Student-activation by employing Student Peer Review with Criteria-Based Assessment

Omid Mirmotahari and Yngvar Berg
Department of Informatics
University of Oslo
Ester Fremstad and Crina Damsa
Department of Education
University of Oslo

Abstract—**Evaluation of student projects by students themselves is highlighted in research as a method to activate the students, and to strengthen their learning and meta-cognition. Assessing others students' achievements can be a way for beginner students to develop insight into the academic criteria and methods in a subject. It can also contribute to their understanding of their own learning process and how they best can make progress with the studies. We used the Canvas functions Rubrics and SpeedGrader to manage a review of 524 student projects. What examined various aspects: how the students experienced this type of learning activity; whether the students' assessments were in line with the teachers' review; in what way and to what extent, reviewing each others' work contributed to their learning; and how data generated through this activity provides information to teachers to improve their teaching. Despite some technical shortcomings and challenges, the experiment generated a series of insightful results.**

Keywords: **Student-activation, Peer Review, Canvas, Constructive Alignment, Feedback, Rubric assessment, Criteria-based Assessment.**

I. INTRODUCTION

The way responsibility for self-learning, meta-cognition, transparency related to learning goals and assessment, feedback, and student activities are affecting working methods, goals and assessment (constructive alignment) (Biggs & Tang, 2007) are highlighted as key elements to ensure quality in higher education (Evans, 2013; Bransford, Brown, & Cocking, 2000; Nicol & MacFarlane-Dick, 2006).

Research results by (McMahon & Thakore, 2006) show that 'constructively aligned' teaching contributes to increased transparency, better learning, more valid and reliable assessment, increased student activation, and increased coherence in learning outcomes. (Morris, 2008) shows that in subjects with aligned teaching students achieve higher average grades, and that there is a strong correlation between the students' sense of achievement and achievement on the exam. (Volante, Beckett, Reid, & Drake, 2010) shows that involvement in the assessment leads to participants taking more responsibility and self-learning management. Students become more involved in the learning and assessment process. (Gibbs, 2006) studies show that it is not the quality of the feedback from the review that increases student activation, but rather that the students know that their assignment is being assessed by

fellow students. It is especially important for the first year students to get a better and deeper understanding of the subject while at the same time undergoing personal development related to general professional competence (Dochy, Segers, & Sluijsmans, 1999; Boud, 2000; Ngar-Fun Liu & Carless, 2006; Hounsell, McCune, Hounsell, & Litjens, 2008; Mirmotahari & Berg, 2017). Each re-assessment provides useful learning outcomes for both the recipient and the one who addresses the task. Students learn what is appreciated in an answer and to look for common mistakes and deficiencies, which helps to give the students a meta perspective on their own understanding and learning (Mirmotahari & Berg, 2018).

The administrative aspect of using reassessment can be time consuming and challenging for the teacher, especially for subjects with over 100 students. There are several online programs that handle the administrative part of submission and assignment of tasks. We have chosen to use Canvas as a Learning Management System (LMS) and use Canvas's own review by Rubrics and SpeedGrader to manage individual assessment for the 562 enrolled students in the introductory course IN1020 at the University of Oslo.

The motivation for carrying out this pilot study has been to answer the following research questions:

(1) In what way do students recognize their fellow students' level of knowledge?
(2) What do students emphasize in their assessments?
(3) How do students experience the use of Canvas LMS?
(4) How data collected through the intervention provides information to the teacher about the students' learning and professional understanding?

In this article, a description of the research method is given in section II, while the results are presented in section III. The article concludes with a discussion of the various findings and results, where we answer the research questions and accompanying proposals for further work.

II. METHOD

The subject IN1020 - Introduction to computer technology is a compulsory first semester subject. As the subject is compulsory for all students admitted to the five study programs at the department of Informatics at the University of Oslo, the student body is a highly heterogeneous group. The students
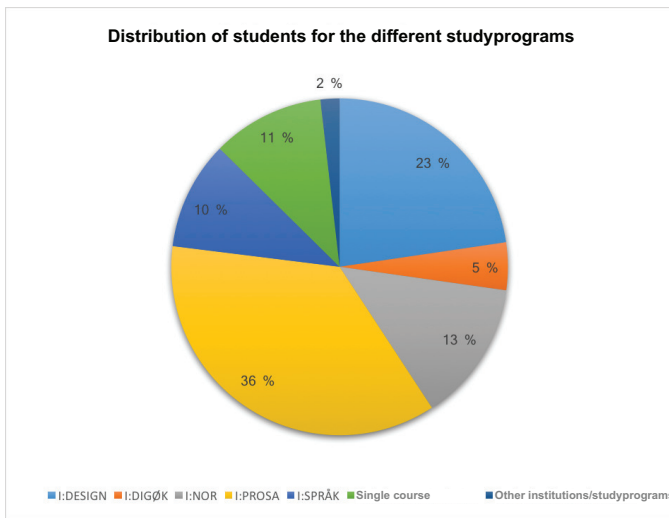
Distribution of students for the different studyprograms

I:DESIGN I:DIGØK I:NOR I:PROSA I:SPRÅK Single course Other institutions/studyprograms

Fig. 1: *The distribution of the 562 enrolled students and their representative study programs. There are very uneven distribution of sex on the various programs, but on average are gender distribution of 70% men and 30% women.*

have a significant differentiated STEM background, due to the admission requirements for the representative study programs. There were a total of 562 students who were enrolled for the subject in the autumn of 2017 and the distribution between the various study programs is shown in Figure 1. The course consists of four focus areas in the fields of hardware, programming (low level), network and security. Each of these focus areas has been taught by different teachers. The teaching in the subject extends over 14 weeks with two hours of lectures, two hours of group tuition per week. In advance of the exam, all students must have passed 3 compulsory assignments. The final grade, passed/non-passed is based only on the final written 4-hour exam. The exam in 2017 consisted of 37 assignments with different weight, but 25% for each of the four focus area.

In this study we focus on the first compulsory submission. This submission is in two parts, the first part being a circuit implementation of a 4-bit multiplier, while in the second part the students review each others project hand-ins. In the first submission they will upload both the circuit design they have implemented in LogiSim, as well as a complementary report describing the method and solution they have used. 524 responses were submitted within the deadline. As part 2 (review), every student receive automatically at least three fellow student answers that they are to assess.

The involved parties in this experiment have been teachers, group/seminar teachers and students. The collection of data is done by using Canvas, questionnaire, course evaluation conducted by the academic committee and qualitative interviews with the students.

A. Canvas layout

We chose to use Canvas for the management of the peer review. We chose to use the built-in features; like Rubrics

and SpeedGrader. Rubrics is a classified form that the teacher sets up as an assessment template. Students will use this form to assess each others project. It was set up so that the students would submit their mandatory assignment only within a given deadline. Then, those who delivered within the deadline will receive at least three fellow students' project to review (after five days). The deadline for the individual review was set to one week. We chose to use a rubric with different weighting and attributes, see Table I. The section is divided into four sections, where the first two parts, paragraphs (1) - (2) and (3) - (5) are assessments based on the submitted circuit implementation and report respectively. Part three and four, (F) and (G), are purely administrative points for the students' assessment of the assignment to approved/non-approved and a grade assessment. These last two points are not included in the score that each student receives. In the analysis that follows, the distribution between the different points will emerge more clearly. After completion of the assessment, students will be able to access the assessments made to review their own tasks.

III. RESULTS

The student's exam results are cross-linked with the results from the first compulsory assignment. Although the exam assessment has been passed/failed, it has been possible to detect what the students have achieved by examining the marks on the exam part that are relevant to this test. The 2017 exam was designed in such a way that it was possible to achieve 25 points on this part in the exam and for convenience to compare, we have used a character distribution based on 91-80-60-49-40% (of 25 points) for ABCDE and F, respectively, below 40%.

36% of the students received the same grade on the exam and from the evaluation by their fellow students, and 46% received $\pm$ 1 grade. This means that a total of 82% of the students already by the first compulsory assignment would knew how they performed in the course. Compared with the 7-day calendar time frame that the students had to carry out a review, 82% reported that it was enough time for the academic assessment, but they think that specific technical Canvas issues were time-consuming. In terms of anonymity, the majority believes that it is important, 17% of students say it is important and 42% believe it is very important. One of the issues the research literature focus on is the student's confidence in the professional competence of his fellow students. In this test, the majority of the students correspond as expected, 76% have no or little confidence in their fellow students knowlegde for assessing this assignment. In particular, this is evident from the comment of student Q:1.

> *"Unnecessary with fellow evaluation, in which case it must be made better demands for reflection and review of own work versus fellow students."*
> *(Student #2476934)*
>
> Q:1

Criteria	Ratings				
(1) Does the implementation work in LogiSim?	Yes		Partly		No
(2) How is the structure of the implementation?	Very clear and high degree of structure.	Good structure.	Has structure, but everything is in a window (no sub-circuits).	Messy.	No structure.
(3) Does the report meet the requirements set in the task?	The requirements are fulfilled and the report contains a conceptual sketch of the circuit as a whole as well as describing sub circuits in detail. The organization is done in a logical and efficient manner that relates important information to each other.	The requirements are met, but could have been better with a better focus	The requirements are met, but the flow is somewhat weak; should have been made some changes to make it easier to read.	Some requirements are not met.	Has significant deficiencies for basic information.
(4) How good is the candidate to use specific terms and convey his solution in the introductory 300-word section describing the solution and other alternative solutions?	Many subject-relevant words are used. Theory and mathematics are used to describe the application in a way that makes it easy to understand the solution. The section is written with good structure and relevant information is used.	Subject-relevant terms are used. Theory and mathematics are used to describe the application, but are written in a way that makes the flow difficult to understand.	Some subject-relevant terms are used and the description isincluded, but not complete. Relevant mathematical concepts are mentioned, but not exploited.	The text is difficult to follow and understand. The description is not complete.	Few or no professional terms are used and the description of the solution is difficult to understand. Irrelevant or no math is used to make the logic concrete.
(5) How well is the relationship between the description of the solution and the solution.	The solution is as a visual drawing of the description. Everything described has been implemented in a well-defined manner.	The solution is relevant, but contains important elements that are not described.			The assignment is solved in a completely different way than described.
(F) Would you have approved this assignment?	YES			NO	
(G) What grade would you like to give this solution?	A	B	C	D	E	F

TABLE I: The table shows the assessment criteria used in Canvas Rubrics.

The compulsory assignment consists of two parts, circuit implementation and report, where both parts will be included in the assessment. The rubric is used for assessing both part and indicate whether the assignment has been passed or not passed. For each row in the rubric, the score is within 1 - 5, except for the criterion (1) where the scale 0-2 is used and the criterion (5) where the scale 0-3-5 is applied. It is possible to earn a total of 22 points. The control in rubric (F) and (G) is not included in the score calculation. It has not been determined by the teacher what the grade limit for approved/failed should be, nor the grade limits. There are several interesting ways to analyze how the students have evaluated each other and what they emphasize in relation to approved/failed and grade. In Figure 2 the distribution of the assessments made with regard to the score (converted to %) for the circuit implementation and the report respectively is shown. The red points represent failed projects, while the blue represents approved. Although the students have not stated the limit for approved and failed, the teacher has defined a limit, this is marked with a green line. As the figure shows, the majority of those who have been approved are within the upper quadrant to the right, while the majority of failed are outside. To characterize the deviations we can look at the figures, Figure 3 and Figure 4. These figures show which part of the task that has been most weighted to give approved or not approved. There is a direct correlation to the fact that the
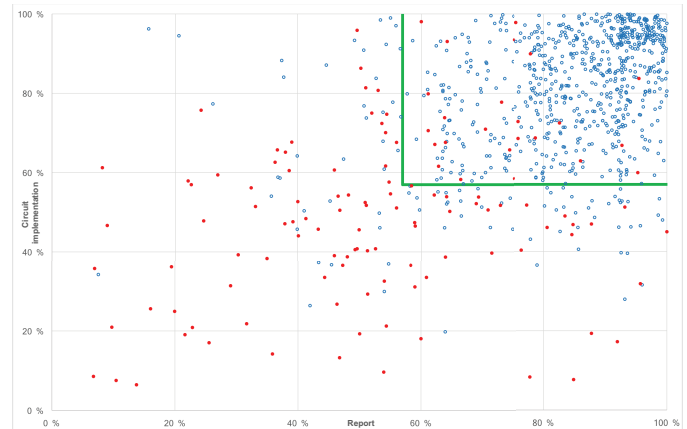

Fig. 2: The blue dots represent approved tasks while red dots represent assignments not approved. Upper quadrant on the right, limited with green lines, represents the amount of tasks that fall under the teacher's requirements for approved assignment, equivalent to 13 points.

circuit implementation works or not, which is the significant determining factor for whether the task is approved or not. Figure 3 shows the overview of the tasks that have been approved, but have a total score below 13 points. As the figure shows, the report is the significant factor, where the impact
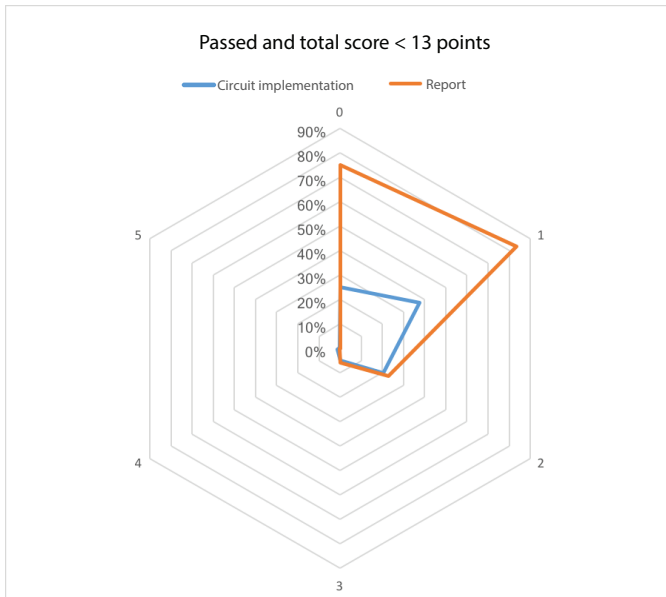
Fig. 3: *The relationship between the score for the circuit implementation and the report in terms in significance. How many student received the representative points (0-5) for the circuit implementation and report respectively. The selection includes answers were the students have considered to be approved but obtained a total score lower than the teacher's limit for passing, which is 13 points.*
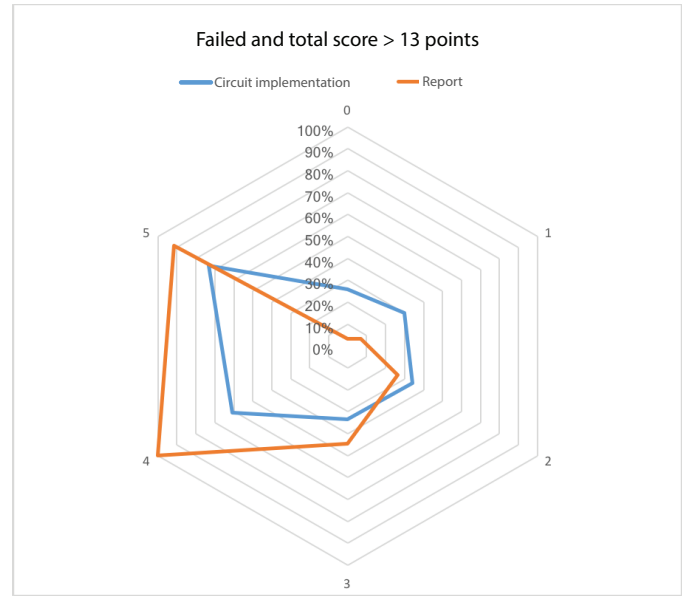

Fig. 4: *The relationship between the score for the circuit implementation and the report measured in significance. The number of students received the representative points (0-5) for the circuit implementation and the report respectively. The number of assignments which students have considered as not passed but overall score is higher than the teacher's limit for approval, which is 13 points, is shown As the graph shows, the report is not significant to get approved or not approved.*

is mainly 0 and 1, while the implementation of the circuit has little impact on the final score. For the opposite case, we see the same trend in Figure 4, where the radar plot shows the results for the not approved assignments, but with a score higher than 13 points. Even though the outcome of the report also fits the top of the score, it is still the implementation of the circuit implementation that overrides the assessment for approved and failed. It is clear that those who have been failed with a score higher than 13 points had very good reports.

Figure 5 shows the deviation of the students' grade compared to the teachers evaluation. The students did not receive the teacher's grades. The green columns, for grades A and C, represent where the teacher has more candidates for the respective grade, while red is where the students have the most candidates. The total deviation relative to the number of answers is 11%. The data shows has an average of 17% that is actually significant in relation to the grade scale, which in turn may suggest a significant difference in final grades. In particular, it is shown in Figure 6 where the plot represents the character distribution given the score that the students have considered. Since this intervention is applied on a compulsory assignment, it is not entirely unexpected that the majority of the students carry out the task with a high score and thus correspondingly to grade A.

A. Experience with the functionality of Canvas

The peer review was executed as an anonymous review in Canvas, but unfortunately all the students submitted their project with their name on the cover page. This limited anonymity even though Canvas had hidden filename and sender. We achieved a partial anonymity, where all the reviews became anonymous to the recipient. Implementing a review of such a large number of participants leads to many practical challenges such as deadline delays, stability in Canvas when many use it at the same time, system crash on local machine, etc. This contributes to frustration both for students and organizers of the project. It's very unfortunate that first year students get such an experience of Canvas that can affect their efficacy and weaken learning outcomes.

The students reported a high degree of frustration when using Canvas. They stated in particular two cases, one regarding to the save button in SpeedGrader, which was often and unpredictably inactive. This led to the students being unsure whether their review had been passed on or not. The other was mainly aimed at the technical functionality of Canvas. Many students wanted to do the assessment over time, but it is not possible to suspend an assessment and reviews can not be edited afterwards. An element that indirectly influenced the students experience when using Canvas was the possibility to set up the assignment to be reviewed by a teacher, this feature was not chosen because of work capacity. Thus, therefore,
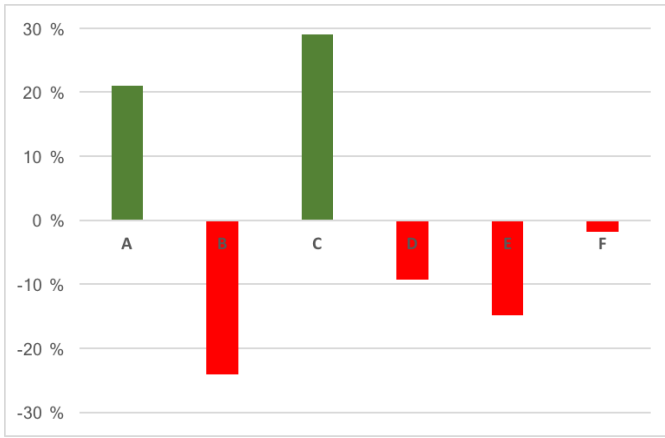
Fig. 5: *The discrepancies between the student's and the teacher's grades. The relative value is positive, green, where the teacher has more approved candidates and corresponding red, where the students have more approved candidates. The total deviation relative to the number of answers is 11%.*

there was no barrier or filter for the students review. We missed the opportunity for the students to give each other a feedback on the assessment they have received, alternatively, to flag reviews that they find inappropriate. We only experienced one review (out of 1572 reviews) that was very inappropriate.

"It was hard to find out how to give feedback to other fellow students. It was a problem with the feedback form, so you have to wait for it to suddenly worked again. Feedback was not saved, buttons disappeared, layout and popup of tables worked bad and that it was hard to find all required functionality"   Q:2

*(*Student #2486411*)*

IV. DISCUSSION

An interesting issue that has emerged as a result of student review is trust in their own ability to make a fair assessment and to provide qualified feedback to fellow students. This is also pointed out by (Mulder, Baik, Naylor, & Pearce, 2013). Likewise, the question of trust is that your own answer is considered fair by fellow students, as (Cheng & Warren, 1997) also points out. We see that the students have limited confidence in each others level of knowledge and thus, initially, are somewhat skeptical about the assessment given by fellow students and in contrast to the assessment by the teacher. This is expected, because the students may have a negative opinion of their own level of competence in the evaluation. Thus, they emphasize to a lesser degree the fellow students' evaluation. We see from the grade distribution that the students are consistently more critical than the teacher and that they are unable to evaluate the answers for a grade A. It is not
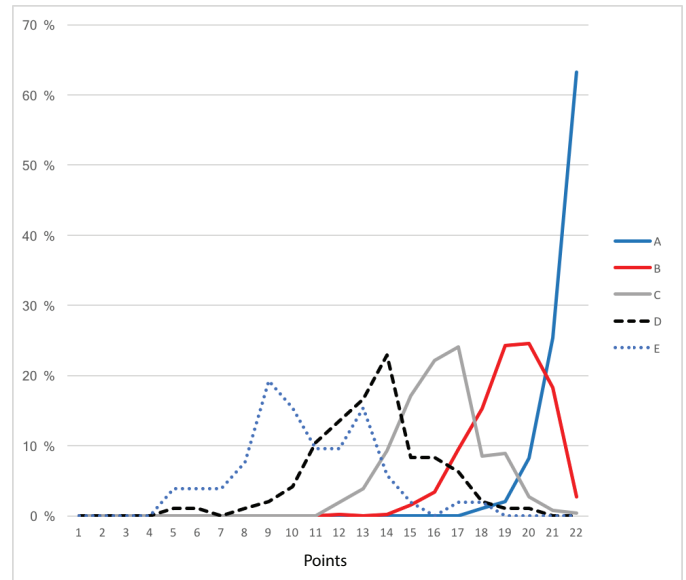

Fig. 6: *The plot shows the character distribution given how the students have assessed each other's assignments in terms of total grades. The distribution between the different grades is by average 19 for B, 16 for C, 13 for D and 11 for E. For example, we see that in the range of 11-15 points, the majority of candidates falls into grade D*

surprising that students lack experience to characterize answers that are very good. We see from the distribution of grades that the students evaluated many of the fellow students to grade B. This may be a result of the fact that many average students evaluated the answers as somewhat better than they themselves perceive as their own level of knowledge. We also see that the percentage of weak grades is higher among the students' evaluation than the teachers evaluation. This may indicate that the students notice specific weak aspects and, to a lesser extent, were able to recognize positive elements in such projects. Although the grades by teachers and students did not show significant deviations, there is no doubt that there is a big difference in their experience of assessing answers.

It is also evident that the students emphasize concrete circumstances, especially if the circuit implementation works or not. This is especially important for approved/failed assessment. Such concrete, and not absolute correct answers, yields a simplified evaluation on the final result given by the students. It also appears that students appreciate and weight the report less, probably due to the lack of effort put into their own report. For future surveys it will be very interesting to look into how those students that had a good report considered fellow students' reports. As a continuation of this experiment, we can implement machine learning algorithms for the textual report and use it to determine if it is possible to determine which factors determine a good and a less good report. Given such a large amount of data used in this study, it will be sufficient for a first-order test of machine learning.

Regarding the process and the experience of participating in

each other evaluation, there is little doubt that this helps the students become more aware of how evaluations are executed and what is important to be aware of in order to improve evaluation of their own work. Awareness about evaluation methodology is important for students in exam situations and to enhance the trust in final grades. This is also evident in constructive alignment regarding the transparency aspect of the learning design.

V. Conclusion

We see that the students have limited confidence in fellow students skills to evaluate and, initially, are more skeptical to the assessment given by fellow students and than given by the teacher. This is somehow expected, because the students may have a negative opinion of their own level of competence in the evaluation of others. We see from the grade distribution that the students are consistently more critical than the teacher and that they are almost unable to evaluate to grade A. It is not surprising that students lack experience and thus avoided the grade A. Furthermore, we see from the distribution of grades that the students gave more often grade B. This may be a result of the fact that many average students evaluate the answers as a bit better than they themselves perceive as their own level of knowledge. We also see that the percentage of weak grades is greater among the students' evaluation. This may indicate that they notice specific weak respondents and, to a lesser extent, were able to recognize positive elements. Although the grades given by teachers and students did not show significant deviations, there is no doubt that there is a big difference in their experience in assessing answers.

References

Biggs, J., & Tang, C. (2007). *Teaching for Quality Learning at University Third Edition Teaching for Quality Learning at University* (Vol. 3th edition (1th edition 1999)). Open University Press. doi: 10.1016/j.ctcp.2007.09.003

Boud, D. (2000). Sustainable Assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, *22*(2), 151–167. doi: 10.1080/713695728

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How People Learn*. doi: 10.17226/9853

Cheng, W., & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education*, *22*(2), 233–239. doi: 10.1080/03075079712331381064

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The Use of Self-, Peer and Co-assessment in Higher Education: a review. *Studies in Higher Education*, *24*(3), 331–350. doi: 10.1080/03075079912331379935

Evans, C. (2013). Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research*, *83*(1), 70–120. doi: 10.3102/0034654312474350

Gibbs, G. (2006). How assessment frames student learning. In *Innovative assessment in higher education* (chap. 2).

Hounsell, D., McCune, V., Hounsell, J., & Litjens, J. (2008). The quality of guidance and feedback to students. *Higher Education Research & Development*, *27*(1), 55–67. doi: 10.1080/07294360701658765

McMahon, T., & Thakore, H. (2006). Achieving constructive alignment: Putting outcomes first. *Aukšojo mokslo kokybè*(03), 10–19.

Mirmotahari, O., & Berg, Y. (2017). Individuell "automagisk" tilbakemelding på skriftlig eksamen. *Nordic Journal of STEM Education*, *1*(1), 287–293.

Mirmotahari, O., & Berg, Y. (2018). Structured peer review using a custom assessment program for electrical engineering students. *IEEE Global Engineering Education Conference, EDUCON*, *1*. doi: 10.1109/EDUCON.2018.8363339

Morris, M. M. (2008). Evaluating university teaching and learning in an outcome-based model : replanting Bloom. *PhD thesis, School of Mathematics and Applied Statistics, University of Wollongong.*, 370.

Mulder, R., Baik, C., Naylor, R., & Pearce, J. (2013). How does student peer review influence perceptions, engagement and academic outcomes? A case study. *Assessment & Evaluation in Higher Education*, *39*(6), 1–21. doi: 10.1080/02602938.2013.860421

Ngar-Fun Liu, & Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher Education*, *11*(3), 279–290. doi: 10.1080/13562510600680582

Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and selfregulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199–218. doi: 10.1080/03075070600572090

Volante, L., Beckett, D., Reid, J., & Drake, S. (2010). Teachers' views on conducting formative assessment within contemporary classrooms. *Annual Meeting of the American Educational Research Association*, 1–23.