To appear on EdMedia Conference Proceedings 2014

A Web-Based Peer Interaction Framework for Improved Assessment and Supervision of Students

Michael Mogessie, Giuseppe Riccardi and Marco Ronchetti
Università degli Studi di Trento, Italy
micheal.mogessie@unitn.it, giuseppe.riccardi@unitn.it, marco.ronchetti@unitn.it

Abstract: One of the challenges of both traditional and contemporary instructional media in higher education is creating a sustainable teaching-learning environment that ensures continuous engagement of students and provides efficient means of assessing their performance. We present a peer-based framework designed to increase active participation of students in courses administered in both traditional and blended learning settings. Students are continuously engaged in attention-eliciting tasks and are assessed by their peers. The framework allows semi-automated assignment of tasks to students. In completing these tasks, students ask questions, answer questions from other students, evaluate the quality of question-answer pairs and rate answers provided by their peers. We have implemented this framework in several courses and run extensive experiments to assess the effectiveness of our approach. We discuss the results of students' surveys of this approach, which, in general, has been perceived as useful in achieving better learning outcomes.

Introduction

Since its early days, the web has shown an immense potential as an educational platform. In his 1997 publication, Owston (Owston 1997) noted that web-based learning could address the problems of higher education if utilised in a manner that would make learning more accessible, promote improved learning, and, in doing so, would at least contain the per unit costs of education.

Seventeen years on, we find that it is still too early to determine if the web is indeed the path to remedying the problems of the traditional classroom. Some, if not all, of the issues that were identified then have yet to be addressed by the online learning approaches of today.

Perhaps the most popular online education platforms today take the form of Massive Open Online Courses (MOOCs), which are open-access online courses that usually rely on video lectures to deliver course content. Although the earliest MOOCs were designed to provide connective online learning advantages (cMOOCs), the more dominant versions are the recent xMOOCs, which are rather an online manifestation of the transmissive instructional design that is deprecated in modern pedagogies (Downes 2008, Rodrigues 2013, Clarck 2013, Cusumano 2014).

The debate as to whether the MOOC model will succeed in fulfilling its promise to transform higher education has continued, with some suggesting an inevitable move towards a business model (Cusumano 2014).

The focus of our work, however, is not on assessing whether MOOCs are the future of online learning but on whether online learning can support learning models in general. We seek to provide an answer to the question 'Can the web promote improved learning?' In particular, we study if peer interaction, coupled with blended learning, can lead to better learning outcomes for students in higher education.

Debates about whether the type of instructional media used in the teaching-learning process plays any role whatsoever in promoting improved learning started long before the inception of the web (Owston 1997, Clark 1983, Clark 1994, Kozma 1991, Kozma 1994). The goal we pursue here is in accordance with the argument that an instruction medium or framework qualifies as a learning tool if it does much more than mere delivery of course content – the answer to the question 'Without it, is it still possible to acquire the required skills and knowledge?' should significantly tip the scale in favour of the medium or framework used to acquire the skills or knowledge in question (Owston 1997).

A recent study has shown that there is a positive relationship between augmenting traditional classroom environments with communication systems such as Electronic Voting Systems (EVS) and improved learning outcomes. Kennedy & Cutts (Kennedy & Cutts 2005) implemented an EVS for a computer science subject at the first year college level and found that there is an association between students' EVS responses throughout the semester and their performance at end- of-semester exams. They underscore that the sooner this association can be identified, the more effective the supervision provided to students will be.

EVS-enabled classrooms usually include sessions for students to engage in short discussions among themselves and with the teacher as part of the feedback process. In the more contemporary classroom settings where the class size is too large to carry out such sessions, the use of EVSs is not common. Moreover, such class sizes limit the degree to which teachers provide supervision to and assess each student.

Assessment and supervision of students in large class sizes is now a problem that is to be addressed from the perspective of automation. Advances in areas such as automatic scoring systems and intelligent tutors reflect this approach (Murray 2003, Leacock & Chodorow 2003, Mohler & Mihalcea 2009).

Researchers have also looked to address this problem by distributing the load on teachers over the class through the use of peer assessment techniques.

However, peer assessment techniques may have limitations. As soon as the task of assessing the performance of a student is delegated to other students, questions regarding the capabilities of the assessors may arise (Kaufman & Schunn 2011). A robust peer assessment solution should therefore have systems in place that factor such issues into the process of scoring in order to reflect the performance level of a student.

Yet another issue may arise in a peer assessment environment is favouritism, bias or collusive behaviour (Pond, Coates, Palermo 2007). If students somehow find a way to give each other positive assessments, for instance by forming factions that are sustained throughout the duration of the course, they could gain unfair positive assessments.

An issue that is not unique to peer assessment is plagiarism (Larkham & Manns 2002, Culwin & Lancaster 2001). In automated peer assessment systems, this may present itself as students submitting questions or answers that have been copied, verbatim or otherwise, from sources on the web (Austin & Brown 1999). A well-designed assessment solution should have the indispensable quality of detecting, if not preventing, such acts of academic dishonesty (Culwin & Lancaster 2001, Austin & Brown 1999, Carroll 2002).

Monitoring student performance and providing continuous supervision is also a challenging task to carry out in classes with large number of students. Teachers often find the demanding load of such classes dispiriting.

The traditional approach to remedying this problem has been for teachers to delegate teaching assistants to carry out laborious tasks such as correcting assignments. Assigning tutors to groups of students is also practiced in many educational institutions.

Depending on the teaching pedagogy in use, and the amount of individual supervision required, an increase in the number of enrolments for a course may have varying effects on learning outcomes and student satisfaction (DiBiase 2004). Consequently, the need for more efficient means of student assessment and supervision may arise.

An automated peer assessment and supervision framework could fit this context. Such a framework could incorporate features that encourage peer-interaction, and could provide the student with immediate feedback upon completion of assignments. Such feedback, for instance, may take the form of a performance level report of the student compared with the rest of the class. Automated analysis of this and other similar reports could help identify concepts the student may have difficulty grasping.

In this work, we present a semi-automated, web-based peer assessment framework that is driven by peer-interaction, with a potential to positively influence traditional and blended learning approaches.

This framework primarily focuses on formative assessment of students (Rowntree 1990). To test its validity, we have utilised the framework in courses with a relatively large number of students. In this paper, we describe the peer assessment framework, its implementation and the results of surveys aimed at validating our approach.

The rest of this paper is organized as follows. In the section that follows, we discuss the state of art in the areas of electronic and peer assessment. We then present the framework, a system we have implemented to test its validity and a report of the experiments and surveys we have conducted. We conclude our discussion by providing a brief summary and identifying work that we intend to carry out in the future.

Assessment

Assessment of learners in traditional face-to-face settings has been there for over two centuries, long enough to evolve into a successful method of measuring students' learning and providing some form of acknowledgment of their progress. Two major purposes of assessment that have been outlined by Rowntree (Rowntree 1990) are:

- Improving ongoing learning and providing support and feedback to learners, and
- Reporting the achievements of learners through an assessment method such as grading or written evaluations.

The first, which focuses on developing the learning of students, is termed as formative assessment, whereas the second, which measures what students have already achieved, is referred to as summative assessment. An assessment method that is composed of both formative and summative parts is termed as continuous assessment (Morgan & O'reilly 1999).

A continuous assessment method is usually structured in such a way that formative methods are used throughout the course, with a possible summative assessment at the end of the course (Morgan & O'reilly 1999).

Although such assessment methods prove successful in classes with a smaller number of students, their appeal quickly fades away as the class size increases. Eventually, instructors find themselves buried in the large bulk of assessment work, and struggling to dedicate much of their time to actual teaching-learning activities.

A number of solutions to the problem of efficiently assessing students of relatively larger class sizes exist today. The remainder of this section discusses some of these solutions.

E-Assessment Solutions

A more realistic approach to coping with large class sizes, e-assessment or online assessment refers to assessment methods that rely on information and communication technologies to automate the bulk of assessment tasks, thereby increasing efficiency and effectiveness of instructors (Joint Information Systems Committee 2007).

Apart from the invaluable decrease in instructor load, e-assessment has additional features virtually absent in traditional assessment methods. These include students' ability to receive instant feedback regarding their performance on a test or a quiz, high reliability and high levels of impartiality.

The large number of e-assessment tools made available since the topic gained the spotlight has led to the issue of interoperability of these systems. The Question and Test Interoperability (QTI) specification was hence established to provide a blueprint for designing and deploying questions and tests across the plethora of e-assessment tools that exist today (QTI, I 2006).

Although e-assessment solutions are mostly used in summative assessment, they could serve a formative purpose in increasing students' awareness of their own level of knowledge of a concept. When used in this manner, e-assessment tools serve as indicators of metacognitive level.

Peer-Based Assessment

Peer-based assessment methods have been widely used in education. Luxton-Reilly (Luxton-Reilly 2009) provides a detailed and systematic review of both generic and domain-specific peer assessment tools in use by educational institutions. Below is a brief description of those that we find most relevant.

PRAISE (Peer Review Assignments Increase Student Experience)

This is a generic peer assessment tool that has been used in the fields of computer science, accounting and nursing (de Raadt, Lai, Watson 2007). It has been used in introductory programming courses by students coming from different disciplines.

Before distributing assignments, the instructor will specify criteria. Once assignments are distributed, students review details of each assignment and submit their solutions. The system waits for the number of submissions to reach a specified number and assigns review tasks to students. Students then review the solutions of their peers according to the criteria. Once all reviews for a solution are complete, the system checks if all reviewers agree according to the criteria and suggests a mark based on the criteria. If there is a disagreement among reviewers, the system submits the solution to the instructor for moderation. The instructor then needs to decide a mark and confirm the release of the result before a student can see their overall mark for the assignment.

PeerWise

This is a peer assessment tool, which students use to create multiple-choice questions and answer those created by their peers (Denny, Hamer, Luxton-Reilly, Purchase 2008). When answering a question, students are also required to rate the quality of the question. They also have the option comment on the question. The author of a question may reply to a comment that has been submitted by the student who rated the question.

PeerScholar

This is another peer assessment tool that was initially designed for an undergraduate psychology class. It aims to improve writing and critical thinking skills of students (Paré & Joordens 2008). First, students submit essays. Next, they are required to anonymously assess the works of their peers, after which they have to assign scores between 1 and 10, and write a comment for each of their assessments. Students are also allowed to rate the reviews they have received.

As mentioned earlier, in order for peer-review to achieve its desired goal, it has to tackle well issues of academic dishonesty such as favouritism, bias or collusive behaviour (Kaufman & Schunn 2011, Pond & Coates & Palermo 2007). Such limitations may weaken the effectiveness of peer-review even further when the practice is intended for summative assessment. Peer-review instead becomes a more valuable tool if applied as an enabler for metacognitive reflection: we believe it should be used for formative rather than summative assessment.

Peer Instruction

Peer Instruction pursues this direction – it is a classroom setting in which instructors provide students with in-lecture multiple choice questions, after which the students vote for the correct answer using Electronic Voting Systems (EVS), also known as clickers. Students then discuss their answers with their peers before committing to an answer. The instructor then displays the correct answer and provides an explanation (Simon & Kohanfars & Lee & Tamayo & Cutts 2010, Fagen & Crouch & Mazur 2002).

Peer Instruction has been implemented in several courses across a wide range of institutions throughout the globe. Survey results reflect the consensus that having students discuss with their peers after providing answers to conceptual questions helps them understand the concepts better. This has been deduced from the observation that the percentage of correct answers increases in the second round students are asked to answer the same or another isomorphic question (Smith & Wood & Adams & Wieman & Knight & Guild & Su 2009, Brewster & Fager 2000).

Study has also shown that communication systems such as EVS may serve as indicators of expected student performance (Kennedy & Cutts 2005). This is an example of assessment that is neither purely summative – its goal is not to obtain a final evaluation of the student – nor purely formative – students do not directly learn from their errors. It is rather a diagnostic approach coinciding with the notion of early intervention – it aids in early discovery of challenges students may face in grasping concepts so that appropriate supervision is provided in a timely manner.

The Framework, the System and its Validation

The Peer-Interaction-Based Framework

Traditional classroom teaching models often struggle to create an environment that promotes student engagement and motivation (Brewster & Fager 2000). We therefore asked how we could engage students, and how we could elicit active involvement from remote students in the case of online or partly online courses. Our proposed approach was to motivate them to regularly interact with their peers by taking part in peer-assessed homework activities. Our hypothesis was that being actively involved in such activities would encourage them to keep up with the pace of the course and to study regularly in a deeper, more engaging manner.

To achieve our goals, we decided to involve students in a question-answer-evaluation loop that would include some game elements and in which they would be the lead actors.

Every week for the duration the course, students are required to ask a question regarding a topic covered during the most recent lectures. Questions may vary from typical assessment queries to requests for clarification and enquiries that would require deeper insight into a theme. When submitting their question, students are required to tag their question using at least two keywords.

Next, each question is assigned to five students at random. Students then rate the quality of the questions they have been assigned on a scale of 1 to 5 across three dimensions – *interestingness*, *difficulty level* and *relevance*.

After the question evaluation phase completes, the teacher selects a subset of the questions. In doing so, the teacher has the option to perform basic filtering using the provided tags as well as simple string matching in order to identify similar questions. Moreover, the teacher may use the average quality of the questions, as perceived by the students and computed from the values of the aforementioned dimensions, to select questions with a specific level of quality.

At present, there is no restriction on how this average is computed. This gives the teacher the freedom to experiment with different formulas to vary the levels of quality of the selected questions.

This approach may help assess the strength of the classroom within the first few weeks of the course. Further analysis may also provide insight into the performance distribution of students in the classroom.

The number of selected questions often falls in the range between one-third and one-fourth of the total number of submitted questions. Selecting such fractions ensures that, in the next phase of prompting students to answer a question, each question will be assigned to three or four students. Each student is then randomly assigned one of the selected questions and asked to provide an answer.

When the answer submission phase completes, every student is assigned a question with a set of three to four answers that were provided by their peers in the previous task. They are asked to vote for the best response. If they believe there is no best response to the question, they have the option to indicate so. They can also comment on the question-answer set, or provide justification if they believe there is no best response. It is worth mentioning here that the entire set of processes is anonymous.

A score is then calculated based on the votes cast by the students. At present, it is simply the total number of votes a student has earned for their response. We acknowledge that it could be based on more sophisticated scoring algorithms. The game aspect of this approach is that students can compete for a classroom-wide ranking. To make the competition more rewarding, the teacher may decide to provide some form of incentive. For example, the top third of students at the end of the course could receive bonus points.

At the end of each cycle, all questions and answers, together with the votes for the answers, are published for students to review.

The System

To implement this framework, we built a web-based system composed of three major components – a task-based peer assessment and data collection component, a basic answer-scoring component, and a basic progress-tracking and feedback-provision component. The system includes a module that levers the authentication infrastructure of the University of Trento (UNITN) to identify students enrolled in the course.

The system allows assigning students a set of weekly tasks. Students are required to complete the first task – asking a question – in a relatively short time, two days at most.

At present, the instructor selects the subset of questions to be used in later tasks. This is a bottleneck that makes it difficult for the system to scale. However, it is done in this manner because in the first experiments we preferred to maintain a certain degree of control. In subsequent experiments, we plan to migrate towards a semi-automated, scalable approach in which natural language processing and machine learning techniques will aid the instructor in selecting questions in a much efficient manner.

The system handles the task of randomly assigning the selected questions to students. In doing so, it guarantees that a student will not be assigned the question they have submitted, and does its best to balance the distribution so that all questions are assigned to the same number of students.

After the question-answering phase completes, the system assimilates questions and answers to formulate multiple-choice questions and randomly assigns them to all students.

The system has a progress-reporting feature that presents each student with basic statistics regarding their activity. Information such as the number of tasks completions, the number of questions from the student that have been selected by the instructor, the number of votes the student has earned and the current highest number of votes earned in the class is reported in the homepage of the student. From such report, the system is able to identify students who are less active and may require close supervision, with little effort.

Our preliminary assessment of the data generated by student activities shows that it is possible to assess the progress level of each student. The size of this important data shows significant growth as the course advances. It is only logical that one should consider applying advanced data analysis techniques to obtain information that would be invaluable in tracking the progress of students.

The entire cycle of asking and answering questions, and evaluating students' answers takes one week to complete and repeats over the 12 weeks that constitute a UNITN semester. All student-student interactions remain anonymous but comprehensive data about the activity of each student is stored in our database, using which we perform detailed statistical analysis.

From a pedagogical point of view, we believe that this approach encourages students to stay engaged and in sync with the course, and to perform further investigation using additional sources such as the Internet.

The Experiment and Results of the Survey

For the purpose of this study, we used the system in three courses, two of which also provided students with video-lectures. Students of those two courses had the option to follow the course in-class or remotely, through the video-lectures over the Internet. Lectures were recorded and published with the LODE system (Ronchetti 2012). To avoid mixing effects, these two courses did not use a flipped-classroom methodology. All three courses were offered at the bachelor and master levels of computer science at UNITN. They were Informatica Generale I (IG), a first year bachelor course, Programmazione Android (PA), a third year bachelor course, and Web Architectures (WA), a course for first year master students.

(Tab. 1) provides a concise view of the data we have collected. #Stud is the number of students, #Q is the total number of questions, #SQ is the number of selected questions, #A is the number of collected answers, and #V is the number of votes.

COURSE	#STUD.	#Q	#SQ	#A	#V
IG	222	870	286	857	846
PA	120	948	224	578	768
WA	40	147	32	139	149

Table 1: Enrolment and data generated for each course

At the end of the courses, we asked students to complete a questionnaire with 13 five-point Likert items, 1 multiple-choice, multiple-answer question and 3 open-ended questions. More than half of the students responded – 124 out of 222 (55%) for IG, 80 out of 120 (66%) for PA, and 23 out of 40 (57%) for WA. Considering the size of the statistical samples and under the assumptions of a simple random sampling of the population, a sample proportion of 50%, and a finite population, the respective margins of error at 95% confidence level are 6%, 6% and 13%.

While 80% of the students of both bachelor courses attended the lectures regularly, only 66% of the graduate students were regular attendees. 20% of the graduate students followed the course online, using the lecture-videos.

In the remainder of this section, we will summarise the results of our study. Responses with values of 4 and 5 are deemed positive whereas those with values of 1 and 2 are considered negative. The "neutral" percentage of value 3 is the difference between the sums of the other two percentages and 100%. Only positive and negative responses have been included in the tables that follow.

A subset of the questions focused on acceptance and perceived effectiveness of the peer assessment framework. We asked students if they had found the framework useful, if the pace of the course had pushed them to work harder, if they had been compelled to study deeper and if, as a result, they had become more attentive in class.

The main result emerging from the analysis of the responses is that students believed the framework we implemented was in fact useful. Interestingly, there is an apparent correlation between the 'maturity' of the students and the perceived usefulness of the peer assessment framework – its appraisal grows when moving from first year students (IG) to graduate students (WA), with third year undergraduates (PA) in between. The following tables show the results. The margins of error are indicated with a ± sign for each percentage value.

	IG	***PA***	***WA***
YES	45% ± 6%	57% ± 6%	87% ± 9%
NO	18% ± 5%	37% ± 6%	4% ± 5%

Table 2: "Asking questions and responding to them was useful"

	IG	***PA***	***WA***
YES	28% ± 5%	40% ± 6%	65% ± 13%
NO	32% ± 5%	31% ± 6%	9% ± 8%

Table 3: "Using the peer assessment framework increased my attention in class"

	IG	***PA***	***WA***
YES	39% ± 6%	49% ± 6%	83% ± 10%
NO	30% ± 5%	27% ± 6%	4% ± 5%

Table 4: "The peer assessment framework pushed me to follow the course more regularly than I would have"

	IG	*PA*	*WA*
YES	43% ± 6%	49% ± 6%	69% ± 12%
NO	24% ± 5%	25% ± 6%	9% ± 8%

Table 5: "The peer assessment framework pushed me to be study deeper"

	IG	*PA*	*WA*
YES	48% ± 6%	39% ± 6%	78% ± 11%
NO	22% ± 5%	30% ± 6%	4% ± 5%

Table 6: "The peer assessment framework helped me prepare better for the exam"

For all three courses, students who believed the method was useful vastly outnumber those who believed otherwise. In particular, there is a perception that the usefulness was demonstrated in the drive to study deeper, which led to better preparation for the exam. Also, they followed the course more regularly than they would have otherwise. Here, however, the signal is much weaker for first year students.

Overall, it appears that there was only a marginal degree of increase in attentiveness in class for undergraduates, where the difference between "YES" and "NO" is less than 10 percentage points. The opposite is true for graduate students, where "YES" wins with a gap of over 50 percentage points. For the item summarised in (Tab. 9), "answering questions was easy", the negative response percentage point is almost three times as much as the positive response percentage point. This, however, is not directly attributed to the approach itself. In fact, it may support the claim that students at the graduate level tend to ask more profound questions. For all other Likert items, positive response percentage points are higher for graduate students.

Another part of the questionnaire focused on the "how". We tried to understand if the provided course materials were helpful and to what extent externally available material was used in completing question-answering tasks. We also asked if answering questions was a challenging task, and if the peer assessment framework imposed a work overload. The results are reported in the following tables.

	IG	*PA*	*WA*
YES	70% ± 5%	60% ± 6%	91% ± 8%
NO	14% ± 4%	10% ± 4%	9% ± 8%

Table 7: "The study material available on the course web site was useful to respond to the questions"

	IG	*PA*	*WA*
YES	74% ± 5%	95% ± 3%	100%
NO	13% ± 4%	0%	0%

Table 8: "Material found on the Internet was useful to respond to the questions"

	IG	*PA*	*WA*
YES	35% ± 6%	56% ± 6%	13% ± 9%
NO	22% ± 5%	14% ± 4%	35% ± 13%

Table 9: "Answering questions was easy"

	IG	*PA*	*WA*
YES	57% ± 6%	71% ± 6%	78% ± 11%
NO	14% ± 4%	10% ± 4%	13% ± 9%

Table 10: "The amount of work required by the peer assessment framework was adequate"

	PA	*WA*
To follow the course remotely	20% ± 5%	27% ± 12%
To recover classroom lessons that I missed	41% ± 6%	68% ± 13%
To review some concepts I had not understood well	58% ± 6%	73% ± 12%
To find questions to ask	28% ± 6%	18% ± 10%
To respond to questions	34% ± 6%	36% ± 13%
To evaluate responses	11% ± 4%	18% ± 10%
I never used the lecture videos	22% ± 5%	14% ± 9%

Table 11: "What did you use the videos for?"

In all cases, students used resources provided for the course but also acquired additional information from other sources, with the latter more often the case. We consider this one of the advantages of the framework, as students do not solely rely on the relatively small subset of resources provided by the teacher, but venture into other realms to find answers and clarifications and to learn more. Again, this trend becomes more pronounced as students mature.

As stated earlier, responding to questions was relatively easier for the undergraduate students. This most likely results from the ability of graduate students to ask more original and less trivial questions.

We were sceptical about the amount of workload imposed by our approach but were later reassured as the majority of students believed it was reasonable.

Finally, we asked students about how they used the video-lectures. Because video-lectures were not available for the first year undergraduate course IG, only the results for the other two courses are reported in (Tab.11).

The large majority of students used videos – mostly to review specific topics. The use of videos in connection with completion of the tasks was also significant. Approximately one-third of the students used the videos in responding to questions. A small percentage also used the videos to find clues about what questions to ask. Yet a smaller percentage also used them in evaluating responses.

Discussion and Conclusions

The results show that students welcomed the approach, and that there were measurable benefits in terms of perceived quality of the proposed framework and of better involvement of the students in the course. Also, this approach pushed students to search learning material outside the course "walls". There appears to be a positive correlation between student maturity and the effectiveness of the peer assessment framework.

Despite a few minor problems, the system we built performed well and allowed us to gather valuable data, and create corpora of large sets of questions, answers and evaluations.

As has been mentioned when describing the system, at present, the instructor carries out the selection of questions. Work is already underway to implement a semi-automated question selection feature that utilises natural language processing and machine learning techniques to group and rank similar questions. This will enable the instructor to select and assign a subset of the questions in a much scalable and efficient manner.

An interesting path we plan to pursue is identifying whether there are trends or patterns in the ways students perform throughout a course. We intend to utilise machine-learning techniques to study historical student activity data to determine if a student's performance level can be predicted at different stages of the course.

We acknowledge that some of the limitations of other peer-assessment techniques may not be absent in our approach. Plagiarism, for example, could potentially undermine the effectiveness of our framework. Nonetheless, considerable advance has been made in detecting anti-plagiarism (Kakkonen 2010). We therefore intend to employ state-of-the-art techniques to detect such academic dishonesty. Regarding the issue of collusive behaviour, we believe that the large number of students, coupled with anonymity and random task assignments, can serve as a deterrent to such dishonest behaviour.

We are currently testing the validity of our framework in another undergraduate level computer science course. In the semesters that follow, we intend to extend our experiments with the framework to courses from other departments and disciplines within UNITN.

References

Austin, M. J., & Brown, L. D. (1999). Internet plagiarism: Developing strategies to curb student academic dishonesty. *The Internet and higher education, 2*(1), 21-33.

Brewster, C., & Fager, J. (2000). Increasing student engagement and motivation: From time-on-task to homework. *Portland, OR: Northwest Regional Educational Laboratory*.

Carroll, J. (2002). A handbook for deterring plagiarism in higher education. *Oxford Centre for Staff and Learning Development*.

Clark, D. (2013). From cMOOCs to xMOOCs and why the difference matters. *Learning Shrew blog*. Retrieved April, 29, 2013.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of educational research, 53*(4), 445-459.

Clark, R. E. (1994). Media will never influence learning. *Educational technology research and development, 42*(2), 21-29.

Culwin, F., & Lancaster, T. (2001). Plagiarism issues for higher education. *Vine, 31*(2), 36-41.

Cusumano, M. A. (2014). MOOCs revisited, with some policy suggestions. *Communications of the ACM, 57*(4), 24-26.

de Raadt, M., Lai, D., & Watson, R. (2007, November). An evaluation of electronic individual peer assessment in an introductory programming course. *In Proceedings of the Seventh Baltic Sea Conference on Computing Education Research-Volume 88* (pp. 53-64). Australian Computer Society, Inc..

Denny, P., Hamer, J., Luxton-Reilly, A., & Purchase, H. (2008, September). PeerWise: students sharing their multiple choice questions. *In Proceedings of the Fourth international Workshop on Computing Education Research* (pp. 51-58). ACM.

DiBiase, D. (2004). The impact of increasing enrollment on faculty workload and student satisfaction over time. *Journal of Asynchronous Learning Networks, 8*(2), 45-60.

Downes, S. (2008). MOOC and mookies: The connectivism & connective knowledge online course. *eFest by Ellumninate, Aukland, New Zealand. Available online at: http://www. downes. ca/presentation/197*.

Fagen, A. P., Crouch, C. H., & Mazur, E. (2002). Peer instruction: Results from a range of classrooms. *The Physics Teacher, 40*(4), 206-209.

Joint Information Systems Committee. (2007). Effective Practice with e-Assessment: An overview of technologies, policies and practice in further and higher education. *Joint Information Systems Committee*.

Kakkonen, T., & Mozgovoy, M. (2010). Hermetic and web plagiarism detection systems for student essays—an evaluation of the state-of-the-art. *Journal of Educational Computing Research, 42*(2), 135-159.

Kaufman, J. H., & Schunn, C. D. (2011). Students' perceptions about peer assessment for writing: their origin and impact on revision work. *Instructional Science, 39*(3), 387-406.

Kennedy, G. E., & Cutts, Q. I. (2005). The association between students' use of an electronic voting system and their learning outcomes. *Journal of Computer Assisted Learning*, 21(4), 260-268.

Kozma, R. B. (1991). Learning with media. *Review of educational research, 61*(2), 179-211.

Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational technology research and development, 42*(2), 7-19.

Larkham, P. J., & Manns, S. (2002). Plagiarism and its treatment in higher education. *Journal of Further and Higher Education, 26*(4), 339-349.

Leacock, C., & Chodorow, M. (2003). C-rater: Automated scoring of short-answer questions. *Computers and the Humanities, 37*(4), 389-405.

Luxton-Reilly, A. (2009). A systematic review of tools that support peer assessment. *Computer Science Education, 19*(4), 209-232.

Mohler, M., & Mihalcea, R. (2009, March). Text-to-text semantic similarity for automatic short answer grading. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 567-575). Association for Computational Linguistics.

Morgan, C., & O'reilly, M. (1999). *Assessing open and distance learners*. Psychology Press.
Murray, T. (2003). *An Overview of Intelligent Tutoring System Authoring Tools: Updated analysis of the state of the art*. In Authoring tools for advanced technology learning environments (pp. 491-544). Springer Netherlands.

Owston, R. D. (1997). The World Wide Web: A technology to enhance teaching and learning?. *Educational researcher*, 27-33.

Paré, D. E., & Joordens, S. (2008). Peering into large lectures: examining peer and expert mark agreement using peerScholar, an online peer assessment tool. *Journal of Computer Assisted Learning, 24*(6), 526-540.

Pond, K., Coates, D., & Palermo, O. A. (2007). Student experiences of peer review marking of team projects.

QTI, I. (2006). Question and test interoperability.

Rodriguez, O. (2013). The concept of openness behind c and x-MOOCs (Massive Open Online Courses). *Open Praxis, 5*(1), 67-73.
Ronchetti, M. (2012, September). LODE: Interactive demonstration of an open source system for authoring video-lectures. *In Interactive Collaborative Learning (ICL)*, 2012 15th International Conference on (pp. 1-5). IEEE.

Rowntree, D. (1990). Teaching through self-instruction: How to develop open learning materials. *London: Kogan Page*.

Simon, B., Kohanfars, M., Lee, J., Tamayo, K., & Cutts, Q. (2010, March). Experience report: Peer instruction in introductory computing. *In Proceedings of the 41st ACM technical symposium on Computer science education* (pp. 341-345). ACM.

Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science, 323*(5910), 122-124.