Peer Assessment System for Modern Learning Settings: Towards a Flexible E-Assessment System

Mohammad AL-Smadi[1], Christian Guetl [1,2] and Frank Kappe[1]

[1] Graz University of Technology, Graz, Austria

[2] Curtin University of Technology, Perth, WA

Abstract—**According to the rapid changes in our life styles and in order to cope with the new requirements for modern learning settings and activities, several applications have been developed to provide our modern society with these learning settings. E-assessment as a main part of e-learning has been affected by these new settings and new aspects such as peer assessment have become more and more important. In this paper, we will present a computer-assisted peer assessment system that can be used to improve the learning process. An overall architecture will be presented and an experiment has been conducted. First findings will be discussed and future work will be mentioned.**

Index Terms—**Computer-Assisted/Based Assessment**; **E-learning; On-line Testing; Peer-assessment; Self-assessment.**

I. INTRODUCTION

Our modern life has affected our societies to be more modern and Global. New life settings and needs have appeared in the 21st century. Our learning settings and systems have been struggling to cope with these changes and challenges. Therefore, new and modern learning styles, settings and resources have been adopted to satisfy our society needs and to help people to improve their skills as well as their expertise to cope the rapid changes in their societies [1]. The learning process has been changed from being repetitive to a new form of learning based on understanding, independency, learners' empowerment and skills improvement [2]. The learning theories have been changed from being associative and behavioral to be more cognitive and constructive, where the measurement have evolved from being scientific measurement (separated from the instruction activity) to have a new culture of assessment (where measurement and instruction have been integrated) [3]. New age of information has appeared where information and communication technology plays a main role in education and learning society. The thing that has addressed the requirement for new skills such as: *cognitive competencies*, *meta-cognitive competencies*, *social competencies* and *affective dispositions* (e.g. perseverance, internal motivation and self-efficiency) [1]. Consequently, new forms of assessment such as: self-, peer- and co-assessment have been implemented to achieve the demandable goals and objectives of the learning process**.**

Peer-assessment has gained its importance from its emphasis on the importance of making the student an important part of the assessment process not only as assessee but also as assessor where students and tutors collaboratively work together in the assessment model [4]. Rather than supporting the learner-centered model, peer-assessment may decrease staff load and time consumed on the assessment process as well as it may develop certain skills for the students such as, communication skills, self-evaluation skills, observation skills and self-criticism [1].

This paper focuses on aspects of peer assessment activities and how a computer-assisted approach can support as well as improve the assessment procedure and the learning process. A web-based prototype has been developed for implementing an enhanced peer assessment procedure and an experiment has been performed. To this end, the rest of this paper is structured as follows: Section 2 outlines related work for peer assessment activities. Section 3 describes the enhanced peer assessment procedure, and the experiment setup. Section 4 shows the architecture of the peer-assessment system. The experiment results and findings are discussed in section 5. Section 6 outlines conclusions and future work.

II. RELATED WORK

Reviewing the domain of peer-assessment, peer-assessment is not new, it can be referred back to a long time of history, where George Jardine the professor in the University of Glasgow from 1774 – 1826 prepared a pedagogical plan that included some peer-assessment methods and advantages [5]. Peer-assessment has been defined as "an arrangement for the peers to consider the level, value, worth, quality or successfulness of the products or outcomes of learning of others of similar status" [6]. From this definition, you can notice that peer-assessment is not a method for measurement but it is a source of assessment that can be utilized within a framework side by side with other methods [7].

Several tools have emerged since the beginning of the 21st century. Some of those tools are part of computer-based assessment systems that implement the peer-assessment methods [8]. The earliest reported system to support peer-assessment developed at the University of Portsmouth, "*The software provided organizational and record-keeping functions, randomly allocating students to peer assessors, allowing peer assessors and instructors to enter grades, integrating peer- and staff-assessed grades, and generating feedback for students*" [9]. One of the first systems with the peer-assessment methods was a tool for collaborative learning and nursing education based on

multi-user database, which was called MUCH (Many Using and Creating Hypermedia). In the same period a Macintosh application has been developed which has implemented a peer-review process for an assignment that has been reviewed by two peers ([9]; [10]; [11]). In the late 1990s, NetPeas (Network Peer Assessment System) has been implemented, and Artificial Intelligence (AI) has been used to develop the tool of Peer ISM that combines human reviewing with artificial ones ([12]; [10]; [13]). Computer-assisted-peer-assessment systems has also affected by the revolution of Word Wide Web (WWW), several web-based system have appeared later on. An example of the first reported web-based system was a web-based tool for collaborative hypertext authoring and assessment via e-mail [14]. Other systems such as, a web-based system for group contributions on engineering design projects [15], the Calibrated Peer Review (CPR) which was introduced in 1999 [16], the Peer Grader (PG) as a web-based peer evaluation system [11], The Self and Peer Assessment Resource Kit (SPARK) which is an open-source system designed to facilitate the self and peer assessment of groups [17], The computerized Assessment by Peers (CAP) is another example [8]. Further examples such as, OASIS which has automated handling for multiple-choice answers and peer assessment for free-text answers, The Online Peer Assessment System (OPAS), which has some abilities for assignment uploading and reviewing as well as groups management and discussions [18], An improvement for this system was introduced in Web-based Self and Peer Assessment system (Web-SPA) to avoid the lack in determining standards, methods of scoring and the workflow of the assessment process [19]. Recent examples of peer-assessment developments are, the enhanced open-source implementation of WebPA system which was originally developed in 1998 [20], as well as the Comprehensive Assessment of Team Member Effectiveness (CATME) system which assesses the effectiveness of team members contributions [21].

III. EXPERIMENT SETUP

The experiment has performed an e-learning activity as part of the course "Information Search & Retrieval (ISR)" at Graz University of Technology in the winter term 2008/2009. The experiment was conducted in a controlled environment in the computer lab with a supervision of the course lecturer. A web-based Peer-ASsessment System (PASS) has been used by the students to participate in the experiment. The same system has also been used by the tutors in the evaluation process of the students' candidate answers. The experiment details are as follows:

- Introductory talk (10 minutes): at the beginning of the experiment a short introduction had been given by the ISR course lecturer about the subject-domain as well as the assessment in general and the peer assessments as an emerging form of assessment. The importance of knowledge acquisition and knowledge assessment in modern learning settings was discussed briefly. The learning objectives behind this experiment were mentioned. The lecturer also stressed on the importance of the students performance during the experiment and clarified that the performance will be given 10 points as part of the overall grade for both the online test and the online peer assessment session of 5 points each.

- Online learning session (45 minutes): "Document Classification" as one of the main topics of ISR course has been chosen to formulate the online learning material of the experiment [22]. The material language is English and it has been extracted from Wikipedia [23]. The material is formulated out of four web-pages and an introduction one, where the students were allowed to access and navigate between them as well as a set of further readings hyperlinks related to the subject-domain.

- Online testing session (15 minutes): The knowledge that had been gained by the student from the last session was assessed in this session. An English test language of five questions has been deployed for the students as a web-based assessment system. During this session the students were not allowed to access any course materials. The test items varied in their complexity, the first questions was a definition question; the second was an enumeration, where the third and the fifth were asking for a concept explanation while the fourth was an abbreviation one. For each of the fifth questions a short-free answer and a confidence value out of 10 (0: very poor; 10: very good) had to be provided. The confidence value is used to evaluate the level of maturity for the student answer (self-directed assessment) see Fig. 1.

- Break (15 minutes).

- Online reference answers preparation (15 minutes): During this session, the students were asked to prepare reference answers for the questions 1, 2 and 5 with a confidence value for their estimation of their answers quality. Differently from the last session the students were asked to access the course content and other useful materials to help them in preparing the reference answers.

- Online peer assessment session (45 minutes): in this session the students used the reference answers they had prepared in the last session to evaluate and peer-assess their peers candidate answers from the online test session. Every student had to evaluate around 30 randomly selected answers for questions 1, 2 and 5 as well as 15 calibrated answers provided by the course teacher. For each answer, the students were asked to mark the answer by special marks for highlighting, underlining or changing to italic. Underlining some parts of the answer means that they are correct, where highlighting them means that they are wrong, and changing them to italic means that they are irrelevant. Input-boxes for missing parts of the answer and additional notes were provided for the students to write into them. A grade should also be provided by the student for the answer from "0" (very poor) to "10" (very good). Buttons were used to represent the candidate answers, they all yellow at the begging and once the student evaluates one of them its button color becomes green. Fig.2., shows an example of the evaluation of answers during the peer-assessment phase.

- Experiment questionnaire (10 minutes): the students were asked to fill in a questionnaire that diagnoses their impressions about the assessment activity of its three parts self-directed, online test and the peer-assessment one, the usability of the web-based as-
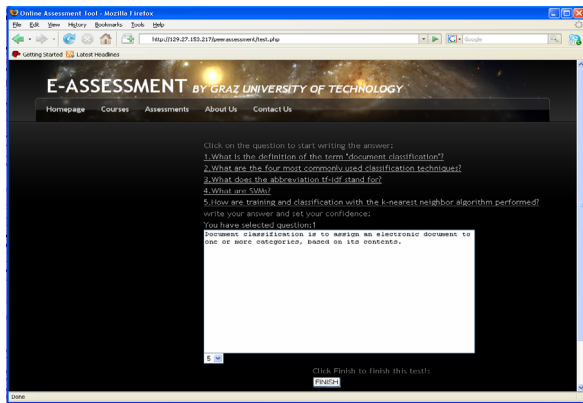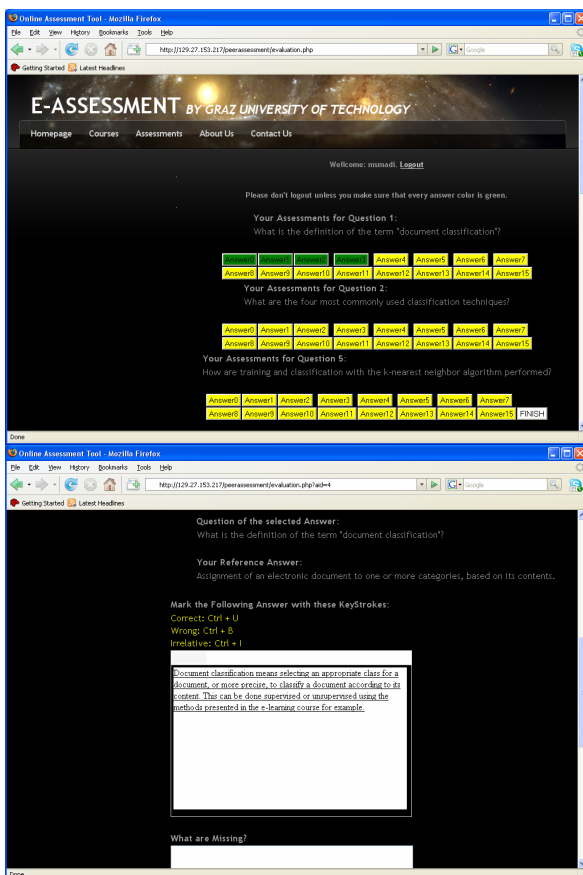
Figure 1.   Online test from PASS.

Figure 2.   Peer-assessment using PASS.

sessment prototype and their suggestions for further enhancements and notes.

- Results delivery: as part of later on feedback provision the students' answers and performance has been analyzed and a final grade has been sent to them by e-mail.

In order to compare the students' peer-assessment results with a reference gradings, a set of tutors had participated in the experiment. The tutors' peer-assessment process was as follows:

- Experiment Introduction: an e-mail had been sent to all the tutors, in which a brief introduction about the experiment goals and procedures had been presented.

- Reference answer preparation: the tutors were asked to use the course content and other related materials

to prepare reference answers that they will use later on in the evaluation process.

- Online peer-assessment: in this step, all the answers from the students (test and reference answers for the five questions) had been evaluated by the tutors. The same marking and grading facilities of highlighting, underlining and changing to italics of some parts of the candidate answers were possible. As well as the possibility of adding notes and missing parts of the candidate answers.

A group of 27 students enrolled at the course of ISR. The students were separated into two groups 12 for the first group and 15 for the second. All of them had participated in the experiment. 14 (51.9%) of the students were taking part in the course as a bachelor program, where 13 (48.1%) were masters' students. 3 (11.2%) were females and 24 (88.8%) were males. The average age of the students was 26.5 years old with a minimum age of 22 and a maximum one of 37. The tutors were a group of 5 PhD students at the IICM (Institute for Information Technology and Computer Media) of Graz university of Technology. All of them were males and has a master degree of computer science.

IV.   SYSTEM ARCHITECTURE

Fig. 3 demonstrates the overall architecture of PASS. PASS has three main modules: user management module, test management module and results analysis and feedback module. PASS has been developed using PHP programming language [24] and MySQL database [25]. The system is built on top of Apache Tomcat Servlet (TOM-CAT) [26] and applies the MVC (Model-View-Control) approach [27]. PASS has been developed as part of a flexible e-assessment system project [28]. The three PASS main modules are:

- User Management Module: from its name, this module handles the authority levels of the systems' users. According to the diversity of the systems' users we have identified three main roles, Administrators' role, teachers' role and student's role. Other roles like parents and decision makers can be easily constructed using this module. This module also handles the login/logon processes based on the users that have been created and the roles that they belong to.

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

- Results Analysis & Feedback provision Module: this module computes the final grades of the different assessment activities that took place during this experiment. Dedicated results' analysis and mining is conducted in this module to support students, teachers and other related decision makers with a valuable feedback.
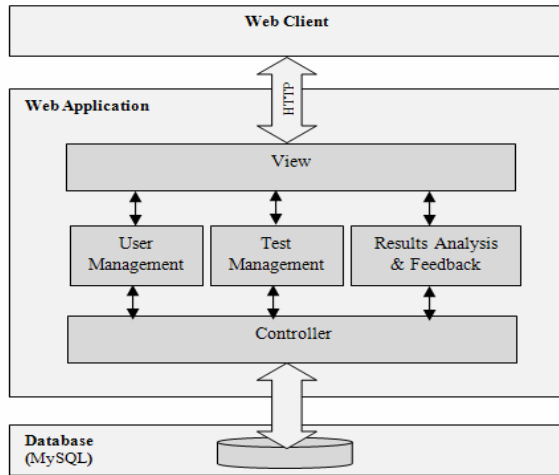
Figure 3. PASS Architecture.

V. RESULTS ANALYSIS

A. Tutors Phase Results

With response to the diversity of experience that the tutors have, it has been decided to use the weighted mean instead of the arithmetic mean to compute the reference marks for the candidate answers. Table 1 shows the tutors experience represented in weights. All of the tutors are PhD students in computer science (CS) and doing well. Some of them have advanced knowledge in information retrieval (IR) as well as in assessment activities (AS). The weights given to the tutors have been decided based on the tutors experience as well as the arithmetic mean of tutors grading where a grade value of 5 represents the reasonable mean of a scale between 0 and 10. The cross correlations values from Table 2 support that by noticing that the best correlations of assessment results were between (T1, T4) and (T2, T5) for the all test items.

Table 2 outlines the cross-correlations of the tutors' grading results as well as their cross-correlations with the weighted mean values of their grading of candidate answers. For all of the test items the cross-correlation values vary between 0.499 (T2, T3) and 0.833 (T1, T4) by a mean value of 0.67 ($\sigma = 0.24$). For test Item 1 which asks for a definition, the cross-correlation values are between 0.555 (T1, T2) and 0.865 (T1, T4) by a mean value of 0.71 ($\sigma = 0.22$). A better situation for test Item 2 which asks for an enumeration, the cross-correlation values are between 0.701 (T3, T5) and 0.949 (T1, T2) by a mean value of 0.83 ($\sigma = 0.18$). For test Item 3 the cross-correlation values are the worst while it asks for an explanation of a concept, they are between 0.126 (T2, T3) and 0.755 (T1, T4) by a mean value of 0.44 ($\sigma = 0.44$). The same findings can be found in the literature where the variance between the tutor's correlation values depends on their experience as well as on the complexity of the assessment task [29]; [30]; [31].

In order to investigate the results, the absolute error of the tutors' individual grading values and the weighted average of their grading is computed. As shown in Table 3, the absolute error for all of the test items is between 2.18 ($\sigma = 1.88$) as worst result and 1.12 ($\sigma = 1.14$) as best result. Similar to the cross-correlation findings, for test item 1 the absolute error varies between 2.30 ($\sigma = 1.84$) as

	Experience				Grading	
	CS	IR	AS	Weight	Mean	σ
T1	☑	☑	☑	2	6.04	3.62
T2	☑	☑ ☑	☑	2	3.86	4.0
T3	☑	☑	☑	1	7.31	3.39
T4	☑	☑ ☑	☑ ☑	2	6.14	3.64
T5	☑	☑ ☑	☑ ☑	3	4.21	3.69

		T1	T2	T3	T4	T5	WMW
All Test Items	T1	1.000	0.706	0.618	0.833	0.686	0.903
	T2		1.000	0.499	0.567	0.707	0.844
	T3			1.000	0.633	0.514	0.699
	T4				1.000	0.645	0.856
	T5					1.000	0.887
	WMW						1.000
Test Item 1	T1	1.000	0.555	0.764	0.865	0.645	0.891
	T2		1.000	0.716	0.613	0.642	0.788
	T3			1.000	0.826	0.675	0.893
	T4				1.000	0.771	0.941
	T5					1.000	0.857
	WMW						1.000
Test Item 2	T1	1.000	0.949	0.816	0.845	0.813	0.963
	T2		1.000	0.799	0.757	0.766	0.923
	T3			1.000	0.688	0.701	0.839
	T4				1.000	0.724	0.888
	T5					1.000	0.915
	WMW						1.000
Test Item 3	T1	1.000	0.392	0.303	0.755	0.640	0.866
	T2		1.000	0.126	0.311	0.534	0.554
	T3			1.000	0.318	0.180	0.353
	T4				1.000	0.570	0.836
	T5					1.000	0.896
	WMW						1.000

	All Test Items		Test Item 1		Test Item 2		Test Item 3	
	Mean	Σ	Mean	σ	Mean	σ	Mean	σ
T1	1.12	1.14	1.83	1.42	0.45	0.62	1.07	0.80
T2	2.04	1.86	2.02	1.83	0.63	0.90	3.48	1.52
T3	2.18	1.88	1.93	1.36	0.88	1.27	3.72	1.77
T4	1.38	1.31	1.62	1.30	0.85	1.50	1.69	0.96
T5	1.65	1.69	2.30	1.84	0.69	1.57	1.95	1.23

worst result and 1.62 ($\sigma = 1.30$) as the best one. The best case can be seen in test item 2 which reflects the simplicity of the assessment activity done by this item as an enumeration item where the absolute error is between 0.88 ($\sigma = 1.27$) and 0.45 ($\sigma = 0.62$). Test item 3 has not only lower cross-correlation but also higher absolute error values between 3.72 ($\sigma = 1.77$) and 1.07 ($\sigma = 0.80$).

B. Students Phase Results

In order to compare the student's peer-assessment performance with the tutor's reference marks, the arithmetic mean of student's evaluation results per candidate answer has been used, the absolute error of that arithmetic mean and the tutor's reference marks (as weighted average of their grading for the same candidate answer) has been computed. For all the three test items (Q1, Q2, Q5) the arithmetic mean of absolute error is quite low with 0.98 ($\sigma = 0.95$). For the three test items individually, test item 1 has the lowest arithmetic mean of absolute error with 0.54

(σ = 0.62). Test item 2 has a value of 0.86 (σ = 0.49), where test item 3 has the highest value with 1.54 (σ = 1.27). The highest value in the case of item 3 reflects the complexity of the assessment activity done by this item as an explanation one.

The correlation between the arithmetic mean of the students' evaluations per candidate answer and the tutor's reference grading as weighted average of their grading for the same candidate answer, is quite high with 0.90 for the whole three test items, 0.97 for test item 1, 0.96 for test item 2, and 0.86 for test item 3. The findings show that the level of agreement represented by the correlation between the students' evaluations and the tutors reference grading depends on the complexity of the assessment activity represented by the assessment item in our case. Fig. 4 represents a scatter plot for the tutor's reference grading in comparison with the students peer assessments for the three test items sorted in ascending order by the tutor's reference grading values.
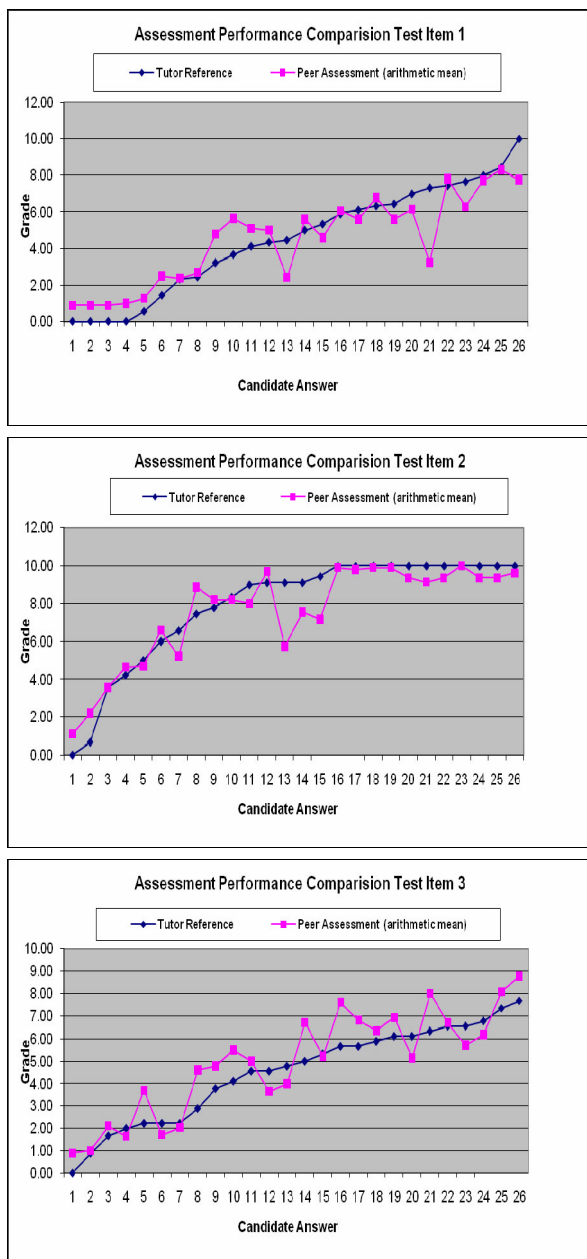

Figure 4. Students Peer-assessment Performance.

C. Students Questionnaire

In this section the results gained from the students' questionnaire will be analyzed and presented. As mentioned earlier, the students' questionnaire was used to diagnose the impressions and opinions of the students about the overall experiment. Matters such as, students' knowledge acquisition, students impressions about the online peer assessment and the usability of the tool were the main sections of the questionnaire.

From the students' point of view, their basic knowledge in the subject before the experiment was with a mean value of 2.56 (σ = 1.4) where ("0" represents fully disagreement and "5" represents fully agreement). The knowledge gained from the online learning phase was with a mean value of 3.65 (σ = 1.05), where preparation of reference answers has supported the students to get better knowledge in the subject domain with a mean value of 3.26 (σ = 1.29). The knowledge that they had gained from the peer assessment task was with a mean value of 3.07 (σ = 1.24), rather than these two tasks, the task of candidate answers evaluation had supported the students to get better understanding of the subject details with a mean value of 3.56 (σ = 1.22). Furthermore, students had used the course content during the peer assessment task with a mean value of 2.52 (σ = 1.74). Fig. 5, shows the results for the students' self estimation of knowledge acquisition from the overall experiment. The self estimation of students' knowledge acquisition has been discussed in several researches ([29]; [32]; [33]).

By analyzing the students' impressions on the peer-assessment as part of a modern learning settings, students like peer-assessment as part of the learning activity with a mean value of 2.74 (σ = 1.51), where they recommend it to be part of computing their performance grading with a low mean value of 1.56 (σ = 1.45), as well as to be considered as part of the future learning settings with a low mean value of 1.85 (σ = 1.32). These results are presented in Fig. 6.

To get better idea about the usability of the tool, students were asked in the questionnaire about their impressions on the tool functionalities and usability. According to the questionnaire, the students' impressions on the overall tool was with a mean value of 2.56 (σ = 1.25), where their opinion about the online test phase was with a mean value of 2.63 (σ = 1.21) and their impression of the pear-assessment part was with a mean value of 2.33 (σ = 1.36). We also asked them about their expectations about the maximum period of time in minutes for the peer-assessment parts and their suggestions were with a mean value of 45.33 minutes (σ = 28.31), where the time for this part in the real experiment was 45 minutes. Fig. 7, shows the students' impressions on the usability of PASS.

The identified problems and room for improvements according to the students can be summarized as follows: adding time indicator for the different phases of the experiment, some enhancements for the design of the peer-assessment phase, e.g. decreasing the page scrolling, using another color for the answer button of the currently evaluated answer, progress information for the number of questions have been answered or answers have been evaluated out of the total number of questions and answers, using a scale of (0-5) instead of (0-10) to grade the candidate answer in order to simplify the choose of the grade, "wysiwyg" editor for the answers in the online assessment

phase and candidate answers preparation is missing, as well as, they complained of the number of candidate answers to be evaluated in an controlled environment within a specific time.In the other side, they liked the whole idea of peer-assessment as a new way for learning and according to one of the students he argued that "by such way of learning I can compare my answer with others and get different point of views about the answer". Others stressed on the benefit of the repetition caused by evaluating several answers for the same question on the overall learning and understanding of the question. Some of the students also liked the possibility of marking parts of the candidate answers during the peer-assessment phase by (underlined, bold, italic) for (correct, wrong, and irrelevant).

VI. CONCLUSION AND FUTURE WORK

In this paper, we have presented an overall architecture for a Peer-ASsessment System (PASS) for modern learning settings. A prototype for PASS were developed, and used to conduct an experiment during the course of "Information Search & Retrieval (ISR)" at Graz University of Technology in the winter term 2008/2009. The experiment consisted of four main phases: online learning phase, online test, reference answers preparation and online peer-assessment of candidate answers. PASS has been enhanced with marking possibilities to facilitate the evaluation process of candidate answers. PASS has been developed with the possibility to mark parts of the students candidate answers during the peer-assessment phase by formatting them to *underlined, bold, italic* according to the correctness of that part as *correct, wrong, and irrelevant*. 27 students participated in the four phases of experiment, as well as 5 tutors participated in the third and fourth phases. The tutors had to grade the candidate answers collected from the students using the same marking facilities provided by PASS.

The students were asked to fill in a questionnaire about their impressions on the overall experiment. According to this questionnaire, students gained new knowledge during the four phases of the experiment; they also recommended using such modern types of assessment as parts of the learning process; and they also suggested different enhancements based on their impressions on the usability of the system.

The reliability of the peer-assessment results was analyzed through an enhanced experiment procedure. The level of agreement between the student's peer evaluations and the tutor's reference grading values varies according to: the complexity of the assessment task (represented by the test items), the experience of the individuals, as well as the motivation and attitudes. Experiment results showed for students as well as tutors the highest level of agreement was for simple assessment tasks such as definitions and enumeration answers, where the level of agreement was fair with more complex assessment activities such as concept explanation answers.

The weighted average has been used to enhance the tutor's grading values as they have different levels of experience. The average of the absolute error between the tutors weighted average grading values and the student's average marks for each candidate answer has been used to evaluate the performance of the students in the peer-assessment task.

For future work, the reliability of the assessment results will be further enhanced by analyzing the tagged sections from the candidate answers as correct, incorrect and irrelevant. The feedback activities will be also improved to provide both students and tutors with valuable information about the peer-assessment procedure. The web-based peer-assessment prototype will be improved according to the usability and functionality findings, as well as to the recommendations from the students and the tutors.
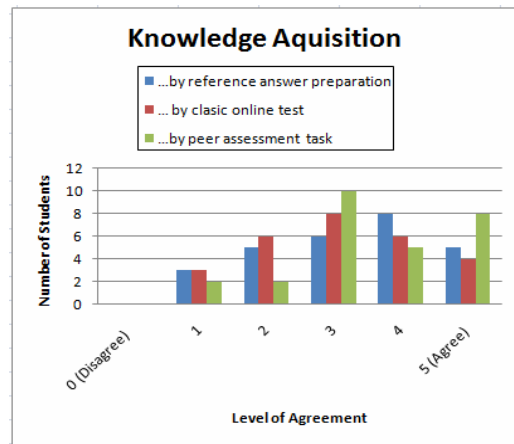

Figure 5. Students' self estimation of knowledge acquisition in peer-assessment experiment.
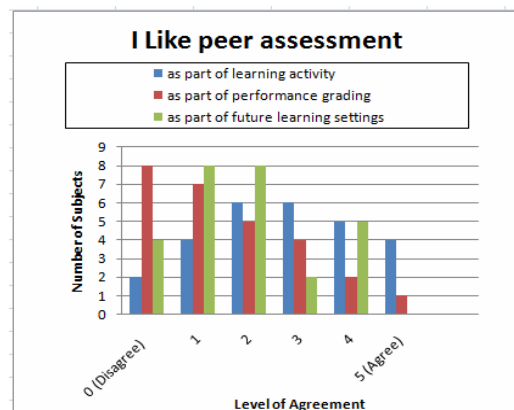

Figure 6. Students' Impressions on peer-assessment in modern learning settings.
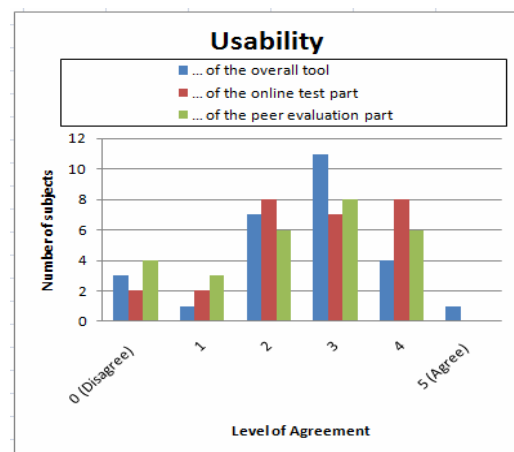

Figure 7. The students' impressions on PASS usability.

REFERENCES

[1] F. J. Dochy; & L. McDowell, "Introduction. Assessment as a tool for learning". *Studies in Educational Evaluation, 23 (4)*, 279-298, 1997. doi:10.1016/S0191-491X(97)86211-6

[2] J.D. Bransford; A.L. Brown; & R.R. Cocking; (Eds.), *How People Learn: Brain, Mind, Experience, and School. Expanded Edition.* Washington DC: National Academies Press, 2000.

[3] L. Shepard,. "The Role of Assessment in a Learning Culture". *Educational Researcher, 29 (7)*, 4-14, 2000.

[4] P. Orsmond, "Self- and peer-assessment: guidance on practice in the biosciences". In *Teaching Bioscience Enhancing Learning Series*, eds S. Maw, J. Wilson, and H. Sears, pp. 1-47 Leeds, UK: The Higher Education Academy Centre for Bioscience, 2004.

[5] K. Topping. "Self and Peer Assessment in School and University: Reliability, Validity and Utility". In Mien Segers, Filip Dochy and Eduardo Cascallar (Eds.), *Optimising New Modes of Assessment: In Search of Qualities and Standards*, Springer Netherlands, 55-87, 2003.

[6] K. J. Topping; E. F. Smith; I. Swanson; A. Elliot, "Formative Peer Assessment of Academic Writing Between Postgraduate Students". *Assessment & Evaluation in Higher Education*, Vol. 25, No. 2, p. 150-169, 2000. doi:10.1080/713611428

[7] G. Brown,; J. Bull, and M. Pendlebury. "What is assessment?". In *Assessing Student Learning in Higher Education.* London: Routledge, 1997.

[8] P. Davies, "Peer-Assessment: No marks required just feedback? Evaluating the Quality of Computerized Peer-Feedback compared with Computerized Peer-Marking". *In Cook, J and McConnell, D (eds), Communities of Practice, Research Proceedings of the 10th Association for Learning Technology Conference (ALT-C 2003)*, 8-10, Sept 2003, Sheffield, UK.

[9] E. F. Gehringer, "Electronic peer review and peer grading in computer-science courses", Proc. of *the Technical Symposium on Computer Science Education*, p. 139-143, 2001.

[10] R. Rada,; S. Acquah,; B. Baker, & P. Ramsey,. "Collaborative Learning and the MUCH System". *Computers and Education, 20(3)*, 225-233, 1993. doi:10.1016/0360-1315(93)90022-B

[11] E. F. Gehringer,. "Strategies and mechanisms for electronic peer review". In Proceedings, *Frontiers in Education Conference*, Vol 1., F1B/2 - F1B/7, 2000.

[12] S. Bull,; P. Brna,; S. Critchley,; K. Davie,; C. Holzherr,. "The Missing Peer, Artificial Peers and the Enhancement of Human-Human Collaborative Student Modelling". *In proceedings of AIED 99*, 269-276, 1999.

[13] C. C. Tsai,; S. S. Lin,; & S.M. Yuan,. "Developing science activities through a networked peer assessment system". *Computers & Education, 38 (1-3)*, 241-252, 2002. doi:10.1016/S0360-1315(01)00069-0

[14] T. Downing,; & I. Brown,. "Learning by cooperative publishing on the World-Wide Web". *Active Learning 7*, 14-16, 1997.

[15] E. A. Eschenbach,; & M. A. Mesmer,. "Web-based forms for design team peer evaluations". *American Society for Engineering Education 1998*, *Annual Conference and Exposition*, Session 2630, 1998.

[16] (CPR)™. Calibrated Peer Review, last retrieved Feb. 5th, 2009, from http://cpr.molsci.ucla.edu/

[17] M. Freeman,; & J. McKenzie,. "SPARK, a confidential web-based template for self and peer assessment of student teamwork: benefits of evaluating across different subjects". *British Journal of Educational Technology, 33(5)*, 551-569, 2002. doi:10.1111/1467-8535.00291

[18] S. Trahasch,. "Towards a flexible peer assessment system". *In Proceeding, Information Technology Based Higher Education and Training (ITHET 2004)*, 516-520, 2004.

[19] Y. T. Sung,; K. E. Chang,; S. K. Chiou,; & H. T. Hou,. "The design and application of a web-based self- and peer-assessment system". *Computer & Education, 45 (2)*, 187-202, 2005. doi:10.1016/j.compedu.2004.07.002

[20] WebPA, http://webpaproject.lboro.ac.uk. Last visited, 5th. February. 2009.

[21] M.W. Ohland, M.L. Loughry, R.L. Carter, L.G. Bullard, R.M. Felder, C.J. Finelli, R.A. Layton, and D.G. Schmucker, "The Comprehensive Assessment of Team Member Effectiveness (CATME): A New Peer Evaluation Instrument," *Proceedings of the 2006 ASEE Annual Conference*, Chicago, Illinois, June 2006. Information about CATME may be found at http://www.catme.org.

[22] ISRDC. "Document Classification". *Online lecture as part of the course ISR, Graz University of Technology,* last edited Nov 25th, 2007, last retrieved Sep. 29th, 2008 from http://www.iicm.tugraz.at/isr/exp

[23] WPDC. "Document classification". *Wikipedia*, last edited Sep. 8th, 2008, last retrieved Sep. 5th, 2009 from http://en.wikipedia.org/wiki/Document_classification

[24] PHP. Official website, general-purpose scripting language suited for Web development, last retrieved Feb. 5th, 2009, from http://www.php.net/

[25] MYSQL. MySQL. Official Website, Microsystems, Inc., last retrieved Feb. 5th, 2009, from http://www.mysql.com/

[26] TOMCAT. Apache Tomcat. Official website, The Apache Software Foundation, last retrieved Oct. 5th, 2009, from http://tomcat.apache.org/

[27] MVC. Model–view–controller, last retrieved Oct. 5th, 2009, from http://en.wikipedia.org/wiki/Model-view-controller

[28] M. AL-Smadi, C. Gütl. "Past, Present and Future of e-Assessment-Towards a Flexible e-Assessment System". *In Proceeding of ICL2008*, Villach , Austria, 2008.

[29] D. Magin, A. Churches,." What do students learn from self and peer assessment?" *In Proceedings, EdTech'88 Conference, Australian Society for Educational Technology*, 27-29 September 1988, last retrieved June 24th, 2009 from http://www.ascilite.org.au/aset-archives/confs/edtech88/magin.html

[30] M. Sullivan; M. Hitchcock,; & G.L. Dunnington. "Peer and Self Assessment during Problem-Based Tutorials". *The American Journal of Surgery,* 177 (March 1999), 266-269. doi:10.1016/S0002-9610(99)00006-9

[31] M. Ward,; L. Gruppen,; & G. Regehr,. "Measuring Self-assessment: Current State of the Art". *Advances in Health Sciences Education, 7 (1)*, 63-80, 2002. doi:10.1023/A:1014585522084

[32] D.M.A. Sluijsmans,. "Student involvement in assessment.The training of peer assessment skills". *Unpublished doctoral dissertation, Open University of the Netherlands, The Netherlands.* 2002.

[33] P. McLaughlin,; & N. Simpson,. "Peer assessment in first year university: How the students feel". *Studies In Educational Evaluation, 30 (2)*, 135-149, 2004. doi:10.1016/j.stueduc.2004.06.003

AUTHORS

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

Christian Guetl is with the Institute for Information Systems and Computer Media, Graz University of Technology, Austria, He is also Adjunct Research Professor at School of Information Systems, Curtin Univerity of Technology, Perth, WA. (e-mail: cguetl@iicm.edu).

Frank Kappe is the head of the Institute for Information Systems and Computer Media, Graz University of Technology, Graz, Austria (e-mail: frank.kappe@iicm.edu).