A Reflection on the Introduction of a Peer and Self Assessment Initiative

Kath MacDonald
School of Health Sciences, 
Queen Margaret University Edinburgh
kmacdonald@qmu.ac.uk

Abstract

This paper discusses the introduction and evaluation of peer and self assessment elements into two nursing modules, (one undergraduate, (UG) and one postgraduate, (PG). 40 UG nursing students and 22 PG education students participated in a Peer and self assessment exercise within their respective modules. Students evaluated the process by questionnaire and made recommendations for future modules. 32 out of 40 (80%) undergraduate & 20 out of 22 (91%) post graduate students returned the questionnaires. Over two-thirds of each group recommended introducing peer assessment into their modules.

Students saw peer and self assessment as motivational, encouraging learning, and a fair and truthful method of assessment. Weaknesses were seen as possibility of personal bias, students unsure about their assessment skills and may disadvantage quiet students or those with English as a second language. Both groups felt peer feedback would further their own learning. Both groups scored their peers at the higher end of the range of scores. Peer and self assessment is seen positively by UG and PG students and is perceived to motivate, facilitate learning, and be fair. Consideration must be given to less dominant group members, as quiet students do not necessarily produce less effort. Peer and self assessment may help students develop sustainable skills which can be used in the workplace, such as judgement and assessment of self and others.

Keywords: peer assessment, self assessment, sustainable learning, reflection.
Introduction

Assessment has been defined as serving several purposes: ranking and grading students, maintaining standards, a means of providing feedback and enhancing student learning and growth, (Rowntree 1987, Boud & Falchicov 2006). Assessment is also seen as a means to developing key sustainable skills that can be utilised in professional careers, such as judgement, and assessment of self and others, (Boud & Falchicov 2006, Quality Assurance Agency (QAA) 2006). A suggested means of achieving these skills is through the use of peer and self assessment practices, (Falchicov & Goldfinch 2000). These practices have been adopted globally in a range of disciplines including social sciences, languages and engineering (Topping, Smith, Swanson & Elliot 2000).

This paper aims to critically evaluate the introduction of peer and self assessment elements into an undergraduate (UG) and Postgraduate (PG) module within the faculty of Nursing in a Higher Education Institution. Both modules are currently summatively assessed by awarding a group mark based on a collective activity.

Student feedback about equity of work within groups and the author’s personal experience as a post graduate student on one of the modules prompted this initiative. It was felt that adding in peer and self assessment elements may, a) increase student engagement, b) be more representative of group processes; (involvement and commitment of group members). c) help develop sustainable skills such as assessment and reflection.

The literature on the use of self and peer assessment will be reviewed. Studies on peer and self assessment are reported separately where possible. However where studies incorporate both elements these are reported together. It is not within the scope of this paper to discuss group assessment, as this is a well established practice within the institution. Students are familiarised with theories of group working (Belbin 1981) early in their studies, particularly in those programmes which result in professional outcomes (e.g. nursing and allied health) where working in groups is standard practice. Critical reflection as a tool for learning is supported by several key authors: (Schön 1987, Kim 1999, Taylor 2000, Johns 2002, Fulbrook 2004, Fowler 2007) and has facilitated innovation in practice (Markham 2002). Reflection provides a forum where learners can discuss issues, deconstruct these and learn from them. Thus reflection (the author’s and the participating students) underpins this educational initiative.
Literature Review

Self Assessment

Boud (1995) and Race, Brown and Smith (2005) contend that students have always self assessed and are continually assessing on an informal basis. Upon submission of work students have already made a judgement on their own performance. Boud, (1995 p5), defines self assessment as; “the involvement of students in identifying standards and/or criteria to apply to their work and making judgements about the extent to which they have met these criteria and standards”.

Self assessment is said to be influenced by gender, with female self-rated scores lower than males, (Langan et al 2008). Females also report feeling more stressed by self assessment than males, (Pope 2005, Harlen & Deakin-Crick 2002). There is also variance in self assessment across academic ability, with lower achievers self-rating more highly than higher achievers, (Boud & Falchicov 1989, Lejk & Wyvill 2001, Papinczak, Young, Groves & Haynes 2007).

Boud and Falchicov’s now dated (1989) meta-analysis of self assessment, found confusion between what was actually being assessed; students’ behaviour (i.e. the process of learning) and the end result (the product.) Pope suggests that studies examining the process –typically group work involvement and commitment- are typically matched to formative assessment as they are difficult to measure numerically, whereas product based assessments, such as the production of a poster or report, lend themselves more to summative assessment. This was the case in the two modules under discussion. A combination of both process and product driven assessments may help to produce self reflective learners (Pope 2005), thus the intention to assess both and reward them summatively.

Increasing student engagement through self and peer assessment

learning which subsequently increases student motivation. Falchicov echoes these advantages and adds that student involvement also helps provision of feedback, development of communication skills and saves teachers’ time. However Cowan’s seminal paper (1988) opposes the view that self-assessment is time saving as development of assessment skills is complex and more time needs to be given to engaging students with the learning criteria in order to be effective assessors. This could also be said to be true of peer assessment as the skills which need to be learned are similar.

Race (2002) supports the view that peer assessment works better when students are engaged in the setting of performance criteria. However, Sluijsmans, Brand-Gruwel, Van Merrienboer and Martens (2004) suggest that despite training in this area, students are more inclined to reward lower level skills (e.g. rewarding content of presentations rather than analysis of data). They recommend that training needs to be of longer duration in order to align peer and faculty results.

**Bias in peer assessment**

The potential for bias in friendships in relation to peer assessment has been suggested, (Magin 2001). Whilst the literature does not support this, (Falchicov & Magin 1997, Magin 2001) there is a strong body of literature (cited in Magin 2001) contending that those who are being assessed report concerns that friendships may influence peer marks. Conversely, peer assessment is said to weed out “free riders” who may go unnoticed by lecturers but are more easily identified in group processes by peers who grade them accordingly, (Bourner, Hughes & Bourner 2001, Elliot & Higgins 2005). There is also evidence to support collusion in peer marking (Rafiq & Fullerton (1996); and a suggestion given to avoid mark fixing is to operate peer marking confidentially in exam-like conditions.

**Reliability of peer assessment**

Whilst most studies favour tutor assessment as the gold standard in trying to prove reliability of assessments, Race contends that multiple assessments cause regression to the mean implying that tutor assessment may not be infallible. Stefani (1998) describes a “proliferation of studies in the in the 1980’s and 1990’s giving complex
statistical data on the comparisons of student and tutor derived marks” p343. What was more important she argued was not quantitative fixation, but the development of “student centred active learning” p344. Similarly Falchikov (2005) argues that students and teachers interpret findings in the context of their own experiences- which are often different; hence agreement between the two sets of marks might not be the desired standard for validity and generalisability of peer assessment. Liu and Carless (2006) suggest that resistance to peer assessment by academics is driven by a perceived disruption in power balance as a well as a concern for reliability. The issue they suggest is not peers sharing responsibility with academics for grading, but about helping students to define learning outcomes and the criteria for assessment through feedback rather than grading. However students are motivated by grades and numbers (Knight 2006) and this creates tensions between the philosophical idealism of assessment and the real world which is driven by success as measured by grades and the avoidance of failure.

**Conflicting evidence in support of peer and self assessment**

Despite the many advocated advantages of peer and self assessment, the evidence to support both forms of assessment as reliable and rigorous forms of assessment is not conclusive, with studies reporting a poor correlation between self, peer and tutor assessed marks, (Sullivan, Hitchcock & Dunnington 1999, Papinczak et al 2007, Langan et al 2008). However Pope’s study contradicts these findings with peer and self assessment highly correlating with faculty awarded marks. Falchicov and Goldfinch’s (2000) meta-analysis comparing peer and teacher marks also support agreement of marks on peer assessment. Falchicov and Goldfinch suggest that peer and self assessment differ somewhat in that peer assessment operates in a public domain whilst self assessment is a more solitary activity, which may not involve comparisons with others. This raises the issue of the nature of the setting of assessment (i.e. confidentially or publicly).

Despite these conflicting findings, there is almost universal agreement from the literature that peer and self assessment contribute to student performance and foster deep learning, through an increased commitment to the process and reflection on one’s own performance and that of others.
Regardless of what the research findings suggest about gender issues, bias, product versus process, summative versus formative assessment and engagement with assessment criteria, the one constant would appear to be that increased engagement with assessment augments deep learning, empowerment and self awareness; some of the skills necessary for sustainable learning in the real world. It was with these potential outcomes in mind that the author embarked on this study.

**Methods**

Two modules were chosen for this pilot initiative (Figures 1 and 2). Both modules have an element of assessed group work and are facilitated by a lecturer. Both modules are given a group mark as part of the summative assessment. Currently no mark is awarded for an individual’s contribution to the group. Self and peer marking elements of assessment were introduced as a proportion of the overall group mark, (Figures 3 and 4). Ethical approval was not deemed necessary for the initiative as the process was debated through the University’s Quality procedures and was considered to be good educational practice rather than research.

**Figure 1. UG Case study**

<table>
<thead>
<tr>
<th>3rd year undergraduate nursing module</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Problem based learning around three case studies, 4 groups of 10 students, n = 40).</td>
</tr>
</tbody>
</table>

**Aims:** To use problem based learning in the investigation of nursing and interdisciplinary approaches to recovery, rehabilitation, anticipatory care and long term conditions. Contribute to group work through the process of problem based learning, using and developing skills of exploration, appraisal and presentation.

**Assessment:** The assessment currently comprises two parts:

- Group portfolio of learning based on one of three packages, (cardiac, neuro, oncology), (60% weighting, tutor assessed)
- Individual reflective account on the student’s journey through the module as part of a group (40% weighting, tutor assessed).

The new initiative incorporated a self and peer element into the group mark. Marking criteria for the peer and self assessments were predetermined and were given to students at the outset of the module. A preparatory session was held to introduce the concept of peer and self assessment and discuss the implications. All students in year 3 are made aware that 20% of the total year mark contributes to their degree classification in the following final year. This module constitutes a third of the year 3 work and so the summative assessment would play a small part in informing their degree classification.
Figure 2. PG Case study

Postgraduate Education module within Faculty of Nursing
(2 groups of 11 students, n=22).

Aims: To enable the student to understand the processes of curriculum design and development. Function as a member of a curriculum planning group and take responsibility for own contribution

Assessment: Currently 2 part assessment:
Group based module where students design and present an educational programme (written) and defend this (oral) at a mock validation event, (60% weighting, tutor assessed).
Individual written account on the process (40% weighting, tutor assessed).
New initiative would incorporate peer and self element into the group mark

Tools

Unstructured open questionnaires were distributed to each group at the end of the modules (Figure 5) and were collected on the last day of each module.

Figure 3. Process for peer and self-marking

- You will be asked to objectively mark each of your peers and yourself on a scale of 1-10, (1-very poor, 10-excellent).
- This should not reflect subjective issues such as personal feelings towards other group members etc.
- If you award a score of less than 3 you MUST justify this in the comments section of the marking grid. (see examples)
- This process will be done confidentially and without conferring with other group members.
- An average mark is then allocated based on the mean of the scores awarded to each member.
- This mark is divided by the mean group mark and results in a peer assessment factor (see example below)
- The peer assessment factor is multiplied by the portfolio mark to give a peer adjusted group mark. This is effectively your individual mark for the group project and counts as 60% of the total assessment.

Example

You are awarded scores of 8,7,7,6,7 by your colleagues and you give yourself an 8
Your mean score is 7.2,
The group mean score is 6.5
The peer assessment factor is 7.2 ÷ 6.5 = 1.1
The group’s poster score is 60%

Your score will be 60% x 1.1 = 66%
Figure 4. Peer Marking Criteria – Exemplar

<table>
<thead>
<tr>
<th>Marks of 7-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actively contributes to the work of the group (in class and online) with ideas and evidence: shares new and relevant information which complements that brought by others, justifies with references, uses a variety of sources, listens to others, encourages participation of others</td>
</tr>
<tr>
<td>Presents and communicates information clearly &amp; concisely</td>
</tr>
<tr>
<td>Regularly contributes to group work (discussion and written work) and is reliable in completing assigned work</td>
</tr>
<tr>
<td>Regularly attends the group sessions (few absences &amp; group kept informed or absences negotiated)</td>
</tr>
<tr>
<td>Demonstrates enthusiasm and commitment in undertaking different group roles (e.g. chair, scribe, group member)</td>
</tr>
</tbody>
</table>

**Mark of 4-6**
all of the above but does not meet the standards in 2 or 3 categories

**Mark of 3 or below**
poor contribution and lack of active involvement towards group work
poor presentation and communication of information to the group
poor attendance at group sessions
demonstrates lack of enthusiasm and commitment to the group

Figure 5. Module Evaluation Form

- What were the strengths of the module?
- Limitations of the module?
- How could this module be improved?
- Can you please comment on the pilot peer marking system: strengths/weaknesses
- Do you think this should be incorporated into the module? Y / N
  - Please give reasons for/ against
- Would you like anonymised feedback (students’ comments) on the justification of the peer marks? Y/ N
- Please comment on why you think this would be useful/unhelpful

Data Analysis

Thematic content analysis was used to identify themes from the open questionnaires. The analysis comprised a combination of a-priori information based on the questions in the evaluation form, as well as an inductive process whereby emergent data was identified. The main a-priori themes clustered around key strengths and weaknesses and suggestions for improvement. Other emergent themes such as “confidentiality” and “feeling under pressure” are also discussed.
Both groups were analysed separately however there were many commonalities across groups (and some differences). Peer review was employed to confirm the credibility of the data, with a fellow lecturer cross checking the data at each stage of the process.

Results

32 out of 40 students in the UG group and 20 out of 22 in the PG group returned questionnaires, (response rates 80%, 91% respectively).

Main themes: strengths

- Feedback from peers can be used to further student learning, (both groups).
- Peer assessment is a good motivator for equity of work (both groups).
- Peer and self assessment is seen as a fair and truthful form of assessment (both groups).
- Peer and self assessment allowed personal assessment of others in response to their effort (both groups).

Main themes: weaknesses:

- Both groups worried about the risk of personal bias in the marking process.
- The undergraduate group felt it unfair that a group mark contributed to their degree classification mark.
- Both groups were concerned that peer assessment would disadvantage quieter students.
- The undergraduate group expressed concerns about confidentiality in the marking process.
- The undergraduate group expressed concern about the weight of responsibility which peer assessment subjected them to.
Discussion

Both groups made positive comments about the peer assessment being a good system (UG 38%, PG 30%) with comments about it being fair and just. However, both groups also felt there was potential for personal bias in the peer marking system; (UG 28%, PG 25%). This is in keeping with Magin’s (2001) work termed the “reciprocity effect”. Despite the contention that students perceive a potential for bias, Magin found only a one per cent variance accounted for reciprocity in a study of 169 matched pairs of students.

Whilst it is difficult to draw conclusions regarding the reciprocity effect in these two studies due to the small number of subjects in the sample, a quarter of students in the two groups perceived this as a potential problem. In practice, however, only 15% of the UG group and 9% of the PG group actually awarded the same marks to each of the group members. This may not be seen as a reciprocity effect, as it could be explained by a lack of engagement in the process or a reluctance to over- or under-mark anyone in the group, however it may suggest bias is a perceived problem rather than an actual one. However it is difficult to draw conclusions given the small numbers.

Both groups felt that peer assessment was a good motivator for making people work within groups (UG 22%, PG 10%), and that peer marking identified the amount of individual effort in groups which may be unseen by tutors, (UG 22%, PG 20%). Several students commented;

peers have greater insight into work effort than tutors and peer mark reflects that

This is in keeping with Bourner et al’s work (2001) who found that the aspect of group work least favoured by students was dealing with “passengers”. Peer assessment may give students some form of control in awarding less marks for less effort to these individuals, or alternatively high marks to students who were perceived as working harder. This comment was reflective of many of the students;

weeds out those who do work unseen by lecturers, but seen by group members
The UG’s commented on the pressure they felt in awarding marks to others and on relying on others to mark them, illustrated in this comment;

*Find it difficult to rely on others to give me a mark*

This may reflect their unfamiliarity with the process and also their perceived lack of confidence in their abilities (Ballantyne, Hughes & Mylonas 2002). This group are much younger than the Masters level students who one could argue are more self assured and experienced and thus feel more confident marking their peers.

The UG group were much more concerned about the confidential aspect of completing the forms than the PG group. Both groups were given a classroom in which to complete the forms but the UG students were unhappy about sharing the room with others in case they were asked to discuss the marks. This issue was not raised amongst PG group, who filled out their marks independently but in the same room and at the same time as their peers. This difference may again be explained by a maturity in the PG group, who may have been more prepared to self disclose.

Both groups felt the feedback from their peers would be useful to further their own learning (UG 69%, PG 80%). This has been described by Carless, Joughin and Mok (2006) as “feedback as feed forward” p. 396, which purports that feedback needs to be given at a time when students can use it to undertake current and further work. This was the case in the PG assessment with most of the group requesting that comments from their peers (which were used to justify the mark allocated) be fed back in order to inform their reflective learning accounts.

Both groups felt that those with English as a second language and quieter students may be disadvantaged by peer marking, as they were concerned that good communication did not always equate to good quality of work and commitment to the task. Several students commented;

*Doesn’t suit quieter members of the group as they may get marked down despite effort*

There is evidence to support this effect, known as the “decibel effect” (Beaman 1998), where the loudest and most dominant students gain more marks, and supports the need
for lecturers to encourage students to evidence their marking objectively and consider what is meant by “effort”.

The UG group were concerned that a peer mark was contributing to their overall degree classification (22%), perhaps reflecting their uncertainty about the rigour of and confidence in this system. However the research suggests that peers tend to score each other higher than their tutors (Papinczak et al 2007, Langan et al 2008) so this may have worked in their favour towards their degree classification.

**Reflection on Personal Learning**

The motivation for introduction of peer and self assessment was to reward effort related to the *process* of learning, as in the author’s view, the product was already being assessed rigorously through the summative marking of the group piece of work and the individual written accounts. The peer assessment criteria were given to the students at the outset of the module, rather than the students designing and having ownership of them. The criteria were based on staff’s perception of effective group working and not the students’. On reflection this may not have resulted in deep learning as perhaps there was less engagement by the students than if they had designed the criteria themselves. This has since been addressed and students now design the criteria, facilitated by a lecturer. It is hoped this will help students to deeply engage in the learning process. It may also help to make explicit to students whether it is process or product that is being assessed and what the minimum standard should be. This will involve engaging with the Institution’s attributes of performance, more closely.

Finally in order to minimise the “decibel effect”, facilitators need to debate with students what constitutes effective group working as there are many ways to contribute apart from verbally; for example through online discussions, which may suit quieter students and prevent them from being disadvantaged in the peer marking system. This may have resource implications as this process will require space in an already tight curriculum to engage students in setting criteria which are measurable and meaningful.
**Limitations of the initiative**

This is a small educational initiative with a small sample size, which cannot be generalised to a larger population. However the findings concur with some of the wider literature and have resulted in deep experiential learning for the author.

Closed questionnaires analysed together may have given more quantifiable results. However this may have limited the richness of information elicited from open questions. No comparison was sought between peer marks and lecturer marks, which may have been a useful exercise to assess reliability of marks. However, facilitators were asked to rank the students according to performance in the groups and these rankings strongly agreed with the peer marks, suggesting agreement between the two.

Whilst there has been previous discussion in this paper as to whether agreement is actually the main outcome- as opposed to deep learning by students- it is still reassuring for the author to find that peers and tutors would appear to have similar judgements.

**Conclusion and next steps**

The findings of this educational initiative concur with some elements of the reviewed literature, and have highlighted issues which may have implications for further practice. It is proposed to take peer and self assessment forward as a summative element of both modules with students being facilitated to set their own assessment criteria. The underpinning philosophy of sustainable assessment aims to help student nurses and teachers to develop skills of judgement, reflection and the capacity to assess themselves and others in the world beyond the HEI contributing to active engagement with lifelong learning.

**References**


