

Can ‘Passed with Distinction’ as a New Grading Scale Favour the Transition towards Formative Assessment?

Christian Jørgensen¹ and Helen Bråten²

¹*Department of Biological Sciences, University of Bergen, Bergen, Norway.*

²*NOKUT, Norwegian Agency for Quality Assurance in Education, Lysaker, Norway.*

ABSTRACT: A rationale behind formative assessment is that repeated assignments with feedback are better suited for developing skills higher up in Bloom’s cognitive taxonomy. While teaching to facilitate active learning is becoming more and more common at higher education institutions in Norway, assessment has changed less and summative exams at the end of each term still dominate. Here, we focus on the relationship between formative assessment and the grading system. Norwegian law allows only two grading scales: the ECTS scale A-to-E plus F(ail); and pass/fail. We present arguments that pass/fail grading may free instructor time for formative assessment with more feedback, enhance collaboration and sharing among students, and reduce stress and anxiety. We also argue that letter grades can distract from desirable learning at several levels: through surface prepping before exams, by disincentivizing interdisciplinarity, by unbalanced effort allocation from teachers, and by presenting developing personal characteristics as fixed in time. However, with only pass/fail as alternative outcomes, there is a real risk that especially surface learners aim to crawl over where the fence is lowest, and thus aim for the minimum threshold of acceptable performance. Good students furthermore report that they want to see their hard work reflected in the grade, which is related to letter grades being perceived as helpful by employers. One may further argue that scales with multiple alternative grades, such as A-F, may unconsciously bias assessment towards content knowledge. We argue that a logical consequence is that a simple grading scale with a reward for outstanding performance could strike a good balance between grading effort and learning benefits, and could thus favour the transition towards formative assessment. We thus suggest considering a three-level scale, “fail/pass/pass with distinction”, where a distinction is awarded to e.g. the 15% best performers.

1 TOWARDS FORMATIVE ASSESSMENT

Higher education aims to be a transformative period during a young person’s most productive years. As educators, we should use the methods that best improve each student’s competence. There has traditionally been a focus on teaching practices that favour learning (Freeman et al. 2014), while it has been more difficult to do assessment and grading in ways that similarly enhance student learning (Raaheim 2016). It has been noted how assessments often are rigid and detached, despite students using assessment as an important landmark to guide their learning (Biggs and Tang 2011). This often causes significant backwash on the learning phase, whereby students try to learn what they think they will be tested on, with the result that deep subject learning is hindered rather than favoured (Biggs and Tang 2011, chapter 10). Biggs’ concept of ‘constructive alignment’ is an attempt to overcome these challenges by deciding on assessment tasks and assignments that test the intended learning outcomes before designing the learning situations that the instructor will facilitate (Biggs and Tang 2011). A similar philosophy is integral to Dee Fink’s integrated course design, where one begins by defining a ‘culminating project’, what the student should be able to *do* after the course, and aligns assignments, feedback, and teaching to reach that goal (Fink 2013).

There are thus two dominant views of designing assessment in higher education. One is to have a final exam to sum up what the student has learned; this is referred to as summative assessment. The second view emphasises assessment as a way of learning, referred to as formative assessment. Assessment is part of the learning process and aims to guide the student towards mastering the expected learning outcomes. Formative assessment is based on the view that only by *doing* can the student really reveal competence, and feedback from an instructor or assessor can inform the student about how to improve.

The shortcomings of traditional, summative assessment has become clearer with the large student cohorts of the knowledge society. This has been epitomized by Biggs and Tang (2011) with the stereotype students Susan, who is academically inclined, driven by curiosity and thus has an inner motivation for deep learning, and Robert, who is at university to get a job for which he needs a diploma

with decent grades, with the implication that he focuses on surface learning prior to exams. Biggs and Tang (2011) warn that these should not be labelled as a good and a poor student, both are necessary for the knowledge society, and it is expected from the societal mission of higher education that both student types succeed. Rather, the pedagogical challenge is to make Robert behave as Susan naturally would, and it has been argued that constructive alignment and formative assessment are effective ways to achieve that.

Despite virtually everyone agreeing with Biggs, Fink, Raaheim (2016), and others that higher education needs to aim for formative assessment, whereby students repeatedly demonstrate their skills and receive feedback on how they can improve, the adoption of formative assessment in Norway progresses relatively slowly. Why? Are the goals poorly defined? Are teachers lacking the competence or resources? Are regulations too restrictive? In this essay we focus on the hurdle represented by grading, and discuss whether a new grading scale could favour the transition towards more formative assessment in Norway.

2 EFFECTS OF GRADED SCALES

Reasons for having graded scales are that the potential for a better grade motivates students to work harder, that the grade provides feedback about own performance, and that grades are useful when employers are sorting job applicants (Kohn 1994). Proponents of simpler pass/fail scale often note that it will reduce stress and anxiety (Bloodgood et al. 2009, Spring et al. 2011), increase collaboration and sharing, encourage broader perspectives and questioning, focus motivation towards the subject (Butler 1988), and require less instructor effort. Letter grades are better for sorting and some feel motivated by them. Although it is often stated that a grade is a form of feedback, it is also easy to argue that a grade is too simple and too aggregated to constitute informative feedback (Kohn 1994). See table 1.

Table 1. Some pros and cons of the two grading scales pass/fail and letter grades. Sources: Butler (1988), Kohn (1994), Bloodgood et al. (2009), Biggs and Tang (2011), Frich et al. (2014), UHR (2015), Sindre (2018).

Property	Pass/fail	Graded scales
<i>Collaboration</i>	Easier to make students collaborate and share, which are skills valued by employers.	Grades often set up competition among students and disincentivize sharing and collaboration.
<i>Feedback</i>	Higher threshold for pass induces need for feedback underway.	The letter grade is standardized feedback, but too late?
<i>Motivation</i>	Students may aim to cross where the bar is lowest, thus aiming for barely passing. Others report that motivation for the subject itself may increase.	A scale of grades may motivate students to aim for their best performance, but for many, motivation from grades is shallow and short-lived.
<i>Deep learning</i>	With less focus on grades, students can direct more of their effort towards deeper understanding.	Grades receive more focus and may favour exam backwash and shallow learning.
<i>Prioritization</i>	When taken simultaneously, pass/fail courses may receive less effort than courses with letter grades.	With limited time, students tend to prioritize courses where marginal effort can improve the final grade.
<i>Stress</i>	Pass/fail has been shown to reduce experienced anxiety.	Graded scales may induce stress and anxiety in some students.
<i>Sorting</i>	A diploma with mostly pass/fail does not expose student quality so employers may set up own tests to sort applicants.	Employers use grades to sort applicants and may favour applicants with graded diplomas. Is the future Big Data?
<i>Resources</i>	Grading can be quicker and free instructor time for other activities.	A graded scale may require detailed attention and significant time.
<i>Complaints</i>	Complaints are limited to students who failed. A complaint committee has only two outcomes to choose between.	With more grades, more students are between grades, could expect more complaints. A complaint committee must do more thorough evaluation.

3 GRADES AS SIGNIFICANT DISTRACTION: GOAL DISPLACEMENT

Just like poor assessment practices, grading may set up false incentives whereby nailing a good grade becomes more important than mastering the learning outcomes. For Robert, the grade is more important than the subject, which from the perspective of the teacher serves as a goal displacement: instead of mastering the learning outcomes Robert focuses on mastering the exam. Although perfectly rational behaviour in a system defined by grades, it is a distraction from the larger goal of higher education: that Robert should master the discipline with its knowledge, skills, and competencies that will be required of him afterwards.

There is reason to believe that many more students than Robert can be distracted by grades when they compose the elective parts of their study program. Solutions to many of society's current challenges involve interdisciplinary research, and from a systems perspective on higher education one could expect clear benefits if more students had training across several disciplines. Strategic, or even unconscious, grade planning can steer students towards choosing courses similar to competence they already have so that the expected grade is better. If they choose to broaden their horizon with coursework from unfamiliar disciplines, the resulting grade may not look so good on the diploma (Schwartz and Sharpe 2011).

More worrisome is that grading that is fair, repeatable, and where complaints end up with the same assessment tend to ask factual questions low in Bloom's cognitive taxonomy. When testing higher cognitive skills, process becomes more important than final outcomes (Green and Emerson 2007), and the grade achieved will depend more on the evaluator's expert opinion and maybe differ with a new committee. This attitude is not only widespread among teachers who want to avoid the administrative burden of grade complaints, but has also influenced institutional regulations and administrative routines and recommendations.

Equally worrisome is that grades also can distract the teacher and the institution to invest large resources in grading although it has very little effect on learning. For big courses, more than half the instructor time can be used for marking and grading, with feedback to students often being limited to the letter grade itself, presented with no justification or feedback that could offer a learning opportunity for the student. What if this instructor time was used for discussing with students, giving them feedback on their level and understanding, what they should read, and how they should practice? From a systems perspective it seems fair to ask whether the resources invested in grading can be justified, or whether that instructor time could have been better used to engage students in e.g. active learning experiences. That this question is rarely asked is not so surprising: it has been noted that 'Grading is one of the least liked, least understood and least considered aspects of teaching' (Green and Emerson 2007), although every teacher has to do it and with little or no formal training.

Grades can even distract long after they have been awarded. When asked, Norwegian employers list interdisciplinarity, studies that include practical experience, and prior work experience as more important or as important as good grades (Støren et al. 2016). The list of desirable competences ranks collaboration and ability to learn independently at the top (Støren et al. 2016), but cooperative skills are rarely reflected in grades. More philosophically, is it really fair that your performance on one day in your teens or tweens is visible on your diploma for the rest of your life, although you may change profoundly, both personally and professionally? Have you had the experience that you learned a lot from an exam that went wrong? And if you are about to hire someone aged 35, which do you value higher: the grades on the person's higher education diploma or their subsequent work experience? Maybe it is easier to illustrate the absurdity of ever-lasting grades with a thought experiment outside academia: Imagine that you got a letter grade on your driver's licence. It might determine how fast you are allowed to drive, or what insurance premium you need to pay. For how many years would it feel fair that the grade affects you? Is the day you took your driver's test still representative for you driving skills?

4 GRADING IN NORWAY

The Norwegian law for higher education §3-9 allows only two grading scales: the ECTS scale A-to-E plus F(ail); and pass/fail. Ca 1 million grades, of which 85% are letter grades, are awarded in Norway annually (UHR 2015), with pass/fail typically used for practical courses, for methodological courses, higher-level courses with small groups of dedicated students, or for a course where nuanced grading is

difficult because it combines disciplines or relies on process (but see discussion below of medical education in Norway).

One Norwegian curiosity is the strong student rights regarding grade complaints. Without having to provide a reason or justification, students can ask for a reassessment of their grade, by a new committee. The administrative burden of this, on administrative personnel, the teacher, evaluators, and the complaint committee, serves as a disincentive for formative assessment where the final grade may often be a composite of multiple assignments. Situations where a grade changes after complaint give the impression that assessment routines are random or of poor quality, and there is pressure from students, administration, university leadership, and politicians for grades to be repeatable almost down to mathematically formulated rules. An unintended consequence of this desire is that multiple choice tests and exam questions that test factual knowledge rather than higher cognitive skills and competences become preferred, as they are easy to mark and likely to end up with the same grade by a new committee.

5 PASS/FAIL GRADING IN NORWEGIAN MEDICAL EDUCATION

In Norway, the choice between letter grades and pass/fail has received considerable attention in medical education. In the mid-1990s, medical studies in Oslo, Trondheim, and Tromsø replaced letter grades with pass/fail throughout the six-year education (Anonymous 2012). At the same time, pre-clinical and clinical studies were integrated, and Problem-Based Learning was emphasized as a guiding pedagogical principle (Handal 2002). Although evaluations concluded that the intended goals had been achieved regarding integration of knowledge across sub-disciplines, depth of learning, and collaborative skills (Wyller and Brodal 2006), the University of Oslo reintroduced letter grading from 2016 (Anonymous 2016). The justification for reintroducing letter grades was now that most medical education in Europe is letter-graded, including the diplomas of half the doctors presently working in Norway (Frich et al. 2014). One problem which to our knowledge has received little attention in this debate is that many countries do not use grade E, such that ABCDF is common in many European countries and the US, and in the US many universities do not even use D. An overview om US medical schools revealed that equal proportions were using two levels (pass/fail), three levels, four levels, or five levels (ABCDF) for the basic science training (Bloodgood et al. 2009). From that viewpoint, Norwegian grades that extend all the way into E appear poor, and there are fewer of the good grades A and B simply because our grading scale is longer.

Pass/fail grading is used for the first two years in many medical educations in USA, and studies have shown that pass/fail led to improved student wellbeing without compromising performance, particularly in introductory courses (Bloodgood et al. 2009; Spring et al. 2011). Unlike the Norwegian system, the US has national standardized medical tests on which students receive grades that can be used by employers when hiring. University studies thus prepare students for those tests, but are not expected to certify or sort doctors as is the case in Norway.

6 IS THERE A NICHE FOR A NEW GRADING SCALE IN NORWAY?

That Norwegian has settled on just two grade systems, pass/fail versus letter grades with five options plus fail, is a historical contingency and should be reconsidered if there are sound pedagogical alternatives. Internationally, there are several scales in use with fewer than five grades. A majority of countries with letter grades use only ABCD, although some have adopted the US system with pluses and minuses on some or all of those grades. Of particular interest are grading scales that are pass/fail, but with distinction awarded to the top candidates. Pass with distinction, pass with honours, and similar three-level grading scales are used in several countries; in Sweden ‘godkänd’ and ‘väl godkänd’ is one such scale and it dominates at several of the bigger universities.

Compared to pass/fail, ‘pass with distinction’ will retain a premium for the top performers, and may thus incentivise an effort beyond merely passing. Active sharing and helping of others could even be specified in the learning outcomes and thus the basis for what may earn the distinction. Although studies question the motivating effect of grades (Butler 1988), most instructors likely hold beliefs that grades make students work harder. From a pragmatic viewpoint, one could expect that many instructors could switch to a ‘pass with distinction’ scale and reap many of the benefits of pass/fail. With ‘pass with distinction’, grading will still be fairly easy compared to a full letter scale, one could expect fewer complaints and simpler complaint processes. A courser grading system is also better suited to evaluate

higher cognitive skills, where the focus often is on process rather than the final answer (Green and Emerson 2007).

The three most important functions of grades are often reported as feedback, motivation, and sorting, and it can be argued that 'pass with distinction' can achieve these as well as or better than letter grades do, but with less effort and while measuring performance at higher cognitive levels. A coarser scale is also more compatible with the composite assessments with feedback and improvement that formative assessment often amounts to.

Allowing 'pass with distinction' in Norway has been raised at several occasions, but because it requires a change of the law a thorough discussion has been postponed, pending initiative from the right body at the national level (UHR 2015). Since the law for higher education is currently being revised, maybe now is the right timing to introduce this grading scale in Norway. Pedagogically, having more options can't hurt. Because grading is so intertwined with assessment and teaching through as in constructive alignment, there are likely pedagogical situations where neither pass/fail nor letter grades would fit the course, but pass with distinction could. Introducing a new option could also trigger local debates about grading and assessment, and increased awareness could in itself lead to more formative assessment.

REFERENCES

- Anonymous. 2012. *Rapport fra arbeidsgruppen Oslo 96+*. 31 pp. Universitetet i Oslo, Det medisinske fakultet, Oslo, Norway.
- Anonymous. 2016. *Eksamenskvalitet og innføring av gradert karakterskala ved profesjonsstudiet i medisin*. 18 pp. Universitetet i Oslo, Det medisinske fakultet, Oslo, Norway.
- Biggs J, Tang C. 2011. *Teaching for quality learning at university*, 4th ed. Open University Press, Maidenhead, UK.
- Bloodgood RA, Short JG, Jackson JM, Martindale JR. 2009. A change to pass/fail grading in the first two years at one medical school results in improved psychological well-being. *Academic Medicine* **84**: 655-662.
- Eldridge G. 2015. How do grades impact intrinsic motivation? *The International Educator*. [<https://www.tieonline.com/article/1688/how-do-grades-impact-intrinsic-motivation->]
- Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP. 2014. Active learning increases student performance in science, engineering, and mathematics. *PNAS* **111**: 8410-8415.
- Fink LD. 2013. *Creating significant learning experiences: An integrated approach to designing college courses*. 2nd ed. Jossey-Bass, San Francisco, CA.
- Frich J, Lundin KEA, Os I. Karaktersystemet – avveining mellom ulike hensyn. *Tidsskrift for Den Norske Legeforening* **134**: 14-15.
- Green KH, Emerson A. 2007. A new framework for grading. *Assessment & Evaluation in Higher Education* **32**: 495-511.
- Handal G. 2002. Utdanning av medisinerne – det pedagogiske perspektivet på Oslo96. *Tidsskrift for Den Norske Legeforening* **122**: 1931-1933.
- Kohn A. 1994. Grading - the issue is not how but why. *Educational Leadership*.
- Krathwohl DR. 2002. A revision of Bloom's taxonomy: An overview. *Theory into Practice* **41**: 212-218.
- Frich J, Lundin KEA, Os I. 2014. Karaktersystemet – avveining mellom ulike hensyn. *Tidsskrift for Den Norske Legeforening* **134**: 14-15.
- Raaheim A. 2016. *Eksamensrevolusjonen*. Gyldendal Norsk Forlag, Oslo, Norway.
- Schwartz B, Sharpe K. 2011. Do grades as incentives Work? Why "smart" incentives can never be smart enough. *Psychology Today* [<https://www.psychologytoday.com/us/blog/practical-wisdom/>]
- Sindre G. 2018. Karakterer – fins de fortsatt i 2030? NTNU Læringsfestivalen, Trondheim, Norway.
- Spring L, Robillard D, Gehlbach L, Simas TAM. 2011. Impact of pass/fail grading on medical students' well-being and academic outcomes. *Medical Education in Review* **45**: 867-877.
- Stan E. 2012. The role of grades in motivating students to learn. *Procedia Social and Behavioral Sciences* **69**: 1998 – 2003.
- Støren LA, Carlsten TC, Reiling RV, Olsen DS, Arnesen CÅ. 2016. Arbeidsgivers vurdering av nyansatte med høyere utdanning og fagskoleutdanning - underveistrapport, første delrapport. *NIFU Arbeidsnotat* 2016:16.
- UHR. 2015. Retningslinjer for karakterskalaen bestått/ikke bestått. Universitets- og høyskolerådet, Oslo, Norge.
- Wyller VB, Brodal P. 2006. Oslo 96 – ti år etter. *Tidsskrift for Den Norske Legeforening* **126**: 2088.