Complex Learning Resources Integrated with Emerging Forms of e-Assessment: an Empirical Study

Mohammad AL-Smadi, Margit Hoefler, and Gudrun Wesiak
Institute for Information Systems and Computer Media
Graz University of Technology
Graz, Austria
msmadi@iicm.edu; gudrun.wesiak@uni-graz.at

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

Abstract- The emergence of Web 2.0 and the influence of Information and Communication Technology (ICT) have fostered e-education to be more interactive, challenging, and situated. As a result, learners felt empowered when they are engaged in collaborative learning activities and self-directed learning. New forms of integrated assessment to support students when it comes to use complex-learning resources (CLR) within learning activities have become highly required. This paper discusses an empirical study about emerging forms of assessment such as, automated assessment, peer-assessment, and group-assessment integrated with CLRs in self-directed and collaborative learning. The first findings show that students were intrinsically motivated towards this approach. Moreover, automatic and peer-assessment supported the students to achieve their learning goals.

Keywords-component; automated assessment; e–assessment; Self-directed learning; peer-assessment.

I. COMPLEX LEARNING RESOURCES WITH INTEGRATED E-ASSESSMENT

Recently, learning settings of learner-centred practices have become more dominant. A new culture of assessment of integrating assessment to CLRs to address requirements of assessing aspects such as cognitive, meta-cognitive, social, and affective aspects have arisen [1]. This research aims to investigate: (G1) the students perception towards the use of CLRs integrated with emerging forms of e-Assessment during self-directed learning activities, and the applicability of using flexible and interoperable education tools in one complex learning resource, moreover (G2) students motivation and attitudes concerning assessment forms such as, automated assessment, self, peer-assessment, and assessment rubrics. Finally, (G3) the students preferred learning style when it comes to use CLO.

The rest of this paper is organized as follows: section 2 explains the study design and analysis, and section 3 concludes the results and reflects on the research goals and hypotheses.

II. STUDY DESIGN AND ANALYSIS

A. Method

The study has been conducted as part of scientific research course, in which three phases where taken by students and ran along the entire course. The course has been delivered in distance learning settings and participants got to know their partners within the study activities.

1) Participants

In this study 12 students had participated, for 5 of them the course was mandatory and the rest participated as life-long learners. 8 participants are males and 4 females with an age range of 22 and 41 years old (M = 32, SD = 6.53). With respect to education level, 3 of them hold a Bachelor degree and 8 holds a Master degree, and 1 has a PhD. degree.

Only 6 students finished the entire study as the course was mandatory for 5 of them. One student participated in all the three phases but s/he did not finish the requirements of phase 3. Two students finished phases 1, and 2 and three students only participated in phase 1.

2) Apparatus and Stimuli

The course material and tests have been provided online via in house developed system with respect to the architecture proposed in [2] for self-directed learning with automatically created tests - using automatic question creator - and based on the e-assessment framework discussed in [3]. Moreover, the tool named “Co-writing Wiki” [4] has been integrated to the system based on Single Sign-On (SSO) approach [5] and used by participants in the third phase of the study to collaboratively solve a problem.

a) Pre-questionnaire

This questionnaire was provided at the beginning of the study and investigated information on demographic data, pervious experience in group work and collaborative learning, general attitudes on self, peer-assessment after [6], and motivational aspects towards using CLR enriched with automatic assessment for self-directed learning.

The section of attitudes concerning self-, peer-assessment has been adapted from the work of [6] to investigate the following four scales of motivation: intrinsic motivation scale measures the student’s motivation doing the peer-assessment activity for its own sake, just out of pleasure, e.g. “In a peer-assessment activity I liked opinions from peers because I got more ideas.”, extrinsic motivation scale measures the student’s motivation doing the peer-assessment activity in order to get approval from the teacher and a good grade, e.g. “In a peer-assessment activity I think the opinions of my work from teachers were more important than those...
from peers.”, evaluating scale measures the confidence of the students in evaluating their peer’s work, e.g. “In a peer-assessment activity I found the strengths of my peers work when I reviewed it.”, and receiving scale measures how students can handle their peer’s assessment in order to recognize their own weaknesses, e.g. “In a peer-assessment activity I recognized my weakness when I got comments from peers.”. Moreover, Answers were given on a 5-point Likert scale (“strongly disagree” - “strongly agree”), so that students could state their level of agreement or disagreement.

In order to investigate the participants motivation towards the course in general and the study phases in particular, a section adapted from [7] has been added based on the following three motivation scales: Intrinsic Goal-Oriented scale measures the students’ intrinsic motivation regarding the course, for instance: “I prefer course material that arouses my curiosity, even if it is difficult to learn” A high value on this scale would mean that the students are doing the course for reasons such as challenges and curiosity, Extrinsic Goal-Oriented scale deals with the extrinsic motivation of the students, e.g. “Getting a good grade is the most satisfying thing for me right now” A student is extrinsically motivated when s/he is rather interested in rewards or good grade than in the task itself, and Task Value scale is about the learning task itself, i.e. how important, interesting, and useful the task and the task material are for the students. More interest in the task should lead to more involvement in one’s learning. To give an example, one item out of this scale is: “I think I will be able to use what I learn in this course in other courses”. Answers were given based on a 5-point Likert scale as described above.

b) Intermediate questionnaire

This questionnaire was provided after the second phase of the study – self-directed with automatic formative assessment - (see procedure section for more details) to investigate aspects such as, quality of learning material and tests, preferred learning style, emotional aspects, and tools usability. Regarding the learning material quality a scale of (“very bad” (1) - “very good” (5)) was used. Students were asked how often they had taken a test based on a scale of “never” (1), “seldom” (2), “sometimes” (3), “often” (4).

Regarding the “usability of the learning scenario” we used the System Usability Scale (SUS) [8] which contains 10 items and a 5-point Likert scale to state the level of agreement or disagreement (e.g. “I think that I would like to use this system frequently”).

The learning style of ‘elaborating’ or ‘repeating’ has been investigated in order to find out if the students’ learning process is rather superficial or aims at a deeper understanding. For this section, items developed by [9] have been translated into English (e.g. item regarding the elaborating learning style: “In my mind I try to connect what I have learned with already known issues concerning the same topic”, item regarding the repeating learning style: “I try to learn the content of scripts or other notes by heart”). The answers were also given based on a 5-point Likert scale.

To figure the participants emotional state during the second phase, the emotional scale developed in [10] has been used. This scale includes 12 items and measures emotions related to learning new computer software as follows: Happiness: (“When I used the tool, I felt satisfied/excited/curious.”), Sadness: (“When I used the tool, I felt disheartened/dispirited”), Anxiety: (“When I used the tool, I felt anxious/insecure/helpless/nervous”), and Anger: (“When I used the tool, I felt irritable/frustrated/angry”). For this section answers followed a scale of: “None of the time”, “Some of the time”, “Most of the time” or “All of the time”.

c) Post-questionnaire

At the end of the third phase, this questionnaire was provided to participants. Aspects such as, task difficulty and learning effort - in terms of hours, attitudes towards the group-assessment with rubrics as part of Co-writing Wiki, Co-writing Wiki usability and participant’s emotional state when they have used it, and further comments and suggestions.

Moreover this questionnaire investigates the participants motivation during the study three phases and their perception about their peer’s motivation as well. For instance students were asked “How motivated were you according to the following tasks?”: reading the contents, working with the self-directed tool, testing myself with questions, writing the essays, working with the Co-writing wiki, planning a study, group assessment activity, and filling in the evaluation questionnaires. A scale of (“absolutely unmotivated” (1) - “very motivated” (4)) has been used to get the participants answers.

3) Procedure

As mentioned before the study procedure has three phases:

a) Phase 1: Introduction to Scientific Research

In the beginning of this phase, students were asked to answer the pre-questionnaire. Moreover, they have been provided an introductory learning material about scientific research course in general, how to plan a study, and experimentation design and analysis. Nevertheless, information about assessment scheme as well as description of the study phases and requirements they need to achieve. Moreover, they have been asked to take a summative test based on automatically created questions from the provided learning content using the scenario discussed in [2].

b) Phase 2: Selected Topics on Experimentation Planning

Students have been grouped by the instructor into 6 groups - two members each - based on their interest in the course (i.e. mandatory of 3 groups and volunteer of 3 groups). After that an online learning material covers scientific research has been provided using the developed system. The content has been divided into two main categories experimentation design and experimentation analysis. For each of them, 6 articles have been delivered. Each group member was requested to select one article from both categories different than the ones selected by his peer within the same group. In order to avoid members from the same group selecting similar articles they have been asked to use the discussion forum from the Co-writing Wiki to agree
on their selections, and to edit the group main page on the Co-writing Wiki with their selections. Moreover, participants introduced each other using the forum, and selected their articles based on their interest.

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

After that, students were asked to write two essays - 1000 words per article summarizing the topics in his/her selected articles using the Co-writing Wiki. The using the peer-review features provided in the Co-writing Wiki, group members could provide feedback on their peers’ essays and learn their topics consequently.

Finally, the essays content-per group - have been used to create automatically a test - for each group - using the self-directed system. Taken this test from group members require them to be aware of peers’ topics. Moreover, students were asked to answer the intermediate questionnaire after this phase.

c) Phase 3: Experimentation Planning

In this phase, groups have been given a problem based on this research question “Is there a difference between ‘Facebook’ users and ‘non-Facebook’ users concerning their sport activities?” Then they were requested to collaboratively plan a study using the Co-writing Wiki to solve the problem, peer-assess other groups studies using online assessment rubric has been designed for this purpose, and provide feedback. Accordingly, each group will receive feedback from other groups based on peer-assessment as well as they learn from others’ ideas.

The students had to write a method section by which they describe how they would investigate this research question. The students were asked to write maximally 4-5 pages in total (max. 2500 words). Furthermore, they did not have to provide any introduction with related research (although this would be mandatory in a real scientific paper). Instead, they only focused on the design of the study and gave some ideas how the analysis could be performed. Group’s final products after peer-assessment and enhancement phase were evaluated by the instructor and detailed feedback has been provided for each group. After all phases of this study have been finished the students were asked to answer the post-questionnaire.

B. Results Evaluation and First Findings

This section reports the results analyzed from students’ answers on the three 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, and using metrics of M1.1: Ratings for functionality/usability of the tool itself, frequency of use. M1.2: Ratings for emotional aspects while using the tools, and M1.3: suggestions and comments based on open questions.

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

Moreover, the students were asked about “what they like about the three types of tests”. Results have shown that the different types of questions helped them getting an overview about the topics. Furthermore, some students also stated that the sub-section and post-tests supported them in observing their learning progress. However, the tests were criticized as they focus factual knowledge.

With respect to the tool usability, the average SUS score based on students responses is 66.88, where the SUS scale gives a score within a range of 0 and 100. According to [11] this score can be considered as “OK” having that the complexity of the learning scenario and the use of multiple tools in a flexible and interoperable way within the same learning scenario. Moreover, with respect to what the students liked about the tool, students stated that they were in favor of the simplicity of the tool and the division of the content into meaningful modules. Furthermore, the students liked the consistency and the possibility to have an overview of the learning progress and their own test results. On the other hand, students mentioned that session time-out was short. Some also complained about the slow interface. Regarding the Test Module within the self-directed tool, some students criticized the difficulty to navigate to different questions.

Regarding comments and suggestions for improvement, some students would prefer a faster responding system and a faster navigation.

With respect to usability of the Co-writing Wiki itself, an average SUS score of 52.08 has been computed. Moreover, almost all students stated that the Co-writing Wiki is easy to use. They also were in favor of the ability to discuss per topic, per page and creating and modifying pages. In addition, they mentioned that the tool was always available and consistent. However, some students complained about the usability of the Co-writing Wiki and its slowness. The students also mentioned that they were not aware of all available functions. It was also annoying for some of them self-assess their contributions. They also mentioned some editing problems, especially when this content has been
Before using the tool, intrinsic motivation regarding peer-assessment activity and using metrics of angry. Students seldom felt consciously happy, sad, anxious, or angry. By interpreting the mean values, it can be assumed 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 ($M = 1.54$, $SD = 0.31$). The results are similar to Co-writing Wiki, the results from a 4-point rating scale showed that the students felt equally happy ($M = 1.72$, $SD = 0.65$), sad ($M = 1.33$, $SD = 0.41$), anxious ($M = 1.42$, $SD = 0.34$), and angry ($M = 1.61$, $SD = 0.53$). Since a one-sample Kolmogorov-Smirnov Test showed that the data for all four emotions are distributed normally ($p$-values range between 0.257 and 0.69), a one-way ANOVA for repeated measures was performed. With $F = 0.874$, $df = 3$, and $p = 0.47$ the results show no significant difference among the four types of emotion. By interpreting the mean values, it can be assumed that the students seldom felt consciously happy, sad, anxious, or angry. By interpreting the mean values, it can be assumed that the students seldom felt consciously happy, sad, anxious, or angry.

2) $H2$: 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 style of the students, and using metrics of M2.1: Ratings of students’ extrinsic and intrinsic motivation regarding peer-assessment activity before using the tool, M2.2: Ratings of students’ extrinsic and intrinsic motivation regarding the course and its tasks before using the tool, M2.3: Ratings of students’ group-assessment activities and M2.4: Ratings regarding the learning styles.

With respect to M2.1, the student’s motivation concerning the peer-assessment, a comparison of the mean values ($t(11) = 5.99$, $p < .01$) shows that the student’s intrinsic motivation ($M = 3.75$, $SD = 0.51$) is significantly higher than their extrinsic motivation ($M = 2.65$, $SD = 0.48$). Thus, the students would participate in assessment for its own sake and out of pleasure and not just for getting a good grade or approval from the teacher. It can be assumed that the student’s first aim was to learn something out of the course and that getting a grade does not play such an important role for them. This result stands in accordance with the fact that half of the students participated in the course voluntarily. For instance, students stated that they liked opinions from peers in order to get more ideas ($M = 4.08$, $SD = 0.67$). In contrast, they would not feel that they have learned nothing if they get a low peer score on their work ($M = 1.75$, $SD = 0.75$).

Regarding to M2.2, the results of the student’s motivation regarding the course and its tasks shows that the intrinsic motivation ($M = 3.94$, $SD = 0.53$) is significantly higher than the extrinsic motivation ($M = 2.83$, $SD = 0.79$; $t(11) = 3.43$, $p < .01$). This means that they are interested in the course for reasons such as curiosity and challenge, whereas high grade or rewards were not so important for them. These findings are supported by the results of the task value scale. A mean value of 3.83 ($SD = 0.74$) shows that the students were really interested in the task itself. The task material was also very useful and important for them. Due to their high interest, it can be assumed that this also leads to more involvement in their learning activities.

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

| TABLE I. MEAN RATINGS OF MOTIVATION DURING THE COURSE |
|----------------|----------------|----------------|
| Motivation while: | M (SD) | Md |
| reading the content | 3.50 (0.55) | 3.5 |
| working with the tool | 2.67 (0.52) | 3.0 |
| testing themselves with questions | 2.50 (0.84) | 3.0 |
| writing essays | 3.50 (0.55) | 3.5 |
| planning a study | 3.67 (0.52) | 4.0 |
| using the Co-writing Wiki | 2.67 (1.03) | 3.0 |
| performing group-assessment | 3.00 (0.0) | 3.0 |
| filling in the questionnaire | 3.00 (0.0) | 3.5 |

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

Regarding to M2.3, in phase 3 students were asked to evaluate the work of other groups. Regarding the assessment rubric provided for the group review, the students stated that the assessment rubric was easy to use ($M = 3.67$, $SD = 1.51$). In addition 50 % of the students agreed on the statement that the assessment rubric supported them to effectively review the product of the other groups ($M = 3.17$, $SD = 0.98$).

The students neither agreed nor disagreed on the statements “The assessment rubric provided for the group review supported me to learn more about other group’s topic.” and “Using the rate control (stars) was very helpful to assess the student’s level of mastery based on the rubric criteria.” In addition, the students were asked what they liked regarding this group-assessment. All of the students mentioned that they liked the group-assessment because of the opportunity to see how other groups approached the problem and solved it in order to improve their own products. In the other hand, some students answered that they would have preferred to give textual detailed feedback to state suggestions and improvements instead of providing short feedback by using the assessment rubric.

With reference to M2.4 (see Fig. 1), A comparison of the mean values shows that there is a significant difference between the elaborating learning style ($M = 4.05$, $SD = 0.56$) and the repeating learning style ($M = 3.04$, $SD = 0.82$; $t(7) = 2.71$, $p < .05$). The students prefer the elaborating learning style, which means that their learning process aims at deeper understanding and is less superficial. Concerning
elaborating, for instance the students stated that they try to link new terms or new theories to familiar terms and theories ($M = 4.38$, $SD = 0.52$). In contrast to that, the students said that they do not learn the content of scripts or other notes by heart ($M = 2$, $SD = 1.07$) which would indicate a repeating learning style.

From M2.2 and M2.4 results we can figure the relation between elaborating learning style and deep learning based on intrinsic motivation to participate in the learning activity. The results from M2.2 show that students were intrinsically motivated after the first Phase of the course. So due to their learning style preference, it can be assumed that the students were still intrinsically motivated in the second Phase, where they received the questions during the self-directed learning phase. Thus, the students answered the questions out of pleasure with the aim to deepen their knowledge.

In addition, the students stated that testing themselves with questions often helps them ($M = 3.63$, $SD = 1.50$). This result is in line with the results discussed above. Therefore, it can be assumed that providing self-directed learning courses with the ability to create automatic tests supported the students to achieve their learning goals.

**III. CONCLUSION AND OUTLOOK**

With respect to the study goals, summarizing (G1) findings, it can be assumed that the tools developed to integrate assessment forms to CLRs are user-friendly, and usable because of the satisfactory SUS score the tools have reached. Moreover, the students were in favor of the various functions of the tools and their simplicity. Moreover, they stated that the tools gave them a good overview of their learning progress. For further improvement, a closer look on the questions quality enhancement and on a faster interface should be considered. Moreover, the study shows the applicability of combining interoperable and flexible learning tools in one complex learning scenario.

Regarding students’ motivation (G2), the results show that the students were intrinsically motivated at the beginning of the course. So they were really interested in the course and its tasks, which lead also to more involvement in their learning activities. Moreover, students’ motivation was high during reading content, writing essays, doing the peer and group assessment, working with the Co-writing Wiki in a problem-based learning scenario, and filling in the questionnaires. In addition, testing themselves with automatically created tests and working with the self-directed learning tool also motivated them.

By investigating students’ learning styles, we found out that the students’ learning process aims at deeper understanding and is less superficial. This result is in line with the results discussed above, because intrinsic motivation is an important condition for this learning style. Thus, it can be assumed that students answered took tests out of pleasure with the aim to deepen their knowledge. Besides, students also stated that testing themselves often supported them in their learning process (G3).

**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.

**REFERENCES**


