

Dancing with the Devil: Formative Peer Assessment and Academic Performance

Jen D. Snowball* and Markus Mostert**

**Associate Professor, Department of Economics and Economic History, Rhodes University, Grahamstown.*

***Educational Technologist, Centre for Higher Education Research, Teaching and Learning, Rhodes University, Grahamstown.*

Corresponding author: Prof. Jen Snowball
Department of Economics and Economic History
Rhodes University
PO Box 94
Grahamstown
6140
South Africa
Tel: +27 46 603 7405
Email: j.snowball@ru.ac.za

Dancing with the Devil: Formative Peer Assessment and Academic Performance

Abstract

Peer assessment using assessment criteria can be important in developing active and independent learners, as well as providing more and faster feedback in large classes. In addition, the evaluative, critical stance required by students in order to assess their peers' work encourages the development of higher-order cognitive skills. However, peer assessment does have potential problems, and there is some debate as to the appropriate academic level at which to implement it, the kinds of feedback (technical versus content-related; general versus specific) that are given, and the ways in which students respond. In addition, there is little evidence that peer assessment has an impact on academic performance. This research reports the results on an online peer assessment exercise for a macroeconomics essay conducted in a large (800 student) economics 1 class at Rhodes University. Data was collected from students via a formal evaluation. In addition, a sample of 50 essays was evaluated in terms of the initial submission, peer feedback received, changes made to the final version, and the correlation of peer and final tutor marks. An OLS regression was used to investigate the impact peer assessment participation on marks. Results showed that there were large deviations between peer and final marks, especially for essays whose final mark was a first class or a fail. In general peers gave more useful feedback on technical aspects, such as presentation and referencing, (which were also the categories in which students most often made improvements) than on content. Regression analysis showed that peer assessment participation was not a significant determinant of final essay mark, but that economics ability and English language proficiency were.

Keywords: peer assessment, online, impact, large class

1 Introduction

The potential benefits of peer assessment have been well documented in academic literature, as Topping (2010) points out in a recent special edition of Learning

and Instruction. Topping (Topping, 1998, p. 250) defines peer assessment as “an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status”. For the assessor, peer assessment involves quite high order cognitive skills, like comparing, contrasting and communicating, which can help to consolidate and deepen the assessor’s understanding of the topic. For the assessed, peer feedback related to discussed or negotiated criteria can help students to improve the quality of their work by comparing their work to that of others, encouraging self-reflection or more accurate self-assessment and (if feedback is given soon) avoiding cumulative errors (Topping, 1998). Vickerman (2009, p. 229) suggests that peer assessment can be used as “a strategy for enhancing the diversity of learning experiences, as well as supporting individual student’s learning needs”.

Such benefits are increasingly important in higher education, where class size and student diversity are both increasing, along with calls for increases in the quality of teaching and learning as well as for institutional accountability (George-Walker & Keeffe, 2010). Although there are a great number and variety of case studies on peer assessment in higher education, few explore the causal relationships between variables, or try to quantify the effects of implementing peer assessment in various ways (Van Zundert, Sluijsmans, and Van Merriënboer, 2010). Van Zundert et al. (2010) call for further research into the effects of being an assessor, rather than being assessed.

This article tries to address these issues using a case study of an online peer assessment exercise conducted in a large first year economics class at Rhodes University, South Africa. Building on the work of Hanrahan and Isaacs (2001), the impact of the peer assessment exercise on students is examined, and (using statistical

analysis) on essay marks. A sample of 50 essays is analysed, taking into account the quality of peer feedback received and the changes made.

2 The academic value of peer assessment: a literature review

One of the most exciting proposed outcomes of peer and self-assessment is the idea that students can develop into independent learners, able to gauge their own successes and failures and who are less dependent on lecturers and tutors (Race, 2001). Hanrahan and Isaacs (2001, p. 65) found that even students who claimed not to have improved the quality of the work submitted for peer assessment recognized the benefits of seeing their own work in the context of the work of others. “Given the isolation of some students in diverse modular degree courses, peer assessment may be the only opportunity they have to see the work of other students”. Secondly, students developed empathy with assessors, appreciating how difficult it is to assess work reliably and fairly. Hanrahan and Isaacs (*ibid.*) argue that this is a very important step in helping students to become “independent agents in the community” and autonomous life-long learners. Writing about formative group peer assessment, Pope (2001) agrees that increasing student confidence and independence can be an outcome of peer assessment as long as it is implemented carefully.

Peer assessment has the potential to develop reciprocity and cooperation among students, to encourage active learning and to increase time on task (Chickering & Gamson, 1987). Used formatively rather than summatively, peer assessment is a form of writing responding (Lupton, 2008), which is contextualised in both the disciplinary discourse as well as in the particular topic at hand. Arguably however, the act of assessing has more potential for enhancing learning as it requires the student as assessor to evaluate, that is, to develop “the ability to estimate the appropriateness of a certain

item according to particular criteria” (Pritchett, 1999, p. 33). Representing the highest level of knowledge and intellectual ability in the cognitive domain identified by Bloom (1956) in his well-known taxonomy, evaluation involves high order cognitive skills such as comparing, contrasting and articulating, all of which can help to consolidate and deepen the assessor’s understanding of the topic. It can further be argued that by assessing other students’ essays, students constantly engage in comparing their own work to that of the peers whose work they are marking. This kind of self-reflection has the potential to lead to more accurate self-assessments, which in turn might minimise cumulative error (Topping, 1998). Other advantages of peer assessment include that students see a range of ways to approach the task, including different assumptions, ideologies and writing styles and that staff time spent on marking (grading) assignments can be reduced.

Although peer assessment in large class settings appears to offer a number of pedagogical and organizational advantages, possible negative outcomes also need to be considered carefully. Depending on how it is implemented, self and peer assessment methods can actually take up more time than traditional methods, both for the students and the facilitators (Hanrahan & Isaacs, 2001). It is also possible that students who are not motivated to participate will not benefit (Smyth, 2004) and that even motivated students may find the task difficult and stressful.

In a study of 90 undergraduate students in the UK who were introduced to formative peer assessment for the first time, Vickerman (2009) found a number of such difficulties. For example, although 55% of students agreed that peer assessment had improved their knowledge and understanding of the subject, many commented that they found the process ‘difficult’ and would have preferred tutors to assess the work. However, 71% of students agreed that it helped to develop writing and critical analysis

skills. In explaining the 45% of students who did not agree that peer assessment had increased their content knowledge, Vickerman (2009, p. 227) suggests that this is because it is easier for inexperienced assessors to comment on the technical aspects of a work, rather than on the content.

There is also some debate about the level (year of study) at which it would be appropriate to implement peer and self-assessment. Hanrahan and Isaacs (2001) present a case study of a large (233 students) third year class and argue that the methods are unlikely to be successful at earlier years of study because of students' lack of experience, both in the sense of course content and in gauging the level of performance accurately. On the other hand, Ballantyne, Hughes & Mylonas (2002) suggest that it might be easier to get first year students, who do not yet have preconceived ideas about assessment, to participate.

Sitthiworachart and Joy (2008) used online peer assessment in a large (213 students) undergraduate computer science class to investigate the correlation of marks and comments given by tutors and peers. While they find that the average marks given by peers are higher than those given by tutors, the two are strongly positively correlated. Tutor comments tended to be briefer than those given by peers, who gave more explanations and suggestions. However, guided by the assessment criteria, both peers and tutors commented on similar issues. These findings suggest that, even at undergraduate level, peer marks and comments can be a useful source of feedback.

Countering the idea that peer assessment from a more competent or skilled assessor will be more effective, Strijbos et al. (2010, p. 300) actually find that, while "feedback from a high competent peer was perceived as more adequate than feedback from a low competent peer", students were less likely to respond to feedback from high competent peers by making changes to their work. This suggests that, for peer

assessment to work well, differences in competence levels between peers should not be too great, and that feedback from low competence peers may be just as useful as that from high competence peers.

Strijbos et al. (2010, p. 300), in a study of 89 graduate students in Germany, find that 'elaborated specific' feedback is not necessarily more effective than 'concise general' (or non-directive) feedback, since the former can result in students taking on a passive, dependent role (or seeing the assessor as being too interfering), rather than engaging in a more collaborative engagement around potential improvements. This leads Topping (2010, p. 342) to suggest that, for students who are inexperienced in peer assessment grades or marks are unlikely to be effective, while general non-directive feedback is more likely to be accepted as a "sharing of ideas" and acted upon.

A great deal seems to depend on how peer assessment is introduced to students. Ballantyne et al. (2002) found that students' attitude to the idea of self and peer assessment will have a large impact on their experiences of it, and it should thus be carefully introduced and motivated. Smyth (2004) reported a very successful use of peer and self-assessment to motivate postgraduate students to take formative assessment methods seriously, but emphasized that students needed to be carefully informed at the outset what the purpose and possible benefits of using these methods are.

This is especially important where tutors are involved (mostly in larger classes). Ballantyne et al. (2002, p. 436) point this out, stating that "Tutors need to take responsibility for, feel ownership of, and be prepared to defend the process regardless of what they might actually feel about it ... if staff are not prepared to commit to the process, it is highly unlikely that their students will". As with discussions and collaborative development of criteria, self and peer assessment will take up class and preparation time that could have been spent on content. The trade off between content

and process will have to be carefully motivated and defended if students are to engage fully (Carlson, MacDonald, Gorely, Hanrahan, & Burgess-Limerick, 2000). In their literature review of 26 peer assessment case studies in higher education, Van Zundert et al. (2010) find that training and previous experience of peer assessment increased its effectiveness, as well as fostering positive student attitudes. However, they find little evidence to suggest that PA has a direct impact on academic performance, and suggest that benefits may become more evident (in terms of academic performance) in the long run.

The challenge then, is to implement peer assessment in such a way that as many positive outcomes are achieved as possible while avoiding the negative ones and, at the same time, keeping in mind the practical (time) considerations for both facilitators and students.

3 The economics 1 case study

With over 800 students registered for macroeconomics 102, this was the largest class at Rhodes University in 2009. The course consisted of four lectures (group sizes varying between about 350 to 400 students) and one tutorial per week, with about 20 students in each tutorial group. Assessment included two essays and two tests (summative assessment that counts towards a class mark, making up 30% of the course mark) and 12 tutorial exercises, one to be submitted each week (formative assessment, not contributing to the course mark. Attendance at 80% of the weekly tutorials and submission of the same number of exercises was compulsory.

To add to the challenge of large class sizes, student diversity has also been increasing significantly: In the 2009 eco102 class 53% were women, 71% were black (African, Indian and Coloured students) and only 46% were English first language

speakers (as compared to 66% in 2003). Average pass rates were generally good (76%), but lower (74%) for English second language speakers than for first language speakers (79%). While increasing diversity has advantages, both in terms of redressing inequality and teaching and learning, it does pose teaching and learning challenges, particularly when introducing students to a new set of economics-specific terminology and in developing students' academic writing skills.

In 2009 online peer assessment was introduced for the second essay of the semester. By this time the students had already experienced "off-line" peer assessment with one of the tutorial exercises and had ample time to become familiar with the course website and the learning management system, Moodle. While a number of dedicated online peer assessment systems have been developed recently (Draaijer & van Boxel, 2006), for institutions which have adopted Moodle, the advantage of using the Workshop module is that it is integrated into the course site which students would already be familiar with.

Students were provided with the essay topic and usual instructions and guides (including detailed assessment criteria) as well as a rationale for engaging in the peer assessment. The potential value of peer assessment was made explicit during face-to-face class meetings and on the course site (see Figure 1 below for an example of such rationale, which was provided on the News forum).

Students were also provided with detailed information about the peer assessment process, which included advice on giving feedback using the assessment rubric (based on the assessment criteria) as well as step-by-step instructions with screenshots for navigating the online process. While students were initially prepared in face-to-face lectures as well as on the course site, forums on the site were used for ongoing support and scaffolding, as well as for sharing of developments as they unfolded.

Since the peer assessment replaced one of the weekly tutorial exercises, the time frame for the exercise was limited to a week. The Workshop module in Moodle was used to manage the process, ensuring anonymity and minimising the administrative burden on the lecturer. Students were required to submit their essay to Turnitin before uploading it to the peer assessment exercise. They were then allocated the essays of two peers to assess, using the rubric. Students initially had about 58 hours in which to assess their peers' essays, after which they had 3 days to respond to the feedback by making changes to their work, before final essays were submitted for assessment by tutors. Acknowledging that providing peer assessments required time and careful thought, the assessments themselves replaced the usual tutorial exercise for that week, and could be submitted to tutors instead of the usual content-based exercise.

Due to extremely unfortunate technical problems our installation of Moodle became unstable at the critical time when students had to submit their assessments of their peer's work resulting in the target date for the submissions of peer assessments being postponed by 2 days. Students were then also given the option to do the normal weekly tutorial exercise published in the course handout, rather than participate in the peer assessment exercise. This led to only about half of the class of 800 students participating in the peer assessment exercise. While initially thought to be problematic, this did provide a useful 'control' group which allowed comparisons between the performance of those students who did and who did not participate in the peer assessment exercise.

Essay 2 will incorporate peer assessment. This means that, before you submit your essay for final marking, you will have the opportunity to get the opinions and suggestions of two of your class mates on how to improve it. Most Eco 102 students in previous classes found peer evaluation useful, both because it gives you a chance to improve your work and also because, in commenting on someone else's, you notice common mistakes and get ideas on other ways to do things. As long as you submit your essay to Turnitin first, there is no danger of anyone else in the class copying your work – you have protected your copyright by submitting it to Turnitin. Below is the timetable of how the peer feedback process will work. Note that it is anonymous, so you won't know whose essay you are commenting on and your evaluators won't know who you are.

Figure 1 Rationale and Motivation

Data on student perceptions of the impact of the exercise was collected during the course evaluation survey and consisted of both quantitative, Likert-scale responses, and qualitative, open-ended question responses. Data on student demographics and essay performance was obtained from the university and from the department respectively. A representative sample (in terms of essay marks) of 50 essays was selected and an experienced tutor evaluated the usefulness of the peer assessment received in various categories and its likely impact in terms of the changes made to the essay initially submitted for peer assessment and the essay submitted to tutors for final grading.

4 Results

Student Perceptions

Building on the classification system of Hanrahan & Isaacs (2001) as summarised in the table below, student perceptions of the effectiveness of the peer assessment exercise were analysed by identifying general dimensions and themes within the data that was collected during the course evaluation. The course evaluation questionnaire used several Likert scale closed-ended questions and one open-ended question: *Setting aside the technical problems with RUConnected, did you find the peer assessment exercise for essay 2 useful? Explain why or why not.* Open-ended responses were analysed using NVivo.

465 students responded to the close-ended statement “The peer assessment of essay 2 helped to improve the quality of my essay”. 58% of them agreed or strongly agreed, 27% were neutral, with only 9% disagreeing or strongly disagreeing (6% marked irrelevant’ or ‘don’t know’, probably because they did not participate in the peer assessment exercise).

In the open-ended responses, 56% of students stated that they had found the peer assessment exercise useful. Table 1 compares the general dimensions and themes found by the Hanrahan & Isaacs (2001) study, with those of the present study. Both studies used formative peer assessment for a class essay that counted towards the final mark, so assignment marks depended ultimately on the lecturer, not peers. Both cases were anonymous, but in the Hanrahan & Isaacs (2001) study, the process was conducted off-line (hard copies of essays submitted), while the Rhodes study was online. The Hanrahan & Isaacs (2001) study took place in a third year class of 233 students, while Rhodes study was at first year level with many more students.

Table 1 Comparison between Hanrahan and Isaacs (2001) and Rhodes dimensions and themes

H&I general dimension	H&I themes	Rhodes results
Gained better understanding of marking	Productive self-critique Reinforced marking procedure	Gain realistic sense of value of own work (15%)
Productive	Get more feedback Improve own assignment Develops critical thinking	Constructive feedback from peers (13%) Chance to improve essay to gain higher marks (27%) Identifying mistakes (19%)
Read other's work	Learn what others are doing See good and bad work	Exposure to other perspectives (9%)
Develop empathy	Develop empathy with lecturer/tutor	Gained greater sense of requirements of the essay (4%)
Motivation	Motivated to impress peers	n/a
Problems with implementation	Time-consuming Process not taken seriously Should get feedback for peer assessment Peer assessment too late to be useful	Time-consuming/workload (9%) Complicated process (4%) Process not taken seriously by peers (29%) Intellectual property concerns (26%)
Difficult	Difficult to be objective No experience of marking Unfamiliar with area Own assignment already meets criteria	Lack of confidence to assess (2%) Poor quality of essays submitted (4%)
Discomfort	Uncomfortable having to read peers' papers Peers can be too critical Uncomfortable critiquing others	Lack of confidence in peer assessments received (19%) Misleading/contradictory feedback (11%)

Nevertheless, a great many similarities between student reactions to peer assessment were found. The greatest percentage of students in the Rhodes study who found the exercise useful fell into the 'Productive' general dimension (Table 1). Receiving constructive feedback from peers that enabled students to identify mistakes and improve their own work seems to have been a prominent positive outcome (59% of those who rated the peer assessment as useful did so because of one of these reasons). As with the Hanrahan & Isaacs (2001) study, reading another student's work, gaining a better understanding of what assessors were looking for, and developing a realistic

sense of the value of one's own work were also found to be important. Some positive student comments on these aspects included:

Most definitely! One doesn't always find their own errors with their work as naturally we believe our work to be correct. So this way of assessing each other is quite eye-opening.

Absolutely! Feedback from our peers was constructive and it was good to do that before the final marks. We got to see how others construct their essays and learn from one another.

To a larger extent it was helpful because other new ideas that I didn't think of were suggested and areas that I thought I had done well were magnified more to show the mistakes that could have cost me a lot of marks.

The peer assessment for me was good idea because it led me to get new ideas on how I could improve my essay writing and when marking other students' essays you tend to look at mistakes that you, yourself were making and in that way you improve as a writer.

Of the 47% of students who did not find the peer assessment useful, the biggest problem seems to have been related to the process or to the 'problems with implementation' general dimension under the Hanrahan & Isaacs (2001) study. Fully 68% of Rhodes students who did not find peer assessment useful fell into this category. Themes were very similar to those identified by Hanrahan & Isaacs (2001), such as the process being time-consuming and complicated, and it not being taken seriously by some peers. An interesting difference is that, while Rhodes students seem to have received the peer assessments with enough time to make the suggested changes – a positive outcome of using the online system, which made the process more efficient – they were greatly concerned about intellectual property rights (copying), which is undoubtedly a negative outcome of using the online format. Some mixed and negative student comments on these aspects included:

Yes, because you get constructive criticism from peers. No, because not everyone approaches the assessment with the same mind frame – to help others. Sometimes it is approached as a joke or something you do merely to meet requirements.

Yes, I did find it useful however I feel that it did give some people the chance to leach off the better work of others.

No! It made it difficult for me to complete my essay, this is because some students do not really go through the essay and they mislead you. It takes up too much of my time and a lot of effort.

Not really, it was quite a long process and not everyone managed to follow it properly. I personally found it as a waste of time.

Another interesting difference between the two studies is that Rhodes students did not report either positive peer pressure – that is, motivation to do better work because it would be marked by peers - or negative peer pressure – discomfort relating to having to assess peer's work. In the Hanrahan & Isaacs (2001) study, students reported both positive motivation and feelings of being uncomfortable with assessing or judging the work of other students. While the Hanrahan & Isaacs (2001) study was as anonymous as possible, identifying information being removed from hard copies of the essays, and no one being allocated the essay of someone in their own tutorial group, the Rhodes online system seems to have been regarded as more truly anonymous.

Two further closed-ended questions explored the idea that changing roles from being assessed to being an assessor is itself a valuable learning opportunity (Topping, 2010). 44% of students disagreed or strongly disagreed with the statement that, “The useful part of the essay 2 peer assessment was the feedback I got from other students”; But, 71% of students agreed or strongly agreed that “The useful part of the essay 2 peer assessment was giving feedback to other students as an assessor. This supports the findings of previous research (Snowball & Sayigh, 2007; Topping, 1998, 2010).

Correlation between peer and tutor assessments

In the interest of enhancing reliability the experienced tutor firstly verified the appropriateness or “accuracy” of the final mark in the sample of 50 essays examined. Since the marks of nine of these essays deviated significantly (more than 10%) from what she felt it should be, these “tutor” marks were adjusted (both up and down).

In order to determine the effect of the writing responding provided through peer assessment, the tutor marks for final submissions were compared with the peer marks for first submissions (see **Error! Reference source not found.2**). It is important to note that this is not a direct correlation between peer and tutor marks as students responding to peer feedback would have arguably resulted in the quality of final submissions being better than first submissions.

Table 2 Comparison between tutor marks (for final submission) and peer assessment marks (for first submission).

Tutor mark (for final submission)		Mean peer mark (for first submission)	Standard Deviation (of peer marks)
1	First class 75%+	68.45	10.484
2A	Upper second 70-74%	68.30	29.512
2B	Lower second 60-69%	61.38	14.563
C	Third class 50-59%	59.77	15.416
F	Fail 49% or less	68.00	12.660
Total		64.27	15.557

Table 2 shows that peer marks were not particularly “accurate” as they tended to congregate in the 60-68% range on average, despite the fact that the final marks spanned the complete range of performance from first class to fail. It is therefore possible to deduce that while the final tutor mark of students with lower second passes were matched more closely by those of their peers (for the first submission), those students with first and upper second passes tended to be marked down by their peers

and those students who failed tended marked up by their peers. This suggests that students are generally reluctant to give very high or low marks for their peers' essays. It follows then that students with poor marks may get a wrong impression of the quality of their work and therefore be unmotivated to improve their essays.

Quality of peer marks and comments

The experienced tutor evaluated the quality or usefulness of the peer mark and peer comments in each of the four assessment categories of the assessment grid by rating them on a 5-point Likert scale. The four categories were: (i) subject (coverage, theory analysis and application); (ii) structure and organisation; (iii) presentation and referencing; and (iv) language and other requirements (for example, the number of references cited and length). In terms of the rating, 5 was excellent (accurate mark; comments useful, constructively aligned with assessment criteria) and 1 was very poor (inaccurate mark; comments missing or superficial, incorrect, destructive, does not align with assessment criteria).

3 shows the experienced tutor's average rating (in a scale of 1 to 5) of the quality of the marks and of the comments in each of the assessment grid categories organised according to the class of performance for the final essay. In most of the assessment categories tutor-peer correlation seems to be higher for marks than for comments. The notable exception is the *Presentation and Referencing* assessment category where peers provided the most useful feedback in terms of comments and marks. This might indicate that, as found by Vickerman (2009, p. 227), students who find it difficult to comment on content or conceptual understandings seem to comment on technical issues such as spelling, grammar and referencing.

Table 3 Average feedback quality by final essay marks (class)

	Subject / Content		Structure and Organisation		Presentation and Referencing		Language and Requirements		General Comment
	Marks	Comment	Marks	Comment	Marks	Comment	Marks	Comment	
1	3.545	3.273	4.000	3.30	3.68	4.00	3.864	3.50	4.05
2A	3.500	3.300	2.900	3.30	4.00	3.20	3.000	3.30	2.80
2B	3.308	3.077	3.538	3.50	3.62	3.77	3.308	3.29	3.62
3	3.077	3.077	3.500	3.63	3.50	3.59	3.269	3.41	3.38
F	2.938	3.313	3.250	2.69	3.56	3.88	3.313	2.93	3.49
Ave	3.260	3.180	3.520	3.33	3.63	3.73	3.390	3.31	3.54

Linked to this finding is that the assessment categories in which improvements were most often made in response to feedback received were *Presentation and Referencing* (41% of the sample) and *Language and Requirements* (38% of the sample), while changes in the *Subject/Content* and *Structure and Organisation* categories were much more rare (11% of the sample).

A declining trend in the tutor-peer mark correlation from the first to the fail classes of performance suggests that peer marks of students with first class passes seem to be more appropriate or “accurate” than those with lower marks.

Changes between first and final submission

The experienced tutor was also asked to evaluate changes between the first and final submission (no change, slight or superficial change, some changes, or significant changes); and to ascertain what the effect of the changes (or lack thereof) was on the final submission (essay got worse, slight, some or significant improvement).

In nearly a quarter of cases, no changes were made in response to the peer assessment, 44% of students made slight changes, 22% some changes and only 10% significant changes. This may reflect the lack of confidence in peer assessments at first year level. What we were not able to explore is the extent to which peer assessors

themselves made changes to their own essays in response to the ideas and experiences gained through assessing someone else's work.

In the majority of cases (74% of the sample), there was some measure of improvement in the quality of the final essay submission compared to the initial submission, which is an encouraging result. Nearly a quarter of students, however, made no changes, and in one case the final submission was actually worse.

The experienced tutor was also asked to try to determine the extent to which the changes made could be attributed to the peer feedback received. In a quarter of cases, the peer feedback did not result in any changes being made. For almost half the sample (47%) peer feedback had only a slight effect. In 19% of cases, there was "some" effect and in only 9% of essays did peer feedback have a "significant" effect.

Superficial feedback gave students a false impression about the supposedly high quality of their work, and not surprisingly did not result in any improvement in those essays. This is especially the case with students who received high marks from their peers and consequently did not have any incentive to improve their work. Similarly, this was the case with students who received feedback that was not useful for making improvements, irrespective of the mark they received from their peers.

It was also noted that comments for improving essays were more likely to be serious and positive criticisms that helped to improve the essay, where essay quality was already quite high. While the direction of causality here is unknown, it may be that, where students could see that they were benefitting from reading a good essay, they were more likely to give thoughtful feedback.

Students thus also tended to give higher marks to those essays that gave them good ideas about what to include in their own work (again related to benefits of *being*

an assessor, rather than one being passively assessed). One comment made this explicit: “your essay gave me good ideas of how to improve my own”.

The impact of peer assessment participation on essay marks

In order to examine the potential impact of the peer assessment on student performance, a statistical model was estimated using regression analysis (see Table 4). It was hypothesised that a student’s mark for essay 2 was dependent on ability in economics and demographic variables, particularly english language proficiency. Economics ability was captured in 3 ways: performance in the first semester (a mark out of 100 achieved for the economics 101 exam); performance in the multiple choice questions of the the economics 102 (second semester) exam (a mark out of 50); and performance in the written questions section of the economics 102 exam (mark out of 30). Demographics included were gender (1 if male, 0 if female), english first or second language (1 if first language; 0 if second language), faculty (1 if commerce, 0 other) and citizenship (1 if South African; 0 if other).

The regression analysis are able to identify, holding other variables constant, the direction, magnitude and statistical significance of the impact of each of the independent variables in the model on the dependent variable (mark for essay 2). The model was estimated using the ordinary least squares (OLS) method, and explained 20% of the variation in essay marks. However, only four of the variables were significant in explaining essay performance. These did not include whether the student had been part of the peer assessment. Thus, peer assessment, while useful in other ways, was not found to have any statistically significant impact on essay mark in this case, as was found by the van Zundert et al. (2010) review.

Economics ability was the most important determinant of essay marks: Holding all other variables constant, each 1 mark (2%) increase in exam written question performance (0.62/30), increased essay 2 mark by 1 percentage point (highly statistically significant); each increase of 0.23% in the first semester (eco101) exam, increased essay 2 marks by 1%. Interestingly, performance in exam multiple choice questions had no statistically significant effect on essay 2 marks, concurring with other research (Buckles & Siegfried, 2006; Chan & Kennedy, 2002) that the two assessment techniques (multiple choice versus written response) are testing different aspects of knowledge. Also as found in other studies (Rayner & Cridland, 2009; Smith, 2009), English first language speakers were found to be at an advantage, having, on average, an essay 2 mark 2.2% higher than second language speakers.

Table 4 Regression analysis

Variable	Coefficient
Constant	40.12146
Peer assessment participation (1 if yes)	0.496019
Exam written question mark (out of 30)	0.620534***
Eco101 exam mark (%)	0.235762***
English first language	2.238834**
Gender (1 for male)	-1.593915
Faculty (1 if commerce)	-1.031186
Citizenship (1 for South African)	0.873062
Exam multiple choice mark (out of 50)	0.064093
Adjusted R-squared (goodness of fit)	0.205636
Prob(F-statistic)	0.000000

*** Significant at the 1% level, **Significant at the 5% level.

5 Conclusions

It has been argued that peer assessment can be a productive way to engage students in higher-order learning activities, and encourage them to become more independent learners (Hanrahan & Isaacs, 2001; Pope, 2001). When used formatively, it can also be a valuable tool for improving academic writing in large class settings and for accommodating various learning styles (Vickerman, 2009).

Building on Hanrahan and Isaacs (2001), this research explored the impact of an online formative peer assessment exercise in a large first year class. While many of the same dimensions and themes were identified, the online environment seems to have additional advantages, but also unique challenges. A majority of students in this study (58%) agreed that the peer assessment exercise improved the quality of their essay. However, as found in other research (Snowball & Sayigh, 2007; Topping, 1998, 2010) far more students (71%) reported that *giving* feedback (taking on the role of an assessor) was the most useful part of the exercise, rather than *getting* feedback from peers.

In evaluating the impact of peer assessment on a sample of 50 essays, it was found that the most useful comments and most accurate assessments were given in technical categories (referencing, layout, requirements etc) rather than on more substantial content issues (also found by Vickerman, 2009). Peer marks also tended to be bunched in the 60-69% range, which may discourage students whose essays are actually worth less from making suggested changes. Regression analysis showed that participation in the peer assessment exercise (compared to students who did not participate) had no impact on essay marks, confirming the contention of Van Zudert et al. (2010) that the impact of peer assessment may only be apparent in the long run.

It is thus suggested that formative peer assessment should not include giving a mark, but instead support students in providing more feedback on substantive, content-related issues, and in responding to such feedback.

6 References

- Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing Procedures for Implementing Peer Assessment in Large Classes Using an Action Research Process. *Assessment & Evaluation in Higher Education*, 27(5), 427-441.
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals* (1st ed.). New York: David McKay.
- Buckles, S., & Siegfried, J. J. (2006). Using Multiple-Choice Questions to Evaluate In-Depth Learning of Economics. *Journal of Economic Education*, 37(1), 48-57.
- Carlson, T., MacDonald, D., Gorely, T., Hanrahan, S., & Burgess-Limerick, R. (2000). Implementing Criterion-referenced Assessment Within a Multi-disciplinary University Department. *Higher Education Research & Development*, 19(1), 103-116.
- Chan, N., & Kennedy, P. E. (2002). Are Multiple-Choice Exams Easier for Economics Students? A Comparison of Multiple-Choice and “Equivalent” Constructed-Response Exam Questions. *Southern Economic Journal*, 68(4), 957.
- Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE bulletin*, 39(7), 3-7.
- Draaijer, S., & van Boxel, P. (2006). Summative peer assessment using “Turnitin” and a large cohort of students: a case study. *CAA Conference*. Loughborough University. Retrieved from <http://hdl.handle.net/2134/4559>

- George-Walker, L. D., & Keeffe, M. (2010). Self-determined blended learning: a case study of blended learning design. *Higher Education Research & Development*, 29(1), 1-13.
- Hanrahan, S. J., & Isaacs, G. (2001). Assessing Self- and Peer-assessment: the students' views. *Higher Education Research & Development*, 20(1), 53-70.
- Lupton, M. (2008). Evidence, argument and social responsibility: first-year students' experiences of information literacy when researching an essay. *Higher Education Research & Development*, 27(4), 399-414.
- Pope, N. (2001). An Examination of the Use of Peer Rating for Formative Assessment in the Context of the Theory of Consumption Values. *Assessment & Evaluation in Higher Education*, 26(3), 235-246.
- Pritchett, N. (1999). Effective question design. In S. Brown, J. Bull, & P. Race (Eds.), *Computer-Assisted Assessment in Higher Education*. London: Kogan Page.
- Race, P. (2001). A Briefing on Self, Peer and Group Assessment. resources, . Retrieved March 23, 2009, from http://www.heacademy.ac.uk/resources/detail/SNAS/snas_901
- Rayner, G., & Cridland, C. (2009). Rewriting the first year biology essay: Addressing student diversity through a dialogic approach to assessment practice. In M. A. Funston & B. Zammit (Eds.), "*Knowing and Knowledge*", *the FYE transition subject of study in the Faculty of Arts, Education and Human Development at Victoria University, Melbourne: An institution-wide approach to retaining and supporting first year* (p. 26).
- Sitthiworachart, J., & Joy, M. (2008). Computer support of effective peer assessment in an undergraduate programming class. *Journal of Computer Assisted Learning*, 24(3), 217-231.

- Smith, L. C. (2009). An analysis of the impact of pedagogic interventions in first-year academic development and mainstream courses in microeconomics. *South African Journal of Economics*, 77(1), 162-178.
- Smyth, K. (2004). The benefits of students learning about critical evaluation rather than being summatively judged. *Assessment & Evaluation in Higher Education*, 29(3), 369-378.
- Snowball, J. D., & Sayigh, E. (2007). Using the tutorial system to improve the quality of feedback to students in large class teaching. *South African Journal of Higher Education*, 21(2), 321-333.
- Strijbos, J.-W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*, 20(4), 291-303.
- Topping, K. J. (1998). Peer Assessment Between Students in Colleges and Universities. *Review of Educational Research*, 68(3), 249-276.
- Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction*, 20(4), 339-343.
- Vickerman, P. (2009). Student perspectives on formative peer assessment: an attempt to deepen learning? *Assessment & Evaluation in Higher Education*, 34(2), 221.
- van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, 20(4), 270-279.