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

M. AL-Smadi, C. Guetl and F. Kappe

Institute for Information Systems and Computer Media

Graz University of Technology

Brückenkopfgasse 1, 8020 Graz, Austria

Phone: +43 (316) 873 -5662;5639;5602 Fax: +43 (316) 873 5699 E-mail: msmadi;cguetl;frank.kappe@iicm.edu

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.

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.

II. RELATED WORK

In this paper, we are going to address one of these forms of assessment which is 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 [4]. 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" [5]. 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 [6]. 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 [7]. 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]. Several tools have been emerged since the beginning of the 21st century where some of them are 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 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 (Web-SPA) system to avoid the lake 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 had been performed as an e-learning activity for the course of "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 Assessment system had been used by the students to participate in the experiment which is also used by the tutors in the evaluation process of the students' 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 domain of the subject 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 is assessed in this session. An English test language of five questions is 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 were variable, where the first questions was a definition one, the second was an enumeration, the third and the fourth were asking for a concept explanation while the fifth was an abbreviation. For each of the fifth questions a short-free answer and a confidence value out of 10 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.

- 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 identifying the reference answers.

- Online peer assessment session (45 minutes): in this session the students used the reference answers from the last session to evaluate and peer-assess their 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 pre-prepared optimal answers by the course teacher. For each answer, the students were asked to mark the answer by special tags 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 assessment 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.
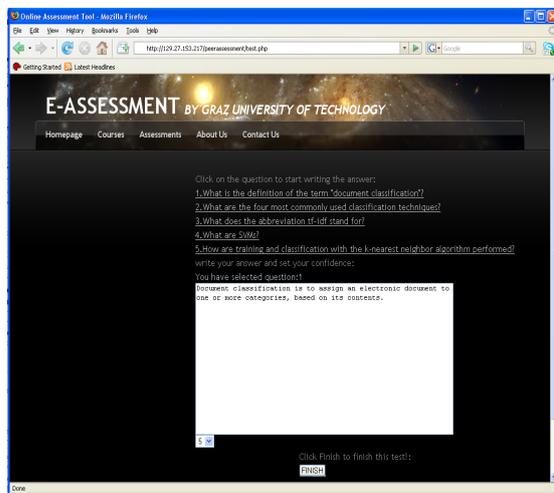

Fig. 1. Online test from PASS.

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 a set o f 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 was taking part in the course as a bachelor program, where 13 (48.1%) was master 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.
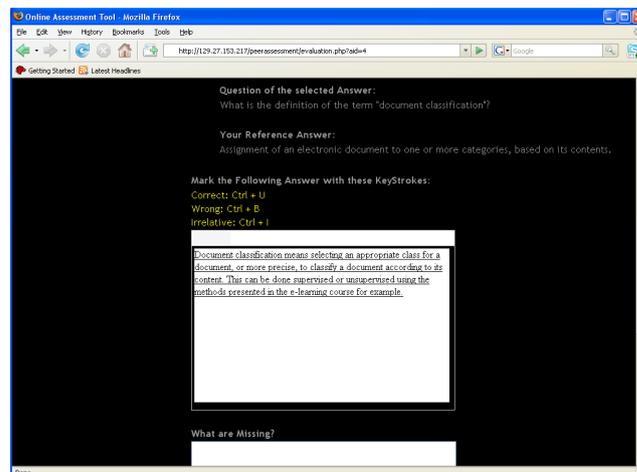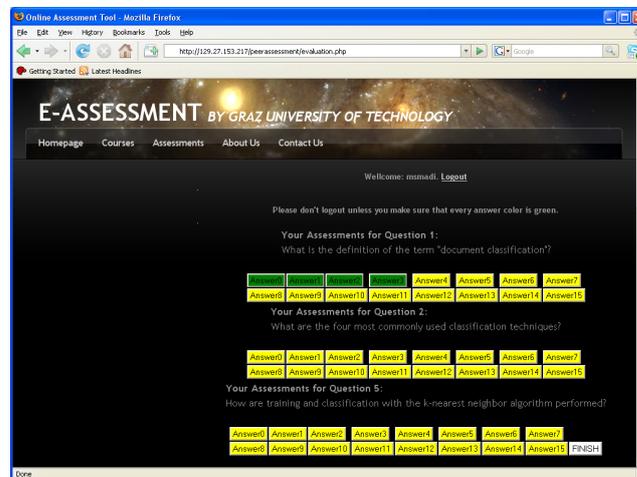

Fig 2. Peer-assessment using PASS.

IV. SYSTEM ARCHITECHTURE

Fig. 3 demonstrates the overall architecture of the Peer-ASsessment System (PASS), which 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 (TOMCAT) [26] and applies the MVC (Model-View-Control) approach [27]. PASS has been developed as part of a flexible e-assessment system project, for further information about this project you can

refer to AL-Smadi and Guetl (2008) [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. For this experiment the course teacher created a new test for an online course of ISR, and selected from a pool of questions a set of items related to this course. The test items were answered by the course teacher as reference answers and had been graded by him too to be compared with the grades given by the tutors and students in the experiment.

- Results Analysis & Feedback provision Module: this module computes the final grads 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.
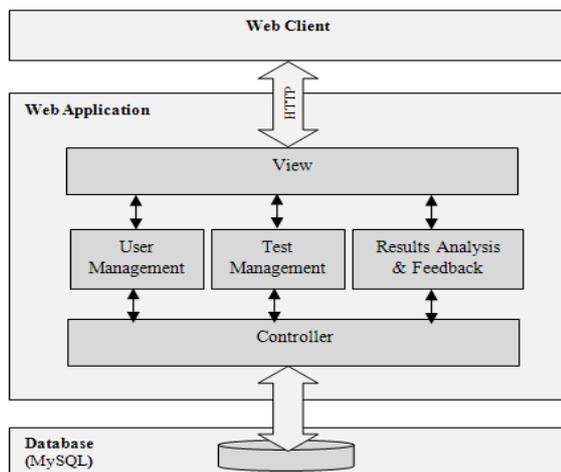

Fig. 3. The overall architecture of the Peer-ASsessment System (PASS).

V. RESULTS & FIRST FINDINGS

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.
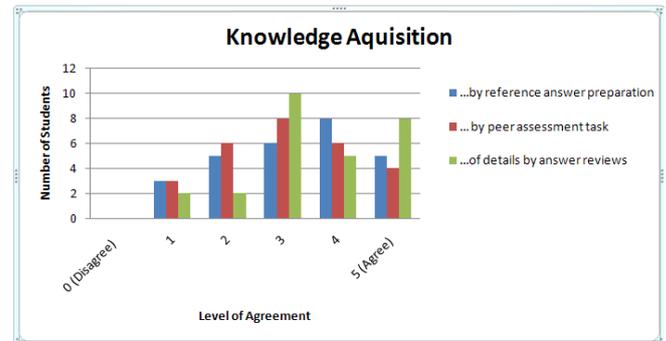

Fig. 4. Students' self estimation of knowledge acquisition in peer-assessment experiment.

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 complete disagreement and "5" represents complete agreement), where the knowledge gained from the online learning phase was with a mean value of 3.65 (σ = 1.05). According to the questionnaire, 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), where 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. 4, 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]; [30]; [31]).
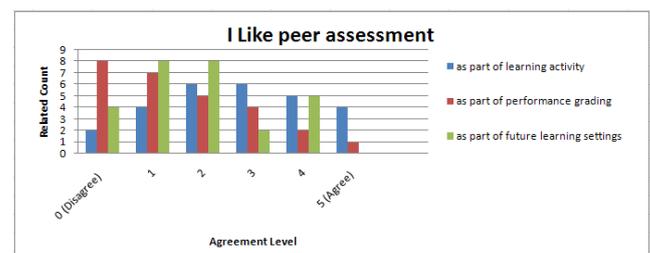

Fig. 5. Students' Impressions on peer-assessment in modern learning settings.

By analyzing the students' impressions on the peer-assessment as part of a modern learning settings and according to the questionnaire, students like peer-assessment as part of the learning activity with a mean value of 2.74 ($\sigma = 1.51$), where they recommend it to be part of the performance grading with a low mean value of 1.56 ($\sigma = 1.45$), or even as a part of the future learning settings with a mean value of 1.85 ($\sigma = 1.32$). Further Details are presented in fig. 5.
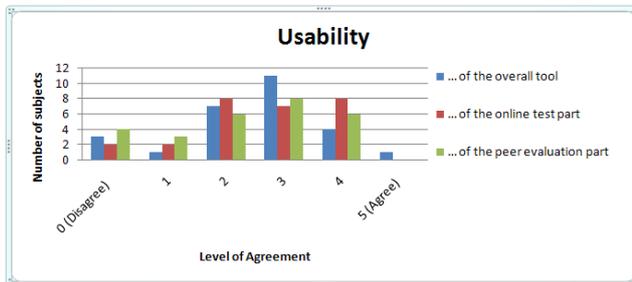

Fig. 6. The students' impressions on PASS usability.

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 ($\sigma = 1.25$), where their opinion about the online test phase was with a mean value of 2.63 ($\sigma = 1.21$) and their impression of the pear-assessment part was with a mean value of 2.33 ($\sigma = 1.36$). 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 ($\sigma = 28.31$), where the time for this part in the real experiment was 45 minutes. Fig. 6, 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 many 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, irrelevant).

VI. CONCLUSIONS & OUTLOOK

In this paper, we have presented an overall architecture for a peer-assessment system. A prototype for this system 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. The software has been designed with marking possibilities to facilitate the evaluation process of candidate answers. 27 students participated in the experiment, as well as 5 tutors participated in the fourth phase to evaluate the answers collected from the students.

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.

For future work, we are going to enhance the system with additional functionalities of feedback and explanations. Furthermore, we want to improve the system's reliability of assessment results. Deep analyses for the collected data such as comparisons between students' grades and tutors ones will be performed. Measuring student's self-directed assessment by analyzing the grades form the online test and reference answers preparation phases will be performed. Lastly, we will develop an improved web-based version of PASS with regards to the survey findings from tutors and students.

REFERENCES

[1]. F. J. Dochy; & L. McDowell, "Introduction. Assessment as a tool for learning". *Studies in Educational Evaluation, 23 (4)*, 279-298, 1997.

[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]. 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.

[5]. 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.

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

[7]. 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.

[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.

[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.

[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.

[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.

[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.

[30]. 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.

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