



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky



PaedDr. Veronika Dropčová

Autoreferát dizertačnej práce

Peer Review and Peer Assessment in Higher Computer Science Education

na získanie akademického titulu *philosophiae doctor*

v odbore doktorandského štúdia: 9.2.3 Teória vyučovania informatiky

Bratislava, 2016

Dizertačná práca bola vypracovaná v dennej forme doktorandského štúdia na Katedre základov a vyučovania informatiky Fakulty matematiky, fyziky a informatiky, Univerzity Komenského v Bratislave.

Predkladateľ: PaedDr. Veronika Dropčová
Katedra základov vyučovania informatiky
Fakulta matematiky, fyziky a informatiky
Univerzita Komenského
Mlynská dolina
842 48 Bratislava

Školiteľ: doc. RNDr. Zuzana Kubincová, PhD.
Katedra základov vyučovania informatiky
Fakulta matematiky, fyziky a informatiky
Univerzita Komenského

Oponenti:

Obhajoba dizertačnej práce sa koná o h
pred komisiou pre obhajobu dizertačnej práce v odbore doktorandského štúdia vymenovanou
predsedom odborovej komisie prof. RNDr. Ivanom Kalašom, PhD.

vo vednom odbore **9.2.3 Teória vyučovania informatiky**

Predseda odborovej komisie:

prof. RNDr. Ivan Kalaš, PhD.
Katedra základov a vyučovania informatiky
Fakulta matematiky, fyziky a informatiky
Univerzita Komenského
Mlynská dolina
842 48 Bratislava

Introduction

Current students' requirements and needs, corresponding to the way they communicate and learn, are different from those students called for 10, 15 or 20 years ago. Young people use web for building vast social networks, for finding and using digital information that are much more complex and flexible than information available in paper world [Ric10] any time before or now.

Nonetheless, differences arising from new technologies, are evident also in other spheres of human life.

Business companies need people, who are not only experts in a particular profession but who also possess different skills, such as to be creative, to be able to give constructive feedback, to learn new things in a short time, etc. (so-called *soft skills*). For this sake, universities are supposed to educate their students in a way that they acquire also these skills sought after by employers. In our opinion, this is a time for introducing Web 2.0 tools in the educational process. To add a new post to the *wikipedia* is not perceived as a difficult action, but to add a *relevant* post that will not be erased by administrators is much more arduous task. Thus this is required to be taught: *working with relevant information*. To write a new blog post is also relatively short procedure, but to build up a huge blog audience needs stamina and right attitude. Hence this is also what needs to be taught: *to produce attractive and interesting work*. Therefore the question is: Why not to teach it with powerful tools such as those offered by Web 2.0?

The main objective of this thesis was to design, implement and validate a peer review or a peer assessment concept as a part of overall students assessment in higher computer science course. Students – participants of this research, enrolled in a Web-design course, were trained in improving of peculiar skills mentioned before with the support of two powerful tools – blogs and peer reviewing. At the beginning, the course project was to implement each student's own blog and also to publish a reasonable number of posts related to the course topic. However, this activity mainly improves student's writing skills. In order to develop her competences in constructive feedback and criticism, each student was asked to review several posts of her colleagues. Later, peer reviewing of blog posts was replaced by peer reviewing of projects. In the last phase of the research, projects were supplied with team-based projects and besides peer review also peer assessment was introduced. Whole peer review process was implemented as a part of the learning management system *courses.matfyz.sk* developed by our students within their bachelor or master theses.

Higher education

In 1987, A. W. Chickering and Z. F. Gamson, came with great, timeless article where seven principles for good practice in undergraduate education are introduced [CG87]. These principles are as follows:

1. *encouraging contact between students and faculty*

As they proclaimed, frequent contact between student and faculty members is *the most important factor* in student motivation and involvement. If student knows a few faculty members and is in contact with them, it can be easier for her or him to stay motivated, keep on working, think about future plans.

2. *developing reciprocity and cooperation among students*

They reminded that good learning is collaborative and social, not competitive and isolated. By sharing one's own ideas and reflecting to others' ideas and their reactions students are trained in thinking and better understanding.

3. *encouraging active learning*

Learning is not a spectator sport – students must talk and write about what they are learning, find relations with past experiences and be able to apply it to their daily lives.

4. *giving prompt feedback*

Students need help in assessing existing knowledge and competence, appropriate feedback on their work and chances to reflect on what they have and have not learned, what they still need to know.

5. *emphasizing time on task*

Students and also faculties need plan of how to allocate their time to reach effective learning.

6. *communicating high expectations*

This principle highlights the necessity of demanding high expectations for everyone – from those, who are bright and well motivated to those who are poorly prepared.

7. *respecting diverse talents and ways of learning*

Students need an opportunity to show how talented they are and in which way and use these talents in their learning.

A few years later, in 1996, A. W. Chickering and S. C. Ehrmann continued in Chickering's previous work and brought ideas how to use technology in harmony with mentioned seven principles [CE96]. According to their ideas, technologies can strengthen interaction between faculty and students and can help shy students to ask questions. Many faculties report that when students know that their finished work will be published on the web, they feel more stimulated and motivated. Moreover, publishing on the web provides a basis for peer evaluation and by this getting more feedback than without technologies.

Blogs in education

Blogs can be used in educational process within different subjects. Blogging as an activity is useful in individual organization of knowledge as well as in receiving feedback and thus comparing different views of things [HK09b]. Since students can also reflect on their own or on their colleagues' experience, blogging is well connected with the theory of constructionism [Rov02]. Blogging can improve critical and analytical thinking and force a blogger to think about how will her topic interpretation be seen by readers [Ora02].

Integration of blogs in programming lectures provides various benefits, as well. They can be used as learning logs with the aim to improve critical reflection of students' work. Blogs can also help in better understanding of topics of the course [Saf08]. Moreover, prospective employers and thesis advisers believe that students need to practice their writing skills [Geh03] and blogs can be an appropriate place for it.

Peer assessment

Peer assessment can be described with following definition: *it is any process where a 'mark', supposedly being some reflection of quality, is awarded by student for some piece of work undertaken by other students.* This definition is based on theories of active learning \cite{piaget1970} and andragogy \cite{cross1981}, but also in social constructionism \cite{vygotsky1964}. It can be described as a form of an assessment developing skills apparent for the highest level of Bloom's taxonomy (the Evaluation level) \cite{bloom1956,reily2009}. According to Topping (\cite{topping1998}), peer assessment is defined as *an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status.* Therefore students engage in reflective criticism of products of their peers and provide them with feedback based on previously defined criteria \cite{falchikov1986}.

Students are required to provide grades and/or reviews to their peers on a product of their work or a performance. They are usually provided with help and support from teachers at the start of using peer assessment techniques but it is reduced over time. There are three standard ways how are criteria for this kind of assessment created: criteria are designed by teacher and students have to accept them; criteria are a product of teacher's and students' agreement or criteria are identified only by students \cite{falchikov2007}.

To establish a universal model of peer assessment is likely difficult - or even impossible task \cite{topping1998}. However, Falchikov and Goldfinch provided a study about peer assessment in higher education where they compared peer and teacher marks in 100 studies \cite{falchikov2000}. Based on this research they compiled a set of seven recommendations to practitioners for implementing peer assessment as follows:

1. if is it possible, do not use a very large number of peers in one assessment group
2. involve students in peer assessment of academic products and processes
3. do not create many individual categories to be assessed by students; it is better to use an overall evaluation with well understood criteria
4. give students an opportunity to become a part of discussions about criteria
5. the design, implementation and reporting of your study has to be well-described
6. use peer assessment in any discipline area in all educational levels
7. a measure of validity should not be built on the proportions of agreement between peers and teachers

Peer assessment in Team-based learning

R. E. Levine emphasized that *students need peer review to feel comfortable that their teammates are contributing their fair share of the group work.* He described five various methods that can be used in peer assessment applied in team-based learning \cite{levine2008}:

1. *Michaelsen method.* Students are expected to assign other members of their teams a score based on their opinion about how much did they contribute to the overall team performance. If there is a six-person team, each member will be supposed to divide 50 points among her teammates (she does not evaluate herself) and usually also to give each of them a qualitative feedback. A final score for each student is then calculated as a sum of received points. However, students are required to discriminate among their teammates. This rule can be very controversial, especially in high-functioning teams where it may lead to gaming the system: students make an arrangement about dividing the points, so in the end each member of a team has the same amount of points. But as Michaelsen pointed out, if gaming the system

is done at the end of a term, it means that all members will probably deserve to have the same score (unlike the situation when it is done at the beginning and after that some students will “loaf”).

2. *Fink method.* Students are asked to divide 100 points among their teammates according to degree each of them contributed to the team work. After that the overall student’s score for the assignment is calculated by summing the scores she got from her peers, dividing it by 100 and multiplying it with a group score for this assignment assigned by a teacher. Learners are also usually asked to provide written peer feedback. In this method, those students whose performance was better than performance of other team members, will get score over 100% and those who were not so active will get score less than 100% of a team score. In contrast with previous method, students do not have to discriminate among their peers. This fact might result in higher students’ satisfaction and their responsible attitude.
3. *Combination of Michaelsen and Fink methods.* In some cases a combination of previous two methods can be used: students assign points to their teammates (number of team members minus one multiplied by ten), but the resulting score is then turned into a percentage. Next step is similar to the Fink method - this score is used as a percent multiplier to modify the student’s final score.
4. *Koles method.* This method is mostly oriented on quantitative feedback. Students are rated in three areas (cooperative learning skills, self-directed learning and interpersonal skills) divided into nine questions where reviewers choose a value to evaluate peer’s skills from following ones: never, sometimes, often, always. Typical skills appearing in this evaluation are: *Asks useful or probing questions* or *Shows care and concern for others*. On the other hand, students are also asked to qualitatively evaluate their teammates by answering the following questions: *What is the single most valuable contribution this person makes to your team?* and *What is the single most important thing this person could do to more effectively help your team?*. The qualitative feedback tends to be more useful than the quantitative one since students are not required to distinguish between the work of their teammates and usually rate them with excessively high numbers.
5. *Texas Tech method.* This method uses mostly quantitative feedback, therefore it is similar to the previous one. Students rate their peers in twelve categories (such as responsibility, humility, preparation for learning activities) on a 5-point scale, where 1 is too little, 5 is too much and 3 is the ideal score. In case a student rates her teammate with the lowest or the highest number, she also has to add a written feedback, otherwise feedback is optional. However, similarly than in previous methods - if students are not required to differentiate between their colleagues, they will evaluate them unfairly positively.

Research problem and the statement of the problem

Blogging has a well recognized potential as a learning activity \cite{mosel2005}. As confirmed by a number of publications (e.g. \cite{mosel2005,razmerita2009,owen2006,homola2009}), together with other Web 2.0 tools it is well aligned with the essential features of educational theories such as social constructivism \cite{von1995}, constructionism \cite{papert1991}, and connectivism \cite{Siemens2005} that perceive learning as a social process in which the knowledge is constructed in cooperation with others using the network technologies. These novel approaches in learning foster creative and collaborative work focused on construction of new knowledge, and they have potential to make the learning process more absorbing and even amusing for the students \cite{popescu2010}.

Encouraged by these findings, we have started to integrate blogging assignments into university courses in 2006 \cite{kubincova2009}. All students’ blogs were integrated into a one portal *blog.matfyz.sk* and every student had an access to it. Topics of blog posts had to be related to the course topics. Students could add comments under the blog articles and by this give feedback each other immediately. Despite initial positive experience connected to various creative realization of the assignment, competitiveness boosting, etc. we realized that combining blogging with organized curricula and integrating it into course work is not a trivial task. Students are not always willing to undergo additional assignments just “for free” without any evaluation reward, while other students may object including such non-traditional assignments into their evaluation. In addition, involving students in active follow up and commenting on their colleagues blogs, which is an integral part of this activity, may be even harder task to accomplish. In our experience ill-organized blogging assignments may easily turn into write-only activity, negating most of their goals.

Aiming to overcome these issues, we have resorted to peer reviewing process, which was found beneficial by other researches in the field \cite{liu2001,lin2001,wu2006,sterbini2012,popescu2011}.

The purpose of this thesis is to describe an assessment strategy design that includes peer review and peer assessment techniques and Web 2.0 tools implemented for one specific computer science course at our faculty -

Web-design course. We have decided for this according to several reasons which can be split into general and specific ones.

In the group of general reasons, there is a need to improve writing skills of students because it seems to be a critical ability not only for writing final thesis, but also for their professional lives (even programmers need to be able to produce some kind of papers or manuals).

Another general reason for choosing our topic is developing of student's competence to provide constructive feedback to her peers since this is useful for both, personal and professional life of each individual.

The last general reason for involving peer review or peer assessment in computer science education at university is a chance for students to get more specific and accurate feedback from their peers. Unlike teachers who have to check and assess every student, students in their peer assessment assignments usually assess only a small group of their peers. Therefore they have more time to write more detailed recommendations for their peers how to process in their work. Moreover, according to our experience, many students understand and accept the feedback from their peers easier than that from the teacher.

Specific reasons are related to our Web-design course since blogging has been its part since 2006. As we mentioned above, although first experience with blogging in our practice were positive, some difficulties have appeared in the next years. First of all - only a marginal part of students engaged and this number continuously decreased next years. According to this fact we decided to find a model that uses blogging assignments and is attractive to students, as well.

Another specific reason for choosing this topic is students' satisfaction with the course. Students tend to criticize non-traditional assignments although they are informed about and provided with positives and advantages that rose from these assignments. This fact often influences students' reactions and consequently negatively affects their evaluation of the course. Therefore, this topic was chosen also with the aim to face the students' complaints related to the course methods.

In conclusion - the research problem of this thesis is peer assessment methods in computer science course in higher education.

Research objectives

The main objective of our research is *to design, implement and validate a peer review or peer assessment concept as a part of overall student assessment in higher computer science course.*

Specific details of the main objective are as follows:

- to define various types of assignments where peer review or peer assessment techniques can be used in a way students are willing to accept;
- to design strategies for the implementation of peer review or peer assessment in these assignments;
- to describe influence of peer review or peer assessment on learning outcomes.

Research questions

In the research design we have decided for, research questions are suggested instead of research hypotheses. Thus the ambition of our research is to answer the following question:

How can an appropriate educational strategy that includes peer review or peer assessment techniques be used in our Web-design course?

This question is divided into several specific items:

- Which of the course assignments are suitable for involving peer review or peer assessment?
- How should students be motivated to engage in peer review or peer assessment activities?
- How are students' learning outcomes affected by their engagement in peer review or peer assessment activities?

Research Timetable

Whole research consists of four core phases:

1. Literature Review

According to the topic of our research we steadily studied numerous literature sources dealing with three basic points. Interest in *the higher education pedagogy* was related to the environment where this research was conducted - the university course. Since the purpose of our research was to join certain types of the *Web 2.0 tools* and the *peer review or peer assessment techniques* in computer science course, studying papers and studies relevant to these two topics was also necessary.

2. 1st round of peer review in Web-design course.

In this phase we used the results of a pilot study conducted on the course Web-design in 2012 (see \cite{diplomovka}, \cite{DIDINFO2013, ICTE2013, ICWL2013}). From this study, where peer review was combined with blog assignments, we learned that peer review was attractive to students, they achieved better grades and exactly defined deadlines helped them to stay engaged in during the whole semester. However, they were not satisfied with conditions for blogging. Based on these findings we decided to change content requirements for blog posts and to give students the possibility to participate less, but still keeping the chance for good grading (fall 2013). Moreover, we tried to define the rules how to find a group of students that can be considered as reliable peer reviewers. This phase was results were presented on several conferences (\cite{DIDINFO2014, ICWL2014, ICALT2014, ICTE2014}).

3. 2nd round of peer review in Web-design course.

After the 1st round of peer review in blogging assignments we had three important results:

- students' engagement was much higher than in years when peer review was not a part of the course;
- students' grades were better than in years without peer review assignment;
- students were still unsatisfied with the course requirements.

Based on this, we decided to move peer review from blog assignment to the project assignment (fall 2014). We had also added a double-check of projects - the first submission of the project was peer-reviewed, then the author could improve her project and only after this she submitted her project in order to get a teacher evaluation.

From our investigation after semester finished we saw that the project outcomes have improved compared to the previous year, when the project assignment at the course was similar, however, the peer reviews were not applied on top of it. The overall grading has also improved. It is important to note, that the assignment and its evaluation were partly restructured with respect to the changes introduced, however, significant improvement was observed in corresponding evaluation categories.

This educational strategy also permitted the students to improve their submissions based on the reviews they received, which proved to be an important point. From an anonymous survey we learned that the students themselves found the new methodology rather useful, especially the opportunity to improve their projects in accordance with the obtained reviews and consequently to get better grading. However, it has to be noted that some students tended to abuse this option, and submitted very unfinished projects for the peer reviewing, since they knew that they can still finish them later on.

We have also found the very strong correlation between the instructors evaluation of the feedback and the students' numeric rating from the peer reviews, which can be considered as proof of the relevance of peer review. More results were presented in \cite{ICSLE2015, DIDINFO2015, ICL2015, ELEOT2015, eai2016}.

4. 3rd round of peer review in Web-design course.

In the last phase of the research a renewed set of rules was established. Students were supposed to design and develop a web application or web site for both desktop and mobile use. They were free to propose their own application site, although they had to meet certain requirements. They worked in teams of 3 - 4 members. Similarly to the previous research phase, the project was split into three rounds.

However, in this year, there were two types of peer reviewing: *peer review of projects*, where a whole team reviewed work of other three teams, and each project was reviewed by three other teams; and *team peer review*, where students reviewed their teammates' contribution to their joint project.

The students' satisfaction strongly improved, while their learning outcomes stayed at the comparable level to the outcomes from previous phases. We have also learned that students appreciated the opportunity to freely choose the project topic. In this phase, we also interviewed several participants and outcomes are very interesting - their opinions varied from each other in most of the discussed issues.

Results from this phase are planned to be published on two conferences: ICL 2016 (an abstract was already accepted) and ICWL 2016.

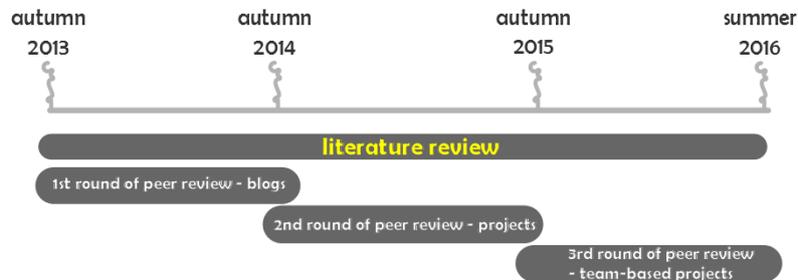


Figure 1: Research timetable

Research Design

After we made a decision what to study, asked specific questions and collected quantifiable data from participants (students), we analysed them by using various statistics tools. Therefore we can say that this research is an instance of the *a quantitative study* \cite{creswell2002}.

For this purpose the *correlation research design* was chosen. This type of design is usually used in case a researcher seeks to relate two or more variables to see if they influence each other, that means when *a need exists to study a problem requiring the identification of the direction and the degree of association between two sets of scores* \cite{creswell2002}. In our case -- we tried to find a relationship between a degree of students' engagement in peer review or peer assessment activities and their learning results from the project assignment, the final exam, a grade they earned and satisfaction with the course.

Two types of correlation research exist: a *prediction* and an *explanatory* design. In the prediction design, a researcher uses variables as predictors, so the aim is to identify variables that will predict an outcome. In the explanatory design a researcher is interested in co-variations of two or more variables - when changes in one variable make changes in the other \cite{creswell2002}. The character of our objectives led us to the explanatory design and moreover, this research met the explanatory design characteristics as follows:

- **we used correlation statistical test for (at least) two variables**
- **we collected data at one point in time** - at the end of a semester. We also used an introduction questionnaire to collect some data about the participants and their relationship to the tools they were going to use, or relationships between them. However, we did not collect data about their study results from previous courses similar or related to the Web-design course, neither we did not care about their performance in other courses they took at the same time or later.
- **we analysed all participants as a single group** - we did not divide the group into categories and we collected scores from the one group.
- **we obtained at least two scores for each individual in the group** - scores as number of submitted articles, number of approved articles, points from exam, points from project, final grade, etc.
- **we used correlation statistical test** - typically we use "standard" correlation coefficient, but we used also Spearman's correlation coefficient and chi-square tests of independence.
- **we made interpretations from the statistical test results** - e.g. we found that *students did not tend to give lower ratings (in peer reviewing) to their friends* as a conclusion from statistics in \cite{ICALT2014}.

Phase 1: Peer review and blogging

Results gained in 2012 from various sources led us to the conjecture that the rules were set too strict, and that a desired level of engagement (i.e., smaller but still rather high) together with satisfying learning outcomes could be achieved even in more relaxed settings, thus improving the overall opinions of the students. We reflected this in the 2013 setting as follows:

- The all-or-nothing rule was relaxed, introducing the scheme: full share of score for 5 articles (13 reviews) approved, half of the score for 3 articles (10 reviews) approved, or zero if less were approved.
- The share of points of the blogging activity was decreased to 17.5%. This was counter-balanced by increasing the share for written examination, resulting into the split: project (25%), blogging (17.5%), written examination (32.5%), and oral examination (25%). Students could now earn the C grading without participating in blogging, or still earn A if they collected one half of the blogging activity score.
- The approval requirements for articles were relaxed, focussing more on topicality and correct formatting of the articles for web publication which is more aligned with the actual content of the course.

Otherwise the setting was kept as in 2012.

Since the beginning of the course in 2007 we faced low students' engagement in blogging. In Fig. 2 the participation of students in blogging over the year is plotted (i.e., the share of those who posted at least one article). During the first five years participation was fairly low (35.79% on average). After the activity was redesigned, this figure immediately jumped to 92.77% in 2012 and further improved to 94.20% in 2013.

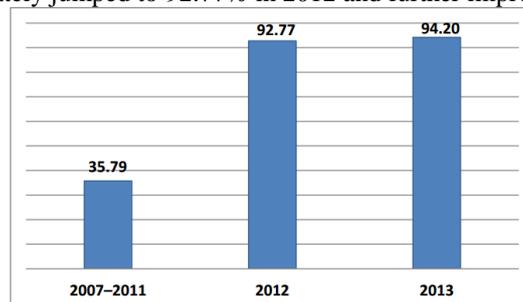


Figure 2: Participation in blogging activities.

Consecutively, in order to verify the possible positive impact of blogging activities on students' learning outcomes, we compare the study results during the years 2007-2013.

The chart in Fig. 3 shows the average grade in different years. The grades scale is from A (excellent, numerical value 1) to Fx (failure, numerical value 6). We observe that since 2007 the grading was gradually worsening, falling as low as 5.03 in 2011, but as soon as the changes were introduced it jumped to 4.10 in 2012 and then 4.06 in 2013, which is also historically the best average grade achieved through the years.

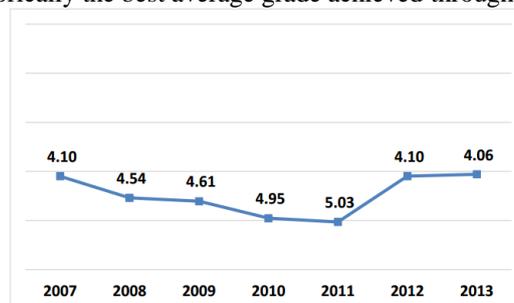


Figure 3: Average grades in 2007-2013.

Furthermore, we compared the achieved exam and project scores with the degree of participation in the blogging activity (the number of articles posted), and in the years 2012 and 2013 also with students success in blogging (the number of articles approved). After the examination of the study evaluation and grading of students in all runs of the course, we observed that the results were notably improved in the years 2007-2011 in which better organization of work and peer reviews were introduced. As the blogging and peer-reviewing activity also contributes to the evaluation, we consecutively analysed two parts of the evaluation which are independent from it - results from the written exam, and from the project assignment. We saw that the results from these activities were remarkably improved in the last two years. When we further compared the results with respect to the students' level of participation and success in the blogging and peer-reviewing activity, we were able to see a certain rising trend with the higher number of blog articles posted, respectively approved by the teachers.

We were also interested to compare the study results between the years 2012 and 2013, in order to evaluate the impact of the changes in the evaluation methodology (the blogging activity was made less obligatory). The change in the average grade was very small, and in fact, it was positive. When we compared the detailed results from the written exam and from the projects, we saw that the average project result fell while average exam

result rose (note that there were also other evaluation categories). This observation is not that significant, compared to the change in overall grading - we can attribute this to the changes between the cohorts each year. Regarding the interpretation of the results presented in this section it has to be noted, that rather than causality, they indicate a correlation between higher engagement in the blogging and peer-reviewing activity and improved learning outcomes. That is, based on this research we cannot claim, that blogging and peer-reviewing directly causes better study results, it may as well be that the students who are more motivated and responsible choose to participate to a higher extent. There are also differences between the abilities and attitudes of each year's cohort, which should be also taken into account. However, what the data clearly show is that the changes which we introduced in the years 2012 and 2013 had an overwhelming impact on the participation, and that the average grading was improved by one full grade compared to the period of 2011 and before. Our faculty conducts an anonymous survey of students' opinions for each course. The course in question was an advanced course, and especially after it was made obligatory in the curriculum it received mixed reviews for several years. Despite the fact that 98 students were enrolled in this year, only 47 of them (47.96%) used this tool to express their opinions. In general we can say that a smaller number of students liked it, while many of them complained that it is too difficult and that blogging is too demanding. Being part of the computer science and applied informatics curriculum, some students frankly admitted that they simply do not like to write.

Phase 2: Peer review and project

Use of peer review in a year 2014 was completely reconstructed. As we mentioned above, during the semester students worked on the project assignment. They were supposed to develop a web application (personal blog with typical blog features), both in desktop and mobile version, and including sufficient amount of meaningful content. This was split into three phases, each of which is evaluated independently:

- Back-end and desktop layout: the back-end is not evaluated (there are other courses which focus on this). The focus is on the quality of the layout, especially: inclusion of all typical blog features in the usual place, clean visual logic and comprehensibility by the user, and robustness.
- Mobile layout and navigation conventions (both desktop and mobile). The focus is especially on: proper development of mobile layout for small screens and touch interfaces, contextual switching between mobile and desktop, and proper implementation of basic navigation conventions on both versions.
- Content quality. The students have to develop some demo content for this phase, which is evaluated w. r. t. web publishing standards and quality criteria (e.g., writing style, proper usage of HTML markup, web content accessibility standards, etc.). This is the most difficult phase, as the number of requirements is quite high, and many students tend to trivialize them (as they prefer to focus on programming).

The project offers the space for hands-on practice for most of the knowledge and skills the students are supposed to master during the course, and it also serves as the primary form of assessment for this practical part. Therefore we put a lot of emphasis on it, and it is our aim to increase the time spent on this assignment as much as possible. In the past runs of the course, the assignment was only evaluated as one final submission. We observed that the students only started to work on the project a few days before the deadline. Several years ago, we split the project into three phases, which resulted in an increase in the time spent working on the assignment, and improved learning outcomes. To further improve this, in 2013 we complemented each of the phases with a peer review round, as they were previously shown to increase students' contact time with the course matter, and to foster experience exchange, and thus improve learning outcomes.

We added a peer review round after each project phase. Each submission in each phase was randomly assigned three reviewers. The review was blind, but not double-blind (i.e., the reviewer knew the identity of the submission authors, but not vice versa).

The reviews were delivered using a structured review form, which was different in each of the three phases. The questions in the form were directed towards the goals of each phase, and they more or less covered the evaluation criteria used by the instructors, but were often differently structured.

The reviewer had to answer each question with a rating ranging from 1 (very poor) to 5 (excellent), and in addition to provide a verbal justification for the given rating, of minimum 100 characters. The rationale behind this minimum limit is to encourage students to give more extensive feedback, e.g., also in cases when they had no objections.

The deadlines for submitting the assignment as well as reviews were strict. Students who missed the submission deadline were not allowed to review in the given phase. The aim of this rule was to avoid the situation in which there were very few submissions but too many reviewers. The reviewing phase took approximately 3-4 days. After reviews were delivered, the authors were given few days for correcting their projects according to the peers' comments. Only after this phase projects were submitted for teacher's evaluation.

This way, each phase of student's project passed double checking before it was evaluated by the teacher - firstly the author checked the project before submitting it for reviewing and then before submitting it to the teacher (so called double control scheme) (see Fig. 4).

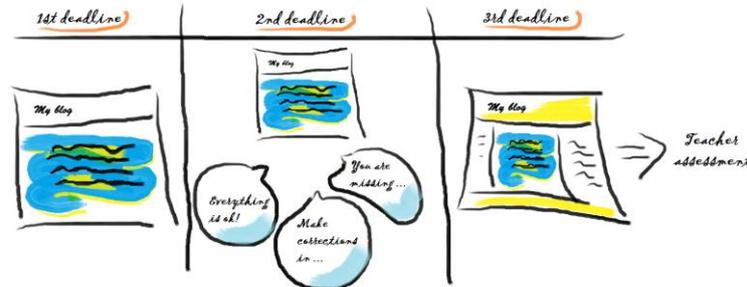


Figure 4: Illustration of double-checking scheme.

To find out the level of students' acceptance of peer review strategies used in the course, we prepared an anonymous questionnaire. Although participation was optional, the questionnaire was completed by 54 students (98.18% of all students who attended the final written exam).

Peer review process could also be beneficial to the reviewer. Therefore we asked students whether they gain some benefits while reviewing others' work. We expected the most popular option would be *I could gain more points*. However, although 61.11% of respondents chose it, there were even more students (68.52%) who stated that they realized shortcomings in their own projects thanks to the reviewing projects of their peers (Fig. 5). Also the fact that only 3.70% of students claimed that peer review was not beneficial to them can be taken for a good result.

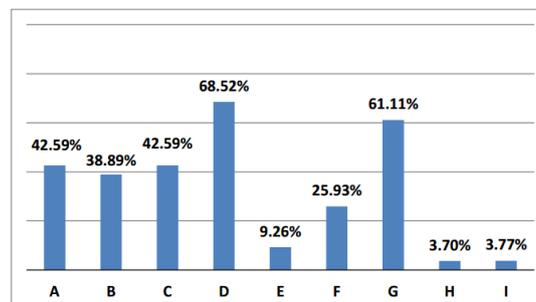


Figure 5: Do you think that you gain some benefits while reviewing others' work?

A: I learned how to test web projects; B: I learned how the project assignment was perceived by my peers; C: I realized how many different types of mistakes can appear on websites; D: I realized shortcomings in my project during peer reviewing; E: I trained my verbal skills; F: I improve in constructive criticism; G: I could gain more points; H: nothing; I: other.

Students' satisfaction and reliability in reviewing

Results from previous section - *Peer review and blogging* – presented in our thesis showed, that despite the better grading and smaller number of students who did not pass the exam and also regardless the fact that they used modern tools in their education, the students overall course rating in the official Student Questionnaire improved only slightly and their dissatisfaction with blogging activities froze at the same level as it was before.

Aiming to oblige students, the course rules were adjusted again in 2014's run. Students were not supposed to blog anymore and the peer review was integrated with the web design project.

The intended effect of all these changes was an improvement of students' attitude to the course which, as we believed, should have been reflected also in course evaluation in the official Student Questionnaire. Aiming to

reveal the reliability of student evaluation of teaching especially for our course we created our own questionnaire including among others also all questions from the official questionnaire.

However, the goal of this part of the research was not only to identify the level of students' satisfaction with the course, but also to verify the reliability of the official Student questionnaire. This motivation was based on other researchers' experience who asked whether this form of evaluation of teachers' work is really trustworthy. Recent studies about these issues show that it is at least disputable and that students do not evidently know what is best for their own learning (\cite{poropat2015}\cite{poropat2009, poropat2014},\cite{dweck2006} and others cited in the thesis).

We were interested in comparison of the outcomes of official Student Questionnaire and the results of our survey whose number of participants approximated 100% of students enrolled in the course. When results from official Student Questionnaire were available, we compared them with all those gained from our survey.

The first of them - treated as the most important - was oriented on students' overall rating of the course quality. Students answered by choosing a value from Likert scale from 1 to 5 (depicted by stars), where 5 is the best. Results from our questionnaire are similar to those gained from the official one for the values 4 (29.73% vs. 29.41% in favour of official questionnaire) or 5 (10.81% vs. 11.76% in favor of our questionnaire). However, there are great differences in choosing values 3 and 1 that can be seen in Fig. 6. While in official evaluation of teaching 21.62% of respondents answered by selecting one star, in our survey it was only 9.80% of them. Moreover, only 16.22% of respondents decided to choose 3-stars option in official questionnaire, but 31.37% of students chose it in our questionnaire.

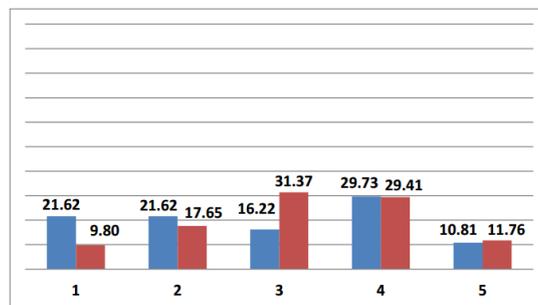


Figure 6: official Student Questionnaire our Questionnaire
Overall rating of the course quality. (Results are in %.)

We have also found interesting differences in absolute numbers of answers to particular questions. The response rate of our questionnaire was almost 100%, the official one proclaimed response rate of this course at the level of 65.51%. Therefore we could conclude that the number of students choosing particular options in all questions in the official Student Questionnaire is supposed to be less or equal to those collected in our questionnaire.

However, some inconsistencies in these answers appeared. We can provide an example of them in the question mentioned above: there were 5 students who marked the one-star option in our questionnaire, but eight students voted this way in the official one. Based on this, at least three people changed their minds between the day when they took part in our survey and the day they completed form for official student evaluation of teaching. Why it happened? We can only guess. It was likely caused by this reason: rating a student gave to the course was influenced by her success in passing the exam. If she failed the exam she evaluated the course worse than she would, if she passed the exam. We can suppose it since students mostly completed the official questionnaire after the exam period was over.

In our opinion, the official Student Questionnaire, also thanks to the possibility to write verbal comments, is a good tool for finding out problems that could not be found by other methods. However, as many other evaluation forms, this one has also its limits. Since it is voluntary and students do not receive any kind of reward for participation, it is impossible to reach very high percentage of engaged students. For example, in last five years the amount of participating students never got over 49% of all students.

Phase 3: Peer review and team-based project

In the last phase of the research a renewed set of rules was established. Students were supposed to design and develop a web application or web site for both desktop and mobile use. They were free to propose their own

application site, although there were some requirements. They worked in 11 teams of 3 - 4 members. The project was split into three rounds as in previous research phase. However since the topic of the project was free (unlike the past, where all projects had to be blogs) there were some differences. Requirements in these phases were set as follows:

- *1st phase: Specification.* Teams were required to consult their project idea with the lecturers and afterwards produce the specification that includes parts as purpose of the application; goals it will support; description of all functionalities it will provide; use case and respective storyboards and expected userbase and two distinct personas.
- *2nd phase: Layout & Functionality.* The goal of this round was to design and implement the layout of the application and all functionality according to the specification for both -- the desktop and the mobile version. Subsequently, the application was expected to be published on the Web. While implementing, team members had to meet demands for design conventions, clean visual logic, robustness, usability and accessibility.
- *3rd phase: Content.* After preparing specifications and implementing the project, teams were supposed to fulfill their applications with suitable content. It had to follow proper web writing and formatting guidelines, be formatted with the most suitable HTML elements, etc.

Student's overall evaluation in this course consisted of the share for written midterm and final test (30%), final oral test (25%) and the project (45%). The share for the project was split between three activities in three rounds. Submission of current project phase in each round equaled 10% of this share and another 5% served for engagement in peer reviewing. However, in this year, there were two types of peer reviewing, as it follows:

- Peer review of projects. A whole team reviewed work of other three teams, while each team work was reviewed by three other teams. Members of each team could make a decision whether they will review other teams work together or these reviews will be a product of individuals.
- Team peer review. In order to fairly evaluate work of each member in a team, we began to use team peer review where students assessed their colleagues in a team with points and reviewed their work answering three predefined questions.

The reviews of team projects were presented in a structured review form, different in each of the three phases - similarly as in previous year. Each phase was related to a different set of goals, therefore the questions in the each form varied. Moreover, all the questions were better explained with extensive hints.

The reviewing team answered each question with a rating of the same range as in previous years (from 1 - very poor to 5 - excellent), and with a verbal justification, as well.

In case of the team peer review the Fink method was chosen for this purpose and students in evaluating work of each of their colleagues answered three questions:

- Contribution summary. Evaluate your colleague's contribution in this round of the project. Summarize what this person has done, and if you are satisfied by his or her contribution.
- Most valuable thing. What is the single most valuable contribution this person makes to your team?
- What can be improved? What is the single most important thing this person could do to more effectively help your team?

The impact of this type of peer review, respectively peer assessment, was firstly used in the second round, while in the very first round we only trained students how to do that. Thus, this was the first time in whole research, when we considered peer assessment in the final student grade.

The deadlines for submitting the assignment, reviews of team works and team peer reviews were strict. Those, who in the given phase missed the deadline for submitting an assignment, have not had an opportunity to review other teams' work. The time for reviewing was set to approximately 3-4 days. Continuously, the teams could correct their projects according to the peers' comments during next few days.

Only after this phase projects were submitted for teacher's evaluation. The team peer reviewing began after this final submission and students had 5 days to review and assess their teammates.

Students' attitude to the peer review or peer assessment slightly differed between this and the previous year. As inputs for this comparison we used the data collected from our questionnaires in 2014 and 2015, both given to the students at the end of the semester. The response rate in 2014 was 98.18%, while in 2015 it was 94.74%.

Although we believe that students are trained in many other professional areas through the peer review, we asked them to define benefits related to the process of peer-reviewing others (Fig. 7). In 2014, we expected the option I could gain more points to be the most often appearing answer, because for many students are points (that help them to pass the course) very important. However, another option - I realized shortcomings in my project during peer reviewing earned even more votes (68.52% for this vs. 61.11% for points). Moreover, results from 2015 varied from those previous in other issues, as well: the most students appreciated an opportunity to realize many different types of mistakes that appear on websites (55.56%), then 41.67% of all students liked peer-reviewing because they realized shortcomings in projects of their teams and the option I could gain more points was ranked as the third most popular together with the option I learned how to test web projects (both were chosen by 36.11% of all students). This is an interesting (and maybe an obvious, as well) observation: if the course is not an obligatory one, students who enrol in will not be so focused on points as students who are enrolled in an obligatory course. Another positive result based on this question responses is that no students chose a negative option nothing.

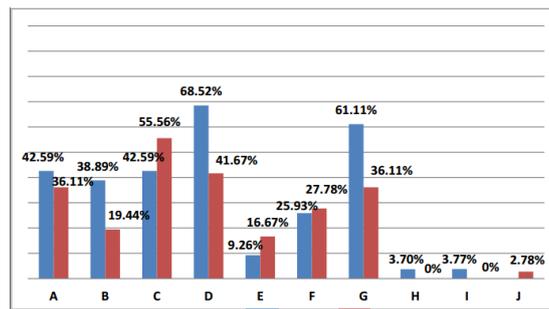


Figure 7: 2014 2015

Do you think that you gain some benefits while reviewing others' work? A: I learned how to test web projects; B: I learned how the project assignment was perceived by other teams; C: I realized how many different types of mistakes can appear on websites; D: I realized shortcomings in my/our project during peer reviewing; E: I trained my verbal skills; F: I improved in constructive criticism; G: I could gain more points; H: nothing; I: other; J: I did not review (available only in 2015)

Team peer review seemed to be more difficult task than the peer review of projects, because participants were required to evaluate not only the products of their colleagues' work, but mainly these students' sense of responsibility in their attitude to the collaborative teamwork. Students were asked to define their points of view about the process of reviewing and assessing their team colleagues. Although the third of the students thought that the team peer reviewing helped the teacher in evaluating individual's work, only about 30% of all students admitted that they liked. Moreover, the same number of students considered the team peer review as an useless activity and 25% of all as an activity that was uncomfortable for them. However, on the other hand, none of the students indicated the fact that they were peer-reviewed by their teammates as uncomfortable for her.

Discussion and conclusion

The intention of this thesis was to answer the question: *How can an appropriate educational strategy that includes peer review or peer assessment techniques be used in our Web-design course?*. In order to answer it in more details, we defined a set of three sub-questions. We will solve them consequently:

Which of the course assignments are suitable for involving peer review or peer assessment?

We tested it on blog posts, individual projects (implementing blogs) and team-based projects (implementing any web application that fulfils requirements). The best results in general were achieved with the last of these assignments - team-based projects. However, as we have already discussed in this chapter, individual projects may also be very suitable, although some changes (in comparison with our conditions) are required.

How should students be motivated to engage in peer review or peer assessment activities?

Although one of the most expected motivation is gathering points that count into final grade, we found another factor that makes peer review attractive to students. By this we mean so-called *double-checking* of projects, thus two checks of the project submission in one phase: the first done by the authors and the second done by the reviewers. Only after the second check and patching reviewed problems the project was submitted for teachers' evaluation. According to this, students had an opportunity to earn higher amount of points than they had without correcting missed mistakes.

How are students' learning outcomes affected by their engagement in peer review or peer assessment activities?

Based on our results, we can conclude that engaging in peer review activities enables students to realize shortcomings in their own knowledge or in their own assignment solutions. This observation is supported by proven significant improvement in students' learning outcomes, such as their results from exams and projects.

Nevertheless, one of the "unwritten" expected goals of this study was to develop a methodology of teaching informatics with new, innovative, modern methods. We focused on a course in higher computer science education and according to the obtained results we achieved this cognition: not every educational strategy is suitable in such courses, but peer review or peer assessment combined with well chosen assignments can improve students' skills in various spheres and encourage them in their future professional lives.

References

- [BF07] David Boud and Nancy Falchikov. Rethinking assessment in higher education: Learning for the longer term. Routledge, 2007 (cit. on pp. 18, 20).
- [BK56] Benjamin Samuel Bloom and David R. Krathwohl. Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. Longmans, 1956.
- [CG87] Arthur W. Chickering and Zelda F. Gamson. "Seven principles for good practice in undergraduate education." In: AAHE bulletin 3 (1987), p. 7.
- [CE96] Arthur W. Chickering and Stephen C. Ehrmann. "Implementing the seven principles: Technology as lever". In: AAHE bulletin 49 (1996), pp. 3–6.
- [Cre02] John W. Creswell. Educational research: Planning, conducting, and evaluating quantitative. Prentice Hall, 2002.
- [Cro81] K. Patricia Cross. Adults as Learners. Increasing Participation and Facilitating Learning. ERIC, 1981.
- [Dwe06] Carol Dweck. Mindset: The new psychology of success. Random House, 2006.
- [Fal86] Nancy Falchikov. "Product comparisons and process benefits of collaborative peer group and self assessments". In: Assessment and Evaluation in Higher Education 11.2 (1986), pp. 146–166.
- [FG00] Nancy Falchikov and Judy Goldfinch. "Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks". In: Review of educational research 70.3 (2000), pp. 287–322.
- [Geh03] Edward F Gehringer. "Building resources for teaching computer architecture through electronic peer review". In: Proceedings of the 2003 workshop on Computer architecture education: Held in conjunction with the 30th International Symposium on Computer Architecture. ACM. 2003, p. 9.
- [HK09a] Martin Homola and Zuzana Kubincová. "Practising Web Design Essentials by Iterative Blog Development within a Community Portal." In: CSEDU (1). 2009, pp. 181–186.
- [HK09b] Martin Homola and Zuzana Kubincová. "Taking advantage of Web 2.0 in organized education (a survey)". In: Proc. ICL. 2009, pp. 741–752.
- [Lev08] Ruth E Levine. "Peer evaluation in team-based learning". In: Team-Based Learning for Health Professions Education: A Guide to Using Small Groups to Improve Learning. Sterling, Va: Stylus Publishing (2008), pp. 103–116.
- [Liu+01] EZ-F Liu, Sunny SJ Lin, Chi-Huang Chiu, and ShyanMing Yuan. "Web-based peer review: the learner as both adapter and reviewer". In: Education, IEEE Transactions on 44.3 (2001), pp. 246–251.
- [LLY01] Sunny SJ Lin, Eric Zhi-Feng Liu, and Shyan-Ming Yuan. "Web-based peer assessment: feedback for students with various thinking-styles". In: Journal of computer assisted Learning 17.4 (2001), pp. 420–432.
- [Mos05] Stephan Mosel. "Self directed learning with personal publishing and microcontent". In: Microlearning 2005 (2005).
- [Ora02] Jo Ann Oravec. "Bookmarking the world: Weblog applications in education". In: Journal of Adolescent & Adult Literacy (2002), pp. 616–621.
- [Owe+06] Martin Owen, Lyndsay Grant, Steve Sayers, and Keri Facer. "Opening education: social software and learning". In: Bristol: Futurelab. Retrieved November 20 (2006), p. 2006.
- [Pia70] Jean Piaget. Science of education and the psychology of the child. Trans. D. Coltman. Orion, 1970.
- [PH91] S. Papert and I. Harel. "Constructionism: research reports and essays 1985-1990 by the Epistemology and Learning Research Group, the Media Lab, Massachusetts Institute of Technology, Ablex Pub". In: Corp, Norwood, NJ (1991).
- [PM11] Elvira Popescu and Laurentiu Manafu. "Repurposing a wiki for collaborative learning-pedagogical and technical view". In: System Theory, Control, and Computing (ICSTCC), 2011 15th International Conference on. IEEE. 2011, pp. 1–6.
- [Pop10] Elvira Popescu. "Students' acceptance of web 2.0 technologies in higher education: Findings from a survey in a Romanian university". In: Database and Expert Systems Applications (DEXA), 2010 Workshop on. IEEE. 2010, pp. 92–96.

- [Por09] Arthur E. Poropat. "A meta-analysis of the five-factor model of personality and academic performance." In: *Psychological bulletin* 135.2 (2009), p. 322.
- [Por14] Arthur E. Poropat. "Other-rated personality and academic performance: Evidence and implications". In: *Learning and Individual Differences* 34 (2014), pp. 24–32.
- [Por15] Arthur Eugene Poropat. "Students don't know what's best for their own learning". In: (2015).
- [RFT09] Ken Reily, Pam Ludford Finnerty, and Loren Terveen. "Two peers are better than one: aggregating peer reviews for computing assignments is surprisingly accurate". In: *Proceedings of the ACM 2009 international conference on Supporting group work*. ACM. 2009, pp. 115–124.
- [Ric10] Will Richardson. "Blogs, wikis, podcasts, and other powerful web tools for classrooms." Corwin Press, 2010.
- [RKS09] Liana Razmerita, Kathrin Kirchner, and Frantisek Sudzina. "Personal knowledge management: The role of Web 2.0 tools for managing knowledge at individual and organisational levels". In: *Online Information Review* 33.6 (2009), pp. 1021–1039.
- [Rov02] Alfred P. Rovai. "Development of an instrument to measure classroom community". In: *The Internet and Higher Education* 5.3 (2002), pp. 197–211.
- [Saf08] Christian Safran. "Blogging in higher education programming lectures: an empirical study". In: *Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era*. ACM. 2008, pp. 131–135.
- [Sie05] G. Siemens. *Connectivism: A learning theory for the digital age*. *International Journal of Instructional Technology & Distance Learning* 2(1), 2005.
- [ST12] Andrea Sterbini and Marco Temperini. "Dealing with openanswer questions in a peer-assessment environment". In: *Advances in Web-Based Learning-ICWL 2012*. Springer, 2012, pp. 240–248
- [Top98] Keith Topping. "Peer assessment between students in colleges and universities". In: *Review of educational Research* 68.3 (1998), pp. 249–276.
- [Von95] Ernst Von Glasersfeld. *Radical Constructivism: A Way of Knowing and Learning*. *Studies in Mathematics Education Series: 6*. ERIC, 1995.
- [Vyg64] Lev S. Vygotsky. "Thought and language". In: *Annals of Dyslexia* 14.1 (1964), pp. 97–98.
- [Wu06] Wen-Shuenn Wu. "The effect of blog peer review and teacher feedback on the revisions of EFL writers". In: *Journal of Education and Foreign Languages and Literature* 3.2 (2006), pp. 125–138.

Author's publications

- [Bej13] Veronika Bejdová. "Budovanie motivácie študentov v elektronickom vzdelávacom prostredí". Master thesis. Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, 2013.
- [BHK14] Veronika Bejdová, Martin Homola, and Zuzana Kubincová. "Blogging in Obligatory Course: A Bitter Victory". In: *Advances in Web-Based Learning-ICWL 2014*. Springer, 2014, pp. 1–10.
- [BKH13a] Veronika Bejdová, Zuzana Kubincová, and Martin Homola. "Blogy vo vyučovaní na univerzite". In: *DidInfo 2013*. ISBN 978-80-557-0527-9. Banská Bystrica: Univerzita Mateja Bela. 2013, pp. 137–143.
- [BKH13b] Veronika Bejdová, Zuzana Kubincová, and Martin Homola. "Can blogging activities and peer-reviewing improve learning outcomes?" In: *Information and Communication Technology in Education 2013*. Ostrava: University of Ostrava. 2013, pp. 37–46.
- [BKH14] Veronika Bejdová, Zuzana Kubincová, and Martin Homola. "Are Students Reliable Peer-Reviewers?" In: *Advanced Learning Technologies (ICALT), 2014 IEEE 14th International Conference on*. IEEE. 2014, pp. 270–272.
- [DHK15] Veronika Dropčová, Martin Homola, and Zuzana Kubincová. "Students' Acceptance of Peer Review". In: *E-Learning, E-Education, and Online Training*. Springer, 2015, pp. 52–59.
- [DHK16] Veronika Dropčová, Martin Homola, and Zuzana Kubincová. "May I Peer-Review Your Web Design Project?" In: *State-of-the-Art and Future Directions of Smart Learning*. Springer, 2016, pp. 243–249.
- [DK14] Veronika Dropčová and Zuzana Kubincová. "Evaluation of peer-review quality in comparison to teacher's grading". In: *Information and Communication Technology in Education 2014*. Ostrava: University of Ostrava. 2014, pp. 61–68.
- [DK15a] Veronika Dropčová and Zuzana Kubincová. "Student questionnaire – Can we trust it?" In: *Interactive Collaborative Learning (ICL), 2015 International Conference on*. IEEE. 2015, pp. 602–607.
- [DK15b] Veronika Dropčová and Zuzana Kubincová. "Peer review pri hodnotení projektov". In: *DidInfo 2015*. ISBN 978-80-557-0852-2. Banská Bystrica: Univerzita Mateja Bela. 2015, pp. 47–50.
- [KB14] Zuzana Kubincová and Veronika Bejdová. "Dokážu študenti spoľahlivo hodnotiť práce spolužiakov?" In: *DidInfo 2014*. ISBN 978-80-557-0698-6. Banská Bystrica: Univerzita Mateja Bela. 2014, pp. 89–96.
- [KDH16] Zuzana Kubincová, Veronika Dropčová, and Martin Homola. "Students' acceptance of peer review in Computer Science course". In: *EAI Endorsed Transactions on e-Learning* 16.10 (Apr. 2016). doi: 10.4108/eai.11-4-2016.151153.

[KHB13] Zuzana Kubincová, Martin Homola, and Veronika Bejdová. “Motivational Effect of Peer Review in Blog-Based Activities”. In: *Advances in Web-Based Learning–ICWL 2013*. Springer, 2013, pp. 194–203.