Closing the communications loop on the computerized peer-assessment of essays

Phil Davies
School of Computing, University of Glamorgan
email: pdavies@glam.ac.uk

The use of self- and peer-assessment is not new to higher education. Traditionally its use has required the complex and time-consuming management of coursework submissions by the tutor, in an attempt to maintain validity and anonymity of the assessment process. In the last few years a number of computerized systems have been developed that are capable of automatically supporting, managing and performing the assessment process. The requirement for student anonymity and the release of the tutor from the process of marking have reduced the ability to develop the iterative process of feedback. This feedback is considered essential in supporting student learning and developing reflective practice. This paper describes the enhancement of a computerized assessment system to support anonymous computer-mediated discussion between marker and marked having previously performed peer-assessment. A detailed description is provided of the integrated assessment process, and an analysis of the use of this anonymous discussion is presented. Anonymous student feedback is presented and analyzed with respect to the perceived benefits of using the system with respect to enhancing the student learning process.

Introduction

This paper describes the use of anonymous discussion within the use of computerized peer-assessment, and reports on student feedback concerning the perceived benefits to the learning process that the use of this system has produced. The merits, and demerits, of peer-assessment have been researched and well documented over a number of years (Boud, Cohen and Sampson, 1999; Dochy, Segers and Sluijsmans, 1999; Falchikov, 1995; Stefani, 1994). Peer assessment provides a means of encouraging students to reflect on the quality of their own work whilst marking the work of their peers. It must be seen to provide a means of enhancing the learning process, not merely reducing a tutor's marking load. This paper is not an attempt to evaluate the use of peer-assessment, but is intended to identify an aspect

of this assessment method that has often been disregarded, namely the possibility of peer discussion on feedback. Discussion on feedback is normally impossible due to the very nature of anonymity that is associated with 'fair' peer-assessment. If anonymity is not present, then external factors can influence the grades/marks produced, as noted by Pond, Ul-Haq and Wade (1995), which 'can be adversely influenced by group collusion, friendship (or hostility) and can allow, and even encourage, domination by individuals'.

The use of computerized systems to support peer-assessment is still in its relative infancy (Davies, 2000; Bhalerao and Ward, 2001; Dennis and Dempster, 2000; Bostock, 2001; Lin, Liu and Yuan, 2001). The de-personalization caused by the use of such systems has obviously reduced the tutor-student contact with response to feedback. It has been noted that learning depends on feedback to the learner (Falchikov, 1995). However, it has also been noted that opportunities for traditional tutor feedback are becoming fewer, it is more and more difficult to maintain levels and quality of feedback (Falchikov, 1995) and detailed feedback on students' work has been substantially eroded (Race, 1995).

Constructive feedback is of utmost importance in enhancing the learning process, and often the student receives a grade or mark which is little use in directly helping a student in the learning process (Miller, Imrie and Cox, 1998), but in order for it to be effective and fair, it needs to be uniform, whereas, as Falchikov (1995) comments: 'tutors may not feel a need to produce detailed formative feedback for students whose grades are satisfactory or of a high standard.'

As noted by Higgins, Hartley and Skelton (2002), whether a tutor provides such feedback or not is no longer an option: it is considered to be essential and often part of the student charter, higher education as a 'service', and it is felt that feedback constitutes part of that service.

It should be noted that, in any argument concerning whether a student is capable of providing constructive comments to their peers, not all tutor comments are necessarily helpful, as reported by Miller *et al.* (1998): 'I am returning this otherwise good typing paper to you, because someone has printed gibberish all over it, and put your name at the top of it' (attributed to a professor from Ohio University).

This form of destructive comment does not leave the student in any doubt as to the quality of their work, but gives little guidance as to how they can improve their work in future. This must be balanced against the opposite feedback of superfluous glowing praise, with no guidance on future improvements that will enable him or her to work on points in need of attention rather than to keep practising points of strength (Brown and Knight, 1994). It may also be proposed that a tutor will provide comments to placate a student, rather than produce controversial comments that could result in the need for further time-consuming consultations.

As previously stated, peer feedback is normally anonymous (Pond *et al.*, 1995), and because of this there is a problem of a student being able to clarify any issues concerning the feedback they receive for their work. For the feedback to be effective, there is a need for clarity, hence dialogue is therefore not simply desirable but, arguably, essential (Brown and Knight, 1994). What is needed is an ability to be able to integrate the feedback process with an enriched learning process for the students (Collis, De Boer and Slotman, 2001).

The enhancement of the Computerized Assessment by Peers (CAP) system (Davies, 2000), as described in this paper, will now permit the marker and marked to consult anonymously concerning the final mark or comments to be associated with an essay.

Methods

One of the main issues to be addressed in this study is whether the students will make use of the discussion element of the CAP system, and whether it will have any perceived impact upon the learning process. In past uses of the CAP system, few students have contacted the tutor to complain about marks or requested further clarification. The issue of whether there have been benefits to the learning process due to the introduction of the discussion element of the assessment process will be difficult to quantify. It will need to be based upon student feedback concerning their personal perceived benefits.

A further pointer as to whether the students have played an active role in the discussion element of the assessment process will be the number of iterations of discussion that take place between the marker and marked.

In order to obtain student feedback on the use of the CAP system, a detailed anonymous questionnaire (Appendix 1) was presented after completion of the assessment process. To maintain ethical rights, the students were requested to agree or otherwise to the use of their quotes in the development of this study via a tick box at the end of the questionnaire.

Assessment process

Assessment should be another episode in learning and it can be argued that a shared understanding of the learning task and the assessment criteria are keys to this ideal. (Stefani, 1998)

The assessment reported in this paper makes use of both self- and peer-assessment of essays within a postgraduate class in Computer Technology at the University of Glamorgan during the Autumn Term, 2001. The CAP system has been used within the university for a number of years to support the peer-assessment of student essays, and to provide a core element to support the ideal of learning through assessment.

In order to provide a framework for the marking of the essays, guidelines were given suggesting what was expected of a first-class degree, an upper second and so on. These guidelines were deliberately left open-ended in order to encourage student subjectivity in the marking process.

The overall assessment process described in this study is worth 35 per cent of the total overall module mark, and possesses the following stages:

1. The students are required to perform a detailed literature review in a specified area of Computer Technology. They are encouraged to use the Web as a source of up-to-date information.

2. The students develop and submit an essay in RTF format via Blackboard's digital drop-box facility.

3. This essay is then self-assessed making use of the CAP system. Detailed comments must be included to justify their allocated marks (Figure 1).

4. Having completed the self-assessment stage, the students must then peer-assess at least eight of their colleagues' essays (automatically randomly selected), again making use of the CAP system (Figure 1). The decision that eight essays needed to be marked was decided upon in order to permit the students to view as many different attempts (varying from excellent to poor in both content and style), without overloading this aspect of the assessment process.
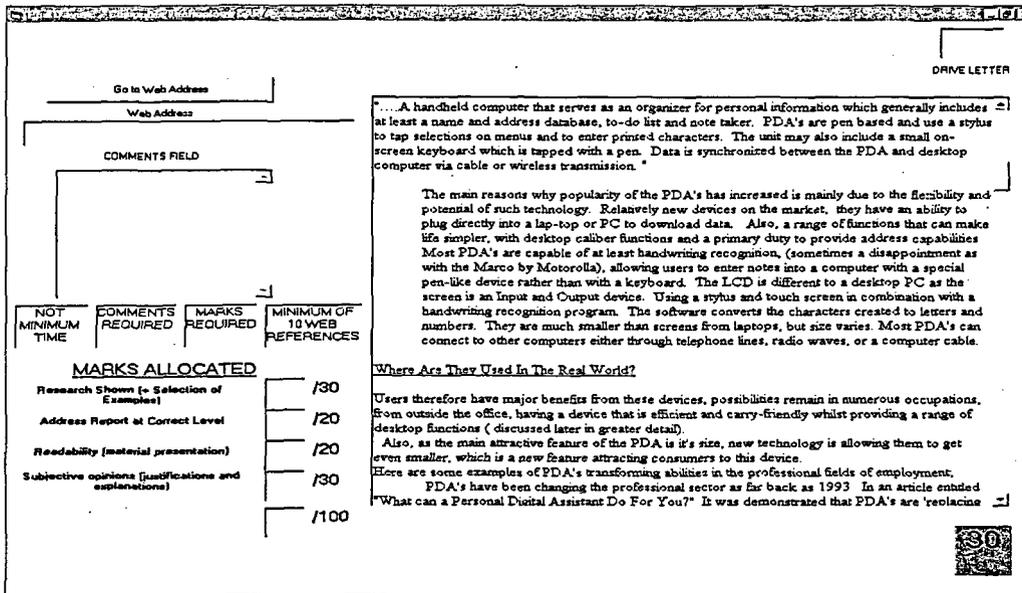
Figure 1: CAP client marking

5. Once this marking stage is completed, the CAP's Feedback Application is used to view the peer comments (Figure 2). It should be noted that the students are only able to view the comments of their peers (not the marks). The decision not to show grades initially was taken to avoid a student merely judging a marking by the grade, not the comments, as the following quotes illustrate:

> simply pay attention to the grade, and seek feedback only when it is perceived to provide 'correct answers' to commit to memory (and only when their grade expectation has not been met). (Higgins *et al.*, 2002)

> pay most attention to their scores or grades, when they get back marked work, and often are quite blind to valuable feedback which may accompany their returned work. (Race, 1999)

6. If the essay owner agrees with the marker's comments they then click 'Accept Result' (Figure 2), and the mark for this particular marking is displayed, and no further discussion can occur with respect to this marking.

7. If the student does not agree with the comments, or requires further clarification of a marker's comments, then they may click 'Reject Comments' (Figure 2). The student
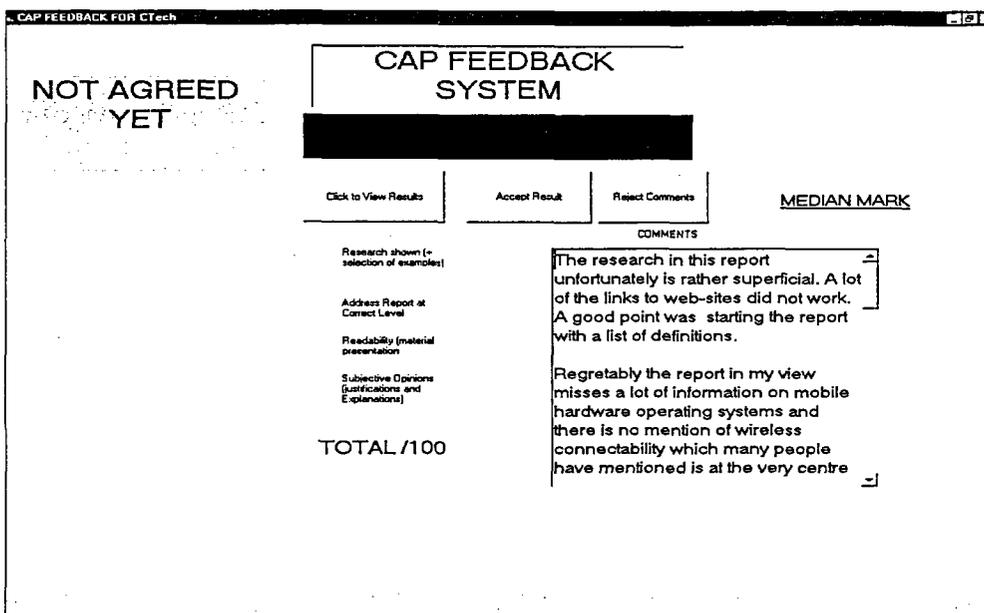
Figure2: CAP feedback to file owner
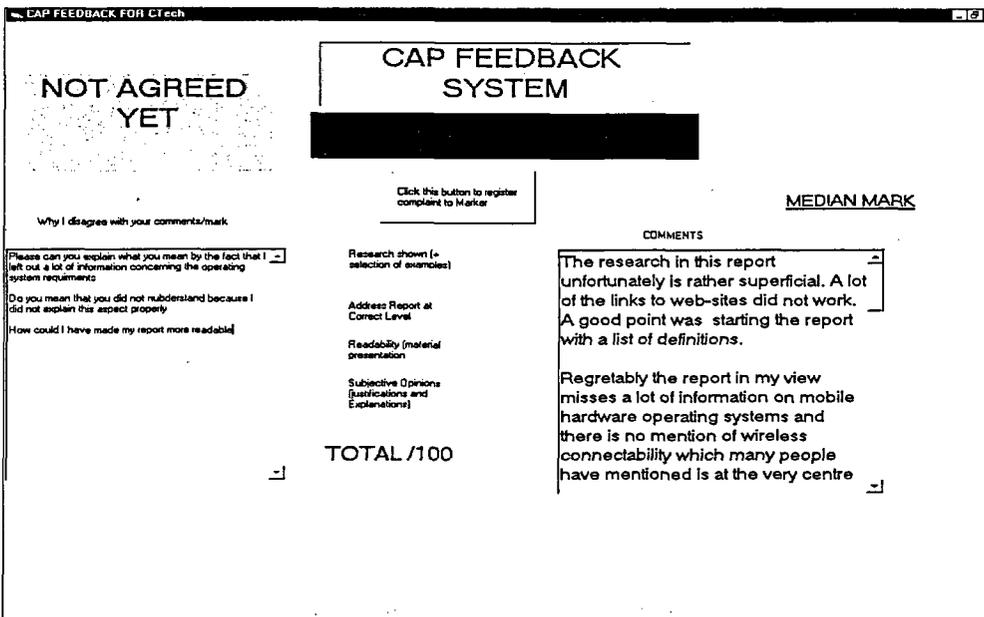
Figure 3: CAP file owner complaint or query

will then need to fill in a request box concerning their query (Figure 3), and this will be passed back to the original marker for them to look at whilst re-marking.

The original marker is now automatically emailed by the CAP system requesting them to re-mark the essay.

8. The marker will run the re-marking application (Figure 4), and this will facilitate the viewing of the essay owner's request, the original essay, comments and marks. The marker is permitted to amend the marks and/or add to the comments based upon the owner's request. Once the marker is content that he/she has satisfied the complaint or query, the revised marks and comments are resubmitted to the system. The owner is automatically emailed, and he/she may review the new comments/answers to their query.
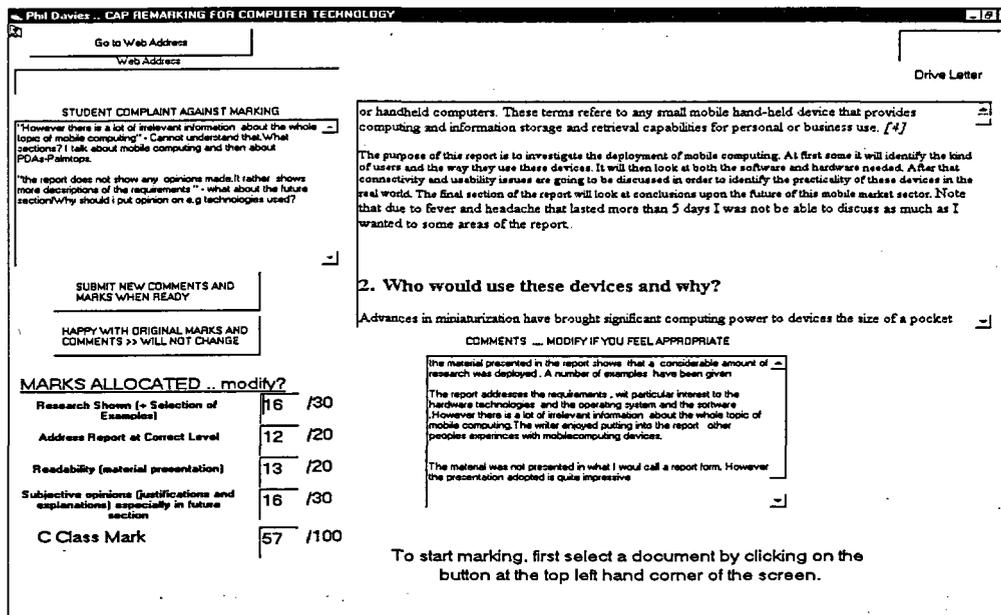
Figure 4: CAP re-marking

This process may continue through numerous iterations until the owner is satisfied and agrees to the comments.

9. A marker may be of the opinion that the original mark/comments are fair. If it is a query from the owner then a reply should occur. If, however, the marker is just asked to re-mark, he/she can simply click on 'Happy with Original', and an automatic email is generated to the owner informing them of the marker's decision.

On viewing the feedback on a document that the 'marker will not change', the mark is automatically displayed along with the comments. If no agreement has been reached between the marker and marked, the tutor will review the marking and grades. The marker will be judged during the 'Marks for Marking' aspect of the assessment as to whether their comments and replies were acceptable or not.

This marking/re-marking process is considered as important as the development of the essay with respect to the assessment process. Because of its importance half of the total marks for this assessment are allocated for this marking process.

A process of compensation is now automatically performed by the system (Davies, 2002), taking into account high and low average markings per marker. This is then used to produce a compensated mean that is used as the grade for the essay. The tutor input to the assessment marking is thus limited to the allocation of the marks for marking.

Marks for marking

Figure 5: CAP marks for marking

In past studies on peer-assessment, if there has been no tangible reward to the student for marking, the quality of the feedback has often been questionable and the result of a highly variable quality (Robinson, 2002). Due to this fact, there is a significant reward to the student for performing the marking in a constructive and formative manner.

The CAP's 'Marks for Marking' application has been extended (Figure 5) to display the communication between the marker and marked. Marks are allocated for:

• Consistency shown in performing the marking process (using the difference between

self-assessed and compensated marks as the Expected Difference Standard). This should be as close to the difference per marking as is possible for each marked essay.

- Quality of the original comments provided with respect to the level of the constructive input, supported suggestions, and analytical skills shown in the subject area.

- Quality of the marker's responses to essay owner's requests for clarification and remarking.

The overall assessment mark is then presented to the student in spreadsheet form, indicating where they have lost or gained marks for the essay and marking.

Results analysis

In this study, 48 out of the 55 students were actively involved in the generation of auto-emails. At the completion of the assessment exercise, 46 out of the 55 students were requested to have performed the re-marking process.

In total there were 244 auto-emails generated; these included requests for re-marking, confirming re-marking had taken place, and also refusing to change comments.

After the initial peer markings there were 440 markings in total for the 55 files. Out of these, 70 per cent were agreed on the first pass of the feedback between marker and marked. Having performed the re-marking process for discussion, the number agreed increased to 80 per cent.

Out of the 20 per cent of the files that remained without agreement, 8 per cent remained in dispute between the marker and marked, 8 per cent were left awaiting re-marking, and 4 per cent were left not agreed due to the file owner not taking an active part in the feedback process.

Out of the 128 requests for re-marking, 15 required a second iteration of feedback, and 3 went to a third iteration of feedback. In total, 75 per cent of these requests for re-marking produced a decision of either agree or no change.

Out of the re-markings that actually took place, 35 per cent of the markings had the original marks changed. This mark change ranged from +15 per cent to –28 per cent. The average change being +3 per cent. There were only 8 instances where the final mark was modified by 10 per cent or more. The –28 per cent change of mark showed an original mark of 88 per cent changed to 60 per cent. However, the comment change did not really justify such a significant alteration. This was noted when awarding marks for marking to the marking student.

The average marks and standard deviations produced for the overall assessment process are shown below:

Self-assessment	69 per cent (Mean)	9.0 (Standard Deviation)
Peer-assessment	63 per cent	8.0
Compensated	62 per cent	7.9

The mean difference of 6 per cent between the self- and peer-assessment corresponds with past uses of this system (Davies, 2002).

Examples of the types of requests that were produced by the owners of the essays to the marker for the re-marking process are shown below:

Please clarify what you mean by . . .

Can you suggest where I can find the information that you suggest in your comments?

I did explain this in my essay at . . .

You do not criticise or praise my work, what do you think of it?

How do you suggest I reference my work so you can find the source of my information?

I'm not English, so don't criticise my grammar.

Analysis of student feedback

This study was the first use of the CAP system at postgraduate level of study. It has been previously used at various levels of the undergraduate programme of study (Davies, 2000; Davies, 2001). The student comments produced within this study provided very similar general feedback compared with these previous studies. The example comments provided below are taken from the student questionnaire (Appendix 1) as returned by 95 per cent of the students who undertook the module's assessment process.

Concentrating on its use in this study, over 86 per cent of the students felt that the use of the assessment process had aided their learning in the module. Some of these supporting comments were:

Helps future work.

Made me realise my self-assessment was too generous.

A selection of comments from the many students who were positive towards the use of the system included:

Become more self-critical and analytical.

Helped me enormously in being more objective about my own work.

Another helpful learning process.

It really helped me being able to do all my marking, remarking and commenting at home over the internet.

Only when I read their comments did I realize how much I had left out.

Insight into what a lecturer looks for.

There were very few negative comments of any significance. The main comment of any real negativity was with respect to the amount of time for the overall assessment process: 'It was very time consuming.' The students reported back that on average the time taken to mark each essay was approximately 40 minutes. This required a considerable effort on the student's behalf, that is, 320 minutes of marking.

When asked whether they had benefited from reading the comments of their peers, over 83 per cent of the students answered in the affirmative. This should be weighed against one

student's comment: 'The best comment I received ended up as my lowest mark.' This variability of views and comments has been noted in past usage of the CAP system, but the added discussion offered by this system appears from the majority of student comments to have made this less of a problem.

Some general comments with respect to the feedback were:

Some were very harsh, but needed to be for me to pay any attention.

Some students really were very constructive.

Initially hated the thought of doing it, but now accept that I've learnt so much from doing it.

When asked whether overall the students were surprised at the general standard of their peers' work, there was quite a number who commented upon the wide range of essays that had been submitted. An example of this was: 'I could have done better when I was in junior school.'

It tended not to be the content of the essays, but the varying styles and referencing that brought forth the greatest number of comments. As many as 62 per cent of the students commented upon how much they had benefited from viewing the links provided by their peers.

A total of 88 per cent of the students expressed that they had found the Web to be an excellent resource for providing up-to-date information in the area of computing. One student did however pose the question: 'Should we really be using the web at MSc level?'

With the study being aimed at more mature students than in the past, the comments concerning the need for anonymity were of particular interest. Over 70 per cent of the students felt that it was essential that anonymity be maintained.

Anonymity helped me feel more relaxed in making my judgements.

It would have undoubtedly reduced my capacity to be wholly objective by introducing inter-personal aspects.

I might well have been intimidated when discussing comments.

If I'd known them I would have probably have given a higher mark.

I would have ignored some classmates' comments due to my low opinion of them as individuals.

This was the first use of the CAP system including the facility for discussion and re-marking. When asked had they benefited from this additional process, 56 per cent of the students answered in the positive, 24 per cent in the negative, and 20 per cent undecided.

Their explanations on questioning made their comments much clearer.

I identified one peer's error in marking, they apologised and corrected their comments.

A few students suggested that they would have liked more anonymous discussion with their markers before they had decided they would not change their mark. Due to the method used of not showing the marks until the comments were agreed, one student complained: 'I

disagreed with him, I discussed it, he changed his comments, and I still didn't like his mark.'

One student highlighted a problem: 'They said that they'd changed their mark, but I'm not sure they really did it.' This was investigated by the tutor and was found to be true in the instance in question. There were some very complimentary comments:

Very good and revolutionary.

Interesting to say the least, I actually enjoyed being assessed.

Some students suggested that improvements could be incorporated:

The lecturer is seen to intervene on disagreements.

Students should have marks taken off for not remarking.

This was the case with respect to the marks for marking process, but was not spelled out to the students in advance: 'better if we knew the mark before agreeing.'

This brings into question the students agreeing comments just because they have been awarded a good initial mark from a student (even though this mark could be reduced via the compensation process).

There were two student comments that were found to be of great interest. One student gave himself a mark of 88 per cent for self-assessment. Prior to the peer marking process he requested that he be allowed to lower this mark to 64 per cent. When asked why, he replied: 'It was only when I thought about how I was going to assess someone else's work did I realise that I had not been critical enough in my own essay.' This was a very positive statement with respect to the benefits of self- followed by peer-assessment. It highlights a problem in assessment where students may be unaware of what is expected of them: 'I haven't a clue what I'm assessed on' (Higgins et al., 2002). The second comment came from general discussion with a student concerning the wide range of marks and associated comments allocated by different students: 'Perhaps the range of self-assessment marks has got something to do with the fact that they may have achieved their first degree overseas where marks and grades are different.'

This is a point to be noted particularly with respect to the use of self- and peer-assessment systems, especially in the area of postgraduate study.

Conclusions

This study has overcome the previous problem of limited, if any, feedback via computerized peer-assessment. To placate some negative student feedback, lecturer intervention to mediate on markings left in conflict (8 per cent of the total markings) could be introduced for future uses of the system. A further stage of acceptance of marks following on from comment acceptance has also been suggested. However, this would substantially lengthen the time required for the assessment process.

Statistically, the acceptability of comments in the student essays has increased by 10 per cent (70 per cent to 80 per cent) following the discussion process. In traditional peer-assessment this would not have been feasible if anonymity were to be maintained.

With respect to the perceived learning benefits of utilizing the CAP system, 86 per cent of the students felt that it had improved their understanding of the subject area. Over half of the students (56 per cent) felt that the use of the discussion element of the system had proved a major positive factor in the peer-assessment process. The need for anonymity was confirmed by the fact that 70 per cent of the students when questioned felt that it was essential to maintain anonymity for them to be able to perform the marking process in an objective manner.

Previous negative comments concerning the lack of constructive comments by the marker have been removed by the addition of the discussion aspect of the CAP system. The uptake of student use of this discussion facility has been mixed. Out of the 8 per cent of markings left in conflict, it would have been more pleasing if more iterations of discussion had taken place before the marker refused to change their comments/marks. In future uses of the system, at least one iteration of discussion will have to be completed before the marker will be allowed to refuse to change.

With respect to judging the effectiveness of the discussion element of the peer-assessment, it would have been pleasing to have seen further development in the iterative process of marker/marked questioning. Statistically this has not been the case, with only a small proportion (6 per cent) of the re-markings extending beyond a request-reply stage. Conversely this could also be viewed as a positive reflection of the quality of the comments, in that the need for further clarification was limited.

The comment that appeared consistently in the student feedback was that the marking process had facilitated greater personal reflection. It provided a means of developing the students' objectivity for future essays, and helped to develop an understanding of how to direct an essay to meet the needs of the potential marker.

Overall, the addition of discussion to the peer-assessment process has removed a number of past criticisms of its use as part of the formative and summative assessment process. The student feedback has been extremely positive, and the perceived student benefits with respect to the learning process have been noted.

Acknowledgements

References

Bhalerao, A. and Ward, A. (2001), 'Towards electronically assisted peer assessment: a case study', *ALT-J*, 9 (1), 26–37.

Bostock, S (2001), 'Peer assessment: principles, a case, and computer support', paper at the LTSN-ICS workshop in Computer Assisted Assessment, Warwick University 5–6 April 2001, *http://www.keele.ac.uk/depts/cs/Stephen_Bostock/docs/ltsnicsapril2001.pdf*

Boud, D., Cohen, R. and Sampson, J. (1999), 'Peer learning and assessment', *Assessment and Evaluation in Higher Education*, 24 (4), 413–26.

Brown, S. and Knight, P. (eds.) (1994), *Assessing Learners in Higher Education*, London: Kogan Page.

Collis, B., De Boer, W. and Slotman, K. (2001), 'Feedback for web-based assignments', *Journal of Computer Assisted Learning*, 17 (3), 306–13.

Davies, P. (2000), 'Computerized peer-assessment', *Innovations in Education and Training International*, 37 (4), 346–55.

Davies, P. (2001), 'Must be more than Multiple Choice Tests for it to be academically credible?', in M. Danson and C. Eabry (eds.), *Proceedings of 5th International CAA Conference Loughborough 2001*, 145–9.

Davies, P. (2002), 'Using student reflective self-assessment for the awarding of degree classifications', *Innovations in Education and Teaching International*, 39 (4), 307–19.

Dennis, H. and Dempster, J. (2000), 'Enhancing critical debate through web-based discussion and publishing', *Interaction* 4 (2), *http://www.warwick.ac.uk/ets/interactions/vol4no2/dennis.htm.*

Dochy, F., Segers, M. and Sluijsmans, D. (1999), 'The use of self-, peer- and co-assessment in higher education: a review', *Studies in Higher Education*, 24 (3), 331–50.

Falchikov, N. (1995), 'Improving feedback to and from students', in P. Knight (ed.), *Assessment for Learning in Higher Education*, London: Kogan Page.

Higgins, R., Hartley, P., Skelton, A. (2002), 'The conscientious consumer: reconsidering the role of assessment feedback in student learning', *Studies in Higher Education*, 27 (1), 53–64.

Lin, S. S. J., Liu, E. Z. F. and Yuan, S. M. (2001), 'Web-based peer assessment: feedback for students with various thinking-styles', *Journal of Computer Assisted Learning*, 17 (4), 420–32.

Miller, A. H., Imrie, B. M., Cox, K. (eds.) (1998), *Student Assessment in Higher Education*, London: Kogan Page.

Pond, K., Ul-Haq, R., Wade, W. (1995), 'Peer review: a precursor to peer assessment', *Innovations in Education and Training International*, 32 (4), 314–23.

Race, P. (1995), 'The art of assessing', *The New Academic*, 4 (3), 3–5, and 5 (1), 3–6.

Race, P. (1999), 'Why assess innovatively?', in S. Brown and A. Glasner (eds.), *Assessment Matters in Higher Education*, Buckingham: SRHE and OU Press, 64.

Robinson, J. M. (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), 183–92.

Stefani, L. A. J. (1994), 'Peer, self and tutor assessment: relative reliabilities', *Studies in Higher Education*, 19 (1), 69–75.

Stefani, L. A. J. (1998), 'Assessment in partnership with learners', *Assessment and Evaluation in Higher Education*, 23 (4), 339–50.

Appendix I

List of questions included within student evaluation questionnaire:

- Did you find that performing the CAP marking has aided your learning in the subject area?

- Can you identify any advantages and/or disadvantages that you noted in using the CAP system?

- On average how long did you take to mark each report?

- Did you find that reading the comments of your peers in the feedback was of personal benefit?

- What were your thoughts on the general standard of your peers' work?

- Did you find it of benefit in following other student's Web links within the marking process?

- Did you find the Web a good source of information for this subject area?

- If the marking and discussion processes were non-anonymous, would it have had any effect upon the way that you marked and/or provided feedback?

- Did you find the opportunity to discuss the comments of your peers of benefit to you?

- Have you any suggestions on how the CAP marking process can be improved or modified?