Assessment for/as Learning
Integrated Automatic Assessment in Complex Learning Resources for Self-directed Learning

Mohammad AL-Smadi and
Gudrun Wesiak
Graz University of Technology,
Graz, Austria
msmadi@iicm.edu; gudrun.wesiak@uni-graz.at

Christian Guetl
Graz University of Technology,
Curtin University of Technology,
Perth, WA.
Graz, Austria
cguetl@iicm.edu

Andreas Holzinger
Graz University of Technology,
Medical University Graz
Graz, Austria
andreas.holzinger@medunigraz.at

Abstract—In the so-called ‘New Culture for Assessment’ assessment has become a tool for Learning. Assessment is no more considered to be isolated from the learning process and provided as embedded assessment forms. Nevertheless, students have more responsibility in the learning process in general and in assessment activities in particular. They become more engaged in: developing assessment criteria, participating in self, peer-assessments, reflecting on their own learning, monitoring their performance, and utilizing feedback to adapt their knowledge, skills, and behavior. Consequently, assessment tools have emerged from being stand-alone represented by monolithic systems through modular assessment tools to more flexible and interoperable generation by adopting the service-oriented architecture and modern learning specifications and standards. The new generation holds great promise when it comes to having interoperable learning services and tools within more personalized and adaptive e-learning platforms. In this paper, integrated automated assessment forms provided through flexible and SOA-based tools are discussed. Moreover, it presents a show case of how these forms have been integrated with a Complex Learning Resource (CLR) and used for self-directed learning. The results of the study show, that the developed tool for self-directed learning supports students in their learning process.

Keywords-component; Automatic Assessment; Complex Learning Resource; e-Assessment; Self-directed Learning; Service-Oriented Architecture.

I. INTRODUCTION

Learners grow up with technology dominating most of their life activities. They use technology anywhere, anytime, and they are faced with the challenge of needing to be engaged and motivated in their learning [1]. The emergence of Web 2.0 and the influence of Information and Communication Technology (ICT) have fostered e-learning to be more interactive, challenging, and situated. As a result, learners feel empowered when they are engaged in collaborative learning activities and self-directed learning. The learners are also provided with e-learning systems that maintain their social identity and situated learning experience. Given the different learning styles of students, educators are faced with the challenge of having to develop assessment tools which are required to appraise the students’ learning process. Assessment forms provided in current e-learning activities have to be adapted so that they can foster effective types of learning such as reflective-learning, experiential-learning, and socio-cognitive learning [2].

To this end, this research investigates the following goals: (G1) the applicability and usability of flexible and interoperable educational tools in one complex learning resource, moreover (G2) students’ perception towards the use of a complex learning resource integrated with automatic e-Assessment during self-directed learning activities, finally, (G3) the relation between students’ motivation and their preferred learning style when it comes to use complex learning objects (CLO) enriched with automatic assessment for self-directed learning.

The remaining part of this paper is organized as follows: Section II explains the notion of CLR and discusses the requirements and architecture for the developed tools, Section III explains the study design and analysis, and Section IV discusses the results and reflects on the research goals and hypotheses.

II. COMPLEX LEARNING RESOURCE FOR SELF-DIRECTED LEARNING

According to [3] the Atomic learning object is defined as “the smallest unit of reuse for LOs that may or may not be associated to one or more multimedia contents”, whereas a Complex LO (CLO) is defined as “an LO whose instructional material is an aggregation of Learning Objects. Being an LO, a Complex LO can be treated exactly as any other LO”. Accordingly we define a CLR as a composite didactic resource consists of one or multiple learning objects (either atomic or complex). Accordingly, CLR inherits the features of LO of reusability and interoperability provided by the standards and specifications used to represent LOs.

The CLR developed for this study is a composite didactic resource for self-directed learning that consists of learning materials represented in SCORM [4], enriched with IMS QTI [5] compliant test items automatically created from learning material. Moreover IEEE Learning Object Metadata (LOM) [6] is used to annotate the CLR. The CLR is
provided by the Intelligent Web Teacher (IWT) [7] - a learning management system allowing the definition and execution of personalized e-learning experience tailored on the basis of learners’ cognitive status and learning preferences - based on fully integrated tools and services.

A. Functional Requirements and Architectural Design

The study aims at developing a CLR for self-directed learning enriched with automatically created formative assessments based on textual learning material. The tools have been designed to consider the following functional requirements: (1) formative assessments are created automatically using an automatic question creator tool based on extracted concepts from textual learning material, (2) Standard compliance by using IMS QTI to annotate created questions, (3) a variety of test items – i.e. multiple choice, true/false, fill-in-the-blank, and open ended questions, (4) the creation and fruition of a self-directed course where students are provided with search features to find and select learning material, and (5) a semi to fully automatic interactive approach by which students are allowed to tag learning content and extract concepts, and get automatically created tests for those concepts.

Assessment tools have emerged from being stand-alone represented by monolithic systems through modular assessment tools to a more flexible and interoperable generation by adopting the service-oriented architecture and modern learning specifications and standards. The new generation holds great promise when it comes to having interoperable learning services and tools within more personalized and adaptive e-learning platforms [8]. This generation highly depends on service-oriented architectures (SOA) where its services support federated exchange (information and control), various levels of interoperability (intra-domain and inter-domain), and service composition (orchestration and choreography). The CLR has been developed with respect to the architecture proposed in [9] for self-directed learning with automatically created tests – using automatic question creator tool (AQC) [10] - based on the e-assessment framework discussed in [8] [11]. For the sake of flexibility a web service has been developed in order to interpret, validate, and create QTI-based assessment items and tests. The web service is developed as part of the suggested middleware for tools interoperability and flexible e-assessment system [8]. The web service is then used based on a service-oriented framework for assessment (SOFA) [11] to support modules of items and tests authoring/viewing, and extend the learning platform with the features provided by AQC as discussed in this study.

The developed CLR for self-directed learning has the following competitive advantages:

- **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 provision** enables an easy export and reuse of test items created by the tool [9].

III. STUDY DESCRIPTION AND FINDINGS

The study has been conducted as part of a scientific research course, in which a CLR enriched with automatic assessment has been used to provide a self-directed learning course. The course has been delivered in distance learning settings and participants got to know their partners within the study activities.

A. Method

The entire study comprised three phases, with one questionnaire provided for each phase. Since this article is on self-directing learning, we will only report the method and results from phases 1 and 2.

1) **Participants**

In this study 12 students had participated, for 5 of them the course was mandatory, 7 participated as life-long learners. Eight participants are male and four female with their age ranging between 22 and 41 years old (M=32, SD=6.53). With respect to education level, three students hold a Bachelor degree, eight hold a Master degree, and one has a PhD.

Only six students finished the entire study as the course was mandatory for five of them. One student participated in all the three phases but s/he did not finish the requirements of phase 3. Two students finished phases 1 and 2 and three students only participated in phase 1.

2) **Apparatus and Stimuli**

The course material and tests have been provided online using IWT as a learning management system allowing the definition and execution of personalized e-learning experience tailored on the basis on learners’ cognitive status and learning preferences based on fully integrated tools and services. Nevertheless, the LimeSurvey\(^1\) deployed on our campus server has been used to deliver three questionnaires -

---

\(^1\) [http://www.limesurvey.org/]
one for each phase of the study - to investigate aspects such as, motivation and attitudes, emotions, preferable learning style, and usability.

a) Pre-questionnaire

This questionnaire was provided at the beginning of the study and investigated information on demographic data and previous experience, and motivational aspects towards using CLR enriched with automatic assessment for self-directed learning.

In order to investigate participants’ motivation towards the course in general and the study phases in particular, a section adapted from [12] has been added based on the following three motivation scales: Intrinsic Goal Orientation Scale measures the students’ intrinsic motivation regarding the course, for instance: “I prefer course material that arouses my curiosity, even if it is difficult to learn”. A high value on this scale would mean that the students are doing the course for reasons such as challenges and curiosity. The Extrinsic Goal Orientation Scale deals with the extrinsic motivation of students, e.g. “Getting a good grade is the most satisfying thing for me right now”. A student is extrinsically motivated when s/he is rather interested in rewards or good a grade than in the task itself. Finally, the Task Value Scale is about the learning task itself, i.e. how important, interesting, and useful the task and the task material are for the students. More interest in the task should lead to more involvement in one’s learning. To give an example, one item out of this scale is: “I think I will be able to use what I learn in this course in other courses”. Answers were given on a 5-point Likert scale, so that students could state their level of agreement or disagreement. The rating scale ranged from “I strongly disagree” (1), “I disagree” (2), “neither/nor” (3) to “I agree” (4), “I strongly agree” (5).

b) Post-questionnaire

This questionnaire was provided at the end of the study on self-directed learning with automatic formative assessment (see procedure section for more details) to investigate aspects such as, quality of written content, quality and frequency of received tests, preferred learning style, emotional aspects, and tools’ usability. Regarding the quality of written content and test questions a scale from “very bad” (1), “bad” (2), “ok” (3), “good” (4) up to “very good” (5) has been used. Where students were asked how often they had taken a test the used scale ranged from “never” (1), “seldom” (2), “sometimes” (3) to “often” (4).

Regarding the “usability of the learning scenario” we used the System Usability Scale (SUS) [13] which contains 10 items and a 5-point Likert scale to state the level of agreement or disagreement (e.g. “I think that I would like to use this system frequently”).

The learning styles of ‘elaborating’ or ‘repeating’ has been investigated in order to find out if the students’ learning process is rather superficial or aims at a deeper understanding. For this section, items developed by [14] have been translated into English (e.g. item regarding the elaborating learning style: “In my mind I try to connect what I have learned with already known issues concerning the same topic”, item regarding the repeating learning style: “I try to learn the content of scripts or other notes by heart”). The answers were also given on a 5-point Likert scale.

To assess the participant’s emotional state during the second phase, the emotional scale developed by [16] has been used. This scale includes 12 items describing four emotions related to learning new computer software as follows: Happiness (“When I used the tool, I felt satisfied/curious.”), Sadness (“When I used the tool, I felt disheartened/dispirited”), Anger (“When I used the tool, I felt irritable/frustrated/angry”), and Anxiety (“When I used the tool, I felt anxious/insecure/helpless/nervous”), and Anger (“When I used the tool, I felt irritable/frustrated/angry”). For this section answers followed a scale from “None of the time”, “Some of the time”, “Most of the time” or “All of the time”.

Moreover, an open comment section has been added to this questionnaire to get additional comments and suggestions from the students.

Finally, a section named motivational aspects has been provided to investigate the participants’ motivation during the study. For instance students were asked “How motivated were you according to the following tasks?”. Reading the contents, working with the self-directed tool, testing myself with questions, and filling in the evaluation questionnaires. The following scale has been used to get the participants answers: “absolutely unmotivated” (1), “unmotivated” (2), “motivated” (3), and “very motivated” (4).

3) Procedure

After learning a content, provided by the developed system, on scientific working and taking a test on this content (Phase 1), students have been grouped by the instructor into 6 groups – two members each - based on their interest in the course (i.e. mandatory of 3 groups and volunteer of 3 groups). In this second Phase of the study, students should deepen their knowledge in two main categories: experimentation design and experimentation analysis. For each of them, 6 articles have been delivered. Each group member has been requested to select one article from both categories different than the ones selected by his peer within the same group. In order to avoid members from the same group selecting similar articles they have been asked to use the discussion forum to agree on their selections. Moreover, participants introduced each other using the forum, and selected their articles based on their interest.

Furthermore the self-directed learning course supported students with the ability to test themselves before, during (after sections), and after reading the article. A “TestMe” button has been added to the course player by which the provided learning content is used to automatically create tests based on the students’ preferences. Those created tests could be taken several times in a formative way to get formative feedback about their current knowledge state with respect to the learning material.

At the end of the study students have been asked to answer the post-questionnaire.
B. Evaluation Methods and First Findings

This section reports the results derived from students’ answers on the two questionnaires and tests the study hypotheses as follows:

1) H1: the use of the tools is easy even if the user is a non-expert

In order to test this hypothesis, the following evaluation criteria and metrics have been used:
- **C1.1:** To evaluate the user’s level of satisfaction towards the tools,
- **C1.2:** To identify possible improvements for the tool based on comments and suggestions,
- **M1.1:** Ratings for functionality/usability of the tool itself, and frequency of use (post-questionnaire)
- **M1.2:** Ratings for emotional aspects while using the tools (post-questionnaire)
- **M1.3:** Suggestions and comments based on open questions. (post-questionnaire)

A. With respect to **M1.1:**

Results have shown that 7 out of 8 students who completed the first two phases of the study have taken formative tests during the self-directed learning in phase 2, and one student said that s/he has never took a test because s/he did not have time. Counting the tests which the students took optionally during phase 2, 30 tests were taken in total. Regarding the three different types of tests the students stated on a 4-point rating scale that they seldom took a test before, during, or after reading the topic (pre-test: M = 2.13, SD = 0.64; sub-sections test: M = 2.25, SD = 0.71.; and post-test: M = 2.25, SD = 0.87). However, looking at the actual data, the students called the AQC 6 times for a pre-test and a post-test (maximal twice per person), and 18 times for the sub-sections tests (between 0 and 8 times per person).

With respect to the tool usability, the average SUS score based on eight students' responses is “66.88”, where the SUS scale gives a score within a range of “0” and “100”. According to [16] “The average SUS score from all 500 studies is a 68. A SUS score above a 68 would be considered above average and anything below 68 is below average”. The reference provides a calculator to convert the SUS score into a percentile rank through a process called normalizing. The calculator “takes raw SUS scores and generates percentile ranks and letter-grades (from A+ to F) for eight different application types”. The score 66.88 the CLR achieved indicates that the tool has higher perceived usability than (40% - 50%) of all products that have been tested, and it can be interpreted as a grade C. According to [17] this score can be considered as “OK” having the complexity of the learning scenario and the use of multiple tools in a flexible and interoperable way within the same learning scenario.

B. With respect to **M1.2:**

Concerning students’ emotions during working with the self-directed learning tool, a comparison of the mean values indicate that the students felt equally happy (M = 1.88, SD = 0.80), sad (M = 1.5, SD = 0.60), anxious (M = 1.41, SD = 0.65), and angry (M = 1.54, SD = 0.31). Multiple dependent t-tests also show that with a significance level of p > .05 there was no difference between the emotions happiness (t(sadness) = 0.98; t(anxiety) = 1.44 t(anger) = 1) sadness (t(anxiety) = 0.31; t(anger) = 0.17), anxiety (t(anger) = 0.57) and anger. By interpreting the mean values, it can be assumed that the students seldom felt consciously happy, sad, anxious, or angry.

Linking the emotional state with the tool frequency of use form last section, we can assume that despite the unclear emotional state during the self-directed learning activity, they frequently requested an automatic test with a rate of (twice per student) on pre, and post-tests, as well as (between 0 - 8 times per student) on sub-sections tests.

C. With respect to **M1.3:**

Regarding what the students liked about the tool, students stated that they were in favor of the simplicity of the tool and the division of the content into meaningful modules. Furthermore the students liked the consistency and the possibility to have an overview of the learning progress and their own test results. They mentioned that the course was well organized and they appreciated that the course was online, so that they could work from anywhere. Also, the content itself was described very well and was precise and useful. On the other hand, students did not like that they were logged out after a short period of time (Session time-out was short). Some also complained about the slow interface. Regarding the Test Module within the self-regulated tool, some students criticized the difficulty to navigate to different questions. Regarding comments and suggestions for improvement, they would like to download content and print it directly as a handout. Besides, some students suggested a layout for a better overview. The students would improve the text structure and recommended clearer instructions for the assessment parts.

Moreover, the students were asked about “what they like about the three types of tests”. Results have shown that the different types of questions helped them getting an overview about the topics. Furthermore, they were in favor of the division of the learning material into small modules. Some students also stated that the sub-section and post-tests supported them in observing their learning progress. In the opposite, they were asked “what they did not like”. First of all, the tests were criticized, as in particular they focus on factual knowledge. Additionally, the multiple choice questions were criticized due to the possibility of having low quality distractors.

1) H2: Using the tools has a positive impact on the users’ motivation concerning their learning activities
In order to test this hypothesis, the following evaluation criteria and metrics have been used:

- **C2.1**: To evaluate students’ motivation concerning their learning activities.
- **C2.2**: To identify preferable learning styles of the students, and the impact of students’ motivation on their preferable learning style.
- **M2.1**: Ratings of students’ extrinsic and intrinsic motivation regarding the course and its tasks before using the tool (pre-questionnaire).
- **M2.2**: Ratings regarding students’ learning styles. (post-questionnaire)

### A. With respect to M2.1:

The results of student’s motivation regarding the course and its tasks show that their intrinsic motivation ($M = 3.94$, $SD = 0.53$) was significantly higher than their extrinsic motivation ($M = 2.83$, $SD = 0.79$; $t(11) = 3.43$, $p<.01$). This means that they are interested in the course for reasons such as curiosity and challenge, whereas high grades or rewards were not so important for them. These findings are supported by the results of the task value scale. A mean value of 3.83 ($SD = 0.74$) shows that the students were really interested in the task itself. The task material was also very useful and important for them. Due to their high interest, it can be assumed that this also leads to more involvement in their learning activities.

In general, questions regarding students’ motivation concerning their learning activities during the three phases revealed that they were motivated up to very motivated over the course of the study. Table 1 shows the mean ratings as well as the respective medians in order to take account of extreme values.

<table>
<thead>
<tr>
<th>Motivation while ...</th>
<th>$M$ ($SD$)</th>
<th>$Md$</th>
</tr>
</thead>
<tbody>
<tr>
<td>reading the content</td>
<td>3.5 (0.55)</td>
<td>3.5</td>
</tr>
<tr>
<td>working with the tool</td>
<td>2.67 (0.52)</td>
<td>3.0</td>
</tr>
<tr>
<td>testing themselves with questions</td>
<td>2.5 (0.84)</td>
<td>3.0</td>
</tr>
<tr>
<td>filling in the questionnaire</td>
<td>3.0 (0.0)</td>
<td>3.0</td>
</tr>
</tbody>
</table>

*Note: ratings were given on a 4-pt. scale*

### B. With respect to M2.2:

A comparison of the mean values derived from eight and seven questions for the elaborating and repeating learning style, respectively, shows that there is a significant difference between the elaborating ($M = 4.05$, $SD = 0.56$) and the repeating learning style ($M = 3.04$, $SD = 0.82$; $t(7) = 2.71$, $p<.05$). The students prefer the elaborating learning style, which means that their learning process aims at deeper understanding and is less superficial. Concerning elaborating, for instance the students stated that they try to link new terms or new theories to familiar terms and theories ($M = 4.38$, $SD = 0.52$). In contrast to that, students said that they do not learn the content of scripts or other notes by heart ($M = 2$, $SD = 1.07$) which would indicate a repeating learning style.

From M2.1 and M2.2 results, we can deduce a relation between elaborating learning style and deep learning based on intrinsic motivation to participate in the learning activity. The results from M2.1 show that students were intrinsically motivated at the beginning of the study. Due to their learning style preference, it can be assumed that the students were still intrinsically motivated during the self-directed learning activity. Thus, the students answered the questions out of pleasure with the aim to deepen their knowledge.

In addition, the students stated that if they learn something, testing themselves with questions often helps them ($M = 3.63$, $SD = 1.50$). This result is in line with the results discussed above. So it can be assumed that providing self-directed learning courses with the ability to create automatic tests supported the students to achieve their learning goals. However, the findings are in line with literature. The research of [18] shows that there is an evidence of the influence of intrinsic motivation on learners engagement that leads to ‘deep’ learning through higher level thinking skills and the conceptual understanding. Moreover, the author of [19] highlights the problems associated with extrinsic motivation as it leads to ‘shallow’ instead of ‘deep’ learning.

### IV. Conclusion and Outlook

With respect to the study goals, this section concludes the findings accordingly and provides some look ahead for future work. To this end, summarizing (G1), it can be assumed that the tools developed to integrate assessment forms with complex didactic resources are user friendly. First, we can be satisfied with the functionality because of the satisfactory SUS score the tools have reached. Moreover, the students were in favor of the various functions of the tools and their simplicity. Second, they stated that the tools gave them a good overview of their learning progress. For further improvement, a closer look on the “fill in the blank” question type, which was not easy to solve and work on a faster interface should be considered. Moreover, the study shows the applicability of combining interoperable and flexible learning tools in one complex learning scenario.

Regarding students’ motivation (G2), the results show that the students were intrinsically motivated at the beginning of the course. So they were really interested in the course and its tasks, which lead also to more involvement in their learning activities. At the end of the course, the students were asked about their motivation concerning different learning activities. According to the results, students’ motivation was high during reading content, and filling in the questionnaires. In addition, testing themselves with automatically created tests – i.e. pre-, intermediate based on subsection, and post-tests - and working with the self-directed learning tool also motivated them.

By investigating students’ learning styles, we found out that the students’ learning process aims at deeper understanding and is less superficial. This result is in line with the results discussed above, as intrinsic motivation has a great influence on learner engagement and their learning style. Thus, it can be assumed that students took tests out of
pleasure with the aim to deepen their knowledge. Besides, students also stated that testing themselves often supported them in their learning process (G3).

ACKNOWLEDGMENT

This research is partially supported by the European Commission under the Collaborative Project ALICE "Adaptive Learning via Intuitive/Interactive, Collaborative and Emotional System", VII Framework Program, Theme ICT-2009.4.2 (Technology-Enhanced Learning), Grant Agreement n. 257639. We are grateful to Dominik Kowald for his support in developing the CLR as well as Margit Höfler and Isabella Pichlmair for the support during the planning and analysis of this study.

REFERENCES