover the learner's *affective* state in order to enhance learning outcome [19]. Assessing learners' motivation can give important information about the underlying reasons for their (missing) learning progress, can help to choose an adequate learning object in personalized learning environment, and can help to improve the learning resources. Similarly, the emotional state of a 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).
- **Assessment referencing** concerns the reference point that is used to evaluate a learner's status of knowledge [20]. *Norm-related* referencing means that a student's performance is compared to the performance of peers. If the comparison concerns the individual's actual status with a pre-defined domain, we are dealing with *criterion-related* referencing. Finally, ipsative referencing compares a learner's actual performance with his or her own performance in the past. *Ipsative 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 strategies**: With respect to the persons involved in the assessment, we differentiate between *who is assessed* and who is the assessor. In most cases the assessment concerns an individual. However, a group assessment is also possible, if the learning product was generated by more than one person. Regarding the role of the *assessor*, it can be differentiated between instructor, peers, self, and the system. Usually, the learner is assessed by the instructor. However, assessment forms that are based on CLR also involve students by asking them to either assess their own work (self-assessment) or the work of their peers (peer-assessment). This enables students to develop meta-cognitive skills (e.g. by finding criteria that reflect the quality of their own or their peers' work) and may also facilitate the work for instructors (e.g. [19] [21] [22]). Additionally, in e.g. serious games, system based assessment can be used [23]. 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. Regarding the assessment of motivation and emotion often 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 type**: *Diagnostic assessment* (or pre-assessment) refers to students' knowledge and misconceptions as well as affective status at the beginning of the learning process. The result can, for example, be used for comparisons with a student's cognitive or affective state at the end of a learning activity (ipsative referencing). The most common form of assessment is certainly *summative assessment*, which takes place at the end of a learning activity to check whether a learner has reached the learning goal. In contrast, *formative assessment* is employed during the learning process in order to monitor and improve the learning progress. Formative assessment provides a more valuable outcome for the learning process than summative assessment, because it supports learners in reflecting their learning performance [24].
- **Adaptivity**: e-Assessment has the great advantage that it allows personalized testing, where the choice of the next learning object or test item is adapted to the needs or knowledge of the learner. Thereby, it can be differentiated between macro-adaptivity (concerning the adaptive presentation of learning content and adaptive navigation support) and micro-adaptivity (concerning non-invasive interventions effecting the presentation of learning objects) [25].
- **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 [6]. 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-assessments, computer-assisted assessment (CAA) with fixed or free response formats is 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 strongly depends on the assessment area (e.g. multiple choice items for knowledge tests vs. rating scales for motivational assessments), the assessment strategy (instructor, self, peer or system as well as individual vs. group assess-

ment), the assessment type (formative vs. summative) and last but not least the learning objective.

- **Feedback**: Feedback helps learners to become aware of gaps in their knowledge, skills, or understanding of a topic and can thus change their learning behaviour (e.g. [24] [26]). It should be provided continuously, although not intrusively in a formal or informal way in order to support the learners [27]. Generally, feedback overlaps with formative assessment.

III. MODEL APPLICATION IN A SELECTED COMPLEX LEARNING SCENARIO

To give an example of how the components of the IMA are reflected in a real learning scenario, we chose a self-directed learning course with a collaborative writing assignment. In an online-course on "Scientific Working" students had to study two articles from provided course material in order to collaboratively write an essay about these articles, and to plan a study. For the writing assignments students formed groups of two and used the co-writing Wiki developed by [28] which provides integrated self- and peer-assessments (see [29] for an evaluation of the tool). In order to receive optional automatic assessments during the reading task participants could use the automatic questions creator AQC before, during, and after reading the articles [30] [31]. Furthermore, they were required to fill out one test provided by the AQC at the end of their reading task. To investigate whether students could benefit from the learning environment, questionnaires covering task awareness, motivational and emotional aspects, and usability were sent to the students at three points during the study. A detailed description of the study can be found in [32].

A. Core Methodoloy

The main *learning objective* was to create a learning environment that supports students in self-regulated learning and working collaboratively. 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).

The *complex learning resource* is a self-directed learning course integrated with a collaborative writing assignment. The provided co-writing Wiki ensures that students work collaboratively, its visualization functions support task and social awareness as well as group well-being. Visualization tools include e.g. an actions feed listing all actions taken by the group members and contribution graphs showing how much each group member has contributed since the beginning of the assignment. Additionally it provides self-, peer-, and instructor assessments. Self- and peer assessments are given by means of a five-star rating scale regarding the importance of the last contribution and short comments. To ensure a fair and consistent assessment across all learning groups an assessment rubric with three main categories (references, content, formal aspects) and 13 subcategories designed for scientific writing was provided for the instructor. Each subcategory could be assessed by means of a five-star rating scale and a short comment field. The rubric was also used for group-assessments, in which students had to assess the work of two other groups. The AQC creates tests automatically by extracting concepts and generating questions (true/false, single choice, fill-in-the-blank, open-ended) based on a selected content, in this case the provided articles. Additionally, the generation of questions based on self-extracted concepts is possible. Testing oneself with questions should stimulate the learning process and support students in self-regulated learning.

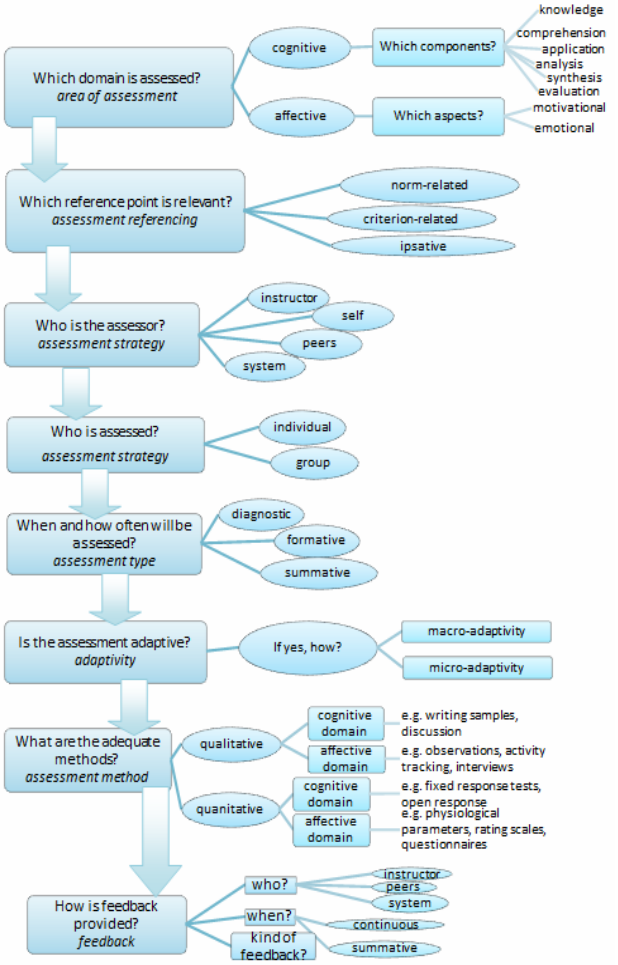

Figure 2. Methodology to design integrated forms of e-assessment

Multiple *forms of assessment* were used. The self-regulated learning environment incorporated short knowledge tests on the content of the provided learning material, which were called by the students themselves. The quality of the collaborative writing assignment was evaluated by means of ratings of importance, short open comments, and assessment rubrics (using the categories literature, content, and style with several subcategories each). The two former types concerned single contributions and were given by students themselves or their group members (self- and peer assessments). Assessment rubrics were used to evaluate the full assignment and were used by the instructor as well as by students for the group-assessments. The eight aspects of assessment outlined in Fig. 2 were covered as follows:

- Assessment area: cognitive competencies were tested on the knowledge level with automatically created questions, affective dispositions by collecting data on students' motivation and emotional status. The collaborative assignment covered the cognitive levels 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 a research design for a given research

question). The level of evaluation is required by the group-assessments.

- Assessment referencing: criterion-related from the instructor's point of view, who had clear standards regarding the quality of a paper and norm-related for the self- and peer-assessments, because students compared their product to their peers' work.
- Assessment strategy I – assessor: short self- and peer-assessments after each change of the collaborative writing assignment; detailed instructor and group (peer) assessments of the final group products; voluntary and required assessment by the system (AQC) for the reading task.
- Assessment strategy II – who is assessed: regarding the reading task (AQC-tests) individuals were assessed, for the writing assignment individual and group contributions were assessed (self/peer and instructor/group assessments respectively).
- Assessment type: formative assessment to monitor and improve students' learning process (self- and peer-assessments, voluntary AQC knowledge tests); summative assessment after the reading task (required AQC test) and the writing assignment (instructor and group-assessment); diagnostic assessment to check students' learning progress (questions regarding students' knowledge concerning scientific working before and after the course).
- Adaptivity: only 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 process. Personalized adaptation of learning content or test-items (e.g. based on students' current knowledge, motivational, or emotional status) was not embedded yet.
- Feedback: summative from instructor, i.e. at the end of the course two tutors gave a detailed individual feedback on the writing assignments; continuous from peers by means of comments integrated in the short peer assessments after each change of the contribution.
- Assessment methods: quantitative and qualitative methods concerning the cognitive as all as affective domains. The three questionnaires consisted mainly of rating scales, which are quantitative, just as the number of correct questions achieved in the AQC tests and the star-ratings given in self-, peer-, and group assessments. Qualitative assessments included open answers in the questionnaires (e.g. regarding improvements of the tool), comments in the self-, peer- and groups-assessments, and a review of the essays by the tutors. For the cognitive domain knowledge tests (AQC) and assessments of the writing assignment were performed. For the affective domain rating scales were used to assess students' motivational and emotional status during the collaborative assignment.

Regarding *evaluation and validation*, the quality of the automatically created questions was evaluated and the impact of the whole tool was validated by investigating students' extrinsic and intrinsic motivation, emotional aspects, learning styles and whether these components had an influence on the learning process.

B. *Inputs to the enriched learning experience*

As far as *educational aspects* are concerned, we investigated students' learning styles by differentiating between the elaborating and the repeating learning style [33] and their relationship to intrinsic and extrinsic motivation. *Psychological aspects* were covered by measuring motivation during the self- and peer-assessments [34], as well as emotions while using the tools [35]. *Technological aspects* in this study concern the co-writing Wiki and the AQC. For the co-writing Wiki, ScrewTurn wiki (an open source wiki using C# and ASP.Net for the front-end presentation layer) has been selected to be enhanced with features to maintain task and social-awareness and group well-being. For assessment in self-directed learning, the AQC was developed to automatically create assessment items based on textual material. *Regarding Standards and specifications*, the co-writing Wiki combines collaborative learning and assessment activities, following the guidelines by [36]. For the AQC IMS-QTI assessment content specifications have been used to represent the created items [37].

C. *Efficiency, effectiveness and quality assurance*

To evaluate the CLR (co-writing Wiki and AQC), students rated the usability of the tools by means of the system usability scale (SUS) [38] and made suggestions for improving the tools. Regarding quality assurance, we planned the study under consideration of the psychological quality criteria objectivity, reliability, and validity.

D. *Adaptivity components*

The described study aimed at investigating the developed tools which where therefore used stand-alone. However, to provide adaptivity in the sense of a learner, knowledge, and didactic model, in the meanwhile both tools were integrated in the Intelligent Web Teacher (IWT) [39], which is a learning management system allowing the definition and execution of personalized e-learning experience tailored on the basis of learners' cognitive status and learning preferences.

IV. EXPERT VALIDATION

As mentioned above, the proposed model was developed in several steps. The original model was validated by an expert from the field of cognitive science, whose main suggestion was to focus more on the assessment part of the IMA. Thus the model was extended by the assessment sub-model as it is depicted in Fig. 2. For a second round of validation, nine e-learning experts from different European universities were asked to validate the model concerning the importance of its components, the accuracy of the relations among the components, and its application and relevance in the field of e-assessment. Additionally they were asked to test and evaluate the two developed tools co-writing Wiki and AQC.

Five experts, two men and three women participated in the study. Their working fields vary from psychological research, information technology, computer science to teaching at university and on high school level. Their research interests also shows a variety of expertise, ranging from media psychology, evaluation, usability, technology enhanced learning, workplace and collaborative learning, community information systems, (mobile) social software, social network analysis, to digital libraries and new product development.

For the validation process, all experts received a detailed description of the model and sub-model, guidelines for the tools, and a questionnaire concerning the quality of the model and tools. Additionally, we provided access to the tools, so that the experts could go through the functionalities and check the tools' usability.

Experts reported that they spent on average 4 hours on reading the Chapter, testing the tools, and answering the questionnaire. The model validation included the 11 items listed in Table I. Levels of agreement were generally stated on a 5-pt. rating scale ranging from (1) "I strongly disagree" to (5) "I strongly agree". For question 9 a 7-pt. scale was used which ranged from (1) "not relevant" to (7) "very relevant". After each question, experts were prompted to comment their rating and if applicable to give suggestions for improvements. Table I summarizes the results.

Overall, the five experts gave medium ratings on the different features of the model. The first three questions covered the accuracy and completeness of the model. The main points of critique concerned the lack of social factors (social entities, social context, group dynamics) and the abstractness of the model. Furthermore, experts missed further kinds of complex learning resources (e.g. problem based and project based learning, mobile technologies or multimedia) and the assessment of relational factors. Thus, the social aspect of learning was integrated more thoroughly by adding social learning to the educational inputs of the IMA model and by differentiating between assessing individuals and groups in the assessment sub-model. With regard to the model's components (questions 4 through 7), the comments given by the experts show that the order of the components and the relationships among the components were not always clear. Again, an illustration by a concrete example was asked for. Here it is important to point out that the components in the general IMA are interrelated in a non- linear way. To highlight the reciprocal influence of the four main components, arrows were added for this part of the model. Regarding the adaptivity components and inputs to the enriched learning experience (red arrows) we slightly adapted the boxes to clarify their areas of influence. Similarly, the order of assessment forms in the sub-model is only one way to proceed, but not meant as linear order. As mentioned in the text, the different forms of assessment influence each other in a non-linear way.

Several times the lack of a concrete example, i.e. the abstractness of the model was mentioned by the experts. To meet this concern, we applied the model in different areas including collaborative and self-regulated learning (Section III in this paper) and storytelling.

The last four questions concern the general quality of the model regarding its relevance for the field of e-assessment, its adequacy for evaluating didactic experiences in adaptive learning systems, and general suggestions for improvements. The model's adequacy was confirmed for the most part. However, one expert suggested the provision of clearer guidelines on how to evaluate didactic experiences. 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. In the same way, 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 relevancy of the model in the field of e-assessment was also rated positively (M = 4.69, SD = 0.89 on a 7pt. scale), with especially emphasizing the elaborated assessment part. As far as the call for more components is concerned, we are aware of the fact that the explicitly mentioned e-assessment components (e.g. CLR, inputs, assessment areas) are only a selected sub-sample of all components existing within the field. However, the model is not meant to be exhaustive but to cover the most important components which are relevant in the field of e-assessment.

Summarized, the expert validation of IMA resulted in the following improvements of the model. First, the social aspect of learning was considered more thoroughly by explicitly integrating it to the model. Secondly, application scenarios for different kinds of complex learning resources were provided to show the special aspects to be considered in a specific scenario. This input is also meant to meet the request for more concrete examples. Furthermore, minor changes were performed, such as the separation of educational and psychological aspects, or some rewording of the model description. Fig. 1 depicts the revised version of IMA.

With regard to the validation of the tools, four experts filled out the questionnaire for the AQC and three the one for the co-writing Wiki. Because this paper focuses on the theoretical model, results are only summarized very shortly. The experts considered the co-writing Wiki for the most part as supportive for students as well as for instructors, especially the visualization tools and the assessment rubric 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. As far as the AQC is concerned, experts confirmed that the tool is a valuable instrument 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.

V. DISCUSSION AND OUTLOOK

The aim of this research was to develop an integrated model for e-assessment (IMA), which meets the challenges of the adaptive e-learning environment build within the ALICE project. The latter combines personalization, collaboration, and simulation aspects wihtin an affective/emotional based approach. The final goal is to provide an interactive, challenging and context aware environmnet that fosters learners' demand of empowerment, social identity, and authentic learning experience. The IMA discussed in this paper, describes an enriched learning experience on an abstract level. It is made up of 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.

TABLE I.
MODEL VALIDATION BY EXPERTS

Question/Statement	Mean[a] (SD)	Comments
1. 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
2. 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
3. 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
4. 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
5. All of the included components are relevant and priorities are set appropriately.	3.80 (0.45)	• •
6. 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
7. The flows are correct.	2.80 (0.84)	• no linear order • different order (text vs. model)
8. 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
9. All in all, how would you rate the integrated model regarding to its relevance in the field of e-assessment?	4.69[b] (0.89)	• emphasize benefit/advantage of this model • add more components • adaptive to underlying system • elaborate and well justified assessment part
10. What would you especially improve regarding the model?	-	• priorities of the model more visible • Skip red arrows → background/context • Integration of relational factors
11. Do you have any further comments?	-	• Focus on individual learning experience, although talking about social interaction and collaboration

[a] 5-pt. rating scales from (1) I strongly disagree to (5) I strongly agree; [b] 7-pt. rating scale from (1) not relevant to (7) very relevant;

The purpose of the IMA is to identify all components that need to be considered whenever an enriched learning experience is developed. However, its core is the aspect of e-assessment, which is no longer a simple task of testing a student's knowledge, but has to consider a wide range of assessment forms in order to give a comprehensive picture of a student's learning process, including cognitive and emotional aspects, individual and social learning, adaptivity, and so on. To give an example of how IMA can be used in practice a case study from the ALICE project has been presented to show how each component of the model was considered in a self-directed learning course (using two tools developed within this context). This first version and application of the model together with the tools developed in this context was validated by a sample of five experts. This expert validation resulted in a few changes of the model, especially regarding the integration of social aspects. The results of the application study and the expert validation show the usefulness of the model regarding the development of e-learning environments with comprehensive assessment procedures.

Future research should include an extension of the IMA to other areas of application and CLR and especially focus on the further development of the assessment sub-model. At the moment, the sub-model gives a comprehensive overview of important aspects that need to be considered when planning e-assessments for CLR. However, to increase the usability of the model, relationships and dependencies between different forms of assessment should be considered. For example, a scenario in which learners do not collaborate does not need a group assessment as strategy. For more convenience of the user (instructor or course developer), automatic suggestions of adequate assessments methods depending on the previously chosen assessment area, referencing, strategy, type, etc. are also conceivable.

ACKNOWLEDGMENTS

This research is supported by the European Commission under the Collaborative Project ALICE (Adaptive Learning via Intuitive/Interactive, Collaborative and Emotional Systems, VII Framework Program, Theme ICT-2009.4.2 (Technology-Enhanced Learning), Grant Agreement n. 257639. We are grateful to the experts for their great support and time effort in validating the model.

REFERENCES

[1] Wesiak, G., Al-Smadi, M., & Gütl, C. (2011). Towards and Integrated Assessment model for Complex Learning Ressources: Findings from an Expert Validation. *15th International Conference on Interactive Collaborative Learning and 41st International Conference on Engineering Pedagogy*, Villach, Austria 26-28 September 2012. 7p.

[2] Bennett, R. E. (2002). Inexorable and inevitable: The continuing story of technology and assessment. *Journal of Technology, Learning, and Assessment, 1*(1).

[3] 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).

[4] AL-Smadi, M., Höfler, M., & Gütl, C. (2011). *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.

[5] Haken, M. (2006). Closing the loop-learning from assessment. Presentation made at the University of Maryland Eastern Shore Assessment Workshop. Princess Anne: MD.

[6] Buzzetto-More, N. A. & Alade, A. J. (2006). Best Practices in e-Assessment. Journal of Information Technology Education, 5, 251-269.

[7] Bartley, J. (2006). Assessment is as assessment does: A conceptual framework for understanding online assessment and measurement. In M. Hricko & S. Howell (Eds.), *Online Assessment and Measurement: Foundations and Challenges.* Hershey, PA: ISP. http://dx.doi.org/10.4018/978-1-59140-720-1.ch001

[8] AL-Smadi, M., Gütl, C. & Chang, V. (2011). Addressing e-Assessment practices in e-Learning Activities: A Review. In S. Barton et al. (Eds.), Proceedings of Global Learn Asia Pacific 2011 (pp. 448-459). AACE. Retrieved from http://www.editlib.org/p/37209.

[9] Pellegrino, J., Chudowsky, N. & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

[10] Almond, R., Steinberg, L. & Mislevy, R. (2002). Enhancing the design and delivery of assessment systems: A four process architecture. *The Journal of Technology, Learning, and Assessment, 1(5) 1-64.*

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

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

[13] 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.

[14] 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.

[15] 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.

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

[17] 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.

[18] Biggs, J. B. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 1-18. http://dx.doi.org/10.1007/BF00138871

[19] Dochy, F. J. & McDowell, L. (1997). Introduction. Assessment as a tool for learning. *Studies in Educational Evaluation,* 23, 279-298. http://dx.doi.org/10.1016/S0191-491X(97)86211-6

[20] McAlpine, M. (2002). *Principles of assessment.* Bluepaper No. 1. CAA Centre, University of Glasgow. Last retrieved Nov. 11th 2012 from http://www.caacentre.ac.uk/dldocs/Bluepaper1.pdf.

[21] 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. http://dx.doi.org/10.1080/02602930500099235

[22] Cano, M. D. (2011). *Students' involvement in continuous assessment methodologies: A case study for a distributed information systems course.* Last retrieved Nov. 11th 2012 from http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5590268&url=http%3A%2F%2Fieeexplore.ieee.org%2Fstamp%2Fstamp.jsp%3Ftp%3D%26arnumber%3D5590268.

[23] Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfield, M. J. Cody, P. Vorderer (eds.). *Serious Games: Mechanisms and Effects.* New York, London:Routledge. 295-321.

[24] Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research & Evaluation,* 8(9). Last retrieved Nov. 11th 2012 from http://pareonline.net/getvn.asp?v=8&n=9.

[25] 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. http://dx.doi.org/10.1007/978-3-540-78139-4_8

[26] Garris, R., Ahlers, R. & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, 33, 441-467. http://dx.doi.org/10.1177/1046878102238607

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

[28] AL-Smadi, M., Höfler, M., & Gütl, C. (2011). Enhancing Wikis with Visualization Tools to Support Groups Production Function and to Maintain Task and Social Awareness. *Proceedings of 4th International Conference on Interactive Computer-aided Blended Learning*, Nov. 2011, Antigua, Guatemala.

[29] Wesiak, G., Al-Smadi, M., & Gütl, C. (2012). Alternative Forms of Assessment in Collaborative Writing - Investigating the relationships between motivation, usability, and behavioural data. *Proceedings of the 2012 International Computer Assisted As-*

sessment Conference (CAA), Southampton, UK. (13p). Last retrieved Nov. 13th 2012 from http://caaconference.co.uk/pastConferences/2012/caa2012_submission_12b.pdf.

[30] 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),* 23-38.

[31] Höfler, M., AL-Smadi, M., & Gütl, C. (2012). Investigating the suitability of automatically generated test items for real tests [12 pages]. *The International Journal of eAssessment, 2(1)* [Online]. Available: Doc. No. 26.

[32] AL-Smadi, M., Wesiak, G., Guetl, C., & Holzinger, A. (2012). Assessment for/as Learning: Integrated Automatic Assessment in Complex Learning Resources. *Proceedings of the Sixth International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS 2012)*. Palermo, Italy. (6p). http://dx.doi.org/10.1109/CISIS.2012.210

[33] Wild, K.-P. (2000). *Learning strategies in academic studies. Structures and conditions.* [Lernstrategien im Studium. Strukturen und Bedingungen]. Münster: Waxmann.

[34] Tseng, S.-C., & Tsai, C.-C. (2010). Taiwan college students' self-efficacy and motivation of learning in online peer-assessment environments. *Internet and Higher Education, 13,* 164-169. http://dx.doi.org/10.1016/j.iheduc.2010.01.001

[35] Kay, R.H., & Loverock, S. (2008). Assessing emotions related to learning new software: The computer emotion scale. *Computers in Human Behavior. 24,* 1605-1623. http://dx.doi.org/10.1016/j.chb.2007.06.002

[36] Macdonald, J. (2003). Assessing Online Collaborative Learning: Process and Product. *Computers & Education, 40(4),* 377-391. http://dx.doi.org/10.1016/S0360-1315(02)00168-9

[37] IMS QTI. *IMS Question & Test Interoperability Specification, Version 2.0 - Final Specification*. Last retrieved March 3rd, 2012 from http://www.imsglobal.org/question/index.html.

[38] Brooke, J. (1996). *SUS: A "quick and dirty" usability scale.* In Usability evaluation in industry. London: Taylor & Francis.

[39] N. Capuano, S. Miranda, and F. Orciuoli (2009). *IWT: A Semantic Web-based Educational System,* IV Workshop of the Working Group on "AI & E-Learning", pp. 11–16.

AUTHORS

G. Wesiak, was with the Institute for Information Systems and Computer Media, Graz University of Technology, Austria (e-mail: gudrun.wesiak@uni-graz.at).

M. AL-Smadi, is with the Institute for Information Systems and Computer Media, Graz University of Technology, Austria (e-mail: msmadi@iicm.edu).

M. Höfler, is with the Department of Psychology, University of Graz, Austria (e-mail: ma.hoefler@uni-graz.at).

C. Gütl, is with the Institute for Information Systems and Computer Media, Graz University of Technology, Austria and with Curtin University of Technology, Perth, WA. (e-mail: cguetl@iicm.edu).

This article is an extended and modified version of a paper presented at the International Conference on Interactive Collaborative Learning (ICL2012), held 26 - 28 September 2012, in Villach, Austria. Submitted, 15 November 2012. Published as resubmitted by the authors 3 December 2012.