Motivating Students and Improving Quality of Learning Using Peer-Reviews

Vadim Ermolayev[1], Natalya Keberle[1], and Sergey Borue[1]

[1] Zaporozhye National University, Zhukovskogo st. 66 69063 Zaporozhye Ukraine
vadim@ermolayev.com, nkeberle@gmail.com, bsu@znu.edu.ua

Abstract. This paper reports about a pedagogical experiment at Zaporozhye National University (ZNU) aiming at improving motivation and learning quality in Computer Science Bachelor programme. The major novelty in teaching and learning practice introduced in the experiment was the use of peer evaluation for the assessment of coursework reports in two disciplines – one in the II-nd and the other in the IV-th year of study. The results were compared to the historical data collected in the previous 3-4 years. Our experiment proved that exploiting students' aspirations for informal leadership and incurred competition constructively is effective and yields some increase in motivation to learn and learning quality. The assessments were also subjectively regarded as more clear and better justified by the students involved in the experiment. A good side effect is also that the students learn the working patterns of the professionals in their field broadly used in academia and industry for making qualitative and unbiased peer evaluations.

Keywords. Motivation, learning quality, peer evaluation, Computer Science, coursework.

Key Terms. Academia, TeachingProcess, Characteristic, QualityAssuranceProcess.

1 Introduction

Recent higher education experience reveals a substantial decrease of the popularity of University education and degrees reflected for instance in the decrease in degree completion rates [1]. Researchers analyzing the reasons for this decrease point out: (i) the rise of pragmatic attitudes to education in life planning among young people; (ii) the trend for devaluation of a University degree as a factor facilitating to employment and career development. As a result and because of the concurrent demographic and economic crises a substantial decrease of interest to quality learning among University students in observed. This observation is supported by the decrease in the student numbers and their grades. Consequently, the employers suffer from a decreased quality of the graduates.

Academia can not remove or relax demographic or economic factors unfortunately. Hence, the only feasible way of keeping academic performance at a competitive level is focusing on the stimuli for their students based on more social than purely pragmatic basics. For example, exploiting the value of informally assessed professional capability and leadership in student groups may be an effective way of stimulating spending more effort in learning.

The research presented in this paper aims at finding out such stimuli for Computer Science students based on their attitude to informal leadership grounded in professional competencies. The idea behind our pedagogical experiment was to place the subjects in an environment which is similar to professional and offer them to peer-evaluate their individual work. Hence, the higher the grades a person gets from his or her peers in such an evaluation – the higher becomes the professional reputation of the person in the group, making him or her informal leader in the group of the peers.

In fact the approach we have taken is not new and has been effectively exploited in the academic world as a peer evaluation mechanism as well as in social networks for forming communities of interest and building social reputation for the individuals in these communities. Such stimuli are qualified as **solidary** (in contrast to **material**) **incentives** [2] i.e. intangible rewards from the act of being a part of a group having coherent interests. In our research we build upon the mechanisms and tool support adopted from the mentioned domains. We involve students in peer evaluation of their individual coursework assignment reports similarly to that of reviewing conference papers. We measure their qualification by: comparing evaluations by peers and instructors – assignment results; and measuring deviations between their individual scorings and the mean values – reviewer competence. The anonymized results are then made available to the group.

We have observed that being an evaluator for the peers' work proved to be a noticeable incentive for the subjects who took part in our pedagogical experiment. Consequently the degree of active involvement and the quality of individual assignment results have increased substantially, in particular for the group in the last year of our Bachelor programme in Computer Science. This observation is backed up by the results presented in Section 4.

The rest of the paper is structured as follows. Section 2 gives a brief overview of the related work in higher education students' motivation. Section 3 presents the set-up of our pedagogical experiment. Section 4 discusses experimental results. Conclusions and plans for the future work are given in Section 5.

2 Related Work

Motivations are denoted as "…reasons individuals have for behaving in a given manner in a given situation" (c.f. [3]). "They exist as part of one's goal structures, one's beliefs about what is important, and they determine whether or not one will engage in a given pursuit'' (c.f. [4]). In academic settings two types of motivation are distinguished – intrinsic and extrinsic. Intrinsically motivated subjects learn for their own sake, because enjoy learning or assess the outcome of the learning process as important for themselves – e.g. [5]. Extrinsic motivation is driven by a desire of getting rewards – from the others; or to avoid punishment. Students motivated extrinsically focus on receiving the approvals – like judgements by lecturers and peers – e.g. [4]. Our approach, though welcoming intrinsic motivation, focuses on obtaining utility of exploiting student's extrinsic stimuli – which proves to become more spread and influential in the current economic settings.

Many authors stress the importance of a skill to maintain and enhance students' motivation as one of the core capabilities of a University lecturer. "A wide variety of theories of learning and teaching recognises motivation as an essential prerequisite for successful learning. The ability to maintain and enhance student motivation is therefore one of the most important skills …, and many publications and training programmes devote considerable space and time to this matter. Applying this theoretical knowledge in practice, however, remains difficult due to the complexity of the concept and the number of different models of motivation available" (c.f. [6]). Our research is focused exactly on the application of motivation stimuli to practice in a Bachelor level Computer Science programme – so the experimental data we have analysed spans across several disciplines taught in the 1-st to 4-th year of the programme at Zaporozhye National University (ZNU).

The mainstream of experimental studies in higher education teaching and learning is centred around using the methodologies of individual subjective assessment by subjects – based on interviews, questionnaires, etc (e.g. [7] to mention just one of many relevant publications). In difference to the mainstream methodology, we exploit the collaborative character that is intrinsic to student collectives and base our approach on well renowned social and peer approaches – this is why peer evaluation is used. Such a method allows us not only to collect and analyse individual judgements, but also to cross-rate the subjects by their own cross-judgements and stimulate healthy competition – thus increasing positive stimuli.

3 Setting up the Pedagogical Experiment

Stimulated by the necessity to seek for a remedy confronting the decrease of interest for quality learning at Universities, we have planned and further conducted a pedagogical experiment at ZNU. We have focused on the individual coursework as one of the important kinds of students' creative activities in which motivation plays a very important role.

Our major objective was to prove a pedagogical hypothesis:

> If students are given an opportunity to act as peer-evaluators of the other students reports, their **extrinsic motivation** to: (a) deliver the coursework; and (b) to perform as good as they can – **will be higher** than among those who do individual work in a traditional way and are graded by their instructor only. Furthermore, the **quality** of submitted reports is expected to **be better**, as students informally compete and cross-evaluate their quality. Finally, the **objectivity of the assessments will be higher**; those will be perceived as fair by the subjects.

For that we have:

- Chosen the disciplines: (a) for which the historical data on the coursework grades existed for several years; (b) for which the complexities of doing coursework assignments were comparable; and (c) the coverage of all four years of our Bachelor programme was even
- Developed detailed assessment forms for inexperienced evaluators offering a clear procedure and set of explicit metrics for coursework report assessment per each involved discipline
- Chosen the student groups comprising the cases when the same group acted as an experimental sample and formerly – a control sample; briefed the experimental groups
- Configured the set of software tools to support the experiment and developed written methodological recommendations for the subjects
- Adopted and adapted simple and effective metrics that allowed measuring the proofs of our research hypothesis

3.1 Pedagogical and Methodological Set-up

The pedagogical set-up of our experiment covers: the choice of disciplines; the preparation of the evaluation forms; and subjects' briefing about the evaluation procedure and tools.

First, we have chosen the disciplines with historical data and good coverage of out Bachelor programme. The choice is summarized in Table 1 – showing that:

- 3+ year historical data on the coursework assignment grades is available
- The disciplines cover all 4 years of study within the programme evenly

The complexity of the assignments, though different per discipline, is comparable as shown in Table 1.

- **Table 1.** Choice of disciplines and complexities of related coursework assignments.

Discipline	Year	2008	2009	2010	2011	Avg
Programming	I	---	100	100	100	100
Algorithms and Data Structures	II	100	150	200	**250**	175
DataBases and Information Systems	III	---	150	150	150	150
Intro to Logical Programming and AI	IV	---	150	150	**250**	183

Legend: numbers in columns 3-7 are coursework complexities. The cells corresponding to our experiment are shaded gray and have bolded numbers.

Grades data for the coursework assignments in Programming (year I) and Databases and Information Systems (year III) form our first and second baseline control datasets respectively.

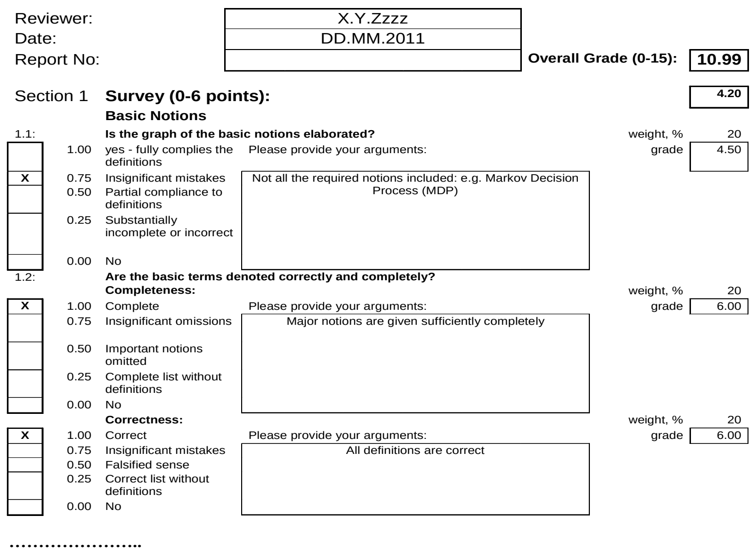
The complexity of the I-st year coursework assignment in Programming has been chosen as basic – represented by 100 abstract points. This coursework contains a survey part on a particular topic and a practical assignment to develop a program solving a given simple problem. The complexity of the coursework in this discipline remains without change for all the 3 years of our observations.

The complexity of the 3-d year coursework in Databases and Information Systems is also static within the period of observation. However, it is 1.5 times more complex as contains several interrelated practical problems in database and IS development using SQL Server software. Another difference is that the subjects for this assignment were the III-d year students whose motivation from one hand and experience from the other hand differ from the ones of I-st year students.

Observations in Algorithms and Data Structures and Introduction to Logical Programming and Artificial Intelligence contain both control and experimental (shaded gray in Table 1) data.

The complexity of the coursework assignment in Algorithms and Data Structures increases from 100 points in 2008 to 250 points in 2011. In 2008 it was very similar in structure to the coursework in Programming – a detailed written presentation of a sorting algorithm studied individually and its practical implementation in a computer program. In 2009 the task of analytically evaluating the computational complexity of the algorithm was added – raising the complexity up to 150 points. In 2010 the task of experimental measurement of the computational complexity and comparing it to the analytical estimation was added – the complexity has therefore increased to 200 points. In 2011 the coursework has been complicated (up to 250 points) by offering a comparative evaluation exercise – the students were tasked to measure the performance of their program and compare to the performance of a program developed by a fellow based on several common datasets containing records of different types.

Secondly, we have developed the evaluation forms for coursework reports in both disciplines. An example of a fragment of an evaluation form is pictured in Fig. 1.


Reviewer: X.Y.Zzzz
Date: DD.MM.2011
Report No:
Overall Grade (0-15): 10.99

Section 1 **Survey (0-6 points):** 4.20
Basic Notions
1.1: **Is the graph of the basic notions elaborated?** weight, % 20
[ ] 1.00 yes - fully complies the definitions — Please provide your arguments: — grade 4.50
[X] 0.75 Insignificant mistakes — Not all the required notions included: e.g. Markov Decision Process (MDP)
[ ] 0.50 Partial compliance to definitions
[ ] 0.25 Substantially incomplete or incorrect
[ ] 0.00 No
1.2: **Are the basic terms denoted correctly and completely?**
Completeness: weight, % 20
[X] 1.00 Complete — Please provide your arguments: — grade 6.00
[ ] 0.75 Insignificant omissions — Major notions are given sufficiently completely
[ ] 0.50 Important notions omitted
[ ] 0.25 Complete list without definitions
[ ] 0.00 No
Correctness: weight, % 20
[X] 1.00 Correct — Please provide your arguments: — grade 6.00
[ ] 0.75 Insignificant mistakes — All definitions are correct
[ ] 0.50 Falsified sense
[ ] 0.25 Correct list without definitions
[ ] 0.00 No

……………………..

Fig. 1. A fragment of the completed evaluation form for the coursework report in Introduction to Logic Programming and AI.

The forms are in fact structured questionnaires covering all the sections of the report and suggesting several weighted Likert scale [8] based metrics covering several aspects that were different for each section. Table 2 contains the lists of the report sections and evaluation questions for both disciplines.

Table 2. Evaluation aspects covered by review forms and scoring weights.

Algorithms and Data Structures		**Intro to Logical Programming and AI**	
Aspect to Evaluate	**Weight %**	**Aspect to Evaluate**	**Weight %**
Section 1: Sorting method and algorithm (0-2 points)		**Section 1: Survey (0-6 points)**	
1.1 Is the description of the family of sorting methods given?	25	1.1 Is the graph of the basic notions elaborated?	20
1.2 Is the algorithm described sufficiently completely and clearly?	50	1.2 Are the basic terms denoted correctly and completely?	40
1.3 Is algorithm stability analyzed?	25	Are the major problems in the field covered?	
Section 2: Software implementation (0-3 points)		1.3 Are the problem statements given?	10
2.1 Is the source code provided?	10	1.4 Is the actuality of these problems explained?	10
2.2 Does the implementation comply with the algoritm described in Section 1?	15	1.5 Are the descriptions of solution methods given?	10
2.3 Are the implementation decisions described sufficiently fully and clearly?	30	1.6 Are the surveyed solution methods compared?	10
2.4 Are the constraints (absence of) wrt input data explained and justified?	10	**Section 2: Solving Problems in Visual PROLOG (0-7 points)**	
2.5 Does the provided software work?	25	2.1 Is **task 2** solved?	10
2.6 Is the source code reasonably commented?	10	2.2 Is **task 3** (traffic expert system design) solved?	30
Section 3: Theoretical estimation of the computational complexity (0-2 points)		2.2.1 Are schematic descriptions of situations at T-shaped crossroad given?	
3.1 Is the estimation the computational complexity given?	50	2.2.2 Are predicates and goals for situations fully described?	
3.2 Is the graphical illustration of the computational complexity given?	50	2.2.3 Are predicates and goals compliant to the schematic descriptions?	
Section 4: Computational experiment – comparison with other algorithms (0-3 points)		2.3 Is **task 4** (traffic expert system implementation) solved?	30
4.1 Are the data sets chosen correctly	50	2.3.1 Are predicates and goals specified correctly?	
4.2 Are the other algorithms for comparison chosen reasonably?	50	2.3.2 Does the developed expert system work?	
Section 5: Experimental assessment of the computational complexity (0-3 points)		2.3.3 Are the implementation decisions documented?	
5.1 Are the rules and programming solution for measuring computational complexity described?	30	2.4 Is **task 5** (traffic expert system refinement) solved?	30
5.2 Is the comparative analysis of the computational complexity given?	30	2.4.1 Are predicates and goals specified correctly?	
5.3 Is the experimental assessment compared with the theoretical estimation?	30	2.4.2 Is the sense of the predicates and structures explained?	
5.4 Is the graphical illustration of the computational experiment results given?	10	2.4.3 Are all possible traffic situations described?	
Section 6: Conclusions (0-3 points)		2.4.4 Does the refined Expert System work?	
6.1 Do the conclusions reflect the results obtained?	50	**Section 3: Conclusions (0-2 points)**	
6.2 Are the references to the adopted components given?	30	3.1 Do the conclusions reflect the results obtained?	50
6.3 Is the report compliant to the template (abstract, contents, references, appendices)?	20	3.2 Are the references to the adopted components given?	30
		3.3 Is the report compliant to the template (abstract, contents, references, appendices)?	20

It has been decided that the overall grade for a coursework report of maximum 20 points is divided in two parts:

- The coursework grade (0 – 15 points) computed as a mean of the three assessments done by two peers and one instructor
- The evaluation grade (0 – 5 points) computed as 5 minus the mean of deviations of the subject's evaluation scores from the mean scores. So, the closer an individual scoring is to the mean scoring in all the evaluation assignments – the higher the resulting evaluation grade is.

3.2 Experimental and Control Groups

Two experimental groups in the II-nd and IV-th year of study have been selected so that the historical coursework grade data was available for them. For comparison, the control data about the grades in the other groups of different years of study and in all four chosen disciplines have been taken into account. The groups for which the control data was accounted for have been further treated as control groups. Table 3 depicts the distribution of the control and experimental groups over the years of study. As could be seen in Table 3, the experimental groups are also control groups but in different disciplines and years of study. So, different ways of comparing the activity and performance in doing coursework assignments arise: the same group in different years; the same group in different disciplines; the same group as experimental and doing the work in a traditional way; etc.

Table 3. Experimental and control groups.

Group No	2008		2009		2010		2011	
	Year of Study	Role	Year of Study	Role	Year of Study	Role	Year of Study	Role
8216			IV	C				
4327	II	C	III	C	IV	C		
4328			II	C	III	C	IV	E
4329			I	C	II	C	III	C
4320					I	C	II	E
4321							I	C

Legend: C – control group; E – experimental group.

At the beginning of the experiment the subjects of our two experimental groups were briefed about: the deadlines; the objectives of peer evaluation; the structure and the content of the evaluation forms; the grades that will be assigned for the reports and for the reviews; the tools they will use in the peer evaluation process.

3.3 Instrumental Set-up

Two procedures have been chosen for evaluation that differed in the used tools. For the experiment with the II-nd year students the workflow based on e-mail exchange and manual supervision has been adopted. For the IV-th year students we have introduced the EasyChair Conference Management System[1] as a tool to manage the process, final ranking and grading. In both cases the structured evaluation forms have been offered to the subjects to be filled out using MS Excel.

4 Results and Discussion

Evaluation process has been organized and executed similarly to peer evaluation of conference papers by programme committee members. Students were invited to serve on the programme committee and the review assignments have been made by the instructors who acted as programme chairs. The results of evaluation have been collected and processed using two different patterns:

– For II-nd year students – collected by e-mail and processed manually using Excel spread sheet as shown in Fig.2 in an anonymized way

	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6	Reviewer 7	Reviewer 8	Reviewer 9	Reviewer 10	Reviewer 11	Reviewer 12	Reviewer 13	Instructor 1	Instructor 2	Average	Deviation - R.1	Deviation - R.2	Deviation - R 3
<Author 1>	7.40													2.60		**5.00**	-2.400	5.000	2.400
<Author 2>		11.79						13.19						11.39		**12.12**	0.333	-1.067	0.733
<Author 3>			8.98						9.80						8.15	**8.98**	-0.003	-0.823	0.827
<Author 4>					1.09					1.53					1.43	**1.35**	0.260	-0.180	-0.080
<Author 5>				12.21							5.98			10.29		**9.49**	-2.717	3.513	-0.797
<Author 6>						13.30						10.98			6.15	**10.14**	-3.157	-0.837	3.993
<Author 7>													0.00		0.00	**0.00**	0.000	0.000	0.000
<Author 8>						10.86							12.18		5.93	**9.66**	-1.203	-2.523	3.727
<Author 9>	6.20											8.61		4.79		**6.53**	0.333	-2.077	1.743
<Author 10>		5.45						6.00						4.60		**5.35**	-0.100	-0.650	0.750
<Author 11>			15.13						13.90					10.01		**13.01**	-2.117	-0.887	3.003
<Author 12>					13.28					13.15				11.38		**12.60**	-0.677	-0.547	1.223
<Author 13>				12.21							9.61				9.15	**10.32**	-1.887	0.713	1.173
Points for Reviews	**3.63**	**4.78**	**3.94**	**2.70**	**4.53**	**2.82**	**0.00**	**4.14**	**4.15**	**4.64**	**2.89**	**3.54**	**3.74**						

Fig. 2. Anonymized review results visible to instructors.

	Reviewer 1	Reviewer 2	Instructor 1	Instructor 2	Average	Review Points	Total
<Author 1>	7.4		2.6		**5.00**	**4.64**	**10**
<Author 2>	11.79	13.19	11.39		**12.12**	**4.15**	**16**
<Author 3>	8.98	9.8		8.15	**8.98**	**3.63**	**13**
<Author 4>	1.09	1.53		1.43	**1.35**	**2.70**	**4**
<Author 5>	12.21	5.98	10.29		**9.49**	**3.54**	**13**
<Author 6>	13.3	10.98		6.15	**10.14**	**4.78**	**15**
<Author 7>		0		0	**0.00**	**2.82**	**3**
<Author 8>	10.86	12.18		5.93	**9.66**	**3.94**	**14**
<Author 9>	6.2	8.61	4.79		**6.53**	**4.14**	**11**
<Author 10>	5.45	6	4.6		**5.35**	**0.00**	**5**
<Author 11>	15.13	13.9	10.01		**13.01**	**3.74**	**17**
<Author 12>	13.28	13.15	11.38		**12.60**	**2.89**	**15**
<Author 13>	12.21	9.61		9.15	**10.32**	**4.53**	**15**

(a) Score table for the II-nd year

No	Author, Title	Scores (0-15)	Average	Decision
1	<Author 1>.LPAI CourseWork Report	13(4),13(4),13(3)	13	ACCEPT
13	<Author 13>.LPAI CourseWork Report	11(3),13(4),13(2),13(4)	12.5	ACCEPT
15	<Author 15>.LPAI CourseWork Report	11(3),13(2),13(4)	12.3	ACCEPT
23	<Author 23>.LPAI CourseWork Report	11(4),7(2),11(4)	10.2	accept?
16	<Author 16>.LPAI CourseWork Report	11(3),11(1),9(3)	10.1	accept?
26	<Author 26>.LPAI CourseWork Report	9(1),9(3),11(4)	10	accept?
17	<Author 17>.LPAI CourseWork Report	9(3),9(2),11(4)	9.9	accept?
18	<Author 18>.LPAI CourseWork Report	9(3),11(4),7(1)	9.8	accept?
25	<Author 25>.LPAI CourseWork Report	9(3),11(3),9(2)	9.8	accept?
14	<Author 14>.LPAI CourseWork Report	7(2),7(3),11(4)	8.8	borderline
27	<Author 27>.LPAI CourseWork Report	9(4),11(4),3(2)	8.6	borderline
22	<Author 22>.LPAI CourseWork Report	7(2),7(3),11(3)	8.5	borderline
24	<Author 24>.LPAI CourseWork Report	9(2),7(3),9(4)	8.3	borderline
21	<Author 21>.LPAI CourseWork Report	7(2),11(3),5(2)	8.1	borderline
28	<Author 28>.LPAI CourseWork Report	9(3),7(4),7(3)	7.6	borderline
3	<Author 3>.LPAI CourseWork Report	3(4)	3	reject
5	<Author 5>.LPAI CourseWork Report	3(4)	3	reject
10	<Author 10>.LPAI CourseWork Report	3(4)	3	reject

(b) Score and rating table for the IV-th year

Fig. 3. Resulting score and rating tables communicated to the students.

– For IV-th year students – collected and processed using the EasyChair installation resulting in a very similar score table as the one in Fig. 2

The authors of the coursework reports have been further notified by e-mail about their results and ratings in the overall list as pictured in Fig. 3. Again, the notifications

[1] http://www.easychair.org/

to the II-nd year students have been manually communicated by e-mail; and the IV-th year students have been notified by the EasyChair.

4.1 Additional Effort for Tutors

As experienced, the additional instructors' effort for organizing and managing the peer review process was substantial.

The major part of their additional work could be qualified as the set-up effort: developing review forms; creating review environments; compiling briefing manuals for the student reviewers; preparing management tables; and configuring the software tools. The result of this effort may however be re-used quite substantially – so the start-up effort may be regarded as an initial investment and neglected in further considerations.

Following two different workflows for the II-nd and IV-th year students implied different management efforts because of using different toolsets. Overall, using EasyChair Conference Management System appeared to be about 3 times less effort consuming than using just e-mail and MS Excel.

4.2 Interpretation of Experimental Results

Table 4 contains the summary of our experimental findings and is structured as follows:

Table 4. Experimental Results.

Discipline	Year		Group No	Avg Score (0-20)	Avg Submission Ratio			Avg Score among Submitted	
	of Study	Calen-dar			No Submissions	Total Stu-dents	Ratio	Factual (0-20)	Aligned by Com-plexity
Programming (PR)	I	2009	4329	5,00	9	23	0,39	12,78	12,78
		2010	4320	10,33	9	15	0,60	17,22	17,22
		2011	4321	4,06	7	16	0,44	9,29	9,29
Algorithms and Data Stuctures (ADS)	II	2008	4327	9,68	21	31	0,68	14,29	14,29
		2009	4328	12,03	21	29	0,72	16,62	24,93
		2010	4329	8,50	10	19	0,53	15,30	30,60
		2011	4320	10,13	13	15	**0,87**	**11,69**	**29,23**
DataBases and Information Systems (DBIS)	II	2009	4327	11,70	14	23	0,61	19,21	28,82
		2010	4328	12,00	18	33	0,55	18,67	28,00
		2011	4329	11,00	12	19	0,63	18,33	27,50
Introduction to Logic Programming and AI (LPAI)	IV	2009	8216	4,47	10	36	0,28	16,10	24,15
		2010	4327	4,76	6	21	0,29	16,67	25,00
		2011	4328	7,14	15	28	**0,54**	**13,33**	**33,33**

- Broad horizontal sections correspond to the data related to one discipline. Two of them are baseline (as explained in Section 3.1) – Programming and Databases and Information Systems. The other two contain both control and experimental data – Algoritms and Data Structures (II-nd year) and Introduction to Logic Programming and AI (IV-th year).
- The Year column informs about the timing attribution of data (years of study and calendar years);
- The Group No column associates the rows to the academic groups. Group numbers may be found similar in several cases – reflecting the availability of both control and experimental measurements for several groups in different years and disciplines.
- The average scores are in fact based on the total number of students in a group which makes it different to the scores in the last two columns computed based on the number of submitted reports.
- Average Submission Ratio is in fact the measure that reflects the motivation of our students to submit their work
- The Factual Average Scores are the averages for the submitted reports, but without balancing them by coursework complexity
- Finally, the rightmost column contains the score averages multiplied by the complexity scaling factors provided in Table 1

Let us explain now how the results given in Table 4 and further interpreted graphically in Fig. 4 prove our research hypothesis.

Firstly, we expected that the introduction of peer reviews as an untraditional way of teaching will increase students' **extrinsic motivation**. This expectation was valid as pictured by the values of submission ratio. Indeed, the ratio of coursework submission in our experiment with the II-nd year students reached the global maximum of 0.87 across all the disciplines. The next lower value was 0.72 which is 15 per cent lower. For the IV-th year subjects the increase in motivation was not that significantly high overall, though very substantial within their year of study. Indeed the reached submission ratio of 0.54 is 1.86 times better than the next lower value of 0.29 in 2010.

Secondly, the quality of submitted reports may have been interpreted as quite average in our experiments: 11.69 in the II-nd year and 13.33 in the IV-th. The registered decrease in scores, compared to the previous year, is: 23.96 per cent for the the II-nd year; and 20.03 per cent for the IV-th year. A compensation for that decrease in quality is twofold:

(i) As the ratio of submissions increased the proportion of the best students (who always submit their work) decreased – so did the average scores. For the II-nd year the ratio increase was 15 percent versus a 23.96 decrease in scores. However, for the IV-th year the increase in submission ratio (86 per cent) substantially outperformed the decrease in average score (20.03 per cent). So, it could be concluded that our approach proved to be effective for the final year students of our Bachelor programme.

(ii) The observed decrease in scores is to some extent explained by the increase of coursework complexity. Indeed, the maximal values of the average scores have been reached in the cases with substantially less complicated coursework assignments – as explained in Table 1. For example, the global maximum of 19.21

corresponds to the assignment weighted 150 points. It is 'outperformed' by the score of 13.33 in our IV-th year experiment because the complexity of the experimental coursework is 250 points. This imbalance is corrected by the values shown in the Aligned by Complexity column of Table 4.

Figure 4 pictures the trends observed in our experiment graphically. The Y-values in Fig. 4(a) are the numbers from the Submission Ratio column of Table 4; while the Y-values in Fig. 4(b) are taken from the Aligned by Complexity column of this table.

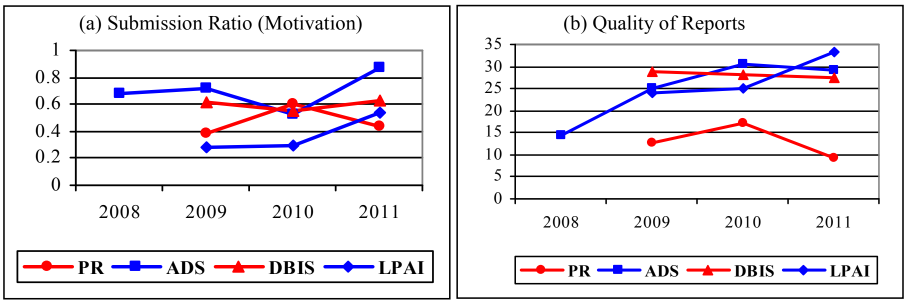


Fig. 4. Graphical interpretation of the increase in motivation and quality of work.

Finally, we hypothesized that the **objectivity of the assessments will be higher** in our experiment. We did not elaborate a proof for that as we did not undertake an experiment for scientifically measuring the objectivity. However, as a very draft estimation, we interviewed our subjects informally. These interviews revealed that the students treat their scores as more clear and objective compared to the previous experience, even if the scores were lower both individually and on average (column Factual of Table 4).

5 Conclusions and Outlook

This paper reported about our pedagogical experiment undertaken for seeking a way of improving extrinsic motivation and learning quality of Computer Science students in our Bachelor programme at ZNU. The increase in motivation has been proven convincingly. Exploiting students' aspirations for informal leadership and incurred competition constructively is effective and attracts people to learning. A gain in the quality of learning was a little bit over-estimated. Indeed, having involved more students in a creative learning activity does not guarantee that the quality of their work increases dramatically by miracle. However, the increase in motivation helped increasing also the quality to some degree – as shown in the previous section.

A good side effect is also that the students learn the working patterns of the professionals in their field broadly used in academia and industry for making qualitative and unbiased peer evaluations.

The results discussed in Section 4 appeared to be positive also for the other colleagues at the department of IT at our University. So, we plan to extend the experiment by covering more disciplines and collecting a broader sample of results in the near future. Among other things, this will allow us basing our work on a statistically representative set of subjects and making our results statistically valid. Finally, we plan to undertake an evaluation of the objectivity of the scoring in our settings.

References

1. Bound, J., Lovenheim, M.F., Turner, S.: Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources. Am Econ J. Appl. Econ. 2(3), 129--157 (2010), doi: 10.1257/app.2.3.129
2. Clark, P. B., Wilson, J. Q.: Incentive Systems: A Theory of Organizations. Administrative Science Quarterly, 6(2), 129--166 (1961)
3. Middleton, J. A., Photini A. Spanias: Motivation for Achievement in Mathematics: Findings, Generalizations, and Criticisms of the Research. J. Research in Mathematics Education, 30(1), 65--88 (1999)
4. Ames, C.: Classrooms: Goals, structures, and student motivation. J. Educational Psychology, 84, 261--271 (1992)
5. Middleton, J. A.: A study of intrinsic motivation in the mathematics classroom: A personal constructs approach. J. Research in Mathematics Education, 26, 254--279 (1995)
6. Ramirez-Iniguez, R., Canton, U.: Understanding Motivation in Large Groups of Engineering and Computing Students. In: Inspiring the next generation of engineers. Proc. Engineering Education (2010)
7. Wong, S.H.S: Motivating students to learn through good and helpful coursework feedback. In: Inspiring the next generation of engineers. In: Inspiring the next generation of engineers. Proc. Engineering Education (2010)
8. Likert, R.: A Technique for the Measurement of Attitudes. Archives of Psychology, 140, 1--55 (1932)