Formative Assessment in Peer Review Settings Online

Agneta Hult and Monica Liljeström
Umeå University
Sweden

Abstract
This paper draws upon findings from an analysis of the written feedback a group of Swedish University students produced as peer reviewers. The study aims at identifying what type of feedback the students provided each other with in order to gain some preliminary insights into if and how peer review preceded by collaborative criteria processing could contribute to learning. The students’ peer assessments were analysed according to current research on formative feedback. The results indicate that the peer review activities online supported student learning.

Introduction
Creating good conditions to support the online learner’s development of the complex skills society expects that university education should result in is a great challenge. Online teachers seldom or never meet their students (an often highly heterogeneous mixture of individuals regarding age, life situations and study backgrounds) face to face, but still have to respond to the demands the current shifts in views on the expected role and direction of higher learning create. These expectations reflect the vision of a new society: “The society of tomorrow will require people who are flexible and able to continue to acquire new knowledge and learn new skills” (Dysthe, 2004, p. 3). Higher education is supposed to support and contribute to this production of self-regulated learners (Steffens, 2006) by helping students develop useful tools for active, lifelong and “sustainable” (Boud, 2002) learning. Additional challenges can be identified in the current trends in Higher Education (HE) in many countries towards extremely fine-grained approaches to measuring student achievement in combination with “a strong social drive to help learners, some with personal histories of spectacular ‘unsuccess’, to obtain qualification” (Sadler, 2007, p. 387).

In this paper we will share some of our experiences from exploring how peer assessment in form of peer review online could function as a strategy to meet the current challenges in HE. We describe how the peer review element was carried out as a foundation to the results and conclusions drawn from performing the empirical study. This study is based on data collected from peer review activities in a message board facilitated in FirstClass for students who participated in a course in a special needs education training program. In this study the students’ feedback to each other is in focus to shed more light on what the implementation of formative assessment in the form of peer review could have meant for the students’ learning. Previous research on feedback was applied in the analysis to
understand: 1.) the nature of the feedback that the students provided for each other; and 2.) the relevance of the feedback as support for learning.

**Background**

The peer review experience reported in this article has its origin in the interest in what strategies online teachers are using to support, monitor, and measure students’ learning and achievements online — an environment where teachers seldom, if ever, have the opportunity to meet their students face to face. What strategies are found useful in education under these conditions? How do these strategies respond to the many challenges in current educational discourse in a time when HE (at least in Sweden) is characterized by raised expectations despite rather limited resources to provide quality education for large, heterogeneous study groups? A study of assessment in 50 online courses in HE in Sweden (Hult, 2008) revealed that the course structure often builds on the principle that the course content is processed through many and frequent smaller assessment tasks on which students receive teacher feedback.

This assessment strategy could be regarded as a practical and useful solution to handle the mixture of challenges in HE. The assessment products created by the students can provide valuable information about every student’s learning process so the teacher can identify what they have achieved and to what degree. But this information can also be used for further monitoring and support of individual student’s learning and carried out through the feedback procedure. Many online students could be at risk if left alone with their studies without this regular feedback, wrestling with how to figure out what to learn and perform to pass the final exams. Even if many online courses offer the possibility to interact with each other, research shows that online students do not always use this opportunity (Rourke & Anderson, 2002). The mass-education system in many countries means that some students might enter HE with a “world view” so different from the views within the academy that there is a risk that they will have great trouble identifying what they are supposed to perform (Bizzell, 1986). Online students’ access to artefacts, staff members and others representing the academy is limited since they seldom or never come to campus, which can make it harder to grasp the expectations derived from the academic tradition to which their performances have to correspond.

The idea to organise courses with assessment as the central method to process the course content could to some degree be justified by research studies on assessment in which a strong relation between assessment and what students learn and how they learn has been identified (Becker, Geer, & Hughes, 1968; Miller & Parlett, 1974). This relation has stimulated many researchers to take interest in the question of how to develop assessment strategies that promotes learning. As a result assessment of learning has since the 1980s been supplemented by assessment for learning (Gipps, 1994) in research and writing as a way of enhancing student learning. Sadler (1989) elaborated on this kind of assessment strategy which he labelled formative assessment. Building on earlier work by
Ramaprasad (1983), Sadler suggested that feedback should offer two things: information about the gap between actual and expected performance; and the means for reducing this gap.

In this aspect the idea to structure a course around frequent assessments seems like a good idea since the teachers’ comments on these assessment tasks can fulfill such function. Research has indicated that different types of formative assessment can reduce the risk of students dropping out of the course and increase the chance they complete the course within the time frame.

However, in order to reduce this gap Sadler also stressed the importance of helping students to appreciate the assessment criteria so they can understand how to direct their learning towards expected learning outcomes. It is questionable that arrangements where teacher feedback on students’ assessments is central always supply the student with such help. Warnings are raised about the recent increased use of formative assessment procedures based on detailed criteria — at worst it could create more teacher dependent students and instrumental, reductionist learning where the intended learning is displaced by procedural compliance (Sadler, 2005, 2007, 2009; Torrance, 2007). Torrance (2007) described the change of focus in assessment and the risk of changing from assessment of learning, over assessment for learning to assessment as learning.

Peer assessment when preceded by students and tutors discussing and interpreting criteria is suggested as a possible strategy to counteract the risk that the use of formative assessment to support learning could trigger instrumental approaches to learning. Summarizing research on peer assessment Topping (2010) concludes that: “Students need to be involved in developing the criteria for assessment in order to feel a sense of ownership and decrease any anxiety” (p. 69). The idea is that this can enhance students’ understanding of the qualities embedded in the academic tradition and expectations represented by the criteria and increase the understanding of how students are meant to direct their learning. Reviewing the work of peers offers the students insight into a variety of approaches to solve the same task and could thereby reduce the risk of promoting instrumentality (McLuckie & Topping, 2004). The conclusion drawn in many previous studies of peer assessment is that engaging students in formative peer assessment sustains the idea of autonomous, independent, and self-directed learners who take responsibility for their own personal and professional development and direct their learning towards successful achievements (Anderson, Howe, Soden, Halliday, & Low, 2001; Bloxham & West, 2004; Boud, 2002; Higgins, Hartley & Skelton, 2002; Lorraine & Stefani, 1998; Macpherson, 1999; O’Donovan, Price, & Rust, 2004).

This background made us interested in altering the strategy with teacher centred formative assessment to be able to explore how peer review could work instead. This resulted in a project in which the peer review element was constructed and implemented. Among expectations in the special needs teaching program (where the peer review element was tested) were that the students should learn to identify problems relevant in the field of special needs education; develop the ability to plan, perform and report a
study in which such problems are focused, and be able to draw conclusions from this about how such problems could be tackled in their future profession. Accordingly we created a model for peer review in which overarching steering documents and course criteria were discussed in relation to scientific research and reporting and the peer reviews were carried out on the students’ work in progress while planning, performing, and reporting a field study.

The students were gathered on campus on three occasions and the peer review idea was introduced at the first campus meeting. Thereafter all the peer review activities were carried out online. The students were prepared for peer review and insightful feedback on each other’s assignments through asynchronous, written communication organised as workshops. The study programme was facilitated in FirstClass, so they were familiar with this platform from earlier courses. The students were already assigned to smaller study groups varying from 4–8 participants. Each group had their own discussion forum.

Besides the pre-planned assessment task to conduct a field study to be presented in the form of a written 10-page research report the students also had to assemble a portfolio with literature and lecture comments. To stimulate participation in the workshop without setting up rules for how many postings students had to contribute we asked students to take into account in their self-evaluation the workshop discussions, and exemplify how they and others had contributed to the learning process.

The students were tutored with questions aimed to challenge their understanding during their discussions in the first workshop period. The workshop began with an initial discussion of course criteria in the light of the Higher Education Act and the Higher Education Ordinance, which could be described as overarching steering documents for all Swedish universities with the official expectations for general (generic skills) and programme specific (professional skills) outcomes of university studies. The ambition was to stimulate the students to interpret and negotiate the meaning of these documents in depth, with the goal to reach a preliminary consensus of what criteria could be valuable to put in use when performing the peer reviews and what qualities these criteria aimed to capture.

The second step was a workshop period where the students individually reviewed two example texts written to correspond with the instructions the students had when writing their own reports. The texts were authored to resemble student essays. Both had strengths and weaknesses to give the students plenty of issues to debate. One text was authored in a purely referential style with a weak ‘author voice’ based entirely on studies already conducted on the research subject but with no clear purpose as to why they were referred to. Text two was written in argumentative style in which the ‘author voice’ was present. It also referred to previous studies already conducted on the research subject but also connected to related issues such as general trends in society and studies made in other fields were also mentioned as a part of the argumentation chain.
The students’ reviews of the articles were processed in the study groups where similarities and differences in reasoning were identified. They were challenged with questions designed to have them more closely scrutinize how the texts were constructed and asked to justify their reasoning in determining strengths and weaknesses in these texts. After this stage they had the option to modify the criteria they had chosen for their peer reviews to correspond with their new insights. Finally the students practiced these (eventually) new insights on reviewing a journal article which was part of the course material, and as a finale had a last discussion about what they had learned from the training in the workshop.

The students then reviewed their peers’ work in progress. The first review was on their preliminary plans for their research study, and the second was on the drafts for the final report of the results from their research studies.

Method

During the 10-week distance course data was collected from the two last session in the workshop where the students reviewed each other’s preliminary plans and the final drafts. The course was called Developmental Work, Leadership and Evaluation and was provided during spring 2008. Their previous study experiences varied in length and content as their professions ranged from preschool teachers to college teachers and in age they ranged from late 20s to over 50. Only three students were men. Due to the enormous amount of data generated by the students’ activity in the workshops during the peer review period we selected four out of nine groups for analysis. Groups had 6–8 members. We picked two groups with a high rate of postings during this period (133 and 126 postings respectively) and two groups with a lower rate of postings (69 each). These postings were not entirely peer review comments on course work, but included among other things social “talk,” practical course matters, and some teacher comments.

In order to understand if and what the feedback these students gave each other could mean as support for learning we consulted previous research on feedback. Peers as a source of formative assessment have been studied a lot with somewhat mixed results (Shute, 2008). The varied results are partly related to the variety in forms of feedback, context and preparation for feedback, learner characteristics, and outcomes measured (Topping, 2010). Studies of reliability and validity of peer assessment suggests that there is not much difference between teachers and peers when it comes to quality of their feedback (Topping, 2003). The reliability then builds on the rather doubtful assumption that teacher assessments were themselves highly reliable. According to Topping (2010) these studies could instead be considered validity studies since they compare peer assessments with those done by professionals, rather than by those made by peers or by the same peers over time. Higher reliability seems to be related to more advanced courses and academic products (writing and presentations) rather than assessment of behavioural skills. Also important for an increase in reliability are the deepening understanding and
sense of ownership that come with discussion, negotiation, and joint construction of assessment criteria with learners (Topping, 2010).

In a review of research on formative feedback Shute (2008) presents suggestions or prescriptions for enhancing learning with formative assessment in four tables under the headings things to do, things to avoid, timing issues, and learner characteristics. Due to space reasons only a selection from each heading is presented here. Things to do include focussing the feedback on the task not the learner, providing elaborated feedback in small enough pieces, and linking feedback clearly to the goals. Among things to avoid are feedback which discourages the learner or threatens his/her self-esteem and progressive hints terminating with the right answer. Concerning timing issues there is a wide support for use of immediate feedback to promote learning, especially for difficult tasks (relative to the learner’s capabilities). In the last table, learner characteristics are combined with type of and timing of feedback. For low-achieving learners the reviewed research suggests that directive or corrective feedback received soon after performance works best. High-achieving learners benefit more from facilitative feedback that challenges them, such as hints and cues.

The first step in analysing the students’ feedback to each other was to find a way to capture the nature of the feedback. We found a model for analysis in a study by Tseng and Tsai (2007) which was based on a framework elaborated by Chi (1996). The categories in this model were interesting since their results indicate a relation between different types of feedback and their support for learning. We found that the categories Reinforcing, Suggestive, Corrective, and Didactic which Tseng and Tsai used were relevant for categorizing the students’ feedback to each other in this first approach to capturing some of the nature of these comments. The categories used in our version are described in Table 1.

These categories were used for identification of the overall balance between the different types of feedback the students gave each other. The principle for determination of coding units was that each time the topic changed a new coding unit started, since the same feedback comment could contain more than one type of feedback. The authors of this paper calibrated their coding principles several times before and during the analysis by selecting concrete samples from the collected data and discussing how to categorise them. In addition to this pre-determined approach we also evaluated their comments in a more qualitative way by paying attention to how the feedback was formulated and received, e.g. if it opened up for discussion and reflection.
Table 1: Analysis Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reinforcing</td>
<td>In different ways reassuring that the performance meets the requirements. “Good clarity and you have pointed out a few aspects which are important for the...” “You have connected to previous research in a relevant way and connected to your own research.”</td>
</tr>
<tr>
<td>Suggestive</td>
<td>Points out that something is incomplete rather than incorrect, and includes suggestions for areas of improvement, thus alerting that there is a problem without telling exactly what the problem is. Such feedback can be in the form of hints. “If I were you I would find out what’s already done and focus on a problem in the school field which they are working with.” “Many references to course literature, maybe some references to the articles as well?”</td>
</tr>
<tr>
<td>Corrective</td>
<td>Points out that something is completely wrong, e.g. design of the report, the content, the usage of theory, references. “The author is called Kvale, not Kavle.”</td>
</tr>
<tr>
<td>Didactic</td>
<td>A more lengthy explanation on errors or inadequate information provided. Lengthy explanations with a lecture tone are taken to direct the peer to be on the right track. “This part of the report is too thin. You need to give the reader more insight in...”</td>
</tr>
</tbody>
</table>

Findings

Although no instructions were given for how many comments each students should contribute, the overall feedback pattern in all study groups indicated a great engagement during the feedback period. All students in each study group formulated feedback to all other group members on the first draft. These contributions could be posted any time during the time the workshop was open to enable flexibility, which means that they all received both almost immediate response (on the same day from some peers) and delayed response (at the end of the workshop week from other peers). The feedback given was entirely focussed on the task and how to improve what the student had written, except for some supportive comments on a personal level, e.g. “you are very clever.”

All groups had during the first work shop elaborated criteria for their peer assessment and most of the comments linked directly to these criteria, to learning outcomes, and to the goals of the assignment.

The analysis of the data collected from four study groups reveals interesting feedback patterns as we can see in Table 2. As is illustrated in this table, the students gave each other a lot of reinforcing feedback. Sometimes these types of reassuring comments seemed to fill a function to ease the stress some of their peers felt about their ability to meet the requirements of the assessment task. But some of these comments also seemed to fill self-reinforcing purposes, as they sometimes gave appraisal and thereafter pointed...
out the similarity with their own work and expressing how comforting it was to see that others used the same strategy/had the same thinking as they did.

Table 2: Feedback Patterns

<table>
<thead>
<tr>
<th>Group</th>
<th>Type of feedback</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reinforcing</td>
<td>Suggestive</td>
<td>Corrective</td>
<td>Didactic</td>
<td>Total</td>
</tr>
<tr>
<td>Groups with lower posting</td>
<td>84</td>
<td>70</td>
<td>20</td>
<td>5</td>
<td>179*</td>
</tr>
<tr>
<td>rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groups with higher posting</td>
<td>172</td>
<td>115</td>
<td>39</td>
<td>16</td>
<td>342</td>
</tr>
<tr>
<td>rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>256</td>
<td>185</td>
<td>59</td>
<td>21</td>
<td>448</td>
</tr>
</tbody>
</table>

*During the analysis we discovered that the number of comments in one group was low because the students in this group lived in the same area and also met face to face to discuss the assignment.

The analysis also shows that the students gave each other a great amount of feedback of a more suggestive nature. Sometimes these types of comments had a negotiating tone as, for example: “This could have been written under the headline ‘implications for further research, or alternatively under some other headline, because what you suggest here is related to that (I think) but I don’t think this was required.”

Some comments had a corrective tone. These types of comments ranged from pointing out misspellings or that a used word was improper to structural issues (e.g., that method reasoning appeared in the findings section of the report). However, as shown in Table 2 this type of comment was used relatively sparsely.

Clearly some students had more confidence than others in how to write a report in a fashion that would correlate to the explicit and implicit expectations of this kind of assessment product. Some of these students’ feedback could clearly be classified as “didactic” to its nature. For example one of them colour marked a segment of a text to illustrate different weaknesses in it and explained in depth how this text could be improved.

In some cases the feedback from others became overwhelming with the result that the receivers of feedback expressed how insecure the feedback made them about their own ability. One student, after receiving mostly didactic and corrective comments, posted an answer in which she said that the feedback she had from her peers made her doubt that she should be partaking in this course as she felt that she did not have the capacity to produce a report that would pass the examination. This student received many lengthy comments of supportive character in which the members in her study group ensured her that they wrestled with similar self-doubt, that they were sure she could perform well enough to pass the examination, and that the feedback given was only suggestions meant
to help her improve her work. They also pointed out that it was up to her to decide what relevance these suggestions had and how she should act upon them.

**Conclusions**

The high activity level during the peer review period indicates that the opportunity to give and receive comments on their work in progress was found to be valuable. Students could have made a minimum of postings since the number of feedback postings each student should present was not regulated by rules.

To understand if the students’ comments to each other were valid we asked the teachers in the course who had met this student group in previous courses and also had carried out this particular course for many years if there was any notable difference in the overall students’ performance. The teachers’ impression was that the student group became more homogenous in the sense that low achievement students seem to meet the requirements for this assessment in higher degree. This could be an indirect indication that the peer review preparation (negotiating criteria, discussing practise reviews) helped the students gain a “deepening understanding and sense of ownership” (Topping, 2010) and a similar understanding of the criteria as their teachers.

Relating to what Shute (2008) identified as valuable in formative feedback, e.g. linking feedback to the goals, it seems that the feedback that these students gave each other had a great potential to promote learning.

Tseng and Tsai (2007) found in their study that reinforcing and suggestive feedback seemed to support the quality of the students’ work, while corrective and didactic feedback seemed to work in the opposite way. As could be seen in our study both corrective and didactic feedback are sparsely used, while the amount of suggestive feedback is almost as large as reinforcing. According to the findings of Tseng and Tsai the students’ feedback patterns would indicate that they enhanced each other’s performance. This was also indicated in the students’ evaluation of the peer review element. Previous analysis of data from this course has shown that the students engaged intensively in the peer assessment activities and that they found this element valuable for their learning (Liljeström, 2008; Liljeström, 2009; Liljeström, Hult & Stödberg, 2008).

One of the risks that has been pointed to with formative and criteria based assessment is that it could trigger students to reductionist learning only focussed on fulfilling limited criteria and details, e.g. formalities like flaws in the references. However, the results from this study indicate that participating in these workshops and peer review did not produce many comments with this approach. As the results have shown the peer review element seems to have stimulated more than just a simple check that the reports fulfilled explicit criteria since there was a richness in discussions both regarding issues that had to do with performing research and reporting its results as well as how they could put these insights
to use when approaching similar tasks in their future profession. This could be an indication of sustainable learning in Boud’s (2002) sense.

References


Agneta Hult, PhD, is senior lecturer at the Department of Education, Umeå University, Sweden. Her research focuses mainly on assessment and evaluation in higher education. Most recently her research has been directed to peer review and collaborative processes in Internet-based courses. She is the head of a research group in the department that
concentrates mainly on research on evaluation and assessment. In her teaching she has been engaged in tutoring on different levels and lecturing on assessment and research methods, foremost qualitative methods. She is also involved in the teacher training program and a member of the board for Umeå School of Education.

*Monica Liljeström*, PhD, is a senior lecturer with an interest in learning and ICT. She works at the Department of Education, Umeå University, Sweden. Her research is aimed at understanding how sociocultural learning theories can be put in practice when organizing online education activities. Dr Liljeström has recently contributed to the growing understanding of collaborative learning online through articles, book chapters, and conference papers in which she has reported the results from her studies of peer review activities in asynchronous text based discussion forums as means to stimulate collaborative learning processes online. Her latest publication is her doctoral thesis: *Learning Text Talk Online. Collaborative Learning in Asynchronous Text Based Discussion Forums* (2010).