Impact of collaborative peer and self assessment on students’ judgment and written communication

Phil Hancock and Gillian Dale-Jones
The University of Western Australia
Keith Willey
University of Technology, Sydney
phil.hancock@uwa.edu.au, keith.willey@uts.edu.au, gillian.dalejones@uwa.edu.au

Judgment and (written) communication are two of the five learning standards identified by the Australian Learning and Teaching Council for Accounting Bachelor and Master Degrees. This paper reports the use of a custom-designed collaborative peer and self-assessment teaching initiative, embedded into the curriculum for postgraduate Introductory Financial Accounting. The study, which is modeled on the one used in the Achievement Matters: External peer review of accounting learning standards project, uses as its base, carefully selected exemplars, written by peers in a previous cohort. Students are required (twice) to assess the quality of the written communication in exemplars (using their judgment), which in turn generates informed debate on the expected standards of written communication and the application of an assessment grid. Assuming that students learn from this process, they then have an opportunity to apply their learning to their own written work.

The study is designed to capture any change in the quality of firstly, the students’ judgment using the online tool Self and Peer Assessment Resource Kit (SPARKPLUS) and secondly, their written communication (using the outcome of an independent grading process). The results indicate a statistically significant improvement in the quality of student judgment in relation to the grammatical, structural and presentation components of written communication. In addition, the quality of their written communication shows a statistically significant improvement in all six components specified by the assessment grid. While there is insufficient evidence to draw any specific conclusions, our observations and student comments suggest that the teaching initiative contributed to this improvement.

Keywords: judgment, written communication, accounting, peer assessment, SPARKPLUS

Introduction

Following the recommendations of the 2008 Bradley Review, the Australian Government announced in May 2009 a new regime for Australian higher education. As a result, in 2009 the Government commissioned the Australian Learning and Teaching Council (ALTC) to undertake a Learning and Teaching Academic Standards Project.

Accounting was chosen as the first discipline in business and in December 2010 the Accounting learning standards were published by the ALTC (Hancock and Freeman 2011). It identifies five threshold learning standards, namely judgment, knowledge, application skills, communication and teamwork, and self-management.

Following the release of the accounting learning standards the Australian Business Deans Council (ABDC) provided seed funding for a project to assess whether accounting graduates met the learning standards. Achievement Matters: External peer review of accounting learning standards jointly led by Hancock and Freeman is seeking to collaboratively develop and implement a national model of expert peer review for benchmarking learning outcomes against the accounting learning standards (Freeman and Hancock 2011). The project uses the online tool Self and Peer Assessment Resource Kit (SPARKPLUS) for peer reviewers to record their decisions.
In this paper we report the use of a custom-designed process of collaborative peer and self-assessment, with the aid of an assessment-criteria grid in a postgraduate accounting unit at an Australian university. The process is modeled on the one used in the Achievement Matters project and SPARKPLUS is also used by the students to record their assessment decisions. The objective of this teaching initiative is to encourage relevant, cognitive gains for postgraduate accounting students within existing levels of resources.

Motivation

In response to the growing shortage of accountants in Australia many universities started offering a postgraduate conversion accounting degree in the 1990s. However, there was widespread criticism of the employability of many graduates from such programs. As a result in 2006 CPA Australia commissioned Dr Bob Birrell to prepare a report on The changing face of the accounting profession in Australia. The Birrell report notes that many of the graduates from postgraduate conversion courses are not working in professional accounting firms because “[although] … such graduates are generally technically proficient and usually possess a strong work ethic, the problem lies with their communication skills” (Birrell 2006, p.16). Birrell, in further work, is quoted as indicating that “more than a third of overseas students who secured visas as Australian-trained accountants had worryingly low English language skills” (Lane, 2009, p.25).

Therefore, the learning standards selected for this study are written communication combined with judgment - it is important that students can judge what is acceptable written communication for professional accountants.

Why use peer learning? To quote Boud et al. (1999),

There are both pragmatic and principled reasons for the current focus on peer learning in university courses. The pragmatic reason is the most obvious. Financial pressure on university funding has led (sic) to staff teaching more students. This has prompted a search for teaching and learning strategies to enable staff to cope without increasing their overall load. Peer learning has considerable promise as it involves maintaining the level of student learning without more input from staff. (Boud et. al 1999, p.415)

Hence the study examines the use of peer learning to develop and assess students’ ability to exercise judgment and to improve written communication skills.

Prior literature

The structure of the literature review mirrors the way in which the teaching initiative was designed – being to capture the benefits and ameliorate the deficiencies (as expressed in the literature) in relation to peer and self-assessment, collaborative learning and the practice of familiarising students with assessment criteria.

Peer and self-assessment

Peer and self-assessment share common features although they are distinct concepts. Boud and Falchikov (1989) define self-assessment as “the involvement of learners in making judgments about their own learning, particularly about their achievements and the outcomes of their learning (p 529)”. Its formative nature has the potential to guide students in improving their own work. Topping (1998) defines peer assessment as “an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the product or outcomes of learning of peers of similar status” (p250).

The benefits of the process of self and peer assessment have been well documented. Quite aside from its recognised ability to improve the operation and fairness of teamwork (Willey & Freeman, 2006b), it is acknowledged that it
1. forces student assessors to have a more conscious understanding of the processes involved in the activity (this benefits even strong students who may perform well through intuition rather than a clear consideration of what constitutes a good or poor piece of work) (Searby & Ewers, 1997),
2. improves the students’ understanding of assessment (Mowl & Pain, 1995)
3. provides students with an opportunity to participate in the assessment process, which increases their motivation (Brindley & Scoffield, 1998)
4. has the potential to assist students to develop important professional skills, such as reflection and critical thinking (Somervell, 1993; Mello, 1993).
5. may be used as a means of producing formative learning oriented feedback to complete the learning cycle and encourage the ongoing development of skills (Willey & Freeman, 2006a).

In more general terms, Willey and Gardner (2010) report also on the capacity of differing types of self and peer assessment activities to engage students and promote learning, while Boud and Falchikov (2006) emphasise the longer term effects of self and peer assessment – linking this form of assessment with the development of lifelong learning skills.

There are some circumstances in which the effects of peer assessment are less favourable or even disadvantageous, for example:

• Issues of validity and reliability of peer assessment are challenged, particularly in relation to essay writing (Topping, 1998; Mowl & Pain, 1995).
• Additionally, student assessors may not be willing to take responsibility for assessing their peers, particularly in small, socially cohesive groups (Topping, 1998; Falchikov, 1995). This effect is particularly accented where the students lack confidence in their competence in assessing their peers, and where the feedback is required to be negative (Ballantyne, Hughes & Mylonas, 2002). Associated with this concern is the effect (which is extremely difficult to detect and monitor) of power relationships between student assessors (Topping, 1998).
• Willey and Gardner (2010) are cognisant of the competitive nature of some students in the learning environment, who would be reluctant to “provide beneficial feedback to tasks that allow for resubmission, for fear of helping a fellow student to exceed their own final grade” (p.441).

The design of this teaching and learning initiative avoids many of these disadvantages because the students are required to assess the work of peers from a previous cohort, not their current peers.

**Collaborative learning frameworks**

Collaborative learning is described by Smith and MacGregor (1992) as an “umbrella term for a variety of educational approaches involving joint intellectual effort by students, or students and teachers together” (p11). In the past two decades, the implementation of collaborative learning, in various forms, has become more mainstream. Willey and Gardner (2012) summarise, in the form of two implementation frameworks, their considerable experience in applying collaborative learning at a tertiary level. Their frameworks engage the benefits of the collective working-memory effect suggested by cognitive load theory, while simultaneously acknowledging the ‘collective ability’ effect. The design of this teaching initiative follows closely the guidance offered by these frameworks.

**Tool used: Criterion grid as a means of gaining a common understanding of required standards of written communication**

The process of students gaining an understanding of the assessment framework is generally considered to have a positive impact on learning. However, more recently, Torrance (2007) documents case study findings that the widespread uptake of formative assessment practices, the key elements of which include ‘feedback focused on helping students to improve and sharing criteria of quality’ (p.282), is not a benign recipe for the development of independent learners. Instead it may lead to instrumentalism - undesirable because it shifts the outcomes from ‘learning’ to ‘criteria compliance’. He calls for ‘divergent’, as opposed to ‘convergent’, formative assessment, which directs the students
towards what they can achieve in an open and exploratory fashion, rather than prescriptive formative assessment which encourages mere compliance with criteria.

O'Donovan, Price and Rust (2001) found that from a student’s perspective, the criteria assessment grid is a constructive means of providing feedback and improving the quality of work, only if it is part of a broader initiative that includes ‘explanation of the standards, exemplars and opportunities for discussion’ (p.74). Sadler (2009) on the other hand is critical of assessment grids and contends that ‘To simply reach for a rubric or construct a scoring key each time a complex work has to be appraised is both impractical and artificial in life outside academe. Equipping students with holistic evaluative insights and skills therefore would contribute an important graduate skill’ (p.178). However, even with holistic assessment the use of criteria to guide the assessor would not be inappropriate and Sadler admits that ‘clearly, there is more work to be done on this important issue of grading complex student responses to assessment.’ (p.178).

The study

The study is designed to capture evidence of any changes in the following:

• The ability of the students to exercise their judgment in assessing exemplars of written work, and
• The quality of the student reports submitted before and after the exercise.

In addition, qualitative and quantitative information was gathered from the students in two further ways - a self-reflection statement along with the final version of their report, and a survey instrument. The initiative is performed in the context of a report-writing assignment in a Masters level introductory financial accounting unit, with an enrolment of forty students – for many, English is an additional language. The unit is taught over a period of thirteen weeks using a weekly face-to-face three hour workshop to deliver the material. For the entire semester, the students are allocated into ten groups of four, to foster a supportive learning environment.

Method

From the students’ perspective, the final result of the learning and teaching initiative is unremarkable - an individually written report based on a topical financial accounting issue, the “recognition of revenue”. However, the process put in place to support the students’ creation of their reports, is what is innovative about this learning and teaching initiative. Design and implementation of the initiative required both significant preparatory work and a clearly defined process during the semester. A description follows.

Preparatory work

In relation to the curriculum

The objective of the initiative is to use peer learning and assessment to improve and assess students’ competencies in judgment and written communication in a collaborative setting. The initiative is seamlessly embedded into the curriculum to enhance the learning experience without reducing the content of the unit and without significantly increasing the work load of the unit coordinator (eg double marking of reports was out of the question). “Revenue recognition” was chosen as the research area and a suitable report assignment was written for this purpose.

In relation to the exemplars

Four (two pairs of) exemplars of mediocre written work from the previous year’s cohort were selected using a pairwise assessment process, as described by Heldsinger and Humphry (2010) to rank ten randomly selected examples.

Those four exemplars were then subjected to a more conventional assessment process using an assessment grid that has been developed over several years, for the purpose of assessing written
In this unit, reports were assessed by two authors of the paper, with one being a peer reviewer in the *Achievement Matters* project and having participated in the national calibration workshops on written communication. Out of six possible criteria, assessors gave exactly the same rating for 20 of the points of assessment. The assessment for the other points differed by only one degree on the scale and discussion reached consensus. This process formed the ‘benchmark’ assessment for the two pairs of (four) exemplars.

### In relation to the SPARKPLUS tool

The SPARKPLUS tool was used to facilitate assessment of the students’ individual assessments of the two pairs of exemplars, with the assistance of a ‘benchmark’ assessment to determine a ‘judgment mark’ to be awarded to each student. At the preparatory stage, the assessment grid criteria were summarized into SPARKPLUS, and each criterion and exemplar’s ‘benchmark’ was not made available to the students.

### Implementation process

This process was described below and summarized in Figure 1.

1. At the start of semester, the students are familiar with the proposed process, supplied with all relevant material (e.g., exemplars), and exposed to a trial-run using SPARKPLUS.
2. The content relevant to the current and previous semesters’ reports is covered by the end of week four. In weeks five and six, homework designed to familiarize the students with content is set.
3. During the non-teaching week, students submit their draft assignment and an individual assessment (via SPARKPLUS) of the first pair of exemplars, using the same criteria that will be applied to their own work.
4. A de-brief is run at the next workshop. Contentious criteria are highlighted to students, who are then given the opportunity to discuss these particular assessments in their groups of four, with the aim of arriving at a single, justifiable group assessment. The group assessments are then collated, and the benchmark assessments for those criteria are revealed to the students.
5. This step repeats and reinforces the assessment process in steps three and four but with another pair of exemplars. This time, the de-brief is more streamlined. Students perform an on-the-spot group assessment of the fifth exemplar before the unit coordinator leads a discussion to review its quality.
6. The students are given a further week to review and resubmit the final version of their reports. They are encouraged to consider and apply what they have learnt from the previous steps. The final version is required to contain a self-reflective statement, explaining how their report has changed from the draft version.
7. The students are given the opportunity to take part in a survey (see Appendix A: Survey instrument) about the initiative.
This aim of this experiment is to evaluate the effect (if any) of the teaching initiative on the quality of the students’ judgment and/or their written reports.

**Did the quality of the judgment change?**

In this instance ‘quality of judgment’ is measured by comparing the student ratings of the exemplars against the benchmark ratings which were established in advance by two of the authors. For each exemplar almost 40 student ratings were collected across each of the six criteria. Exemplars A and B were performed concurrently but before the intervention, while exemplars C and D were performed concurrently, but after the intervention. The six criteria (taken from the grid) are shown here:

<table>
<thead>
<tr>
<th>K1</th>
<th>Knowledge 1</th>
<th>Relevant issues are raised and the discussion is convincing. Are the conclusions valid?</th>
</tr>
</thead>
<tbody>
<tr>
<td>K2</td>
<td>Knowledge 2</td>
<td>Does the work contain reference to, and reflect an understanding and application of appropriate, adequate literature?</td>
</tr>
<tr>
<td>W1</td>
<td>Writing style 1</td>
<td>Is the written work grammatically correct and well structured? (This includes sentence and paragraph structure and logical sequencing.)</td>
</tr>
<tr>
<td>W2</td>
<td>Writing style 2</td>
<td>Is the report well presented and in accordance with the requirements that were stipulated for the assignment?</td>
</tr>
<tr>
<td>W3</td>
<td>Writing style 3</td>
<td>Is the in-text referencing correctly applied?</td>
</tr>
<tr>
<td>W4</td>
<td>Writing style 4</td>
<td>Is the end-text referencing correctly applied?</td>
</tr>
</tbody>
</table>
SPARKPLUS requires the students to enter their ratings on a continuous spectrum, which is evenly sub-divided into categories as below. The students are aware (from a trial run) that the spectrum is a continuous measure.

- WB Well below acceptable
- BA Below acceptable
- AC Acceptable
- AA Above acceptable
- WA Well above acceptable

Table 1 reports a descriptive overview of the students’ ratings compared against the benchmark rating. On average, they initially allocated lower marks for most of the writing style criteria (W2 to W4), although they were more lenient with grammar and structural considerations (W1). After the intervention their ratings were more balanced. The results for the knowledge criteria (K1 and K2) indicate that after the intervention the students become more demanding with these criteria.

Table 1: Average student response: Higher or lower than the benchmark

<table>
<thead>
<tr>
<th>Before/after intervention</th>
<th>Exemplar</th>
<th>K1</th>
<th>K2</th>
<th>W1</th>
<th>W2</th>
<th>W3</th>
<th>W4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>A</td>
<td>Lower</td>
<td>Higher</td>
<td>Higher</td>
<td>Lower</td>
<td>Lower</td>
<td>Lower</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Equal</td>
<td>Lower</td>
<td>Higher</td>
<td>Lower</td>
<td>Lower</td>
<td>Lower</td>
</tr>
<tr>
<td>After</td>
<td>C</td>
<td>Lower</td>
<td>Lower</td>
<td>Higher</td>
<td>Lower</td>
<td>Lower</td>
<td>Higher</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Lower</td>
<td>Lower</td>
<td>Lower</td>
<td>Higher</td>
<td>Higher</td>
<td>Higher</td>
</tr>
</tbody>
</table>

Since the aim of the exercise is to develop judgment, the students’ percentage ratings are themselves of limited interest. The size (not direction) of the gap between the ratings and the benchmark (hereafter termed the ‘difference’) is more representative of the quality of the judgment. In order to eliminate directional effects, the absolute value of the ‘difference’ is used.

If a difference of less than ten percent either side of the benchmark is defined as a demonstration of sound judgment, then Figure 2 gives an indication of the quality of judgment, by exemplar and by criterion. Visually, improvement in judgment appears to occur in relation to criteria W1, W2, W3 and W4.

Figure 2: Percentage of student ratings within ten percent of the benchmark rating
Table 2 shows the means of the grouped results for Exemplars A and B (C and D), which represent the before (after) intervention situations. An improvement in judgment (shown by a decrease in the differences) occurs in relation to criteria K1, W1, W2 and W3, however, a two tailed t-test indicates that these improvements are statistically significant only in the cases of criteria W1, W2 and W3. The unexpected result for W4 arises because of the effect of outliers.

A control test is performed to establish whether the differences for Exemplar A (C) are significantly different from those for Exemplar B (D) – to ensure that there is a degree of internal stability in the ratings differences both before and after the intervention. For the criteria W1 and W2, in which it is thought that an improvement has occurred, it is shown that there is no significant internal inconsistency in the ratings before and after the intervention. However, for criterion W3, the ratings differences for A and B (i.e. before the intervention) are significantly different from each other (as they are also between C and D). This lowers the credibility of the finding reported in the previous paragraph in relation to criterion W3.

Table 2: Means of the differences between student and benchmark ratings before (AB) and after (CD) intervention

<table>
<thead>
<tr>
<th></th>
<th>K1</th>
<th>K2</th>
<th>W1</th>
<th>W2</th>
<th>W3</th>
<th>W4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exemplars A and B</td>
<td>18.53</td>
<td>16.10</td>
<td>15.48</td>
<td>26.46</td>
<td>25.42</td>
<td>14.73</td>
</tr>
<tr>
<td>Exemplars C and D</td>
<td>15.23</td>
<td>19.27</td>
<td>10.99</td>
<td>12.44</td>
<td>15.56</td>
<td>15.08</td>
</tr>
<tr>
<td>Improvement in judgment</td>
<td>Yes</td>
<td>No</td>
<td>Yes*</td>
<td>Yes*</td>
<td>Yes*</td>
<td>No</td>
</tr>
</tbody>
</table>

* Significant at 1%

In summary an improvement in judgment appears to exist in relation to criteria W1 and W2.

Did the quality of the written communication change?

After the teaching initiative, both draft and final reports were assessed by an independent examiner, who specialises in written communication. The examiner used the same assessment grid, however within both knowledge criteria, her focus was on the students’ ability to communicate the information, as opposed to the choice of content selected by the students for presentation. These draft/final reports (although not obvious to the examiner) effectively formed a matched pair sample because each pair of reports was written by the same person.

Table 3: Summary of independent examiner's assessment of draft and final student reports

<table>
<thead>
<tr>
<th></th>
<th>Possible score</th>
<th>K1 5.00</th>
<th>K2 5.00</th>
<th>W1 2.00</th>
<th>W2 1.00</th>
<th>W3 1.00</th>
<th>W4 1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>Mean</td>
<td>2.74</td>
<td>3.22</td>
<td>1.12</td>
<td>0.61</td>
<td>0.53</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>2.80</td>
<td>3.20</td>
<td>0.80</td>
<td>0.60</td>
<td>0.40</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>Std deviation</td>
<td>0.76</td>
<td>1.08</td>
<td>0.49</td>
<td>0.26</td>
<td>0.29</td>
<td>0.28</td>
</tr>
<tr>
<td>After</td>
<td>Mean</td>
<td>3.23</td>
<td>3.87</td>
<td>1.20</td>
<td>0.75</td>
<td>0.73</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>3.00</td>
<td>3.80</td>
<td>1.00</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Std deviation</td>
<td>0.80</td>
<td>0.80</td>
<td>0.51</td>
<td>0.19</td>
<td>0.26</td>
<td>0.29</td>
</tr>
<tr>
<td>Score (After - Before)</td>
<td>0.49</td>
<td>0.65</td>
<td>0.08</td>
<td>0.14</td>
<td>0.20</td>
<td>0.09</td>
<td>1.65</td>
</tr>
<tr>
<td>Number of report pairs</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>Instances of no change</td>
<td>9</td>
<td>5</td>
<td>16</td>
<td>15</td>
<td>10</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>Instances of improvement</td>
<td>23</td>
<td>24</td>
<td>14</td>
<td>16</td>
<td>23</td>
<td>13</td>
<td>32</td>
</tr>
</tbody>
</table>

As shown in Table 3 within the 35 pairs of reports, the total assessment for the final report exceeds the assessment for the draft report in 32 cases. In one case there is no change and in the other two cases, there is a slight decrease in the assessment. However, on average, the results show a significant
improvement between the draft and final reports (for W1 and W4, with 95% confidence, and with 99% confidence for all the other criteria). While these results reflect a clear improvement in their written communication, it is less clear what caused the improvement, since it is not permitted at the university to run a true control group who do not undergo the teaching innovation. The factors which may cause the improvement include (1) the fact that the students are assessed on their final report, but not on their draft report, (2) the students may have had more time to assimilate the material, (3) students undertaking a concurrent communication unit may have been positively influenced by that process and (4) the teaching innovation may have enabled the students to more effectively self evaluate and improve their reports. It is likely that some combination of these factors is at work.

Did the students take the opportunity to self-evaluate their reports?
The report requirements state that “the final version of the report must include a statement, of not more than 250 words, explaining how the self-evaluation process changed your work from Draft one to the Final Version.” The aim of this requirement is to encourage the students to take time to reflect on the process of self-evaluation. If the students find value in the process, then there is potential for this intervention to seed or nurture, the habit of regular self-evaluation - an important component of independent learning.

The free-form reflective statements indicate that the students did take the opportunity to self-evaluate their work. Reference to the teaching intervention was unsolicited, however 62% of the respondents specifically credit it with providing insight into some component of their written work.

How did the students perceive the process?
The responses to the survey (in the Appendix) are scored as follows, in order to perform statistical analysis: Strongly disagree = 1, Disagree = 2, Neutral = 3, Agree = 4 and Strongly agree = 5.

The average rating for each question is summarised in Table 4, where it is seen that students indicate somewhere between ‘agreement’ and ‘strong agreement’ for questions one to six (which concentrate on the exemplar process itself), whereas for questions seven to nine (which concentrate on application of what they may have learned through the exemplar process to improve their work) their average response lies somewhere between ‘neutral’ and ‘agreement’. Interestingly, they credit an improvement in their understanding of the assessment criteria to all three parts of the exemplar process – their individual review (Q4), their collaborative peer review (Q5) and the instructor’s de-brief (Q6), although there is a statistically insignificant preference shown for the effect of the instructor’s de-brief.

Table 4: Summary of average responses to survey (see Appendix A)

<table>
<thead>
<tr>
<th></th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
<th>Q7</th>
<th>Q8</th>
<th>Q9a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average score</td>
<td>4.03</td>
<td>4.31</td>
<td>4.08</td>
<td>4.08</td>
<td>4.03</td>
<td>4.28</td>
<td>3.68</td>
<td>3.92</td>
<td>3.79</td>
</tr>
<tr>
<td>Number of responses</td>
<td>38</td>
<td>39</td>
<td>39</td>
<td>39</td>
<td>38</td>
<td>39</td>
<td>39</td>
<td>39</td>
<td>39</td>
</tr>
</tbody>
</table>

The demographic information collected via the survey enables initial analysis of whether the perceptions of students about the teaching initiative are to some extent dependent on the first language of the country in which they undertook their previous studies. For all of the survey questions, except Q9a, students whose previous study was in a country where English was the first language, rated the effects of the intervention more highly than students whose previous education was in an Asian country.

An equivalent analysis of the survey responses by gender revealed nothing remarkable.

In summary, it would appear that the whole process was perceived by the students to have improved their confidence in assessing written material, their understanding of assessment criteria, and their ability to apply these skills to their written work. In answer to question ten, many of the comments emphasised the challenge and learning that had resulted from the process of having to assess, using specific criteria, another’s piece of written work.
Conclusions

Collaborative classrooms stimulate both students and teachers. In the most authentic of ways, the collaborative learning process models what it means – to question, learn and understand in concert with others. Learning collaboratively demands responsibility, persistence and sensitivity, but the result can be a community of learners in which everyone is welcome to join, participate and grow (Smith & MacGregor, 1992, p.29).

This teaching and learning initiative was challenging. The preparation was considerable, although during semester, the initiative created only a slightly higher workload than the previous year. More particularly, our academic judgment was subjected to close scrutiny, which was daunting - not because the benchmarks were necessarily right or wrong, but simply because justifying and explaining what has almost always been an intuitive process, is confronting. The debrief sessions required serious preparation and planning.

With the benefit of hindsight, one improvement to the process would be to include an example of a ‘good’ piece of writing as one of the exemplars. It would give the students a more secure launching pad for analysis and comparison.

The results support that there were benefits for the students both in terms of their judgment and written communication skills. In a study of this kind, the analysis of quantitative results lends a degree of rigour however the value of the qualitative and anecdotal findings should not be overlooked. The majority of students moved from hesitation and bewilderment through to being more confident about assessing the quality of the exemplars, via a process of enthusiastic engagement. Two remarks, both from strong, mature-aged students are worth documenting – the first said that he wished he had undergone this process before starting employment and the second commented that in all her years of study she had never previously learned the power of learning through participation. It should also be noted that this study has attempted to measure only the short term benefits to the students. It would be most interesting to know if any of the perceived benefits are more permanent.

References


Copyright Phil Hancock, Gillian Dale-Jones and Keith Willey. The authors assign to the TL Forum and not for profit educational institutions a non-exclusive licence to reproduce this article for personal use or for institutional teaching and learning purposes, in any format, provided that the article is used and cited in accordance with the usual academic conventions.

Appendix A: Survey instrument
Questionnaire in relation to the Peer Review exercise
Completion of this survey indicates my consent to participate

Demographic information

Gender: 

Previous discipline(s) of study: 

Country of previous education: 

Level of previous education: 

Years spent in workforce: 

Teaching & Learning Forum 2013
11
### Questions about the Peer Review exercise

<table>
<thead>
<tr>
<th>Please circle your response</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>After the peer review exercise, I feel more confident about assessing written communication than I felt before the process.</td>
<td>SD</td>
<td>D</td>
<td>N</td>
<td>A</td>
<td>SA</td>
</tr>
<tr>
<td>2</td>
<td>After the peer review exercise, I understand the meaning of the assessment criteria better than I did before the exercise.</td>
<td>SD</td>
<td>D</td>
<td>N</td>
<td>A</td>
<td>SA</td>
</tr>
<tr>
<td>3</td>
<td>After the peer review exercise, I understand the expected standards for written communication conveyed by the assessment criteria better than I did before the exercise.</td>
<td>SD</td>
<td>D</td>
<td>N</td>
<td>A</td>
<td>SA</td>
</tr>
<tr>
<td>4</td>
<td>My understanding of the assessment criteria increased as a result of having to complete an individual assessment of the exemplars.</td>
<td>SD</td>
<td>D</td>
<td>N</td>
<td>A</td>
<td>SA</td>
</tr>
<tr>
<td>5</td>
<td>My understanding of the assessment criteria increased as a result of discussing the exemplars with peers.</td>
<td>SD</td>
<td>D</td>
<td>N</td>
<td>A</td>
<td>SA</td>
</tr>
<tr>
<td>6</td>
<td>My understanding of the assessment criteria increased as a result of the de-brief facilitated by the unit coordinator.</td>
<td>SD</td>
<td>D</td>
<td>N</td>
<td>A</td>
<td>SA</td>
</tr>
<tr>
<td>7</td>
<td>The peer review exercise enabled me to improve my written communication assignment.</td>
<td>SD</td>
<td>D</td>
<td>N</td>
<td>A</td>
<td>SA</td>
</tr>
<tr>
<td>8</td>
<td>The peer review exercise improved my judgment in rating against the assessment criteria.</td>
<td>SD</td>
<td>D</td>
<td>N</td>
<td>A</td>
<td>SA</td>
</tr>
<tr>
<td>9a</td>
<td>The peer review exercise has changed the way I would write in future professional contexts (job applications, reports, assignments etc).</td>
<td>SD</td>
<td>D</td>
<td>N</td>
<td>A</td>
<td>SA</td>
</tr>
<tr>
<td>9b</td>
<td>Briefly explain how:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>What were the positive aspects (if any) of undertaking the peer review exercise?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>What were the negative aspects (if any) of undertaking the peer review exercise?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Any other comments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Teaching & Learning Forum 2013