

Don't write, just mark: the validity of assessing student ability via their computerized peer-marking of an essay rather than their creation of an essay

Phil Davies*

University of Glamorgan, UK

This paper reports on a case study that evaluates the validity of assessing students via a computerized peer-marking process, rather than on their production of an essay in a particular subject area. The study assesses the higher-order skills shown by a student in marking and providing consistent feedback on an essay. In order to evaluate the suitability of this method of assessment in judging a student's ability, their results in performing this peer-marking process are correlated against their results in a number of computerized multiple-choice exercises and also the production of an essay in a cognate area of the subject being undertaken. The results overall show a correlation of the expected results in all three areas of assessment being undertaken, rated by the final grades of the students undertaking the assessment. The results produced by quantifying the quality of the marking and commenting of the students is found to map well to the overall expectations of the results produced for the cohort of students. It is also shown that the higher performing students achieve a greater improvement in their overall marks by performing the marking process than those students of a lower quality. This appears to support previous claims that awarding a 'mark for marking' rewards the demonstration of higher order skills of assessment. Finally, note is made of the impact that such an assessment method can have upon eradicating the possibility of plagiarism.

Introduction

Davies (Davies, 2002a) reported a student comment 'I learnt more from marking the work of others than from writing my own report'. If this were to be true then does it negate the need for students to create their own essays? This forms the basis of this study in making use of a computerized peer-assessment environment in order to

*School of Computing, University of Glamorgan, Pontypridd, Mid-Glamorgan CF37 1DL, UK.
Email: pdavies@glam.ac.uk

assess a students' knowledge in a particular subject area, rather than making them write an essay in this area.

The benefits of peer-assessment are well documented with respect to student and staff benefits (Falchikov, 1995b; Boud *et al.*, 1999). Part of the negativity that surrounds peer-assessment is the need for 'proof' that the end mark produced by the peers is equitable with respect to the mark that would have been provided by a tutor if they had performed the marking process (Falchikov & Goldfinch, 2000). Numerous studies have reported back favorable results with regard to the equity of student and tutor marking (Stefani, 1994). The majority of the research previously carried out in the area of peer-assessment has tended to concentrate on the marking aspect of the process. This generation of marks is only part of an assessment process with feedback being of utmost importance (Collis *et al.*, 2001). Some previous studies have attempted to concentrate on the comments provided by the peers rather than merely just the marks (Falchikov, 1995a). Assessing whether the quantification of the quality of peer-comments can also be used as a method for assessing the quality of an essay is a relatively new area of research (Davies, 2003a).

In recent years computerized systems have been introduced in order to provide managed environments to support the use of peer-assessment systems (Davies, 2000; Bhalero & Ward, 2001; Bostock, 2001; Lin *et al.*, 2001; Parsons, 2003) aimed at assessing the quality of essays and reports. Systems are now also being created that are attempting to utilize the inherent benefits of peer-assessment in other areas such as computer programming (Sitthiworachart & Joy, 2003; Lewis & Davies, 2004).

The higher-order skill of evaluation (Bloom, 1956) is an area that would be expected to be assessed and used to identify the 'better' students. The method of peer-assessment detailed in this paper attempts to address this issue, and facilitate a method whereby these 'better' students will be able to demonstrate these skills and score higher. Lin *et al.* (2001) noted that 'high executive students contributed substantially better feedback than the low executive thinkers'. It will be interesting to ascertain whether this maps to their ability in the marking of the work.

Robinson (2002) suggests that as much as a third of the feedback provided by students was 'inadequate'. Therefore students producing such feedback will receive fewer marks than those students who show consistency within both their marking and commenting. Rada and Hu (2002) suggest that students should receive credit for doing good comments. This was also taken account of in the study reported here.

Essays are considered as a means of assessing the subjective skills of a student. However, lecturer marking can be highly subjective, 'Essays are demonstrably the form of assessment where the dangers of subjective marking are greatest' (Race, 1995). If the attribution of marks for peer-marking as presented in this study is to be of any value, then the student's overall subjectivity should be removed (high and low marking and commenting), and their ability in showing consistency of evaluative skills should be what is rewarded.

Due to the experimental nature of this exercise, other assessments were included within the overall summative assessment process, namely objective tests (a series of MCQ tests including confidence testing) (Davies, 2002b) and also the creation and

submission of an essay of their choice (accompanied by a presentation) within the subject area of the module (Networks and Internet Architectures). This ensured that the assessment of the learning outcomes of the module was fully covered. It was decided that in order to attain full student engagement within all aspects of the assessment process, there would be an equal contribution of marks from each of the assessment methods that contribute to the final summative grade.

Background

One of the aims of this work was to replace essay writing with peer assessment. The students were directed to research a particular area of study within a module, and then used this to evaluate the work of a previous student. Part of the study was to investigate how a mark for the demonstration of evaluative skills can be mapped to an actual standard marking scale?

The study was undertaken during the Autumn Term 2003, within the School of Computing at the University of Glamorgan. The assessment was aimed at a cohort of 34 students studying on the Post Higher National Diploma (HND) course of study. This course is designed for students who have previously attained a HND and are using this Post HND course as a bridging year to increase their credits to a level whereby they are able to enter into the final year of studies on the Bachelor of Science Honours course (BSc Hons). The students undertaking the module Networks and Internet Architectures (NIA) were from a broad range of previous named HND awards varying from Network Administration, Computer Studies to Information Systems and Business.

At the start of the module students were presented with a series of questions on computer networking. From this, the area of N-tier architectures was highlighted as one which they had not covered. This area is traditionally covered within the final year module in Distributed Systems and Enterprise Networks (DSandEN). In the academic year 2002–2003, an essay had been set to the final year group in this area of N-tier Architectures. From this, 39 essays were available, each on average had been marked by their peers six times and a compensated peer-mark having been previously derived. Rada et al note (Rada *et al.*, 2002) that ‘Students (as well as teachers) may manifest bias, and a student may unfairly evaluate another student’s work’. By having a number of markings this bias is removed. Also the compensation processes employed minimizes this aspect of the peer-generated grades.

The NIA students were provided with web references on the basics of N-tier Architectures. They were also provided with the same assessment pro-forma that had been given to the DSandEN students. To ensure that they perform the peer-marking process as closely to the DSandEN students, they were able to use the CAP Menu driven marking tool (Figure 1). The DSandEN students had also made use of the anonymous communications facilities of the CAP system (Davies, 2003), however as these students had already graduated the previous year, this facility was not available to the NIA students.



Figure 1. CAP marking via pull-down menu

Methodology

The essays to be marked were from the last year's final year. The compensated peer-generated marks for each of these essays had already been recorded. The marking of these essays had been undertaken making use of the pull-down menu driven CAP system (Figure 1).

Lin *et al.* (2001) note that 'some students complain that holistic peer feedback was often too vague or useless'. By using this menu-driven system for marking, greater specificity can be generated within the marking process. Also students make use of the menu system as a scaffold for their own commenting (Davies, 2003a). Assessment via this tool provides both the ability to mark and comment upon an essay. In this way two quantified values can be generated and compared, i.e. the peer-mark and the feedback-index. Analyses of previous student commenting indicates that while doing the marking, students include free-text comments as well as using the menu driven marking system, therefore the feedback-index used included these comments as well as those generated by the menu system. Use was made of the 'Mark-Up' tool (Figure 2) in order to create feedback-indexes for each essay marking.

The NIA students as part of their requirements within this assessment were asked to mark six essays each. The average feedback-index was created for each essay previously marked by the DSandEN students for comparative purposes (Figure 3).

A compensation process was included to take into account of both high and low commenting/markings, as there is not necessarily a direct correlation between the mark produced and the comments. On one occasion a mark of 90% was allocated by

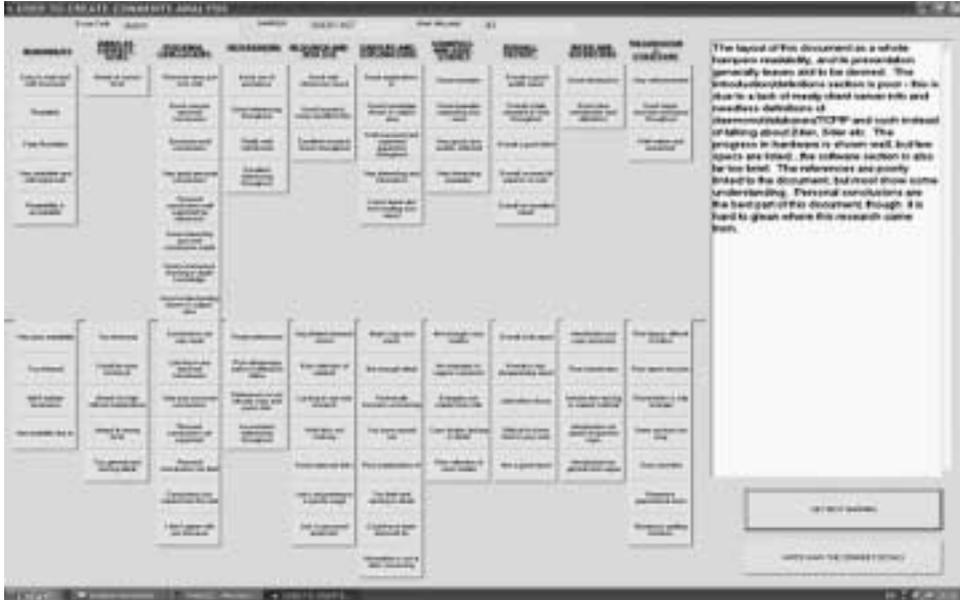


Figure 2. Mark-up tool for quantifying comments

a student, who then went on to heavily criticize almost every aspect of the report (Davies, 2000).

In order to assess the validity of using these feedback indexes, a comparison between the average peer-marks and feedback indexes was performed for these

Working Student	MARK GIVEN	Probability	Level of level of evidence	Personal Conclusion	References	Reasons and use of words	Content Explanations	Examples Case Studies	Overall Report Quality	Introduction/Outline	Report Structure	TOTALS
99054015	58	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	-1
99044358	73	1.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	5
99073548	64	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	-2
99041152	72	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0
99044815	75	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	4
99079906	64	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	-2
99015192	64	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	2
97000399	76	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0
		0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	

0.25 -0.125 0.75 -0.125 -0.625 -0.625 0.375 0.625 0.25 0
Feedback Index 0.75

Figure 3. Creation of feedback index for DSandEN essays

essays. If there is validity within these ‘scores’ for an essay, then they are used as a measure against the peer-marking and commenting produced by each of the post-HND students.

By producing average differences within the markings of a Post HND student and those of the essays, a measure of a student’s ability in evaluating the essays is produced (Figure 4). These deviations in marking and commenting are used to produce a grade for the Post HND student both representing their marking and commenting ability.

To produce an average difference for the marking, the peer mark produced by the DSandEN students has the mark given by the NIA student taken away from it + or – X. The average of the differences can then be calculated, i.e. (X1) + (X2) ... and then subsequently be divided by the total number of markings giving AV. This figure represents the average differences from the perfect 0. However, in doing this no account is taken of the consistency of a student’s marking. Therefore, the absolute differences are calculated from this average and divided by the total markings, i.e. $\Sigma(AV - [\text{difference for each marking}]) / \text{number of markings}$. This numerical value now represents the consistency of the marker. A similar process is performed in order to produce a numerical value to represent the consistency of the marker with regard to the differences of the feedback indexes (associated with comments).

These differences are then mapped via some form of linear structure in order to generate a valid grade for a marker. How the allocation of this absolute grade will be produced from this consistency mark will be presented later in this paper.

Calculate Average Differences

FILE MARKED	MARK GIVEN	WORLD MARK	DIFFERENCE	TIME TO MARK	COMPOUND FEEDBACK INDEX	FEEDBACK INDEX GIVEN BY MARKER	FEEDBACK DIFFERENCE
001	81	83	-2	25	1.000	1	-0.04166
002	81	83	-2	32	1.507	5	-0.06201
003	81	80	+1	26	0.503	-10	+0.02306
004	75	73	+2	36	0.520	4	-0.02280
005	80	78	+2	33	1.000	0	-0.00000
006	81	84	-3	28	0.400	1	0.00000
			-0.2	28.8			-0.04166

Absolute Average Marking Difference: 0.00000 Absolute Average Feedback Difference: 0.00000

Student Number: 00012126 Student Name: Stephen Brown Get Next Student

Figure 4. Quantification of differences in NIA student markings

Table 1. Mapping of non-compensated DSandEN marks to comments

F/Index	-5	-4	-3	-2	-1	0	0	1	2	3	4	5	6	7
	39	46		58	60	60	69	63	67	63	70	73	76	80
		54		45	59		58	68	67	68	69	68	73	
					60		65	67	60	69	70			
					59			60	62		76			
					50			62			75			
								62			68			
											71			
Avg	39	50	0	51.5	57.6	60	64	63.7	64	66.7	71.3	70.5	74.5	80
St Dev		5.66		9.19	4.28		5.57	3.14	3.56	3.21	3.04	3.54	2.12	

Results and analysis

To use the DSandEN results as a ‘marker’ for the NIA students marking, there must be a confidence that these initial results are valid. Of the original essays 10% had been cross-marked and there was a maximum of 5% variation in any of the compensated peer-marks awarded which was considered acceptable. The qualitative assessment of the comments generated within the marking process was not previously generated. Making use of the Markup tool (Figure 2), a feedback index was created for each essay (Figure 3) as marked by the DSandEN students. The correlation of these grades to the marks is shown in Table 1.

However, some students tend to over- or under-mark. This is addressed by using the compensated peer-mark. Table 2 shows the effect of using the compensated feedback indexes for the DSandEN student markings. There is a significant positive correlation between the average feedback indexes and the average peer marks awarded for the essays (Tables 1 and 2). This improves by performing the compensation process (positive correlation of over 0.9).

More importantly the average standard deviations produced for each average reduces from 4.33 to 3.61. This shows that the range of marks within each category

Table 2. Mapping of compensated DSandEN marks to comments

F/In	-5	-4	-3	-2	-1	0	0	1	2	3	4	5	6	7	8	9
	39		54	58	50	60	60	68	63	63	71	76	73			80
	46			58	59		62	69	67	67	69	68	73			
				60	45		65	60	68	71	68		76			
					59		60	67	62	70	69					
								62			75					
Avg	42.5	0	54	58.7	53.3	60	61.8	65.2	65	67.8	70.4	72	74	0	0	80
St D	4.95			1.15	6.95		2.36	3.96	2.94	3.59	2.79	5.66	1.73			

Table 3. Comparison of averages of non-compensated and compensated average feedback indexes

F/Index	-5	-4	-3	-2	-1	0	0	1	2	3	4	5	6	7	8	9
Avg T1	39	50		51.5	57.6	60	64	63.7	64	66.7	71.3	70.5	74.5	80		
Avg T2	42.5		54	58.7	53.3	60	61.8	65.2	65	67.8	70.4	72	74			80

of the feedback indexes is reduced. Also noticeable (Table 3) is the fact that the top essay mark has moved along the linear scale significantly with regard to the feedback received by performing the compensation process. This positive correlation maps well with the previous results (Davies, 2003a). Due to this correlation it is fair to assume that the feedback indexes and the peer-marks produced for the DSandEN essays are true measures of the quality of the work produced, and can be used as controls for the assessment of the NIA students. It was decided to permit the NIA students to perform their marking and commenting of the essays making use of the CAP marking system (with the same menu driven commentings) as had been used by the previous year's DSandEN students.

The average mark for the essays produced and marked by the DSandEN students was 63.52%, with a standard deviation of 8.69. The average mark produced for the same essays, as marked by the NIA students was 58.75%, with a standard deviation of 12.71 (only 5 out of 34 markers on average over-marked). A positive correlation of 0.77 existed between the average compensated marks generated for the essays by the DSandEN students, with the average marks produced by the NIA students. Looking at the feedback indexes generated by the NIA students, on average their feedback was -1.37 (only 8 out of the 34 markers over-commented) compared with the feedback produced by the DSandEN students.

Table 4 shows the results of the NIA markings and comments. On examining the use of the menu-driven comments and the free-text comments, a number of students tended to make use of both facilities integrated together. It was therefore decided that the feedback index produced for each marking should also include these free text responses by again making use of the Markup (Figure 2) Application. Also included in Table 4 are the gradings (0-5) for this and other assessments within the study. Table 4 shows the average absolute differences (as discussed previously) produced by the students for both their marking and commenting.

To allocate a final grade, three assessments were specified. Therefore the differences produced via the peer-marking process by the NIA students needed to be quantified. In trying to allocate marks in a linear manner, the following grading was decided upon (Table 5). In previous years of this module the average mark produced for the students has been between 55-60%. Therefore it was assumed reasonable, as there were no indicators that this year's NIA students were any different to previous years, that a similar average would be produced within this cohort. A range of 0-5 marks was decided upon, and linear scales were determined for each aspect of the assessment based on these previous year's expectations.

Table 4. Results For NIA students in all Assessments (mapped to linear scale 0-5)

Student	Average compensated difference in marking	Grade 0-5 for marking driven)	Average difference (just using menu)	Grade 0-5 for menu comments	Average difference (using free text + menu)	Grade 0-5 for Free-text comments	Difference in feedback menu/free text	Combined grade for marking and free-text commenting/20	Combined grade for marking and free-text commenting %	MCQ range produced 0-5	Essay produced 0-5	Total based upon /5 equal	Results based on MCQ and essay	Average time taken to mark an essay (minutes)
1	4.8	4	2.54	4	2.8	4	-0.2	16	80	3.4	3.3	70.9	66.3	35.5
2	18.2	1	9.5	1	10.3	0	-0.8	2	10	1.6	3.5	37.4	51.1	32.7
3	8.3	3	4.5	3	5.0	3	-0.6	12	60	3.6	2.1	58.4	57.7	54.5
4	6.3	4	4.3	3	5.0	3	-0.8	14	70	2.6	2.6	58.1	52.2	38.3
5	5.7	4	2.1	4	4.0	4	-1.9	16	80	3.4	2.6	66.9	60.4	33.2
6	4.3	4	2.6	4	3.6	4	-1.0	16	80	3.7	3	71.0	66.5	30.2
7	16.0	1	4.2	3	7.3	2	-3.1	6	30	2.6	2.1	41.2	46.8	55.8
8	3.3	5	1.7	5	2.7	4	-1.0	18	90	1.9	2	56.3	39.5	61.0
9	4.2	4	2.5	4	6.1	2	-3.6	12	60	3.4	2.6	60.4	60.6	49.6
10	11.5	3	5.1	3	5.8	3	-0.7	12	60	3.3	2.8	60.5	60.8	35.2
11	10.3	3	3.5	4	4.5	3	-1.0	12	60	3.3	3.1	62.8	64.2	36.3
12	11.9	3	3.2	4	2.8	4	0.4	14	70	2.2	3	57.9	51.9	17.0
13	16.5	1	8.3	1	9.3	1	-1.0	4	20	2.6	2.8	42.3	53.5	41.8
14	8.9	3	5.7	3	6.2	2	-0.6	10	50	1.5	2.1	41.2	36.7	68.6
15	19.5	1	9.1	1	8.9	1	0.2	4	20	3.6	1.8	42.4	53.6	47.3
16	9.7	3	4.7	3	3.9	4	0.8	14	70	1.3	2.3	47.0	35.4	16.0
17	11.0	3	4.6	3	7.6	2	-3.0	10	50	2.6	2.8	52.0	53.1	65.2
18	12.2	2	2.4	4	5.5	3	-3.2	10	50	3.4	3.1	59.6	64.4	33.7
19	9.3	3	5.7	3	8.4	1	-2.8	8	40	2.1	2.3	42.2	43.2	40.3
20	15.2	2	2.2	4	2.8	4	-0.7	12	60	2.8	2.6	56.5	54.7	58.0
21	15.1	2	3.3	4	5.4	3	-2.0	10	50	3.4	2.6	57.2	60.7	40.2
22	9.6	3	4.1	3	5.9	3	-1.7	12	60	2.3	2.9	54.8	52.1	48.5
23	2.7	5	4.8	3	7.9	2	-3.1	14	70	3.4	2.5	62.5	58.7	34.5
24	7.6	4	4.7	3	3.7	4	1.0	16	80	3.4	1.9	61.5	52.3	34.3
25	11.0	3	3.3	4	4.6	3	-1.3	12	60	3.0	2.6	57.5	56.2	33.4
26	10.3	3	3.9	4	4.7	3	-0.8	12	60	1.4	0	29.2	13.8	21.2

Table 4. *Continued*

Student	Average compensated difference in marking	Grade 0-5 for marking	Average difference (just using menu driven)	Grade 0-5 for menu comments	Average difference (using free text + menu)	Grade 0-5 for Free-text comments	Difference in feedback menu/free text	Combined grade for marking and free-text commenting/20	Combined grade for marking and free-text commenting %	MCQ range 0-5	Essay produced 0-5	Total based upon /5 equal	Results based on MCQ and essay	Average time taken to mark an essay (minutes)
27	8.2	3	2.1	4	2.2	4	-0.2	14	70	2.6	3.4	63.3	60.0	35.7
28	10.3	3	3.0	4	3.5	4	-0.6	14	70	2.6	2.9	59.6	54.4	44.0
29	9.2	3	4.2	3	5.9	3	-1.7	12	60	4.0	3.4	69.4	74.1	43.0
30	17.6	1	4.3	3	6.3	2	-2.0	6	30	3.2	3.1	52.4	63.6	41.3
31	6.5	4	7.2	2	6.2	2	1.0	12	60	1.9	0	32.7	19.1	49.5
32	4.7	4	2.2	4	2.0	4	0.2	16	80	1.7	2.1	52.2	38.4	42.7
33	4.0	4	1.8	5	6.2	2	-4.3	12	60	3.7	3.1	65.4	68.0	28.00
34	11.3	3	7.0	2	8.8	1	-1.75	8	40	3.6	2.9	56.5	64.7	81.6
Avg	9.9	3	4.2	3.3	5.5	2.8	-1.2	11.5	57.6	2.8	2.5	54.7	53.2	42.0
St Dev	4.6	0.7	3.2	1.4	4.3	2.1	1.0	5.7	28.3	0.2	0.3	10.1	1.1	32.6

Table 5. Mapping to common assessment scale 0–5

Marks	Marking difference	Feedback difference	Mapping for MCQ and essays
5	<4	<2	90–100%
4	<8	<4	80–89%
3	<12	<6	60–79%
2	<16	<8	40–59%
1	<20	<10	20–39%
0	20 or more	10 or more	0–19%

A comparison of average feedback differences produced the results shown in Table 6.

The overall mark for the peer-marking aspect of the coursework was 2.88/5 with a standard deviation of 1.41 (using menu + free text as the feedback score). As a percentage this evaluates to 57.6% (within expectations from previous years). The mark for the essay and presentation was slightly lower than was expected (50%). However, during the presentations it was noted that the NIA students' ability in developing an essay of their own was in general quite poor. This skill may be one that is developed throughout the course of the Post HND course.

In the past it was considered that the more time taken to mark an essay results in a more detailed and precise marking and commenting. The average time taken to mark the essays was 42 minutes (Table 4). This varied considerably between students with a standard deviation of 32.6 for the times taken. The times taken to mark an individual essay ranged from 10 to 104 minutes. The student with the highest average time for marking was 81 minutes (Number 34) yet he only received 1 for his quality of commenting, whereas the student with the lowest average time for marking 16 minutes (Number 12), received 2.8 for commenting, with both receiving 3 for consistency of marking. From viewing the figures there was no significant correlation between the time taken and the grade awarded for the peer-marking processes.

For a true comparison to be made between the various assessment methods, then they needed to be graded in a consistent manner (Table 5). A composite final grade was produced on the basis of equal weightings of the three methods of assessment. Correlation between essay/MCQ combined grade and the final assessment grade including the peer-assessment was 0.80. This would suggest a good match of the

Table 6. Average results for assessments (0–5)

Method of assessment	Average /5	Std Dev	Avg as %
Score for peer marking	3.0	0.71	
Score for feedback just menu driven	3.29	1.41	
Score for feedback with menu + free text	2.76	2.12	
Score for MCQ	2.8	0.15	56%
Score for essay + presentation	2.52	0.27	50%

students within the grade awarded for the peer-marking process. Table 7 shows the frequency distributions of the number of students within each grading.

In ordering the students via their final grade awarded there is on average a consistency of performance against the three methods of assessment used (Table 8).

If based only on MCQ and essay (50/50 split) and removing the peer-marking process, the final average would have been 53.20% (with a standard deviation of 1.13). This would have had an effect of increasing the final overall average by 1.48%.

Table 7. Frequency distributions of results

	0	1	2	3	4	5
Marking differences	0	5	3	15	9	2
Feedback on menu only	0	3	2	13	14	2
Feedback on menu + free text	1	4	8	10	11	0
Final mark (based on menu)	0	3	3	20	7	1
Final mark (based on text + menu)	1	4	6	17	6	0
MCQ tests	0	7	10	16	1	0
Produce own essay and presentation	2	2	20	10	0	0
Final based on 50/50 MCQ and essay only	2	4	16	12	0	0

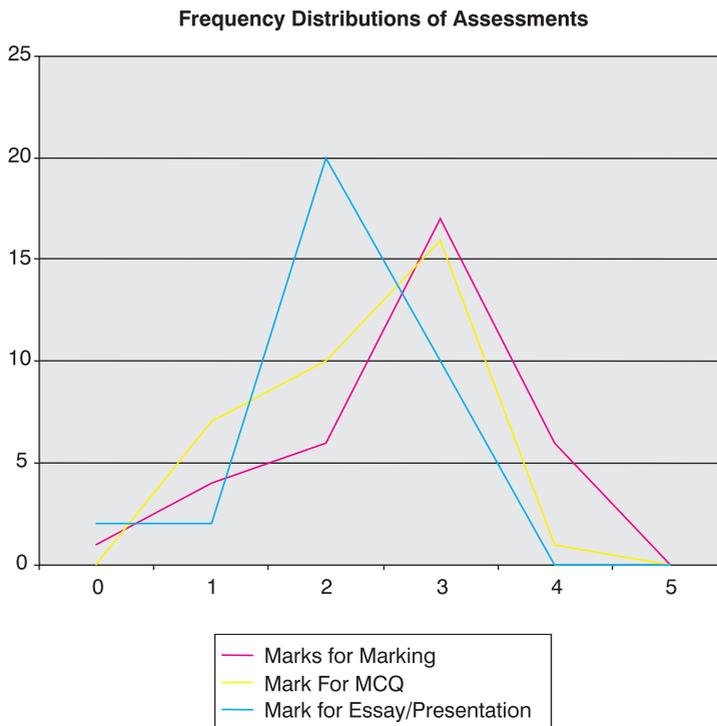


Table 8. Frequency distribution of NIA student results (final grades)

Final overall % mark	Number of students	Average MCQ	Average essay	Average peer-marking
70% +	2	3.52	3.13	4.0
60–69%	9	3.39	2.82	3.33
50–59%	14	2.78	2.67	3.00
40–49%	6	2.28	2.21	1.92
30–39%	2	1.75	1.75	1.75
20–29%	1	1.38	0	3.00

Note: student 20–29% did not submit an essay.

Based upon the final results produced by all three assessment processes, it is interesting to note which students performed best at the peer-marking process, i.e. displayed the greatest improvement by demonstrating evaluative skills. Table 9 indicates that the claim at the outset of this study that the higher order students will be rewarded appears to be vindicated.

The results are affected by the fact that one of the two students in the 30–39% group did not submit an essay. Without this student it would have been –13.69%. Also the student in the 20–29% category did not submit an essay. The results indicate on average a correlation between the results produces across the three methods of assessment and also the rewarding of higher order skills as demonstrated by the ‘brighter’ students.

Student feedback

Seventeen responses were received from the 34 students (50%) who undertook this assessment process. Of the students who replied they all stated that it was the first time that any of them had used any form of peer-assessment. The students were asked whether they had prepared in a different way prior to the marking than if they were going to write an essay themselves. 14 out of the 17 stated they had prepared in exactly the same way as they would have normally. Those who felt that they had prepared in a different manner generally made the point that they had performed

Table 9. Improvement based on performing peer-marking process

Based on all three results	Effect on student’s final % by including peer-marking process
70% +	+4.53%
60–69%	+1.51%
50–59%	+1.82%
40–49%	–2.16%
30–39%	–0.03%
20–29%	+15.45

much more research. They felt that a much better understanding of the subject area was required prior to the actual assessment process taking place. One student made note of the fact that their research continued throughout the marking process and they had often looked up areas that they didn't understand whilst progressing through the marking process. This might result in a problem for this student being able to maintain a standard throughout their marking process.

The students were asked how they felt about marking their peers. A number highlighted 'an apprehension' in undertaking this method of assessment initially, but this lessened throughout the course of the marking process, and a number commented 'it really became enjoyable'. A number of the students expressed their concerns that they 'didn't know how many marks to give or take away'. This feeling of uncertainty was expressed by most of the respondents.

Since these Post HND students are deemed to be at level two of a degree scheme, it was felt that it would be interesting to assess their thoughts on the quality of work produced by a level three student. The common thread of comments related to 'amazed at the wide range of work I had to mark'. The feedback reported back that there were few 'average' essays but they tended to be 'either excellent or poor'.

The students were questioned as to whether there were particular areas of the essays that they had found to be good or bad. The areas that were reported to be good were the use of examples, good grammar throughout, and where students had attempted to provide significant personal conclusions. The main areas that were reported as being poor concerned the referencing of the material. The markers commented on how difficult it had been in many essays 'to find out where they've got their information from'. On being questioned on whether they the marker had a good understanding of the subject area, 15 out of the 17 respondents answered in the positive. Comments such as 'I've a much better understanding than if I'd written my own' and 'I now have the confidence to explain this subject to my classmates' were very positive. The two students who did not feel they had a good understanding both said that their knowledge was still 'very general'.

On seeking improvements to the method used, the common comment from the students was that they required better initial guidance concerning such matters as the marking scheme, how to 'judge a good from a bad piece of work' and what to look for when marking. With regard to the CAP system one student suggested 'the pull down menu comments are inserted in alphabetical order'. Two students suggested that a good essay be provided, with the marks given and also the comments. If this were the case then the subjectivity of the peer-marker may well be weakened.

With regard to not doing an essay of their own, no student felt they'd been disadvantaged in any way. A number felt that they'd improved their knowledge significantly in the subject area because 'really had to learn the stuff, not just write about it'. Two students were pleased that 'the hassle of writing an essay has been removed'. Overall the feedback was very positive towards this method of learning through assessing.

A general comment made concerned the students not knowing 'how will I get my marks'. Even though it was explained up front that they would be judged on their

consistency of feedback and marking, the visualisation of what was considered 'good' and 'bad' was difficult for the students to judge.

A fact that has been noted in the past as a possible negative against peer-assessment is the problem of students not knowing or following the university rules associated with the identification of plagiarism. This is identified in the following 'quite a lot of copy and pasting (plagiarism), but it was referenced'?

Conclusions

This study aimed to assess whether students could be judged on their knowledge in a particular subject area by marking rather than creating an essay. The results indicate that this is possible and that student feedback is positive.

From the analysis of the results it has been noticeable that not all students perform at the same level in the various assessment types undertaken (though the average shows good mapping). This indicates that it is important to assess via a variety of methods to ensure students are assessed fairly.

Removing the ability to include free-text responses within the peer-assessment process removes the need for employing the Markup utility, and hence speeds up the provision of awarding a 'mark for marking'. However, from student feedback and the tutor's analysis of the feedback, there is an indication that the inclusion of free-text responses as well as the menu-driven system encourages students to include emphasis and further subjectivity within their comments. It has also been noted that the interpretation of a comment may differ from student to student. On the menu-driven system the comment 'Overall a fair report' might be deemed to be a negative comment and this interpretation problem is currently being researched, students are now able to develop an augmented database to include their own comments as well as the standard comments. Each comment is given a weighting by the student thus producing a greater degree of subjective based feedback. The results from this research will be reported upon in the future.

One of the main advantages of this form of assessment is the elimination of plagiarism. The students are not able to copy material off the web (as they do not produce their own essays) and can not copy off their peers in the marking process as the essays are randomly selected. An interesting question arises is what procedures should be undertaken if a student in peer-marking identifies plagiarism within a student's essay, when that student may have already been awarded their qualification. From the essays produced and presentations of Post HND students, the author noted that the students lacked the level of skills of expression and referencing that would be expected of a level three student. These skills may well be collected via further experience at a higher level of study and augmented through the duration of this Post HND year of study. If the NIA students had been required to have written an essay themselves in this area of study they may well have received a lower grade. Would this have been a fair reflection on their knowledge in the subject area?

In producing the mappings in Table 5, it is noted that the linear scale 0–5 is used to represent a student's achievement in a particular item of assessment. This is based

upon expectations based upon previous student cohorts. It is questionable whether these mappings could be used across all modules, years, courses, etc. Also from the author's experience it has been noted that different ethnic backgrounds have varying expectations, e.g. a student from one culture may consider 80% poor, whilst another might considered this good.

What this study has shown is the knowledge acquisition process need not be limited to the development of an essay, but can be enhanced via peer assessment. This study highlights the need for a compensated grade. This is illustrated by the example of the original marking and commenting for one DSandEN student, who was awarded a mark for their own work of 71%, but who consistently under-marked by an average of 2% per essay. However, on looking at his commenting, noting the feedback index scales shown previously of -5 to +9, his average under-commenting was -5.16. This means that just looking at his comments would have a very negative response. If their essay was only to be judged on a numeric representation of the commenting then possibly a good essay would receive a very low mark. For these methods of peer-assessment to be utilized then it is clear that the need for the computerisation of the processes is essential for management purposes.

To summarize this study has produced positive results, both statistically and from student feedback. The results do support the fact that the 'brighter' students achieve more by utilizing their higher order skills. The method of assessment reported is capable of augmenting the assessment process, however the author would not make any claims concerning replacing the need for student essay production completely within a course of study. A final question that arises from this study is at what stage of the educational process or at what age group could this method of assessment be deemed acceptable? This is an area that will provide future work before any assumptions may be formulated.

References

- Bhalerao, A. & Ward, A. (2001) Towards electronically assisted peer assessment: a case study, *ALT-J*, 9(1), 26-37.
- Bloom, B. (1956) *A taxonomy of educational objectives: handbook of cognitive domain* (New York, McKay).
- Bostock, S. (2001) Peer assessment: principles, a case, and computer support, paper presented at the *LTSN-ICS workshop in Computer Assisted Assessment*, Warwick University, 5-6 April. Available online at: http://www.keele.ac.uk/depts/cd/Stephen_Bostock/docs/ltsnicsapril2001.pdf.
- Boud, D., Cohen, R. & Sampson, J. (1999) Peer learning and assessment, *Assessment and Evaluation in Higher Education*, 24(4), 413-426.
- Collis, B., De Boer, W. & Slotman, K. (2001) Feedback for web-based assignments, *Journal of Computer Assisted Learning*, 17, 306-313.
- Davies, P. (2000) Computerized peer assessment, *Innovations in Education and Training International*, 37(4), 346-355.
- Davies, P. (2002a) Using student reflective self-assessment for awarding degree classifications, *Innovations in Education and Teaching International*, 39(4), 307-319.
- Davies, P. (2002b) There's no confidence in multiple choice testing, in: M. Danson & C. Eabry (Eds) *Proceedings of the 6th Annual CAA Conference*, Loughborough, July, 119-130.

- Davies, P. (2003) Closing the communications loop on the computerized peer-assessment of essays, *ALT-J*, 11(1), 41–54.
- Davies, P (2003a) Peer-assessment: no marks required just feedback?—evaluating the quality of computerized peer-feedback compared with computerized peer-marking, in: J. Cook & D. McConnell (Eds) *Communities of Practice, Research Proceedings of the 10th Association for Learning Technology Conference (ALT-C 2003)*, University of Sheffield and Sheffield Hallam University, 8–10 September.
- Falchikov, N. (1995) Improving feedback to and from students, in: P. Knight (Ed.) *Assessment for learning in higher education* (London, Kogan Page).
- Falchikov, N. (1995a) Peer feedback marking: developing peer assessment, *Innovations in Education and Training International*, 32(2), 175–187.
- Falchikov, N. & Goldfinch, J. (2000) Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks, *Review of Educational Research*, 70(3), 287–322.
- Lewis, S. F. & Davies, P. (2004) The automated peer-assisted assessment of programming skills, in: P. Chalk (Ed.) *Proceedings of the 8th JAVA and The Internet in the Computing Curriculum Conference JICC8*, January.
- Lin, S. S. J., Liu, E. Z. F. & Yuan, S. M. (2001) Web-based peer assessment: feedback for students with various thinking-styles, *Journal of Computer Assisted Learning*, 17, 430–432.
- Parsons, R. (2003) Self, peer and tutor assessment of text online: design, delivery and analysis, in: J. Christie (Ed.) *Proceedings of 7th International CAA Conference*, July, 315–326.
- Race, P. (1995) The art of assessing, *The New Academic*, 4(3) and 5(1).
- Rada, R. & Hu, K. (2002) Patterns in student–student commenting, *IEEE Transactions on Education*, 45(3), 262–267.
- Robinson, J. (2002) In search of fairness: an application of multi-reviewer anonymous peer review in a large class, *Journal of Further and Higher Education*, 26(2), 183–192.
- Sitthiworachart, J. & Joy, M. (2003) Deepening computer programming skills by using Web-based peer assessment, *4th Annual LTSN-ICS Conference*, Galway.
- Stefani, L. A. J. (1994) Peer, self and tutor assessment: relative reliabilities, *Studies in Higher Education*, 19(1), 69–75.