LONGITUDINAL EVALUATION OF SELF-ASSESSMENT AND PEER REVIEW IN A CAPSTONE COURSE

Peter TÖRLIND[1], Lisa LARSSON[1] and Lars EKLÖF[2]
[1]Product Innovation, Luleå University of Technology, Sweden
[2]Design, Luleå University of Technology, Sweden

ABSTRACT
One crucial part of education is teaching students to critically evaluate and reflect on their work. One way to perform this is through peer review and self-assessment. In this research paper, we present the results of a longitudinal study over five years with 239 students following the implementation and evaluation of peer review and self-assessment. Using qualitative and quantitative analysis, we explore different types of self-assessment, the benefits of incorporating self-assessment into the learning process, and lessons learnt during the years. Results show that students appreciate assessing their own and others' work. The students in the study are very good at evaluating their capabilities, the difference between the self-assessment and teachers' final assessment was about 10%. With a studio-based approach, with formative feedback throughout the process, individual oral and written presentations and support from self-assessment, team feedback and teacher discussions, there is a much higher certainty that students are assessed accurately.

Keywords: Assessment, teamwork, peer learning

1 INTRODUCTION
One crucial part of education is teaching students to critically evaluate and reflect on their work. One way to perform this is through peer review and self-assessment. By peer review, students spend time reading and assessing other people's work, learning to determine what is good and bad [2]. Peer review can also be combined with self-assessment, a critical process that allows students to reflect on their learning and personal growth. It helps them identify their strengths, weaknesses, and areas for improvement, which is essential for personal growth and professional development. Peer review and self-assessment are valuable tools for self-regulated learning because it helps the students to set goals, monitor progress, identify and address areas of difficulty, and adjust strategies as needed. In this research paper, we present the results of a longitudinal study over five years with 239 students following the implementation and evaluation of peer review and self-assessment. Using qualitative and quantitative analysis, we explore different types of self-assessment, the benefits of incorporating self-assessment into the learning process, and lessons learnt during the years.

2 BACKGROUND
A common problem when assessing design teams is that teamwork can introduce free riding, where students rely on their teammates to carry the workload. Also, students in a group may have varying skills or abilities, and those who are struggling may hold the group back, while more advanced students may become frustrated with the pace or level of work. This can be a frustrating experience for other group members and lead to decreased overall performance. For the teacher, it can be hard to assess individual contributions accurately. Students working in a group often tend to specialise in specific areas based on their existing skills and interests. For example, strong vocal students are responsible for presentations, strong writers are responsible for documentation and students that excel in sketching and rendering is responsible for illustrations. As a result, some students may conform to their comfortable roles rather than challenge themselves and develop new skills. So, in the assessment, it is vital that all learning objectives of a course are assessed on the individual level. To evaluate the goals, some kind of

framework is needed. In 2015, the competence profiles [3] for Industrial Design Engineering were developed based on supporting students' understanding of the essential characteristics of an industrial design engineer. The goal was that the framework should help teachers and students understand how and with what quality a specific competence should be developed. The framework consists of eight different categories. Each category is further divided into sub-categories, visualised as a progression chart where the student starts as a beginner and can develop into an expert (an example of progression in communication is available in Table 1).

2.1 Peer review

Peer review between students has several benefits because it gives students constructive feedback from their peers, which can help them improve their writing and critical thinking skills. This feedback is often more relatable and understandable than feedback from teachers or professors. Something that is often missing is that students need to learn how to evaluate others' work and give good feedback. One way of doing this is facilitated peer-review sessions [1], where several students read a text, provide individual feedback and then have to discuss their feedback with others. This encourages collaboration, highlights different views, and creates a shared understanding of good and bad. It is also shown that the actual feedback is not the most crucial part. However, by reading and evaluating others' documentation, students also learn how to improve their own written text [1]. Peer review also empowers students to take ownership of their learning and encourages them to be active participants in the learning process and to take responsibility for their own progress.

2.2 Self-assessment

Self-assessment can take many forms and occurs when students make *"judgements about their own learning, particularly about their achievements and the outcomes of their learning"* [4, p. 529]. Panadero et al. [5] found 20 different categories of self-assessment. The typical self-assessment procedure includes self-reflection, self-evaluation, and self-grading to a more complex form of self-assessment that involve rigorous analysis of strengths and weaknesses in relation to explicit criteria. Self-assessments are a valuable tool for students as it helps them take ownership of their learning and empowers them to make meaningful changes to improve their understanding of design and assess the quality of their own work.

Boud [6] highlights that self-assessment contains two parts, where the first part is often neglected:

- The involvement of students in identifying standards and/or criteria to apply to their work.
- Making judgements about the extent to which they have met these criteria and standards.

To fully embrace the idea of self-assessment, it is vital to teach students the characteristics of good work, or as Boud states: *"It requires them to consider what are the characteristics of, say, a good essay or practical work and to apply this to their own work"* [6, p. 12]. Some general conclusions from the meta-analysis of self-assessment performed by Boud and Falchikov [4] and Pandero et al. [7] show that self-assessment more often agree than disagree with staff marks and that 'good' students tended to underrate themselves compared to staff marks, whereas 'weak' students tended to overrate themselves.

3   METHOD

Boud and Falchikov [4] recommended thirteen requirements for research studies on self-assessment that have been used as a guideline when reporting the study. They also identified a lack of replication with different groups taking the same course in the following years. This study is a longitudinal study of using self-assessment following five cohorts of students (2018-2022 n=239) using the same self-assessment criteria. All student quotes have been translated from Swedish by the authors.

4   IMPLEMENTATION

The implementation of peer review and self-assessment has been done in a third-year capstone design course. The role of the course in the program is to integrate knowledge and skills acquired previously in the program and focus on improving teamwork and interpersonal skills in a product design project. Initial reflections on the development of the course with a focus on academic writing have previously been presented in [1]. In the course, students work in small teams (3-4 students) that go through a design process with four phases. Students know when and what they should deliver at each stage gate, and then it's up to them to decide which methods are suitable for performing the design. After each phase, students present their progress and receive critique during four design reviews. They also produce a 4-

page written Process Memo (PM). The course ends with a presentation and documentation of the final concept. For each phase, a facilitated peer review and formative feedback from the teaching team (oral presentation and written PM) are used to improve the design, presentations, and documentation. In 2015 a facilitated peer-review process was introduced work [1] where students review others' documentation (PM1-3) and give feedback on a draft document, which is then updated before the teaching teams give feedback. Peer review saves time for the coaches, in a typical course, peers do more than 150 reviews before the teachers see the document. Formative feedback is essential because this is where students set the baseline of what is acceptable and what makes documentation excellent. Previous work [1] highlighted that most students appreciate the peer review sessions and believe it has improved the quality of the written documentation. The course evaluation also highlights that students think that the feedback from others is not the essential part; by reading others' documentation, they gain a better understanding of how good documentation is written.

4.1 Individual self-assessment

At the end of the course, students perform a self-assessment on *Communication* (oral, written, and visual), *collaboration* (active contribution to a project team), *create and develop* (Think and act innovatively and Prototype and test) and *Problem solving* (analysis). Students are informed of this self-assessment at the start of the course, previous research [8] [9] has identified the importance of presenting and discussing the assessment criteria with the students before starting the activity to ensure what will be assessed at the end of the course.

Table 1. Example of self-assessment of communication abilities

ORAL COMMUNICATION				
☐ NOVICE	☐ ADVANCED BEGINNER	☐ COMPETENT	☐ SKILLED	☐ EXPERT
Orally present work in a structured way, keeping track of time, and using appropriate aids.	Present ideas and arguments in a confident and persuasive manner	Use a wide range of presentation techniques for different audiences and situations	Quickly and convincingly answer questions and discussions based on discussions with various people	Present in English in a credible manner.

Self-assessment process:
1. *Self-assessment,* where assesses their own competencies and abilities and must describe how they meet the learning objectives (with examples from the course).
2. *Team feedback,* students' self-assessments are then reviewed by their team members, who give feedback on the student's assessments.
3. *Final assessment*, teachers review the assessment and do a final assessment based on students' assessment, team feedback and the interaction they had during the course.
4. *Quality assessment of feedback*, the teacher also assesses the quality of the feedback given to their team members.
5. *Teacher assessment* of the final documentation.
6. *Teacher conference*, teachers present their preliminary assessment of students, comparisons between teams' performance, discussion of outliers, feedback before the final assessment.

4.2 Team assessment of project work and documentation

In 2018 a self-assessment of the final documentation was introduced. In this assessment, the team had to argue why they fulfil the criteria for the documentation (developed by the examiner in the course). The same standards were then used by the final assessment by the teaching team, for details, see [1]. In 2023 we involved students in identifying standards and/or criteria to apply to their work in a documentation workshop three weeks before the deadline for the documentation. Student teams in the workshop had the opportunity to discuss what is essential for product documentation and develop their own criteria. The criteria were written on Post-its and pasted on a whiteboard. After the workshops, each student team had to write a rubric for three criteria.

5   RESULTS

The results are divided into quantitative and qualitative sections based on 239 students from 2018-2022. Six students were removed from the analysis due to missing self-assessment and/or peer review data.

5.1 Quantitative results

Table 2. Overview of self-assessment

	Number of students		Number of teams	Self-assessment		Teacher-Self		StdDev Teacher-Self	
Year	F	M	Total	F	M	F	M	F	M
2018	22	33	14	20,3	20,0	0,6	0,6	2,7	1,9
2019	16	40	14	21,4	21,7	0,6	0,2	1,9	2,4
2020	16	26	11	19,1	20,2	1,2	0,4	3,2	3,4
2021	23	23	12	21,7	21,1	1,3	0,6	2,8	3,3
2022	10	24	9	21,0	21,1	1,5	-0,1	2,3	2,3
Total	87	146	60	20,7	20,8	1,0	0,3	2,7	2,7
		233			20,8		0,6		2,7

From the results of the whole population, we can conclude that students are generally good at assessing their own work. The average difference between teacher assessment and self-assessment is less than 1 point ± 2,7 (Table 2), in agreement with earlier research [4] [10]. If we look at the difference between female and male students, it seems like females underestimate their self-assessment, i.e., the teachers will, in more cases, grade them higher than their self-assessment (Figure 1).
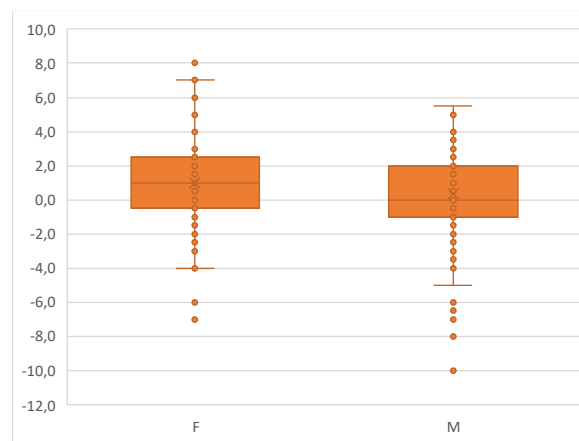

Figure 1. The difference between female and male students

Weak students often overestimate their work, and strong students underestimate their work. Many students with the lowest teacher scores assessed themselves at a much higher performance.
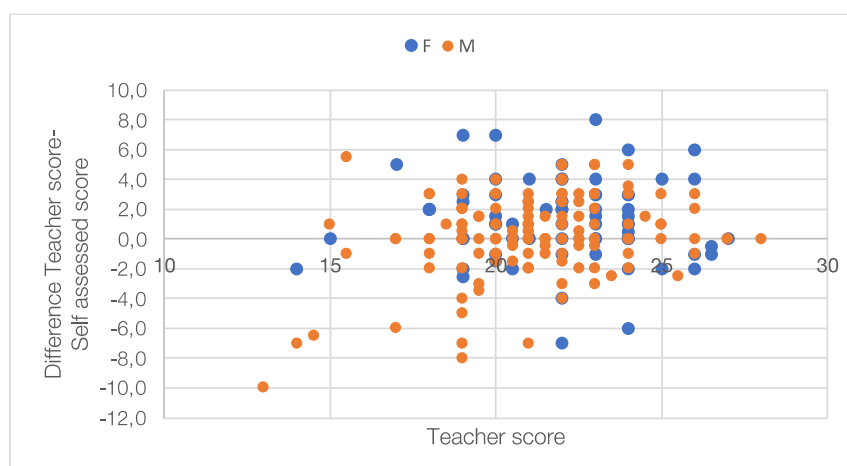

Figure 2. Plot over the teachers' final assessment (x-axis) and the difference between teacher-student assessment

5.2 Qualitative results

Qualitative results were mainly from the student feedback during self-assessment and course evaluations. The feedback on the self-assessments has been crucial in different ways. Team members

often reinforce an assessment, and it is also quite common for team members to highlight personal competencies that students themselves may not be aware of. Finally, it also performs a 'sanity -filter', so students cannot take credit for something they did not commit.

5.2.1 Reinforcement and boost of students' assessment

Team feedback complements and reinforces students' own assessment: *"I was responsible for the presentation in DR2, I didn't use cheat sheets, and I got good critique from the teachers' you have good fluency and contact with the audience' in the presentation I used both illustrations and animations in the presentation. I also brought simple prototypes that I used during the presentation. In the final presentation, I was given the responsibility to present the needs of the users, in my opinion, I also did this professionally, here, I worked on getting a good flow in the presentation, moved and used more body language."*. Student argumentation on oral presentation (2018). Feedback from a team member *" You have chosen to value yourself as competent and I think that is correct, you also made an excellent presentation in DR2 and the final presentation. You have also participated and answered questions in a good way during all design reviews."*. Several students are restrictive and modest over their achievements and make a rather careful self-evaluation; commonly, their teammates raise their self-evaluation. *"You have selected 'advanced beginner' in your rating. I think you were a little too self-critical. You were the only one of us in the group who had tools to make and who made physical prototypes and made really nice ones with the user in focus!"* Feedback Prototyping F to M (2021). Another example *"I think you underestimate yourself in this part. In addition to all the criteria under skilled, you also looked at the members' competence and led the project forward. Definitely reached the level of skilled!"* Feedback contribution to a project team (2019). It was also common from the team feedback with suggestions on how to improve in the future: *"An area of improvement would be to spend a little more time on the initial sketches, as they have great importance for the project going forward. Overall, a great job!"* Feedback visual communication (2021).

5.2.2 Sanity filter

The team feedback works as a sanity filter, so students can't write about things they have not performed. Student A writes about prototyping (2021): *"Time constraints and certain difficulties in terms of construction limited the number of physical prototypes, but I was involved in creating and analysing the prototypes we produced in CAD."*. This was met by team member B *"Student A has been sketching during DR2. He also did the product calculation and some diagrams in DR1. He also made the animation shown at the final presentation. Renderings... yes, possibly 'to some extent'. Haven't seen any rendering."*. Another student in the same team responded, *"Student A has not shown any higher degree of prototyping or testing in this project. He was not involved in role analysis or testing of concepts in the workshop. His contribution to the CAD model probably accounts for 1% of the total time spent on the modelling."*. The overestimation of performance in the lowest-performing students is also visible in Figure 2.

5.2.3 Feedback from course evaluations

It was apparent that most students liked to evaluate their own performance. Comments from students in course evaluations highlight the benefits of having to assess one's own as well as others' performance in order to reflect on each team member's contribution to project work. Another often-mentioned benefit was the value of the self and peer assessments as support in self-development. The task of assessment was perceived as very difficult but fun.

5.3 Creating own criteria

From the workshop, student teams highlighted the following criteria as the most important for the final documentation *Product design 38%* (describing and visualisation of the product and its features), *Desirability 23%* (focusing on user needs) and *Feasibility 19%* (focusing on how the product fulfils critical functions and needs).

5.4 Teachers' assessment of individual work

When evaluating individual students, it is crucial that the teacher meet and interact with the students throughout the course. The course uses a studio-based teaching method, where each teacher follows a small group of students. Through the recurring coach meetings and design reviews, the teaching team monitors students and sees how they act during coach meetings, who understand product and process

aspects, and who interact and answer questions during a design review. The four joint design reviews also allow the teaching team to evaluate team performance and compare teams. Even though most of the student's self-assessments are very close to the teachers' assessments, some students considerably overestimate or underestimate themselves (see outliers in Figure 1). This is discovered through team feedback and frequent meetings between coaches and teams, as well as all students being responsible for an oral presentation and a written document each.

6 CONCLUSIONS

This research presents the findings from a five-year longitudinal study of self-assessment in a capstone course with 239 students. In agreement with previous studies [11], students appreciate the task of assessing their own and others' work. The students in the study are very good at evaluating their capabilities, the difference between the self-assessment and teachers' final assessment was about 10%. It can be difficult to assess individual contributions accurately when running collaborative design projects. Using a studio-based approach, with formative feedback throughout the process, individual oral and written presentations, and support from self-assessment, team feedback and teacher discussions, there is a much higher certainty that students are assessed accurately. For the 2023 edition of the course, the teaching team have decided to redesign the team assessment and follow the advice from Boud [6] to involve the teams in identifying standards and/or criteria for good documentation. The new procedure involves students discussing and iterating high-level objectives and success criteria for their project work before using them for team assessment.

REFERENCES

[1] Törlind P. "Improving Written Communication: Implementation at Industrial Design Engineering" in *Proceedings of the 15th International CDIO Conference*, 2019.

[2] Gibbs G. "Using assessment strategically to change the way students" in *Assessment Matters in Higher Education*, vol. 41, Buckingham, Open University Press, 1999, pp. 41-53.

[3] Wikberg-Nilsson Å. and Törlind P. "Student Competence Profiles: A complementary or competitive approach to CDIO?" in *Proceedings of the International CDIO Conference*, 2016.

[4] Boud D. and Falchikov N. "Quantitative studies of student self-assessment in higher- education: A critical analysis of findings," *Higher Education*, vol. 18, no. 5, pp. 529–549, DOI: 10.1007/BF00138746, 1989.

[5] Panadero E., Brown G. T. L. and Strijbo J.-W. "The Future of Student Self-Assessment: A Review of Known Unknowns and Potential Directions," *Educational Psychology Review*, vol. 28, pp. 803–830, DOI: 10.1007/s10648-015-9350-2, 2016.

[6] Boud D. Enhancing learning through self-assessment, New York: Routledge Falmer, 2003.

[7] Pandero J., Johnsson A. and Botella J. "Effects of self-assessment on self-regulated learning and self- efficacy: Four meta-analyses," *Educational Research Review*, vol. 22, pp. 74-98, DOI 10.1016/j.edurev.2017.08.004, 2017.

[8] Panadero E., Jonsson A. and Strijbos J.-W. "Scaffolding self-regulated learning through self- assessment and peer assessment: Guidelines for classroom implementation," in *Assessment for Learning: Meeting the challenge of implementation*, Springer International Publishing, 2016, pp. 311-326.

[9] Panadero E. and Jonsson A. P. "The use of scoring rubrics for formative assessment purposes revisited: A review," *Educational Research Review*, vol. 9, pp. 129–144, DOI: 10.1016/j.edurev.2013.01.002, 2013.

[10] Rendon L. I. "Validating culturally diverse students: Toward a new model of learning and student development," *Innovative higher education*, vol. 19, pp. 33-51, 1994.

[11] Taras M. "The Use of Tutor Feedback and Student Self-assessment in Summative Assessment," *Assessment & Evaluation in Higher Education*, vol. 26, no. 6, pp. 605-614, DOI: 10.1080/02602930120093922, 2001.