Person-Centered Learning using Peer Review Method

An Evaluation and a Concept for Student-Centered Classrooms

https://doi.org/10.3991/ijep.v8i1.8099

Dominik Dolezal(✉)
University of Applied Sciences Technikum Wien, Vienna, Austria
TGM – Vienna Institute of Technology, Vienna, Austria
dominik.dolezal@echnikum-wien.at

Alexandra Posekany
TGM – Vienna Institute of Technology, Vienna, Austria
Danube University Krems, Krems, Austria

Christoph Roschger, Gottfried Koppensteiner
TGM – Vienna Institute of Technology, Vienna, Austria

Renate Motschnig
University of Vienna, Austria

Robert Pucher
University of Applied Sciences Technikum Wien, Vienna, Austria

Abstract—Using peer assessment in the classroom to increase student engagement by actively involving the pupils in the assessment process has been practiced and researched for decades. This paper analyzes the applicability of peer assessment to exercises at secondary school level and makes recommendations for its use in computer science courses. Furthermore, a school pilot project introducing student-centered classrooms, called "learning office", is described. Additionally, a concept for the implementation of peer assessment in such student-centered classrooms is outlined. The evidence collected suggests that peer review is a viable option for small- and medium-sized exercises in the context of computer science education at secondary school level under certain conditions, which are discussed in this paper.

Keywords—Peer Review, Person-Centered Learning, Student-Centered Classrooms, Learning Office, Computer Science Education, Moodle Workshops

1 Introduction

Peer assessment has been used in the classroom as a method to increase student engagement by actively involving the pupils in the assessment process for decades. In 1983, Carl Rogers, who is seen as the inventor of the person-centered approach as a result of his research in client-centered psychotherapy [1], described a science course at a university involving peer assessment as an "inspiring addition to the documentation of a person-centered approach" and admired the professor's evaluation process in which "the students play a major part" [2, pp. 89–93]. Freiberg built on Rogers' work and provided evidence of the positive effects of applying person-centered principles to the classroom, citing a study by the National Consortium for Humanizing Education (NCHE) and summarizing that students "learn more and behave better when they receive high levels of understanding, caring, and genuineness than when they receive low levels of support" [3]. Motschnig et al. iteratively introduced person-centered principles to a computer science course in higher education, making it the best-rated bachelor-level course of the university excluding courses rated by less than five students [4], which indicates that introducing person-centered approaches such as peer review to computer science courses leads to a positive impact on the students' perception of the course as well as the learning effect.

Dochy et al. performed an extensive literature review [5] and depicted positive effects of peer assessment on students' learning as they become more involved in both the learning and assessment process. Gibbs analyzed reports and studies regarding students' experience of feedback in his article "How assessment frames student learning" in 2006 [6] and found indicators that suggest an increase in student performance if their work is peer reviewed by other students. Gibbs stated that it is not the quality of the feedback which increases student engagement, but rather the instantaneousness of feedback and the fact that it is peer reviewed. He derived eleven "conditions under which assessment supports student learning", six of which address feedback, depicting the importance of the quality, quantity and timing of feedback.

Bauer et al. conducted an empirical study which analyzed students' opinion on online peer review in the context of higher education, implementing a peer review system for a scientific writing course [7]. They concluded that students appreciate online peer review and highlight the importance of the review criteria. A computer science course addressing Unix shell programming in higher education was evaluated by Sitthiworachart et al. [8]. They concluded that students appreciate peer assessment as they realized their own mistakes by looking at the work of others and start thinking about their own work more deeply. In addition, most of the students were satisfied with the mark from the peer assessment. Gehringer used peer review in three computer science classes, an undergraduate one and two graduate ones, and evaluated the students' perception of the peer review method [9]. He found that students perceive peer review as being helpful to the learning process and value the feedback on their work.

However, reviews do not need to be done online: Figl et al. compared online peer reviews with face-to-face peer reviews in 2006 [10], focusing on collaboration aspects. They found out that face-to-face reviews improve communication as they promote discussions, but may be more effort for the students to conduct, which is why they recommend combining both methods. Standl developed a framework of educational patterns to be applied to computer science courses at secondary school level, including the peer check as one of the assessment methods for person-centered learning [11]. He suggests that students assessing each other learn more than students who only get assessed by the teacher. However, he recommends using this pattern primarily for projects as it is a time-consuming task.

This paper builds on the existing literature and includes published research [12]. It analyzes the applicability of peer assessment to smaller exercises in computer science classes at secondary school level. Courses bound to a strict curriculum as an external requirement may require students to learn certain topics through exercises, which is not typical for person-centered classrooms. However, introducing person-centered methods such as peer review to a traditional classroom setting may still provide the benefits of the peer assessment method, which is analyzed in this paper. Standl suggests using peer review after a project phase, which raises the question of whether peer assessment is also useful for small- and medium-sized exercises. Based on the reviewed literature, it seems natural that the advantages of peer review can also be observed when they are used for regular exercises in traditional classrooms.

Rogers questions the traditional school practices characterized by state-designated and prescribed curricula, standard tests and instructor-chosen grades, which reduce "meaningful learning" to an "absolute minimum" [2, pp. 11–21]. In this paper, we outline an approach to student-centered classrooms with individual advancement compliant to a given curriculum as well as the use of peer review as a person-centered learning method. These student-centered classrooms have been introduced within the scope of a school pilot project and are called "learning offices", which is derived from the German word "Lernbüro" [13].

This paper deals with the following research questions and makes recommendations for using peer review for regular exercises in computer sciences courses at secondary school level:

1. To what extent does peer review promote student-centered learning?
2. Is the feedback quality of students comparable to the feedback of a teacher?
3. Do students receive feedback in a timelier manner using peer review?
4. Does grading become more or less transparent?
5. Are reviews by peers a reasonable alternative to teacher assessment?
6. What is a reasonable number of exercises to be assessed by peers?
7. Are students overall satisfied with the peer review method?

2 Methodology

2.1 Test Setup

Two secondary school classes in their 13th year of study, hereafter referred to as "A" and "B", consisting of 29 and 28 students respectively were introduced to the peer assessment method within the scope of the same software engineering course. Both classes had 13 exercises to be assessed throughout the software engineering course. Two of these 13 exercises were assessed by peers, while the other eleven exercises were assessed by one of the two teachers. The students were asked to evaluate those two exercises using an anonymous online questionnaire and compare them to exercises evaluated by the teacher. At the end of the course, they were asked to rate the learning motivation for doing all 13 exercises. The software engineering course used a blended learning concept, i.e. the lessons were supported by an online Moodle course which contained all relevant information, learning material, discussion forums, and completed exercises as well as grading. The two peer reviews have been conducted using the Moodle workshop activity, which allows students not only to upload their solution for the respective exercise, but also to grade each other. A workshop activity consists of five phases [14]:

1. **Setup.** During this phase the instructor describes the exercise and provides instructions for the students. Another essential task of this phase is creating the assessment form: teachers define criteria to be used by the students to grade each other. A criterion has a description and can be graded using points or a scale. One of the most important tasks of the instructor is defining clear and concise criteria for the assessment phase, which should be mutually agreed upon with the students.
2. **Submission.** Students can submit their solutions during the submission phase and, if set, the submission deadline has not yet been exceeded. After the students have uploaded their solutions, the instructor can randomly or manually assign submissions to the students. By default, a random assignment uses at least 5 reviewees per submission. However, this number can and should be adjusted according to the preferences of the teacher and the students. Meeting the submission deadline is crucial as assigning late submissions requires a lot of effort.
3. **Assessment.** During this phase students review the solutions of their colleagues if the optionally configurable deadline has not yet been exceeded. Students use the assessment form for grading, which has been defined by the teacher in the setup phase.
4. **Grading evaluation.** As soon as the workshop activity has been switched to the grading evaluation phase, students can no longer edit their assessments. In this phase the teacher can review the peer assessments and manually override them.
5. **Closed.** When the instructor closes the workshop, the grades as well as the feedback are posted to the students' gradebook.

2.2 Assignments

The students were asked to work on each of the 13 exercises for one to three weeks. The two peer assessments have been conducted using an iterative approach: students' feedback on the first peer assessment and the lessons learned have been discussed in the classroom and incorporated into the second one. The first peer reviewed exercise involved implementing a simple person database using the Java Platform, Enterprise Edition (Java EE), which "provides a standards-based platform for developing web and enterprise applications" [15] used for implementing multitier applications. The second assignment requested students to develop a graphical interface for a route planner using a RESTful API. Representational State Transfer (REST), initially developed by Fielding and defined in his dissertation [16], is a set of design criteria [17] for an architectural style for distributed systems. For this exercise, the students were asked to use the Google Maps Directions API [18] and implement a graphical user interface employing Qt and PySide [19].

After the students read the instructions and review criteria, they were given two weeks to solve the Java EE exercise and one week to implement the RESTful client. After the submission the students were randomly assigned five to six reviewers and five to six reviewees for the peer review. The whole process was designed to be anonymous: the reviewees did not know their reviewers and vice versa. This is untypical for person-centered approaches; however, it helps to reduce prejudice.

2.3 Evaluation

To quantitatively assess the students' attitude towards the peer review method, both classes were asked to voluntarily fill out an anonymous questionnaire to give feedback on both peer assessments and estimate the impact on several factors of learning. They were asked to compare the peer assessment with regular teacher assessment and rate their level of agreement from "disagree"[1] (1) to "agree" (5) on a semantic differential scale, a "generalizable technique of measurement" [20] developed by Osgood et al. to measure the meaning of words. Unlike the Likert scale, which provides labels for each possible option in its originally published form [21], a semantic differential scale only labels the end points and visually indicates a continuous scale to simulate an interval scale [22]. The students rated their agreement with the following statements:

1. I have received more feedback than usual.
2. I have received the feedback in a timelier manner.
3. The overall quality of the feedback was higher.
4. I studied the task's subject matter more thoroughly.
5. I have learned something new from the solutions of others.
6. I think that others learned something new from my solution.
7. Grading was more transparent.
8. I am overall very satisfied with the teacher assessment method.
9. I am overall very satisfied with the peer assessment method.

Furthermore, the students were asked to define their preferred number of exercises to be reviewed by peers: 100%, 75%, 50%, 25%, or 0% of all exercises. Finally, they could give positive feedback and ideas for improvement through two open questions.

At the end of the school year, the students were asked for another rating of all 13 exercises regarding their motivation to learn. They were asked to rate each exercise from "not motivating" (1) to "motivating" (5). The students were shown the results directly after they filled out the form and gave feedback in a final discussion.

The following error bars represent 95% confidence intervals calculated using the t-distribution first described by Gosset [23]. The bars should give an impression of the overall variability and do not necessarily indicate implications for a larger population as the students are not representative of it. The mean value with confidence intervals

[1] The rating scales, feedback questions, and students' quotes have been translated from German.

was chosen over the median value with the interquartile range as the visualizations are more powerful and intuitive. To account for multiple testing during the analyses, the level of significance was adjusted to α=0.005, which is equivalent to a Bonferroni correction for 10 simultaneous tests [24].

3 Results

The following charts show the mean level of agreement with the described statements.

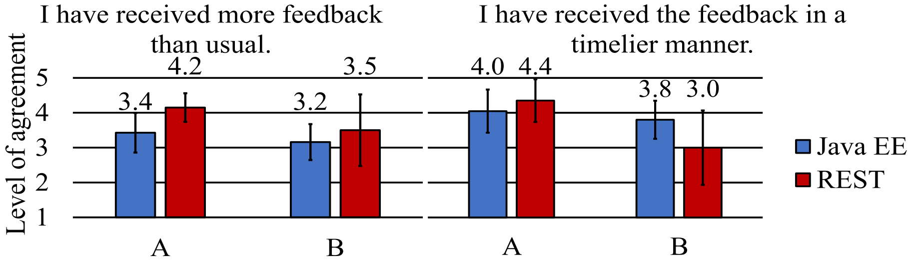

Fig. 1. Feedback quantity and timing

Fig. 1 presents the reported feedback quantity and timing of the first and second peer reviewed exercise. The students agreed in the second iteration with an average of 3.9 that they received more feedback when the exercise was peer reviewed, which is explainable by the number of reviewers: five assessments provide more feedback than a single assessment by the teacher. The measured increase in feedback quantity from the first iteration to the second is explainable by the revised feedback modalities. Written feedback was optional in the first iteration, which led to students giving only ratings, but no suggestions for improvement. This issue was discussed with the students and they agreed on giving written feedback on each submission in the second iteration. The students therefore report that they receive more feedback on peer reviewed exercises than on exercises assessed by the teacher.

The students also perceived a measurable improvement in the timing of feedback with an average level of agreement of 3.9 for both iterations. Receiving feedback on an exercise within one week seems to be faster than the average assessment time required by a teacher. Due to administrative work and late submissions, the second iteration needed two weeks for the grading phase in class B, which explains the drop. Verbal answers regarding what the students liked about the quantity and timing of the feedback were:

“Mostly more feedback than by a teacher”

“Fast feedback”

“Exercises were graded within one week. You don’t have to wait for a month for a feedback.”

“Guarantee that the exercise is graded within one week”

“Feedback within one week”

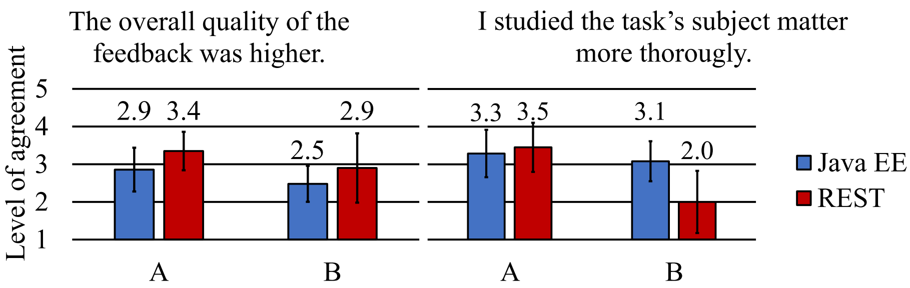

Fig. 2. Feedback quality and student engagement

Fig. 2 depicts the feedback quality as well as the self-reported student engagement. The students were not too pleased with the quality of the feedback in the first iteration as they rather disagreed with this statement. After the mentioned change from an optional written feedback to a mandatory feedback, there was a distinct increase in feedback quality to a total response average of 3.2, indicating that the quality of the feedback by students is indeed comparable to the feedback of a teacher. This suggests that students give reasonable feedback like a teacher would.

The reported student engagement seems to be slightly better than when only assessed by the teacher: the average level of agreement of both iterations was 3.1, whereas the response of class A reached 3.5 in the second iteration. The noticeable drop of student engagement in the second iteration of class B can be explained by analyzing the qualitative feedback of the students: three students who disagreed with this statement wrote that they experienced an unfair deduction of points, resulting in frustration and a drop of student engagement. Such cases should be discussed with the respective reviewer and reviewee to clarify the problem and mutually agree on a fair grading. The teacher's role as a facilitator here is to minimize unfair penalties. Furthermore, the REST exercise was less open and less creative, which seems to have a limiting effect on student engagement. Some answers to what they liked about feedback quality and student engagement were:

"Feedback of others is indeed helpful"
"Receiving hints which are not given by teachers in some cases"
"Altogether, better feedback than usual"
"Finally useful feedback!"
"Detailed feedback"
"Suggestions for improvement"
"More in-class communication about problems and solutions"
"Pupils study the subject matter more in-depth"

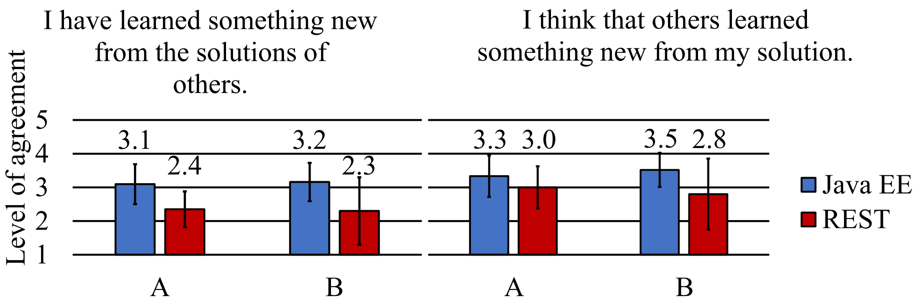

Fig. 3. Learning effect

The results of the evaluation of the own learning effect gained from the reviewing process as well as the estimation of the learning effect of others are shown in Fig. 3. Most of the students reported to have learned something in the first iteration, while they rather disagreed with this statement in the second iteration. The reason again lies within the assignment: while the Java EE exercise allowed many different solutions (free choice of database, user interface, validation etc.) and was more creative, the REST exercise had more predefined elements and contained a suggested user interface as a screenshot. Therefore, the solutions to the REST exercise were similar, which is why students could not really learn new and different techniques. This suggests that peer review is especially suitable for more creative and open exercises. Verbal feedback on what they liked about the learning effect includes:

"You can see what others did better/worse"

"Making sure that everything works on different computers"

"Many hints"

"A reasonable, informative comment. I am happy!"

"I could give other students reasonable feedback, maybe even more than a teacher due to his limited time"

"Grading good and bad solutions (Dos and Don'ts)"

"You learn different coding styles"

"Seeing different approaches"

"Seeing how others solved the exercise"

"You see how others solved the exercise. This improves the learning effect."

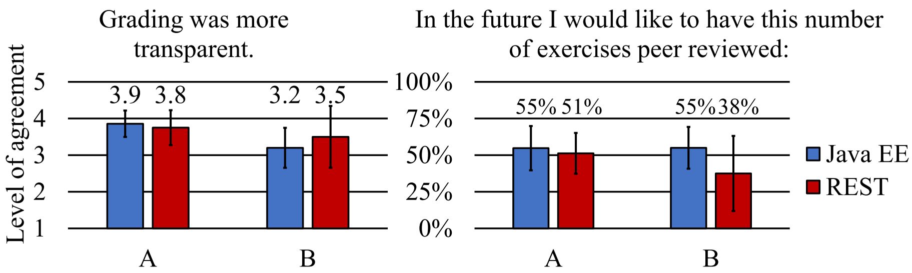

Fig. 4. Transparency of grading and number of peer review exercises

Fig. 4 shows the reported transparency of grading as well as the preferred number of exercises being peer reviewed. Students report that peer assessment follows a more transparent grading scheme than teacher assessment. The average level of agreement on both iterations was 3.6; only about 12% disagreed with this statement. Five reviews seem to give a better estimate of the grade than a single assessment by the teacher as seen by the students.

On average, the students would prefer about half of the exercises (52%) peer reviewed. The noticeable drop in class B can again be explained by the frustration of some students due to reported unfair assessment. Qualitative feedback on what they liked about grading included:

"Criteria clear and understandable"
"You know what to focus on and what is important for grading"
"A wide selection for a precise grading"
"Mainly fair and understandable feedback"
"The criteria were more precise this time"
"More transparent"

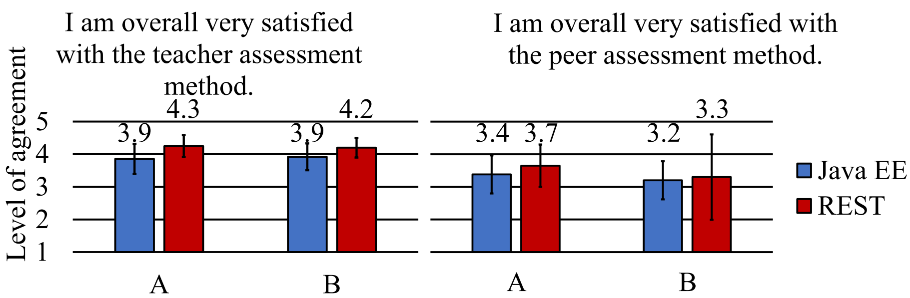

Fig. 5. Satisfaction with teacher assessment and peer assessment

The self-reported satisfaction with teacher assessment and peer assessment is depicted in Fig. 5. The students still like to be assessed by the teacher. Although they report to like the peer assessment method with a total average of 3.4, the students still prefer the opinion of the teacher as an expert. One of the students stated: "Teachers are always the best at grading. It's simple and reasonable." This can also be a compliment to the teacher as they are very satisfied with his or her teaching and grading. However, although students value high-quality feedback from a teacher, peer assessment still seems to trigger higher student engagement, as also suggested by Gibbs' research [6]. Furthermore, the students themselves stated in the second iteration that student feedback is of the same quality as teacher feedback.

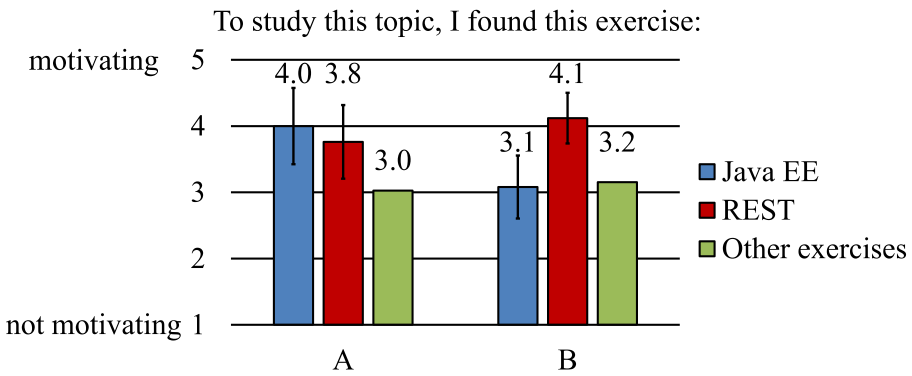

Fig. 6. Learning motivation

Fig. 6 shows the results of the final rating of all 13 exercises regarding their motivation to study the respective topic. In three of four cases, the peer reviewed exercise received a rating notably better than the average of all other exercises. A single factor analysis of variance (ANOVA) [25] shows that the difference between the mean values of all 13 groups is highly significant ($p=1.36*10^{-14}$). A non-parametric Wilcoxon rank sum test with continuity correction [26], [27] against the null hypothesis that the true location shift equals 0 also returns $p=2.20*10^{-4}$. Interestingly, class B rated the second peer reviewed exercise distinctly better than the first one. This seems to contradict their initial feedback right after the peer reviewed exercises were completed as they reported more positive effects in the first iteration. In the discussion following the final rating of all 13 exercises, the students stated that the Java EE exercise was outside the sphere of their interest, which led to a low rating for the Java EE exercise.

4 Correlation Analysis

To better understand the analyzed factors of learning as well as their interactions with each other, a correlation analysis using Spearman's rank correlation coefficient [28] has been carried out. Unlike Pearson's correlation [29], the Spearman correlation is nonparametric and can be used for monotonic functions, including ordinal data. The p-values have been derived from a two-tailed t-test based on the approximation using the t-distribution as described in [30].

Table 1. Spearman correlation coefficients of analyzed learning factors (Java EE exercise)

	Quantity	**Quality**	**Timing**	**Student Engagement**	**Own Learning Effect**	**Others' Learning Effect**	**Transparency**	**Satisfaction Peer Review**	**Satisfaction Teacher Assessment**	**Peer Review Amount**
Quantity		0.56 ($5.0*10^{-5}$)	0.19 (0.202)	0.15 (0.323)	0.42 (0.003)	0.26 (0.084)	0.32 (0.030)	0.45 (0.002)	0.18 (0.225)	0.43 (0.003)
Quality	0.56 ($5.0*10^{-5}$)		0.39 (0.007)	0.41 (0.005)	0.39 (0.007)	0.32 (0.031)	0.29 (0.054)	0.59 ($1.4*10^{-5}$)	0.16 (0.301)	0.54 ($1.2*10^{-4}$)
Timing	0.19 (0.202)	0.39 (0.007)		-0.07 (0.661)	-0.09 (0.566)	0.02 (0.913)	0.07 (0.624)	0.37 (0.010)	0.01 (0.949)	0.38 (0.009)
Student Engagement	0.15 (0.323)	0.41 (0.005)	-0.07 (0.661)		0.31 (0.037)	0.34 (0.020)	0.22 (0.142)	0.35 (0.019)	-0.16 (0.277)	0.44 (0.002)
Own Learning Effect	0.42 (0.003)	0.39 (0.007)	-0.09 (0.566)	0.31 (0.037)		0.39 (0.007)	0.25 (0.100)	0.29 (0.050)	-0.12 (0.434)	0.19 (0.214)
Others' Learning Effect	0.26 (0.084)	0.32 (0.031)	0.02 (0.913)	0.34 (0.020)	0.39 (0.007)		0.18 (0.243)	0.26 (0.077)	0.08 (0.577)	0.30 (0.043)
Transparency	0.32 (0.030)	0.29 (0.054)	0.07 (0.624)	0.22 (0.142)	0.25 (0.100)	0.18 (0.243)		0.39 (0.008)	0.02 (0.874)	0.44 (0.002)
Satisfaction Peer Review	0.45 (0.002)	0.59 ($1.4*10^{-5}$)	0.37 (0.010)	0.35 (0.019)	0.29 (0.050)	0.26 (0.077)	0.39 (0.008)		0.19 (0.208)	0.73 ($1.2*10^{-8}$)
Satisfaction Teacher Assessment	0.18 (0.225)	0.16 (0.301)	0.01 (0.949)	-0.16 (0.277)	-0.12 (0.434)	0.08 (0.577)	0.02 (0.874)	0.19 (0.208)		-0.02 (0.911)
Peer Review Amount	0.43 (0.003)	0.54 ($1.2*10^{-4}$)	0.38 (0.009)	0.44 (0.002)	0.19 (0.214)	0.30 (0.043)	0.44 (0.002)	0.73 ($1.2*10^{-8}$)	-0.02 (0.911)	

Table 1 represents the correlation matrix of the reported levels of agreement with the presented statements of both classes at the time the Java EE exercise was carried out. The matrix shows the Spearman's correlation coefficients between the respective factors as well as the p-values. The table was designed to be symmetric for easier use and simple lookup. The highest significant correlation with r_s=0.73 ($p=1.2*10^{-8}$) has been measured between the satisfaction with the peer review method and the preferred number of peer assessments in class, which is a reasonable connection: the more satisfied students are with this method, the higher the number of reviews they would prefer. In turn, the satisfaction is strongly linked to the reported feedback quality with r_s=0.59 ($p=1.4*10^{-5}$). If students are satisfied with the quality of the received feedback, they also tend to be satisfied with the peer review method itself. Feedback quality is also significantly connected to feedback quantity and the preferred number of peer assessments in class with r_s=0.56 (p=5.0*10^{-5}) and r_s=0.54 (p=1.2*10^{-4}) respectively.

The measured correlations between the preferred number of peer assessments in class with the transparency of grading as well as the feedback quantity are also significant with r_s=0.44 (p=0.002) and r_s=0.43 (p=0.003) respectively, suggesting that not only feedback quality, but also feedback quantity and transparency of grading are crucial factors in favor of the peer review method. The preferred number of peer assessments also significantly correlates with the reported student engagement with r_s=0.44 (p=0.002). Even if less strongly and not significant with this sample size, further factors also seem to impact the preferred number of reviews and satisfaction: the timing of feedback relates to them with r_s=0.38 (p=0.009) and r_s=0.37 (p=0.010) respectively; student engagement relates to satisfaction with r_s=0.35 (p=0.019).

Table 2. Spearman correlation coefficients of analyzed learning factors (REST exercise)

	Quantity	Quality	Timing	Student Engagement	Own Learning Effect	Others' Learning Effect	Transparency	Satisfaction Peer Review	Satisfaction Teacher Assessment	Peer Review Amount
Quantity		0.42 (0.020)	0.31 (0.091)	0.25 (0.181)	-0.05 (0.782)	0.09 (0.655)	0.32 (0.088)	0.23 (0.212)	0.12 (0.529)	-0.04 (0.842)
Quality	0.42 (0.020)		0.18 (0.347)	0.43 (0.018)	0.34 (0.066)	0.24 (0.194)	0.41 (0.023)	0.42 (0.020)	-0.05 (0.799)	0.44 (0.014)
Timing	0.31 (0.091)	0.18 (0.347)		0.51 (0.004)	0.18 (0.350)	0.22 (0.235)	0.38 (0.037)	0.15 (0.419)	0.13 (0.484)	0.17 (0.370)
Student Engagement	0.25 (0.181)	0.43 (0.018)	0.51 (0.004)		0.34 (0.063)	0.40 (0.029)	0.50 (0.005)	0.49 (0.006)	-0.10 (0.607)	0.60 (4.2*10^{-4})
Own Learning Effect	-0.05 (0.782)	0.34 (0.066)	0.18 (0.350)	0.34 (0.063)		0.51 (0.004)	0.26 (0.170)	0.41 (0.024)	-0.09 (0.632)	0.72 (6.1*10^{-6})
Others' Learning Effect	0.09 (0.655)	0.24 (0.194)	0.22 (0.235)	0.40 (0.029)	0.51 (0.004)		0.20 (0.284)	0.41 (0.025)	0.05 (0.791)	0.42 (0.020)
Transparency	0.32 (0.088)	0.41 (0.023)	0.38 (0.037)	0.50 (0.005)	0.26 (0.170)	0.20 (0.284)		0.47 (0.008)	-0.15 (0.425)	0.43 (0.017)
Satisfaction Peer Review	0.23 (0.212)	0.42 (0.020)	0.15 (0.419)	0.49 (0.006)	0.41 (0.024)	0.41 (0.025)	0.47 (0.008)		-0.06 (0.770)	0.70 (1.4*10^{-5})
Satisfaction Teacher Assessment	0.12 (0.529)	-0.05 (0.799)	0.13 (0.484)	-0.10 (0.607)	-0.09 (0.632)	0.05 (0.791)	-0.15 (0.425)	-0.06 (0.770)		-0.33 (0.075)
Peer Review Amount	-0.04 (0.842)	0.44 (0.014)	0.17 (0.370)	0.60 (4.2*10^{-4})	0.72 (6.1*10^{-6})	0.42 (0.020)	0.43 (0.017)	0.70 (1.4*10^{-5})	-0.33 (0.075)	

The correlation matrix of the analyzed factors of learning in the second iteration is shown in Table 2. This time, the highest correlation has been found between the own learning effect and the preferred number of peer assessments with $r_s=0.72$ ($p=6.1*10^{-6}$), which was not the case in the first iteration. This is a new and interesting finding: the correlation seems to be stronger in exercises which have a more predefined solution compared to exercises allowing creativity. This may indicate that the students who did not manage to completely solve an exercise with a predefined solution especially benefit from the peer review method. The second highest correlation has been measured between the satisfaction with the peer review method and the number of peer assessments in class with $r_s=0.72$ ($p=6.1*10^{-6}$), which is again a reasonable connection.

The preferred number of peer assessments is also linked to the student engagement with $r_s=0.60$ ($p=4.2*10^{-4}$), while the student engagement seems to correlate with the perceived transparency of grading with $r_s=0.50$ ($p=0.005$) and the satisfaction with $r_s=0.49$ ($p=0.006$), although the correlation was not significant at the chosen significance level. The transparency of grading may correlate with the satisfaction and feedback quality with $r_s=0.47$ ($p=0.008$) and $r_s=0.41$ ($p=0.023$) respectively, which might be proven with a larger sample size. The links between feedback quality and the preferred number of peer assessments as well as the satisfaction could not be proven to be significant this time. Feedback timing also did not highly correlate with the number of peer reviews or with the satisfaction this time, which could be explained by the fact that students had to wait two weeks for the feedback instead of one, showing the importance of instantaneous feedback. However, the timing significantly correlates with the student engagement with $r_s=0.51$ ($p=0.004$), suggesting that fast feedback may promote student engagement.

To sum up, the satisfaction with the peer review method as well as the preferred number of peer assessments seem to be especially connected with the following factors:

1. Student engagement: students who report a higher student engagement also wish to have more exercises assessed by peers. This correlation proved significant in both iterations. This may suggest that the peer review method promotes student engagement of those students who are satisfied with this method, or that engaged students tend to favor peer assessment.
2. Learning Effect: the learning effect was strongly connected to the preferred number of peer reviews in the second iteration. This indicates that students who learned from other submissions favor a higher number of peer reviews. Peer review may therefore benefit the learning effect.
3. Feedback quality: the first iteration revealed a significant correlation between the quality of feedback and satisfaction with the peer review method as well as the preferred number of peer assessments in class, indicating that peer assessment and feedback quality are strongly connected. The second iteration showed a similar correlation, however, it did not prove significant with the given sample size.

4. Feedback quantity: feedback quantity also correlated with the satisfaction and the preferred number of peer reviews in the first iteration. This may indicate that the number of reviewers is especially important for open exercises.
5. Transparency of grading: the preferred number of peer reviews correlated with the perceived grading transparency in the first iteration. Students perceiving this method as fair also tend to give higher ratings for this method, while students who seemingly experienced unfair gradings do not favor this method. The second iteration also revealed a similar non-significant correlation.

Although the timing of the feedback did not prove significantly linked to the satisfaction and the preferred number of peer assessments, it seems to correlate with the student engagement. Moreover, the second iteration, which involved a grading phase of two weeks instead of one, showed a distinctly weaker correlation between feedback timing and the rating of this method, which depicts the importance of fast feedback.

5 Recommendations

Based on the introduced empirical results, the following recommendations for promoting student-centered learning using the peer review method in computer science courses at secondary school level in traditional classrooms can be formulated:

1. **Qualitative feedback.** An agreement to provide written feedback should be made with the students. By default, giving qualitative feedback is optional in the Moodle workshop activity.
2. **Anonymous feedback.** Although this is unusual for person-centered approaches, reviewer and reviewee should be anonymous to reduce prejudices. This maximizes transparency of grading and prevents "upvoting" and "downvoting". This may sound easy at first glance, but students may be used to putting their name on their submissions.
3. **Fast feedback.** One of the main advantages of peer assessment is fast feedback. In order to be useful, feedback should be given in a timely manner [6]. One week seems to be a reasonable time for computer science exercises.
4. **Black-box testing.** The rating criteria should be formulated in a way that every student – regardless whether he or she was able to solve the exercise – can assess them. Some students and teachers raised the objection that students may not be qualified to assess the assignments of others. However, if the criteria focus on features, students can rate the submissions from a user's point of view.
5. **Transparent criteria and conditions.** The rating criteria as well as the general conditions for giving feedback need to be communicated and agreed upon. This ensures that students use the same rating scale and grading becomes reproducible.
6. **Final grading by the teacher.** Although most of the peer review grades did not need to be changed, it is important to listen to the students when they experienced an unfair grading. The issue should be discussed with the teacher being the facilitator to clarify the problem and mutually agree upon a fair grade.

7. **Shared level of basic knowledge.** Students need a certain level of expertise in order to give good feedback on the solutions of others. If they still struggle with computer science basics, it is questionable whether they are able to thoroughly test a program, even if the criteria are formulated from a user's perspective.
8. **Exercises allowing creativity.** Peer review seems to be especially useful if there are multiple possible solutions for the exercise as the students seem to learn more from each other and are more engaged. This corresponds with the findings of Standl [11], who recommends using peer review for project-based assignments. The more behavioristic an exercise is, the less powerful peer review becomes.
9. **Do not overuse it.** Students report that they would be fine with a peer review on every other exercise. However, the two peer reviews were a refreshing alternative to the teacher assessments in this case. They are still time-consuming, both for the teacher and the students, and could lose their charm if they are overused.

6 Peer Review in Student-Centered Classrooms

6.1 Learning Office

The concept of a learning office or learning atelier features a studying environment supporting students in self-organized individualized learning. A pioneer in the field, Margret Rasfeld, described the learning office by several distinct attributes [13], [31]. Firstly, learning contents are split into well-defined modules which the students work on independently. Self-explanatory materials allow the students to work at their own pace and current individual level. The students learn self-organization as they have to plan ahead, carry out and finish their modules in order to cover the total content of the subject. Secondly, instead of presenting materials, the teachers provide aid in the organization of students' study plans, similar to the "work contracts" in the "experiment" described in Rogers' book [2, pp. 45–56], as well as help them to structure and revise their learning efforts. Instruments for structuring the students' learning are their personal logbooks, learning paths and regular tutorials. Thirdly, while working on contents is individualized and certain situations like oral presentations require working alone, the students are also encouraged to work in groups and tutor each other. Fourthly, in this concentrated working atmosphere not only are the students more aware of their own learning status and goals, but also are the teachers enabled to individually support them. Finally, in order to complete a module, the students have to successfully pass tests. Upon completion of a module, they receive a certificate with detailed feedback and recommendations for building their future work on these acquired skills.

For each subject, the students in the learning office have to complete a predefined set of tasks or exercises. These are typically either marked as mandatory or as optional, where in order to positively complete the course, students have to do all of the mandatory exercises. If this has been achieved, the number of optional exercises solved contribute to the final mark between one and four. Alternatively, in some classes where it is harder to distinguish between mandatory and optional topics, exercises may contain both. In these cases, grading an exercise is not dual but takes into account to what extent the tasks have been fulfilled.

Additionally to exercises, in many subjects there are written examinations for each module which are also taken into the final grade, either at self-chosen individual date within a given timeframe or at a fixed date for all students of a class. Besides serving as a tool for grading, these examinations play another crucial role within the learning office: they propose a timeline for the students indicating a deadline until all tasks of a module have to be fulfilled. As it turns out, such a proposed timeline significantly helps students in structuring their efforts. These subjects usually have the exercises and tasks arranged in a linear order where the topics sometimes strongly build on each other, such as applied mathematics. Some subjects on the other hand provide exercises without a given order, only being dependent on exercises within the same module, but the modules themselves are largely independent.

Grading of exercises and tasks is done in different ways dependent on the subject:

- **In-class evaluation.** In some subjects, the students have to do practical tasks in class and present their solutions to the teacher. They get direct feedback from the teacher, whether the task is finished or whether certain aspects still have to be refined.
- **In-class examination.** Sometimes a task may consist of a written examination about a (sub)module. The student approaches their teacher at the beginning of a lesson and is given a set of questions which have to be solved during the lesson.
- **Off-class evaluation.** Assignments are handed in on the online platform Moodle to be corrected or graded later. This typically includes a first feedback loop by the teacher where errors are pointed out and the student is given the chance to correct them. Only after the second round of handing in assignments, the actual grading process starts.

Exercises may be turned in as hand-written materials, which is typical for languages and several tasks in applied mathematics. An alternative is handing in computer-based exercises on Moodle, which happens more frequently in technical and IT subject.

6.2 Concept for Peer Review in the Applied Mathematics Learning Office

The subject applied mathematics has structured modules based on booklets which provide detailed descriptions, explanations and exercises for the students and help them to create their personal scripts with definitions, descriptions, graphs and solved exercises. These hard-copy booklets are supported by online materials in module-based Moodle courses. In addition to the restructured material of the booklets, the online courses provide explanatory videos created by their teachers which guide students through the examples. Furthermore, students get the chance to check their current learning advancement by trying online exercises on Moodle, which are individually created from a predefined set of questions covering the basic contents of the module. A big advantage is that these online self-checks provide immediate feedback to the students without the fear of feeling embarrassed in front their peers or teacher if they are not yet ready for a test.

Materials and in-class evaluations are manually checked by the teacher who also provides individual feedback to the students on their performance. Unlike the computer science course presented in the previous chapters, students have 4-5 years less experience at school and therefore lack the requirements for a fully student-centered reviewing and evaluation system. Nonetheless, the learning office is designed for students tutoring each other. In the learning environment of the applied mathematics learning office, two usages of an additional peer reviewing process step are about to be implemented for in-class evaluation and off-class evaluation. An important aspect is to apply this method mainly to exercises in a first reviewing process which does not involve grading of the person who receives the feedback.

The person giving feedback is graded for their competence and motivation in providing feedback, to ensure a certain quality on the one hand and to allow them to receive a gratification for their efforts showing a deeper understanding than their peers. Giving peer review requires a higher level of understanding than just solving a problem oneself and allows the students with a deeper understanding to reach this aspect of the subject which could not be gained by just going step by step through their own work.

In addition, for the in-class examination, we can add an additional loop where the students who have already finished this task may share their knowledge with their peers who have just taken their test, explaining to them errors and mistakes pointed out by the teacher during the examination. In this way, the learning office's intended synergies are applied as not only do the students with a deeper understanding provide the others with feedback on their in-class work, but they are also able to reach a higher cognitive level of competence in the subject.

7 Discussion

We analyzed the applicability of the peer review method to small- and medium-scaled exercises in computer science courses at secondary school level to introduce person-centered approaches to traditional classrooms. Based on empirical evidence collected over one year, the following answers to the research questions have been found:

1. To what extent does peer review promote student-centered learning?

Peer review seems to promote student-centered learning if the method is used correctly. The results indicate a clear improvement in feedback quantity and timing as well as student engagement and motivation to learn. In addition, the students liked the peer review method and regarded it as a refreshing alternative to predominating teacher assessments.

2. Is the feedback quality of students comparable to the feedback of a teacher?

Yes. Students report that feedback quality is indeed comparable to the feedback of a teacher if written feedback is given.

3. Do students receive feedback in a timelier manner using peer review?

Yes. The students received feedback on their assignments faster.

4. Does grading become more or less transparent?

The students stated that grading became more transparent, which is explainable by the higher number of persons who assess the submission. Furthermore, teachers have to pay special attention to the definition of the rating criteria for a peer review, so that the grading scheme of exercises is sufficiently transparent.

5. Are reviews by peers a reasonable alternative to teacher assessment?

Yes. Peer review seems to be a reasonable alternative to teacher assessment in computer science courses. Nevertheless, some constraints need to apply in order to make it a useful tool for teaching and assessing.

6. What is a reasonable number of exercises to be assessed by peers?

Students report that about every other exercise could be reasonably peer reviewed. However, peer review should not be overused. The exact number of exercises depends on the type of exercise.

7. Are students overall satisfied with the peer review method?

Yes. Students are satisfied with peer review as an assessment tool. However, they report that they still value the high-quality feedback of an expert.

8 Conclusion

To conclude, the overall feedback on the peer review exercises was very positive. The students reported that the quality of the feedback by students is comparable to the feedback of a teacher if written feedback is provided. The students received the feedback faster and they valued that grading was more transparent, because they received more than one grading. In addition, the teacher needs to pay special attention to the rating criteria. Peer review seems to be a reasonable alternative to teacher assessment; about every other exercise could be peer reviewed according to the students' feedback.

The correlation analyses revealed a strong and significant correlation between the preferred number of peer assessments and student engagement, which could indicate that this method promotes student engagement of students who are satisfied with the peer review method. The learning effect was highly correlated with the preferred number of peer reviews in the second iteration and seems to be higher in exercises with a predefined solution, which makes this method particularly attractive for applied mathematics. Feedback quality and feedback quantity were strongly connected to the satisfaction with the peer review method as well as the preferred number of peer reviews in the first iteration, which may indicate that feedback quality and quantity is especially important for open exercises. The transparency of grading was strongly connected to the preferred number of peer assessments in the first iteration. Although the timing of the feedback did not prove significant in both iterations, it correlated with student engagement in the second iteration.

Overall, the students were satisfied with this method. However, the students reported that they still value the high-quality feedback of a teacher, which can be a compliment to the teacher as they are very satisfied with his or her teaching and grading. It was found that some additional constraints such as open assignments as well as obligatory and fast feedback need to apply to make peer review practicable and reasonable for small- and medium-scaled exercises in traditional classrooms. We plan to further investigate peer review with different subjects and its use in student-centered classrooms to follow up on our current findings.

9 Acknowledgment

We would like to thank our students for their participation in the two peer reviews as well as their openness and willingness to share their experiences and feedback with us. Without their help, we would not be able to carry out our research in the field of education and pedagogy.

10 References

[1] Rogers, C. R. (1959). A Theory of Therapy, Personality and Interpersonal Relationships as Developed in the Client-centered Framework. Psychology: A Study of a Science, 3, pp. 184-256

[2] Rogers, C. R. (1983). Freedom to Learn for the 80's (2nd edition). Columbus, Ohio: C.E. Merrill Pub. Co.

[3] Rogers, C. R. and Freiberg, J. H. (1994). Freedom to Learn (3rd edition). Columbus, Ohio: C.E. Merrill Pub. Co., p. 253

[4] Motschnig, R., Sedlmair, M., Schröder, S. and Möller, T. (2016). A Team-Approach to Putting Learner-Centered Principles to Practice in a Large Course on Human-Computer Interaction. 2016 IEEE Frontiers in Education Conference, pp. 1-9 https://doi.org/10.1109/FIE.2016.7757576

[5] Dochy, F., Segers, M. and Sluijsmans, D. (1999). The Use of Self-, Peer and Co-assessment in Higher Education: a review. Studies In Higher Education, 24 (3), pp. 331-350 https://doi.org/10.1080/03075079912331379935

[6] Gibbs, G. (2006). How assessment frames student learning. Innovative Assessment in Higher Education, pp. 23-36

[7] Bauer, C., Figl, K., Derntl, M., Beran, P. P. and Kabicher, S. (2009). The Student View on Online Peer Reviews. SIGCSE Bulletin, 41 (3), pp. 26-30 https://doi.org/10.1145/1595496.1562892

[8] Sitthiworachart, J. and Joy, M. (2003). Web-based peer assessment in learning computer programming. 3rd IEEE International Conference on Advanced Technologies, pp. 180-184 https://doi.org/10.1109/ICALT.2003.1215052

[9] Gehringer, E. F. (2000). Strategies And Mechanisms For Electronic Peer Review. 30th ASEE/IEEE Frontiers in Education Conference, vol. 1, F1B/2-F1B/7 https://doi.org/10.1109/FIE.2000.897675

[10] Figl, K., Bauer, C. and Mangler, J. (2006). Online versus Face-to-Face Peer Team Reviews. 36th ASEE/IEEE Frontiers in Education Conference, pp. 7-12 https://doi.org/10.1109/FIE.2006.322469

[11] Standl, B. (2013). Conceptual Modeling and Innovative Implementation of Person-centered Computer Science Education at Secondary School Level. Doctoral thesis, University of Vienna, p. 118

[12] Dolezal, D., Motschnig, R., Pucher, R. (2017). Peer Reviews In Person-Centered Classrooms: Computer Science Education At Secondary School Level. 20th International Conference on Interactive Collaborative Learning, pp. 233-242
[13] Koppensteiner, G., Lepuschitz, W., Leeb-Bracher, U., Hollnsteiner, K. and Merdan, M. (2017). Educational Programs for Fostering Entrepreneurial Skills and Mindset. 20th International Conference on Interactive Collaborative Learning, pp. 1716-1724
[14] MoodleDocs: Using Workshop (February 25, 2017). Accessed December 6th, 2017. https://docs.moodle.org/32/en/Using_Workshop
[15] Gupta, A. (2013). Java EE 7 Essentials. Sebastopol, California: O'Reilly Media, Inc., p. 1
[16] Fielding, T. R. (2000). Architectural Styles and the Design of Network-based Software Architectures. Doctoral thesis, University of California, Irvine
[17] Richardson, L. and Ruby, S. (2007). RESTful Web Services. Sebastopol, California: O'Reilly Media, Inc., p. 80
[18] Google Maps APIs: Directions API – Getting Started (June 7, 2017). Accessed December 6th, 2017. https://developers.google.com/maps/documentation/directions/start
[19] Qt Wiki: PySide (December 2, 2016). Accessed December 6th, 2017. https://wiki.qt.io/PySide
[20] Osgood, C. E., Suci, G. P. and Tannenbaum, P. H. (1957). The Measurement of Meaning. Urbana, Illinois: University of Illinois Press, p. 76
[21] Likert, R. (1932). A Technique for the Measurement of Attitudes. Archives of Psychology, 22 (140), pp. 5-55
[22] Tullis, T. and Albert, B (2013). Measuring the User Experience (2nd edition). San Francisco, California: Morgan Kaufmann Publishers Inc., p. 18
[23] Student (1908). The Probable Error of a Mean. Biometrika, 6 (1), pp. 1-25 https://doi.org/10.1093/biomet/6.2-3.302
[24] Hochberg, Y. (1988). A Sharper Bonferroni Procedure for Multiple Tests of Significance. Biometrika, 75 (4), pp. 800-802 https://doi.org/10.1093/biomet/75.4.800
[25] Fisher, R. A. (1934). Statistical Methods for Research Workers (5th edition). Edinburgh, Scotland: Oliver and Boyd, pp. 198-272
[26] Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. Biometrics Bulletin, 1 (6), pp. 80–83 https://doi.org/10.2307/3001968
[27] Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. New York City, New York: McGraw-Hill, pp. 75-83
[28] Spearman, C. (1904). The Proof and Measurement of Association between Two Things. The American Journal of Psychology, 15 (1), pp. 72-101 https://doi.org/10.2307/1412159
[29] Pearson, K. (1895). Note on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London, 58 (347-352), pp. 240-242 https://doi.org/10.1098/rspl.1895.0041
[30] Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). Numerical Recipes in C: The Art of Scientific Computing (2nd edition). New York City, New York: Cambridge University Press, p. 640
[31] Education Innovation Lab: Werkzeugkiste Lernbüro (2015). Accessed December 6th, 2017. http://www.htl.at/fileadmin/content/Individualisierung/ARGi/VNT_2015/toolkit-LB-Werkzeugkiste_LB.pdf

11 Authors

Dominik Dolezal is a researcher and lecturer at the University of Applied Sciences Technikum Wien, Austria. He is also a school teacher at the TGM – Vienna Institute of Technology, Austria. His interests lie in the field of computer science education, person-centered learning, software development and user experience design.

Alexandra Posekany is a researcher and biostatistician at the Danube University Krems, as well as lecturer at the Universities of Applied Sciences Technikum Wien and Campus Wien, Austria. She is also a school teacher at the TGM – Vienna Institute of Technology, Austria. Her interests lie in the field of mathematical and statistical education, applied statistics and machine learning.

Christoph Roschger is a researcher and lecturer who obtained his PhD in computer science from the Vienna University of Technology in the field of mathematical logic, game theory, and semantics of natural language. After lecturing at the University of Applied Sciences Technikum Wien, Austria, he now works at the TGM – Vienna Institute of Technology, Austria, and has already taught several subjects in the learning office.

Gottfried Koppensteiner is the head of the IT-department at the TGM – Vienna Institute of Technology, Austria, where he founded the learning office. He is also an active researcher at the Practical Robotics Institute Austria (PRIA) working on robotics in education as well as in research. Gottfried obtained his PhD from the Vienna University of Technology in the field of industrial automation.

Renate Motschnig is head of the Educational Technologies research group at the University of Vienna, Austria. She holds a double assignment to the Centre for Teacher Education and the Faculty of Computer Science. She (co-)authored more than 170 scientific articles and four books. Her interests lie in the field of person-centered learning in combination with software engineering.

Robert Pucher is head of the Department of Computer Science at the University of Applied Sciences Technikum Wien, Austria. He studied and wrote his dissertation at the Graz University of Technology and has been working for the University of Applied Sciences since 1994.

This article is a revised version of a paper presented at the ICL2017 conference held in Budapest, Hungary, 27-29 September 2017. Article submitted 19 December 2017. Resubmitted 11 January 2018. Final acceptance 04 February 2018. Final version published as submitted by the authors.