Grading systems, features of assessment and students’ approaches to learning
Lars-Owe Dahlgren, Andres Fejes, Madeleine Abrandt-Dahlgren and Nils Trowald

lars-owe.dahlgren@liu.se, andreas.fejes@liu.se, madeleine.abrandt.dahlgren@liu.se

Published in Teaching in Higher Education, 14(2), 185-194.

Address for correspondence
Department of Behavioural Sciences and Learning
Linköping University
581 83 Linköping
SWEDEN

Introduction
In recent years, considerable efforts have been devoted to harmonising the systems of higher education throughout Europe. The core arguments are that this would facilitate student and staff mobility and permit free movement of academic graduates. Ultimately, the aim is to make Europe a leading region in higher education. The Bologna process comprises several steps, e.g. the introduction of a so-called Diploma Supplement to facilitate comparisons between degree documents. Furthermore, a new system has been proposed to facilitate the comparability of study programs as well as the students’ achievements. This is called ECTS, European Credit Transfer System. ECTS, thus, comprises two parts; a system with credits to describe the duration of the program, and a 7-step grading system to indicate the performance of the students (EU 1999). It is this latter part of ECTS that will be focused on in this article.

In the discussion on a common grading scale, there is an idea that such a scale will increase the comparability between students since each country will use the same grading system (a 7-step grading scale). In some countries, the changes are not extensive, as they already had grading systems close to seven steps. In other countries, e.g. Sweden, the changes are more extensive as some educational programs today only use pass/fail (Ministry of Education 2004; 2005). Even if the implementation of the ECTS grading system has been discussed to some extent in the academic community (cf. Karran 2004; Löfgren 2006), the relation between grading systems, assessment and learning has been more or less neglected. Grading systems are closely related to assessment due to the fact that the latter forms the foundation of the former. Grading and assessment can thus not be discussed separately, particularly not if the aim is to understand the impact of the grading system on students’ learning.

Our main argument in the article is that the structure of grading systems has an impact on the way assessments are carried out, which, in turn, has an impact on the quality of students’ approaches to learning. It is well known that assessment has different functions; as a means of control and of enhancing learning (Kvale 1973; 1990; Trowald 1997). An assessment may emphasise the control aspect, as in an evaluation in relation to a set of intended learning outcomes, or it may emphasise the learning aspect and the potential for learning embedded in the assessment. These different foci would consequently bring about differently designed assessment tasks. It is also likely that a grading system requiring detailed differentiation between students would emphasise the control aspect of the assessment. There is convincing research evidence that assessment has a
far-reaching impact on student learning. In short, it has been shown that assessment affects students’ choice of what to focus on in the literature and their approaches to the learning task as well as the retention of knowledge. (Snyder 1968; Becker et al. 1968; Miller and Parlett 1974; Trowald & and Dahlgren 1993; Wiand 1998; Fejes et al. 2005; Dahlgren and Fejes 2005).

As a consequence of the endeavour to harmonise higher education in Europe, a debate around learning objectives and assessment criteria is emerging. The formulation of criteria for assessment is logically related to the structure of the grading system in use, and thereby has the potential of being as important as the nature of the assessment tasks in terms of the impact on student learning. Review articles on assessment in higher education point to the fact that there seems to be a consensus in the literature about the importance of articulating and using criteria for assessment in relation to grading (Woolf 2004; Sadler 2005). The core of this line of arguing is that articulated assessment criteria make the relationship between assessment and the outcome of learning explicit, and thereby enhances the students’ learning and the faculties’ reflection and prevent judgments from being based on intuition rather than rationality. It has, however, been shown that putting this into practice could be problematic. Experiences from the UK show that standardised criteria for assessment seem to be difficult to apply in university teaching. There does not seem to be any corresponding consensus among university teachers about what is meant by criteria for assessment, and what implications such criteria would have (Sadler 2005, Tan and Prosser 2004). What does seem to matter when it comes to the feasibility and relevance of assessment criteria is that their meaning is negotiated and discussed between teachers as well as between teachers and students (Rust et al. 2001; O’Donovan et al. 2001). In such cases, it has been shown that assessment criteria can improve student learning. (Rust et al. 2003). A few practice-oriented studies show successful results in articulating, negotiating and using assessment criteria also for what seems to be harder to capture – generic academic skills such as creativity (Cowdroy and De Graaff 2005) or critical reflection (Fisher 2003). Drawing on the above-mentioned research, it is reasonable to assume that there is also a relationship between judgement criteria and students’ learning. This relationship constitutes another link in the chain of relationships between grading systems, assessment task characteristics, judgement criteria and students’ approaches to learning.

In this article, the aim is to further investigate the relationship between three of the links in this cause-effect chain; the structure of the grading systems and the content and structure of assessment tasks as well as the students’ approaches to learning. The empirical findings to support our arguments will be derived from a survey study that has been carried out among students in different study programs in higher education in Sweden.

The empirical study
A questionnaire comprising 14 fixed-response and two open-ended questions was distributed to 402 students at Swedish institutions of higher education. The sample comprises students in 13 different study programs. They were distributed over 8 universities and university colleges. Of the participating students, 125 were enrolled in programs with a pass/fail assessment. There were 277 students in programs with graded assessments. The grading system in the latter group may vary somewhat from fail, pass and pass with distinction with up to 5-step grading scales. The internal dropout rate varies between questions but never exceeds five per cent of the respective groups. It was not possible to match students from identical programs, but with different grading systems,
in the study. This is due to the fact that e.g. all engineering programs in Sweden apply a five-step grading scale. This fact is of course calls for caution when generalising the results from the present study. The students in the two groups, however, represent a variety of professional programs as well as liberal arts studies. This fact makes it less likely that the experiences expressed by the students co-vary with e.g. field of study. We may, however, not exclude the possibility that parts of the differences may be explained by other factors than the structure of the grading systems applied in the different programs. The pass/fail grading group comprises students from teacher education, physical education, music (teachers), agriculture, veterinary medicine, dentistry and physiotherapy. The multi-step grading group comprises students from theology, law, business studies, linguistics, music (church musicians) and engineering (two different programs from two different institutions).

Because of the somewhat uneven sizes of the pass/fail and the multi-step scale samples, percentages as well as absolute numbers will be given. The results will be presented in three different sections; relationship between grading systems and features of assessment; assessment and student approaches to learning; grading systems, assessment and learning. For reasons of space, only a sample of the questions in the survey will be presented here.

Relationship between grading systems and features of assessment
The first three questions are designed to uncover the relationship between grading system and assessment. Question number one asks about the extent to which the emphasis in assessment is one important or unimportant parts of the course. The response alternatives have been reduced to three intervals, each comprising three steps. The data are given in table one below.

Table 1. Assessment emphasising important or unimportant parts of the course.

<table>
<thead>
<tr>
<th></th>
<th>Unimportant 1-3</th>
<th>Average 4-6</th>
<th>Important 7-9</th>
<th>Total numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass only</td>
<td>4 (3.4%)</td>
<td>29 (24.6%)</td>
<td>85 (72.0%)</td>
<td>118 (100%)</td>
</tr>
<tr>
<td>Pass with distinction</td>
<td>22 (8.2%)</td>
<td>118 (43.7%)</td>
<td>130 (48.2%)</td>
<td>270 (100%)</td>
</tr>
<tr>
<td>Total numbers</td>
<td>26 (6.7%)</td>
<td>147 (37.9%)</td>
<td>215 (55.4%)</td>
<td>388 (100%)</td>
</tr>
</tbody>
</table>

Chi-square = 19.2. p<. 001

The differences between the groups are substantial. A vast majority of the students in pass/fail grading systems describe their assessments as focusing on important parts of the course, whereas a little less than half of the students in graded systems do this.

Question number two asked the students to indicate the relative distribution of assessment tasks with regard to whether they required factual knowledge or understanding. Informants were asked to allocate numbers totalling 100 per cent for the different kinds of knowledge, respectively. The results are given in table two.
The response distribution for this question is somewhat unexpected. Although there is a weak trend showing that the pass/fail only students rate their assessment tasks as emphasising understanding somewhat more than the students in the graded programs, the difference is nevertheless not statistically significant.

Question number three brings up features of the assessment tasks once again. Here, the students are asked to indicate whether the assessment items emphasise what in the question is defined as reproduction of knowledge i.e. whether it is sufficient to answer a question by quoting a textbook or applying standard procedures for problem solving. On the other hand, the emphasis may be on production of knowledge, i.e. combining or synthesising knowledge or problem-solving strategies to suit new situations and problems. The respondents were asked to allocate percentages totalling 100 per cent for the two kinds of knowledge. The results are shown in table three.
of the assessment task towards focusing on more unimportant parts of the course, demanding more reproduction of knowledge, and the assessment seem to be more focused on factual knowledge than understanding. In the next section, we will turn to the question of how assessment influences the student’s approaches to learning.

Assessment and student approaches to learning
As argued, multi-grading systems influence the construction of the assessment tasks in specific ways. In this section, we will focus on how the assessment tasks influence the student’s behaviours.

The fourth question asks the students whether they think that pass/fail grading would promote good learning and encourage co-operation between students. The response alternatives were Yes, No and Hesitant. The results are summarised in table four below.

Table 4. Would pass/fail grading promote good learning and co-operation between students?

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>Hesitant</th>
<th>Total numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass only</td>
<td>70 (57.4%)</td>
<td>10 (8.2%)</td>
<td>42 (34.4%)</td>
<td>122 (100%)</td>
</tr>
<tr>
<td>Pass with distinction</td>
<td>74 (27.2%)</td>
<td>105 (38.6%)</td>
<td>93 (34.2%)</td>
<td>272 (100%)</td>
</tr>
<tr>
<td>Total numbers</td>
<td>144 (36.5%)</td>
<td>115 (29.2%)</td>
<td>135 (34.3%)</td>
<td>394 (100%)</td>
</tr>
</tbody>
</table>

Chi-square = 47.7, p<.001

Students in pass/fail programs are convinced that their own grading system promotes good learning and student co-operation. Quite a few of the students in multi-step graded programs, however, also share this opinion. A third of the students in both groups are hesitant with regard to this issue.

The fifth question deals with the priority given to previously given assessments tasks when preparing for the assessment. The percentages for the two groups are given in table five.
Table 5. Priority given to previous assessment tasks when preparing for the assessment

<table>
<thead>
<tr>
<th>Low priority 1-3</th>
<th>Average 4-6</th>
<th>High priority 7-9</th>
<th>Total numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31 (26.5%)</td>
<td>35 (29.9%)</td>
<td>51 (43.6%)</td>
<td>117 (100%)</td>
</tr>
<tr>
<td>Pass with distinction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 (6.4%)</td>
<td>80 (24.4%)</td>
<td>156 (69.9%)</td>
<td>268 (100%)</td>
</tr>
<tr>
<td>Total numbers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63 (16.3%)</td>
<td>115 (29.9%)</td>
<td>207 (63.6%)</td>
<td>385 (100%)</td>
</tr>
</tbody>
</table>

Chi-square = 13.8, p<.001

When preparing for the assessment, the approaches employed by students in multi-step graded programs comprise the study of previous assessment tasks to a greater extent.

Question number six was formulated: How important is the upcoming assessment when you plan your studies? Mark your answer in the scale below. 1=unimportant and 9=totally decisive. The distribution of answers is given in table six. The original 9-step scale has been reduced to three intervals labelled unimportant, average and important.

Table 6. The importance of the assessment when planning the studies.

<table>
<thead>
<tr>
<th>Unimportant 1-3</th>
<th>Average 4-6</th>
<th>Important 7-9</th>
<th>Total numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 (13.9%)</td>
<td>45 (36.9%)</td>
<td>60 (49.2%)</td>
<td>122 (100%)</td>
</tr>
<tr>
<td>Pass with distinction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 (5.4%)</td>
<td>68 (24.6%)</td>
<td>193 (69.9%)</td>
<td>276 (100%)</td>
</tr>
<tr>
<td>Total numbers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 (8.0%)</td>
<td>113 (28.4%)</td>
<td>253 (63.6%)</td>
<td>398 (100%)</td>
</tr>
</tbody>
</table>

Chi-square = 17.8. p<.001

Students in programs with pass/fail grading pay less attention to assessment when planning their studies than do their colleagues in programs with multi-step grading systems.

In this section, we argue, that there is a relationship between the assessment task and the student’s behaviour. Students in programs with multi-step grading systems think to a greater extent than students in programs with pass/fail grading that previous assessments are important as a support to prepare for their assessment and that the assessment is important for how they plan their studies. In the next section, we will discuss the relation between grading systems, assessment and learning.
Grading systems, assessment and learning

The last example from the questionnaire to be reported is whether the assessment is an opportunity for learning. The response alternatives are Yes, No and Hesitant. The results are given in table seven.

Table 7. The extent to which assessment is a further opportunity for learning.

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>Hesitant</th>
<th>Total numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass only</td>
<td>90</td>
<td>14</td>
<td>20</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>(72.6%)</td>
<td>(11.3%)</td>
<td>(16.1%)</td>
<td>(100%)</td>
</tr>
<tr>
<td>Pass with distinction</td>
<td>114</td>
<td>87</td>
<td>72</td>
<td>273</td>
</tr>
<tr>
<td></td>
<td>(32.5%)</td>
<td>(30.7%)</td>
<td>(36.8%)</td>
<td>(100%)</td>
</tr>
<tr>
<td>Total numbers</td>
<td>204</td>
<td>101</td>
<td>92</td>
<td>397</td>
</tr>
<tr>
<td></td>
<td>(51.4%)</td>
<td>(25.4%)</td>
<td>(23.2%)</td>
<td>(100%)</td>
</tr>
</tbody>
</table>

Chi-square = 48.0, p<.001

The observed chi-square value is very high, indicating a substantial difference between the groups. Students in programs with pass/fail grading to a greater extent experience assessment as an additional opportunity for learning than do the students in programs with multi-step grading systems.

If we summarise all the seven questions, we can see that there is a clear relationship between the grading systems, the assessment tasks and the students’ learning. In programs with a multi-step grading system, the assessment tasks focus more on unimportant parts of the course, on factual knowledge and reproduction of knowledge. We argue that these kinds of assessments influence the quality of the students’ learning. Instead of promoting a deep approach to learning, they promote a surface approach to learning (Marton et al. 1999). Such a statement is further supported when summarising questions four to six. As can be seen, a multi-step grading system, to a greater extent than a pass/fail grading system, steers the students towards focusing more on the assessment task in itself when planning their studies than on learning. Further, such a system to a lesser extent promotes cooperation between students. The last question supports the conclusions mentioned above as students from programs with a multi-step grading system to a lesser extent than students in a program with pass/fail grading experience the assessment as an additional opportunity for learning.

Discussion

In this article, we have proposed a chain of relationships between grading systems, assessment, judgement criteria, and students’ approaches to learning. Even though we were not able to established matched groups as regards field of study, for reasons mentioned above, we still claim that the results show empirical evidence supporting the hypothesis that different kinds of grading systems co-vary with the kind of assessments presented to the students. The assessment, in turn, has a direct influence on the students’ ways of preparing for the assessment. In brief, a multi-step graded assessment appears to co-vary with an emphasis on reproductive objectives, focusing not only on essential parts of a course, but also to a large extent on less important parts. Almost three quarters of the students exposed to pass/fail only grading, experience assessment
as a further opportunity for learning, whereas just under a third of their colleagues in graded systems have such experiences of the assessment. Moreover, assessments in themselves, particularly previous assessments, play a greater role for students in graded systems than for those exposed to pass/fail systems.

What remains to be investigated at an empirical level is the character of assessment criteria in relation to different grading systems. The study presented here provides indirect evidence of existing differences depending on which grading system is applied.

These findings are important against the backdrop of changes taking place and the debate in Europe about harmonising the higher education systems. The contemporary discussion can be viewed as containing a tension between conflicting demands on the feasibility of criteria for judgement. The need for comparability between countries in Europe requires general and broadly applicable judgement criteria.

The Bologna Working Group on Qualification Framework (2005) published a report giving recommendations and suggestions for an overarching ‘Framework for Qualifications of the European Higher Education Area’ (EHEA). A thorough investigation on the national arena of what qualification on different levels means and the formulation of descriptors and learning outcomes for these levels are recommended. It is pointed out that descriptors for a European framework by their nature must be general since they must be applicable over a broad range of disciplines, and accommodate national variations. For practical reasons, such criteria must be short and easy to understand. In the discussion about learning outcomes and judgement criteria, it is essential to also consider their relationship with the prevailing grading system.

There is research evidence showing that in order for judgement criteria to be applicable in assessments, they need to be locally situated within a particular field of knowledge, and that they need to be articulated by and mutually negotiated between the actors who do the assessing and those who are assessed (Rust 2001; O’Donovan et al. 2001; Wilmott 2003; Currie 2000). If judgement criteria are too general, there is an obvious risk that they will either be applied more or less in a mechanical way or be subject to different interpretations by different teachers, which will in turn jeopardise the possibilities of comparing students’ performances. The discussions about judgement criteria must also include grading systems and, as our results show, grading systems are not neutral in relation to students’ approaches to learning. It is reasonable to assume that multi-step grading systems bring with them an attenuation of the examiners’ attention to be able to discriminate between performances on a number of different levels. Consequently, the critical question about what constitutes features of an approved and a non-approved performance, respectively, is paid less attention. Furthermore, judgement criteria designed for fine-graded discrimination of performances may also emphasise quantitative rather than qualitative aspects of knowledge. Students are not stupid; they may quickly learn how to comply with such criteria when taking examinations.

Apart from the differences regarding judgement criteria observed between the two groups of students, it is also obvious that they report on fundamental differences concerning the type of assessment tasks they are facing. In line with the discussion above, one might hypothesise that to inform grading in a multi-step system the single assessment task must permit a corresponding grading. Hence, there is a risk that the task itself invites students to focus on quantitative rather than qualitative aspects of the response.
Moreover, it is our conviction, based on previous research as well as the empirical study reported here, that the implementation of a 7-step grading system may compromise the quality of students’ learning to such an extent that we should refrain from recommending it. During the course of their studies, all students adjust to and comply with the kind of grading system to which they are exposed. The claim made here is that it is well documented that there is a causal relationship between the nature of the grading system, significant features of the assessment tasks and the students’ subsequent approaches to learning. Since the latter has previously been found to determine the nature of the outcomes of learning, we could also state that, in view of a long chain of cause and effect relationships, grading systems are among the most influential factors determining the quality of higher education, if not all kinds of education. Previous research has mainly emphasised the relationships between the content and structure of assessment tasks and students’ approaches to learning. In an indirect way, this relationship has also affected the nature of the learning outcomes. As a result of this kind of research, enormous efforts have been made to inform and educate staff in higher education in the field of examination. The time has come to include the structure of grading systems more systematically in this area of research. With a contribution from the present study, we know that multi-step grading systems generate assessment tasks as well as judgement criteria that force students to adopt approaches to learning that jeopardise a deeper understanding of the course content. From a political perspective it is, of course, attractive to facilitate mobility of students and staff in higher education as well as academic professionals. Academic staff would certainly be willing to support this idea as well. It is, however, a fact that in this case, measures taken on the political level have severe repercussions on the inner life of institutions of higher education. Since the quality of higher education seems to be at stake, academics must also have a say on this issue. Examination in higher education as well as in all parts of the education system is a highly interdependent system of grading, assessment tasks, judgement criteria, students’ approaches to learning and features of the learning outcomes. However, this phenomenon is only partly understood. One way of deepening our understanding of examination is to try to place our partial knowledge into a larger context of knowledge, for the benefit of quality in higher education.

References


Currie, J., Harris, P. and Thiele, B. 2000. Sacrifices in greedy universities: are they gendered? Gender in Education 12, no. 3: 269-292


---

1 Lars-Owe Dahlgren is Professor of Education at Linköping University and a visiting professor at Karolinska institutet (medical university), Sweden. His field of research is higher education and he has published extensively within this field. He is currently leading a research project focused on students’ transitions from higher education to working life.


3 Madeleine Abrandt Dahlgren is Professor of Education at Linköping University, Sweden. Her field of research is higher education. Her publications include a variety of books and articles, particularly on student-centred educational design, problem-based learning in different academic contexts, cross-cultural learning in a web-based environment. A recent research focus is students’ transitions from higher education to working life.

4 Nils Trowald is a former senior lecturer at the Swedish university of Agriculture. He is now retired from his position.