Different forms of assessment in computer science and electrical engineering courses and their relationship with grades

An empirical study from the University of Stavanger

E. B. Abrahamsen[1 *], J. T. Selvik[1], J. T. Kvaløy[1] and P. Jimarkon[1]

[1] University of Stavanger, Norway

*Corresponding author. E-mail: eirik.b.abrahamsen@uis.no

Abstract: Various studies have focused on the relationship between grades and the different forms of assessment used in higher education. To contribute to these discussions, in this article we have analysed the relationship between form of assessment and grade in computer science and electrical engineering courses at the University of Stavanger. These are common courses in modern universities, not only in Norway, and represent a combination of skill- and competence-oriented courses. As a basis for examining the differences between grades and forms of assessment, we focus on grade distributions. In addition, attention is also paid to the differences in the proportion of students with the best and the lowest grades, for different forms of assessment. We show that there is a difference between forms of assessment and average grade. Furthermore, the proportion of students with the best and the lowest grades is largely affected by the form of assessment.

Key words: form of assessment, grades, computer science, electrical engineering

1 Introduction

In Norway, an increasing focus on the quality of education and study in higher education is particularly seen in the Norwegian White Paper, *Quality Culture in Higher Education* (Meld. St. 16, 2016-2017), which declares that the status of educational activities should be strengthened, and that teaching competence should be given more weight than it is today. In continuation of the report on quality, the government has asked educational institutions to develop systems that value high-quality teachers, promote academic careers and improve the status of education, as well as placing more weight on the students' voice. See also Abrahamsen et al. (2020).

The Norwegian White Paper points out that high-quality teaching is crucial for achieving the goal of high-quality education in the study programmes. Norwegian universities and colleges have also developed, or are in the process of developing, schemes for the validation of highly qualified teachers; most of these systems have a clear focus on raising the quality of students' learning through various input factors (Abrahamsen et al., 2020). See also Meld. St. 16 (2016-2017).

The current Supervision Regulation also requires both active participation by students in their own study situation and a clearer student focus in the education programmes (Supervision Regulation, 2017). The requirement for the active role of students is in accordance with and based on *The Standard and Guidelines for Quality Assurance in the European Higher Education Area* (ESG, 2015).

The Supervision Regulation (2017) also stipulates that institutions must review their study programmes, to ensure a good connection between learning outcome descriptions, teaching and learning activities and forms of assessment. The choice of form of assessment and its connection to learning outcome descriptions is of special interest, as we know that the assessment form is often considered to greatly influence the students' study behaviour, e.g., as an effect of previous exams being used extensively in the teaching and preparation for the examination. As such, the assessment has a key role beyond the evaluation of student achievement, i.e., certification. Assessment is also an instrument for student learning (Norton et al., 2013; Gibbs, 2006). Based on the combination of the certification and learning aspects, some assessment types are seen as more beneficial to the students, and there is a general perception that, overall, students perform better in some forms of assessment than in others (Carless, 2015, p.9; Bridges et al., 2002). See also argumentation in Section 2.

Following the assumption that some exam forms are more beneficial to the students, we question how much this influences the grading in numbers. Specifically, we question how the average grade, the fraction of students achieving a high or low grade, and the overall grade distribution vary between different assessment forms. The objective is to study the influence of selecting the written exam as a default form for the assessment.

To answer these research questions, in this article we have analysed the relationship between form of assessment and grade, with a focus on computer science and electrical engineering courses at the University of Stavanger. These types of courses are often practical in nature and represent a combination of skill- and competence-oriented courses. This study contributes to increased insight and provides an appropriate basis for ensuring a good connection between learning outcome descriptions and forms of assessment, which can be regarded as our attempt to strengthen the administration of assessment in computer science and electrical engineering courses at the University of Stavanger.

The remainder of this article is structured as follows. In Section 2, we present a review of the fundamentals of assessment literacy. In Section 3, the data collection and analysis are presented. Then, the results are presented in Section 4, while discussions are provided in Section 5. Finally, in Section 6, we draw some conclusions.

2 Assessment Literacy

Assessment is one of the most complex and important areas in higher education. It provides the basis for teachers to make inferences about students' progress and to make judgements about their performance. Practices of assessment and evaluation differ substantially across higher education institutions, depending on purposes, standards and policies. To align their practices with the policies involved, teachers must develop their competence, knowledge and skills regarding 'assessment literacy.'

Assessment literacy is defined by Pastore and Andrade (2019, pp.135-136) *as* 'an interrelated set of knowledge, skills, and dispositions that a teacher can use to design and implement a coherent and appropriate approach to assessment'. The study adds that assessment-literate teachers are competent at utilising information to inform their assessment practice, such as: in the selection and development of assessment tools; the administration of assessment; and the management of reporting and communication issues.

Previous studies have shown that assessment literacy is impactful for successful assessment. Day et al. (2018) stress that both frequent and continuous assessments contribute directly to the affective outcomes, cognitive processes and behaviour of students. A study by Millet (2018) investigated grading reliability by measuring the correlation between the average grade for a class and the students' average grade point average and found that lenient grading was associated with lower grading reliability. Grade inflation is a noteworthy concern in assessment, according to Mostrom & Blumberg (2012), who put forward that grade distributions should be addressed and documented carefully, so that the teacher can distinguish between undesirable grade inflation and desirable grade improvement. For a more specific and practical purpose, a study conducted in Norway by Chirumamilla et al. (2020) investigates: 1) perspectives of cheating during three types of written examination: paper exams, bring your own device e-exams and e-exams using university-owned devices; and 2) perceptions about the effectiveness of some typical countermeasures against cheating across these examination types. In a more recent study (St-Onge et al., 2022), the COVID-19 pandemic led to an investigation of faculty adaptations in assessment. This highlighted both the challenges and opportunities for enhancing assessment literacy and implementing sustainable e-assessment in higher education.

There is a focus in the literature on sustainable assessment linked to continuing learning, referring to 'assessment that meets the needs of the present without compromising the ability of students to meet their own future learning needs' (Boud, 2000). The fundamental idea is that assessment also has a role in students' learning beyond the course. According to Boud and Soler (2016), it reflects a 'move from assessment of learning to assessment for learning'. It implies a greater focus on the learning aspect compared with the evaluation aspect of the assessment. For learning, the assessment has a key role in promoting the active engagement of students (Rawlusyk, 2018).

Despite the sustainability argumentation and the societal benefits of learning-focused assessments, there are, in higher education practice, strong mechanisms linking academic success to completion and high grades, for example the grade requirements students must satisfy if they are aiming to undertake a PhD. In Norwegian education, as an example, there is a high focus on learning objectives, and courses are run with the expectation that students meet these objectives and complete the coursework. In practice, the main role of the assessment is shifted towards primarily evaluating students against the objectives. This would be somewhat in line with what Biggs and Tang (2011) refer to as 'constructive alignment,' where assessment and teaching activities are aligned with the learning outcomes for the student. In the end, the student will be measured against the learning objectives. Consequently, as Haggis (2003) points out, if students adopting a surface strategic approach seem to be successful in terms of good grades, there is no strong incentive for them to make significant changes. Students might then maintain a surface strategic approach to achieve high grades, not being in favour of deep learning (Entwistle, 2000).

From these studies, it is common to see that understanding our own practices in assessment in higher education enhances teaching effectiveness and student learning. Reference can also be made to the 'washback' (or 'backwash') effect, which describes the influence the assessment has on the teaching and learning activities leading up to the assessment, including preparations for the exam (see, e.g., Ramsden, 2003, p.182; Cheng & Curtis, 2004). Overall, assessment literacy enables teachers to provide precise feedback and tailor instructional strategies based on data. In principle, it should ensure fair, unbiased and transparent assessments and an accurate measurement of learning outcomes. Additionally, it supports curriculum development and institutional accountability, contributing to continuous improvement and the meeting of accreditation standards.

As an instrument to rank how well students have achieved the learning objectives, the assessment necessarily has a quality standard to meet. From a certification perspective, it should not matter to the evaluation result, then, which exam form is selected. However, this is not necessarily the situation. Millet (2018) studied the relationship between students' grade point averages (GPA) and the grades they received from a specific course. The correlation between the GPA and the course grade is a measure of 'grading reliability,' while the difference between the average course grade for the class and the GPA is a measure of 'grading leniency' (ibid, 2018). These two measures were studied using data from more than 50,000 course sections, showing that lenient grading is associated with lower grading reliability.

Several studies have explored the relationship between forms of assessment and grades, to confirm or deny whether the assessment form matters to the result. For example, Day et al. (2018) reviewed 88 articles and examined the relationship between intermediate assessment and student grades, showing, in particular, a positive effect of corrective feedback. However, this positive effect is not always a result of constructive feedback or motivation. Mostrom and Blumberg (2012), point to 'grade inflation' as one effect that might seem to be a 'grade improvement.' As Boretz (2004) points out, the grade distribution might not fully reflect how well students perform. There might be mechanisms leading to students receiving higher grades than they deserve. Hence, it is important to be aware of why grades have shifted (ibid, 2004). For example, for some assessment types, such as oral exams, there might be bias, challenging consistency or fairness. One such bias is the gender bias in student evaluations, where male instructors might be given a more positive evaluation (MacNell et al., 2015). Such effects might be used as an explanation if better grades are given in project work or similar forms of assessments than in written examinations.

3 Data Collection and Analysis

The data on which this study is based was collected from FS (Common Student System), which is a study administration system developed for universities and colleges in Norway. In this system, one can find, inter alia, information on when a student started their education, which study programme they belong to, which courses they have taken, and the grades they have received, as well as the form of assessment that has been used for each individual examination.

As a basis for our study, we have collected information about form of assessment and grade from all computer science and electrical engineering-related courses at both bachelor and master levels that were delivered in the Faculty of Science and Technology at the University of Stavanger in the period from 2016 to 2019. Grades for re-sit exams, for bachelor and master theses and from courses with pass/fail evaluation are excluded from this study. Due to COVID-19 effects, we do not use data from 2020 or thereafter.

Since academic staff at the University of Stavanger do not have access to FS, the data was received by us from a member of the administrative staff responsible for FS at the Faculty of Science and Technology at the University of Stavanger. The information received for each completed exam included course code and course name, year, semester, assessment tool, grade, name of study programme, the student's study level (bachelor and master), and the course's study level. The courses study level is divided into four classes, 100 level for first year courses, 200 level for second year courses, 300 level for third year courses and 500 level for master courses. No personal data was received, i.e., we could not trace individual students.

In total, 9,731 grades from students completing evaluation in courses with a letter grade assessment have been included in our study. Among these registrations, different forms of assessment have been used, as described in Table 1. For practical reasons, we have divided the various forms of assessment used into the seven main categories shown in the table. The categorisation was carried out by two of the authors of this article. Each of the authors first individually assessed all the assessment tools to try to find a suitable way to group them. Then, in the case of disagreement, the two authors discussed their categorisation and reached consensus.

Table 1. *Categories and assessment tools used in all computer science and electrical engineering courses.*

Category of assessment	Assessment tools
Category 1: Written exam	1. A written exam 2. A written test 3. Written 4. Written exam 5. Written test
Category 2: Written exam and project/exercise	1. Written exam and project report 2. Submission of work, project report and written exam 3. Assignments and written exam 4. Project report with presentation and written exam 5. Project report and written exam 6. Written exam and assignment(s) 7. Written exam and programming project 8. Written exam and exercise 9. Written test and exercise
Category 3: Oral exam	1. Oral exam
Category 4: Oral exam with project work	1. Oral exam and report 2. Report and oral exam
Category 5: Oral exam and written exam	1. A combination of oral exam and written exam
Category 6: Project work, hand-in assignment or similar	1. Submission of work (parts and whole) 2. Lab exercise and other exercises 3. Assignment 4. Programming project 5. Written project report
Category 7: Project work, hand-in assignment or similar accompanied with oral presentation	1. Programming project and oral exam 2. Project work with oral presentation 3. Project task with oral presentation 4. Project assignment with oral presentation

As a basic test of whether there is a relationship between grade and form of assessment, we conducted chi-squared tests. We also present the data with tables and figures that illustrate the distribution of grades for different categories of assessment forms. Furthermore, to assess if the factors class size, course level, and whether the student was admitted on a computer science/electrical engineering study program or not, could explain differences in grade distribution we did various regression analyses adjusting for these factors, while also taking into account dependencies within courses and year. More precisely we used mixed effects logistic regression models with random intercept for course and year, and with form of assessment, class size, course level and student type (computer science/electrical engineering or not) as covariates. In these models we used three different outcome variables: Whether the grade was A or B versus worse, whether the grade was A, B or C versus worse and whether the student passed (grade A-E) or failed. Tests with a p-value <0.05 were considered statistically significant. The statistical analyses were performed in IBM SPSS 25 and R version 4.3.3.

4 Results

To start studying the relationship between form of assessment and grade for the computer science and electrical engineering courses, we first present a table showing the grade distribution for each of the seven main categories of forms of assessment that have been used in this study. See Table 2. In addition, information about the number of students that received each of the grades for each form of assessment is also provided. The average grade is calculated based on the information given. Grade A is given a score equal to 5, B is equal to 4, etc. The total number of grade registrations is 9,731.

The grade distributions given in Table 2 are visualised with a stacked bar chart in Figure 1. Based on the information in Table 2 and in the stacked bar chart in Figure 1, we see a clear tendency for the grade distribution to change as a result of the form of assessment used. The chi-square test of independence shows that the relationship between grade and form of assessment is highly statistically significant, with a p-value $<10^{-6}$.

If we merge the first two categories (written exam; and written exam and project/exercise), the next three (oral exam; oral exam and project work; and oral exam and written exam) and the last two (project work, hand-in assignment, or similar; and project work, hand-in assignment, or similar with oral presentation), we also reach the same conclusions. The relationship between the three categories of form of assessment and grade is statistically significant, with a p-value $<10^{-6}$.

The grade distributions of these three main categories of forms of assessment are given in Table 3 and further visualised in Figure 2. In the following, these three categories are referred to as: Category I: Written exam, Category II: Oral exam, and Category III: Project work.

It is important to be aware that, in the present study, we had a large dataset and then even small differences become statistically significant. That is why we should not place too much emphasis on the results being statistically significant. It is better to focus on effect differences (differences in distributions, in average grade between the different forms of assessment, etc.).

From the results shown in Figures 1 and 2, we see that there are important differences in grades for different forms of assessment. To get some insight into whether the differences in grades between assessment forms can be explained by different assessment forms being used in courses of different sizes and/or at different stages during a study program, and whether different proportions of students from the computer science/electrical engineering study program versus students from other departments could explain some of the differences seen in grades obtained with the different assessment forms we also did several regression analyses for grade outcomes adjusting for these factors. We ran mixed effects logistic regression models for the proportion of students getting A or B versus worse, A or C versus worse and pass (A-E) versus fail, and compared the results for assessment category when adjusting and when not adjusting for the factors listed above. For all these models we found a significant effect (p-value $<10^{-6}$) of assessment category after adjusting for these factors. In Table 4 the results for grade A-C versus worse are reported, the results for the other outcomes are similar. The results show that the differences between assessment categories are a bit reduced when adjusting for class size, course level and student type, but not a lot.

Table 2. *Grade distribution for different categories of forms of assessment. The data is collected from the Common Student System.*

WRITTEN EXAM							
Grade	A=5	B=4	C=3	D=2	E=1	F=0	
Number of students	767	961	1725	814	617	771	5655
Grade distribution	13.6%	17.0%	30.5%	14.4%	10.9%	13.6%	
Average grade	2.67						
WRITTEN EXAM AND PROJECT/EXCERCISE							
Grade	A=5	B=4	C=3	D=2	E=1	F=0	
Number of students	205	553	560	166	12	0	1496
Grade distribution	13.7%	37.0%	37.4%	11.1%	0.8%	0.0%	
Average grade	3.52						
ORAL EXAM							
Grade	A=5	B=4	C=3	D=2	E=1	F=0	
Number of students	2	7	9	2	0	1	21
Grade distribution	9.5%	33.3%	42.9%	9.5%	0%	4.7%	
Average grade	3.29						
ORAL EXAM AND PROJECT WORK							
Grade	A=5	B=4	C=3	D=2	E=1	F=0	
Number of students	6	28	13	7	1	0	55
Grade distribution	10.9%	50.9%	23.6%	12.7%	1.8%	0.0%	
Average grade	3.56						
ORAL AND WRITTEN EXAM							
Grade	A=5	B=4	C=3	D=2	E=1	F=0	
Number of students	31	100	195	50	19	8	403
Grade distribution	7.7%	24.8%	48.4%	12.4%	4.7%	2.0.%	
Average grade	3.12						
PROJECT WORK, HAND-IN ASSIGNMENT OR SIMILAR							
Grade	A=5	B=4	C=3	D=2	E=1	F=0	
Number of students	589	531	483	118	43	24	1788
Grade distribution	32.9%	29.7%	27.0%	6.6%	2.4%	1.3%	
Average grade	3.80						
PROJECT WORK, HAND-IN ASSIGNMENT OR SIMILAR WITH ORAL PRESENTATION							
Grade	A=5	B=4	C=3	D=2	E=1	F=0	
Number of students	111	141	34	22	1	4	313
Grade distribution	35.5%	45.1%	10.9%	7.0%	0.3%	1.3%	
Average grade	4.04						
							9731

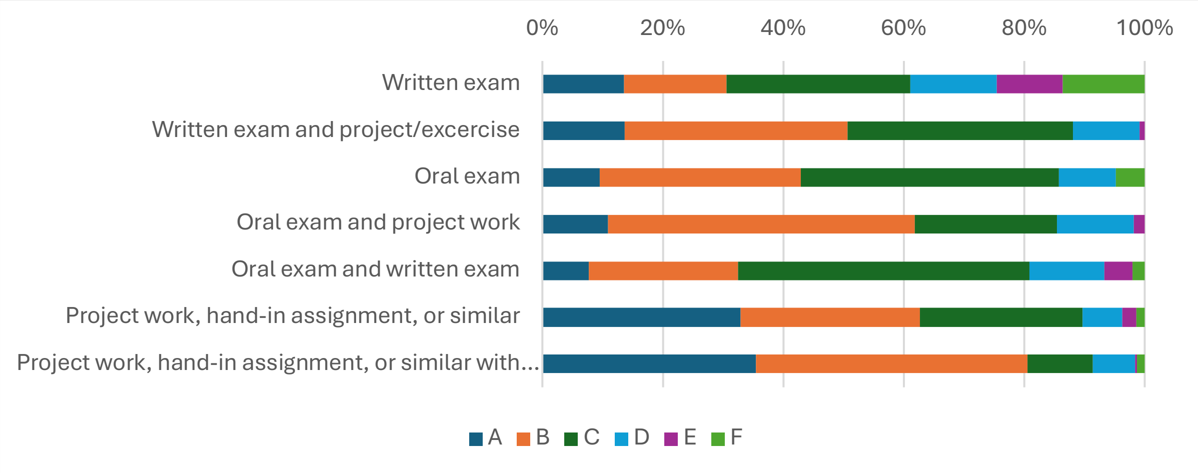

Figure 1. *Stacked bar chart showing grade distribution for the seven main categories of assessment form used in this study.*

Table 3. *Grade distribution for the three main categories of forms of assessment.*

CATEGORY 1: WRITTEN EXAM							
Grade	A=5	B=4	C=3	D=2	E=1	F=0	
Number of students	767+205= 972	961+553= 1514	1725+560= 2285	814+166= 980	617+12= 629	771+0= 771	5655+1496= 7151
Grade distribution	13.6%	21.2%	32.0%	13.7%	8.8%	10.8%	
Average grade	2.85						
CATEGORY II: ORAL EXAM							
Grade	A=5	B=4	C=3	D=2	E=1	F=0	
Number of students	2+6+31= 39	7+28+100= 135	9+13+195= 217	2+7+50= 59	0+1+19= 20	1+0+8=9	21+55+403= 479
Grade distribution	8.1%	28.2%	45.3%	12.3%	4.2%	1.9%	
Average grade	3.18						
CATEGORY III: PROJECT WORK							
Grade	A=5	B=4	C=3	D=2	E=1	F=0	
Number of students	589+111= 700	531+141= 672	483+34= 517	118+22= 140	43+1= 44	24+4= 28	1788+313= 2101
Grade distribution	33.3%	32.0%	24.6%	6.7%	2.1%	1.3%	
Average grade	3.84						

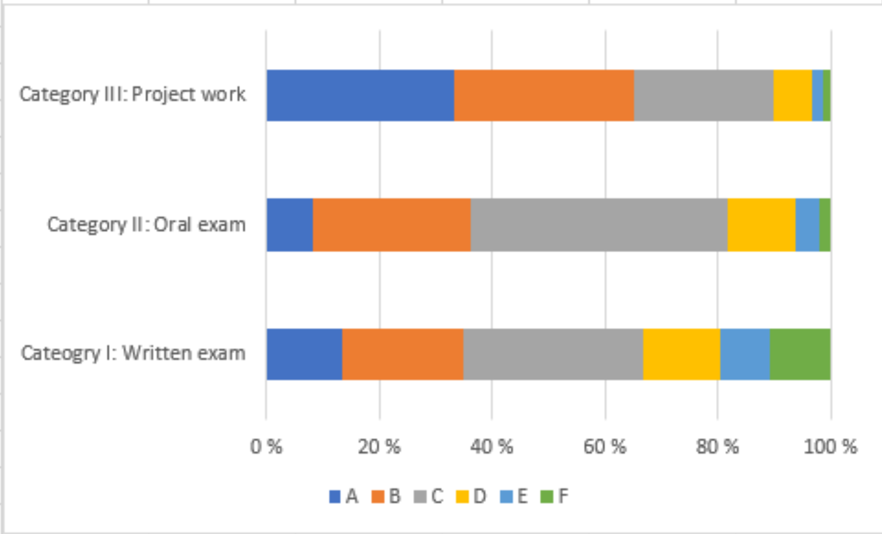

Figure 2. *Stacked bar chart showing grade distribution for the three categories of form of assessment used in this study.*

Table 4. *Effect of assessment category for the chance of getting grade A-C vs worse. Results reported for unadjusted model (considering only assessment categories) and for adjusted model, adjusting for class size, course level and student type (computer science/electrical engineering or not).*

Category of assessment	Unadjusted odds ratio (95% CI)	Adjusted odds ratio (95% CI)
Category 1: Written exam	Reference category	Reference category
Category 2: Written exam and project/exercise	4.7 (4.0-5.6)	3.2 (2.7-3.8)
Category 3: Oral exam	3.8 (1.3-16.4)	3.3 (1.1-14.8)
Category 4: Oral exam with project work	3.8 (1.9-8.6)	3.8 (1.9-8.7)
Category 5: Oral exam and written exam	2.7 (2.1-3.5)	3.0 (2.3-3.9)
Category 6: Project work, hand-in assignment or similar	5.5 (4.7-6.5)	3.9 (3.2-4.6)
Category 7: Project work, hand-in assignment or similar accompanied with oral presentation	6.8 (4.6-10.3)	6.0 (4.1-9.1)

5 Discussion

With reference to the results presented in the previous section, there are strong indications that assessment form strongly influences the grade that students achieve. In the discussion below, we focus on i) number of responses in the study, ii) average scores (grades), iii) maximum and minimum scores, and iv) overall distributions.

Starting with the number of responses used to establish the distributions, 58.1% refers to written exams, while another 15.4% refers to the use of written exams supplemented with the use of project work or exercises, i.e., in total, around three quarters of responses relate to the written exam assessment. This suggests that, even though several computer science/electrical engineering courses could have a strong skill orientation, it can be claimed that the students are widely tested using a traditional competence-oriented assessment form. The inclusion of project work nuances this, but the point is that these are courses which, in principle, could have seen a higher fraction of non-written forms, based on the learning objectives being of a more practical nature, compared with some of the other science disciplines. That around 75% of the assessments lean towards this direction suggests that the key learning objectives, such as programming skills, can be well evaluated through written assessment forms. However, it can also be claimed to be a matter of how the written exam is formulated, and it may be challenged by the motivation of the lecturer designing the exam, i.e., that it may, for example, be more comfortable for the lecturer to design and evaluate written exams.

The main result in Table 2 is the difference in grade distributions. Now, these results must be read in relation to the number of responses, but there is clearly a much lower average grade for written exams, compared to that of all the other types of assessments. Also, when referring to the main categories in Table 3, the average grade for written exams comes out below the other forms. In fact, the assessments involving project work seem to have an average score one grade higher. There are obvious reasons for this. One is contributions from cooperation with others. Project work could be performed in groups as teamwork with support, where there is a more flexibility in terms of time. There is also a relationship between how the project is defined (objective) and the quality of supervision received before completing the report or product. In contrast to written exams, project work typically is also much about the process, how to make students learn by doing, and how to achieve an acceptable result. In many situations, students have a stronger sense of ownership of project work, making it more motivating to work on, consequently leading them to spend more time on it. From such a perspective, it could make sense that some perform better and obtain a higher score – but not on such a scale.

Further reasons for the better results for project work assessment are related to when in the study programme these assessment forms are used and how the low performing students fall away. In project courses, the students who understand, during the course, that they will not be able to produce a good result may choose not to deliver any project work and withdraw from the course. Or, they deliver project work but get a bad mark and then withdraw from the course before the exam. These students are then not counted in the grade statistics. Furthermore, courses with project work assessment are most common in the later parts of study programmes, while written exams are most common in the first parts. A selection will then have taken place, with the weaker students, who did not do well in the early written exam courses, having fallen out of the study programmes. Moreover, the early courses are often courses which are mandatory for a wide range of study programmes, while the later courses are more study programme-specific. All things discussed above might contribute to the differences seen in grade distributions. Notice, however, that we still found substantial difference between the assessment categories after adjusting for the two last of these factors.

Another interesting observation made from the score distributions presented in Tables 2 and 3 is how high the score for project work is. An average score of 3.84 suggests that most of the students are performing at a high level. Besides the factors mentioned above, it could be that some students find some theoretical courses harder and are not doing as well in courses which do not contain project work, but it is more likely that there are other mechanisms at work, such as evaluations not being as strict as those for written exams. A C-grade should represent an ‘average’ of the student population, so that when about two thirds of students are awarded A and B, there is reason to question the quality of the evaluation criteria and the application of these.

Looking at the maximum and minimum scores, the point indicated above – that there is a high fraction of A and B scores – is evident, but so is the fact that almost no students are performing poorly. When there is a written exam, around 20% get an E or a F, while, if there is an assessment of project work, then only around 3% are at this level. The same tendency is also found for oral exams, where around 6% get an E or F. It seems that, once the examiner knows who is submitting the product, the chance of getting a C or better is high. Again, it challenges the quality and the application of the evaluation criteria.

The above point can also be seen from the overall distributions, where the mode of distributions for project work is an A (33%), with B as a close second (32%), while, for written and oral exams, the mode is C for both: 32% and 45%, respectively. Although oral and written exams are somewhat skewed towards Bs, in these programmes project work assessment is clearly giving the best grades.

One could argue that project work and oral exams offer students a better opportunity to communicate their skills and competence. However, the assessment is intended to test what the students have learned, and their scores should not depend on whether they are tested in one format or another. Assuming that the assessments are properly designed to test the learning objectives, the students should not under-perform in courses with written exams.

One assumption is that the quality of pre-work or student activities leading to assessments which are not in a written exam format leads to higher quality, but there can be no basis for such a claim. The assessment must be aligned towards the learning objectives, and, when basically the whole class gets a top score – and this is a trend – then this is something to do with the didactic design; there are reasons to assume something is wrong with the level of the curriculum, the assessment design or the evaluation criteria.

6 Conclusions

Published studies from, for example, Bridges et al. (2002), have already indicated that the type of assessment influences the distribution of grades. Such a finding is supported by the evidence and results from the study presented in this paper, where the assessment scores are considerably higher when the students are tested via project work or oral exams, compared with written exams.

Almost 10,000 grades from computer science and electrical engineering courses at the University of Stavanger in Norway show that, in courses with a project work assessment form, the grades are significantly better, compared with the alternative forms. In courses with the written assessment form, around 14% of students achieve an A, while the majority receive a grade around C. For project-based assessments, 33% of students receive an A, and for oral exam, 8% achieve an A. This suggests that the assessment form matters but also that there is improvement potential regarding the design of project work and oral exams, which seems to have grade distributions skewed too much towards good grades. From the distribution of project work and oral exams, the picture seems to be that the students are exceptionally good, which can be disputed by the fact that the majority of written exams have rather 'average' scores; selection effects in terms of when in the study programme the various assessment forms are used and how drop out is recorded can explain parts of this.

A conclusion is that assessment forms which are not written could need stronger assessment guidance, to ensure that test design and evaluation are more in accordance with the actual level of the students, to not give the impression that average students are best in their class. This study is our effort to understand our teaching context, by making explicit the relationships between assessment types and grades. By doing this, we increase our level of assessment literacy, in terms of the standards and management of assessment. We stress once again that the practice of assessment literacy needs to be addressed, and the results of studies such as this one can inform what needs to be prioritised and improved.

Acknowledgements

We would like to thank Atle Jakobsen at the University of Stavanger, for his helpfulness in extracting the data from the Common Student System (FS). We would also like to thank two anonymous reviewers for their suggestions and recommendations, all of which have contributed to improving the paper.

References

Abrahamsen, E. B., Selvik, J. T., Moen, V., & Kvaløy, J. T. (2020). Om sammenhengen mellom motivasjon, studentaktiv undervisning og konstruktive tilbakemeldinger. En studie fra Universitetet i Stavanger [On the relationship between motivation, student-active teaching and constructive feedback. A study from the University of Stavanger]. *Uniped*, *43*(04).

Biggs, J. B., & Tang, C. S. (2011). *Teaching for quality learning at university: what the student does.*(4th ed.) Maidenhead: McGraw-Hill; Society for Research into Higher Education; Open University Press.

Boretz, E. (2004). Grade inflation and the myth of student consumerism. *College Teaching, 52*(2), 42–46.

Boud, D. 2000. Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151–167.

Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education, 41*(3), 400–413. http://dx.doi.org/10.1080/02602938.2015.1018133

Bridges, P., Cooper, A., Evanson, P., Haines, C., Jenkins, D., Scurry, D., Woolf, H., & Yorke, M. (2002). Coursework marks high, examination marks low: Discuss. *Assessment and Evaluation in Higher Education, 27*(1), 35–48. ISSN 0260-2938.

Carless, D. (2015). *Excellence in university assessment: Learning from award-winning teaching*. (1st ed.) London: Routledge. https://doi.org/10.4324/9781315740621

Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In: Cheng, L., Watanabe, Y., & Curtis, A. (eds.): *Washback in language testing*. (1st ed.). New York: Routledge: https://doi.org/10.4324/9781410609731

Chirumamilla, A., Sindre, G., & Nguyen-Duc, A. (2020). Cheating in e-exams and paper exams: The perceptions of engineering students and teachers in Norway. *Assessment & Evaluation in Higher Education*, *45*(7), 940–957.

Day, I. N., van Blankenstein, F. M., Westenberg, M., & Admiraal, W. (2018). A review of the characteristics of intermediate assessment and their relationship with student grades. *Assessment & Evaluation in Higher Education*, *43*(6), 908–929.

Entwistle, N. (2000). Promoting Deep Learning through Teaching and Assessment: Conceptual Frameworks and Educational Contexts. 1st Annual Conference ESRC Teaching and Learning Research Programme (TLRP), University of Leicester, November 2000. http://www.tlrp.org/acadpub/Entwistle2000.pdf

ESG. (2015). *Standards and guidelines for quality assurance in the European Higher Education Area (ESG).* Brussels, Belgium.

Gibbs, G. (2006). How assessment frames student learning. In: *Innovative Assessment in Higher Education*. Routledge.

Haggis, T. (2003). Constructing images of ourselves? A critical investigation into 'approaches to learning' research in higher education. British Educational Research Journal, 29(1), 89–104

MacNell, L., Driscoll, A., & Hunt, A. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, *40*(4), 291–303https://doi.org/10.1007/s10755-014-9313-4

Meld. St. 16. (2016-2017). *Kultur for kvalitet i høyere utdanning* [Quality culture in higher education]. Ministry of Church, Education and Research.

Millet, I. (2018). The relationship between grading leniency and grading reliability. *Studies in Higher Education, 43*(9), 1524–1535.

Mostrom, A.M., & Blumberg, P. (2012). Does learning-centered teaching promote grade improvement? *Innovative Higher Education, 37*(5), 397–405.

Norton, L., Norton, B., & Shannon, L. (2013). Revitalising assessment design: What is holding new lecturers back? *Higher Education, 66*(2), 233–251.

Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three dimensional model. *Teaching and Teacher Education, 84,* 128–138. doi:10.1016/j.tate.2019.05.003

Ramsden, P. (2003). Learning to teach in higher education. (2nd ed.) London New York: RoutledgeFalmer.

Rawlusyk, P. (2018). Assessment in higher education and student learning. *Journal of Instructional Pedagogics, 21*.

St-Onge, C., Ouellet, K., Lakhal, S., Dubé, T., & Marceau, M. (2022). COVID-19 as the tipping point for integrating e-assessment in higher education practices. *British Journal of Educational Technology, 53,* 349–366. https://doi.org/10.1111/bjet.13169

Supervision Regulation. (2017). *Forskrift om tilsyn med utdanningskvaliteten i høyere utdanning* (studietilsynsforskriften) [The Norwegian study supervision regulation]. https://lovdata.no/dokument/SF/forskrift/2017-02-07-137. Downloaded 18.05.2021.