REVIEW IN COMPUTERIZED PEER-ASSESSMENT

WILL IT HAVE AN EFFECT ON STUDENT MARKING CONSISTENCY?

Phil Davies

Review in Computerized Peer-Assessment. Will It Have an Effect on Student Marking Consistency?

Phil Davies
Department of Computing & Mathematical Science
Faculty of Advanced Technology, University of Glamorgan, South Wales email: pdavies@glam.ac.uk

Abstract

This short paper details work in progress that identifies an extension to the CAP Peer Assessment System that permits students to review the marks and comments of essays they've marked, having been allowed to view the comments of others who have also marked these particular essays.

The development of a compensation process that takes into account high and low markers is also discussed and whether the introduction of this review stage negates the necessity for this compensation process in the overall peer-assessment process.

Also presented is a review of the system in automatically allocating a 'mark for marking' that relates directly to the quality of the marker's work in both supplying marks and comments that match the quality of the marked essay.

Background

Over the past seven years the CAP (Computerized Assessment by Peers) has been used as a tool to support the peer-assessment of both essays and multimedia presentations. This tool over this period of time has evolved from a basic marking tool that replicates traditional peer-assessment (Davies, 2000), to include anonymous communications between marker and marked (Davies, 2003) and the inclusion of menu driven comments and weightings to take into account subjectivity of the marker and automatic creation of a mark for marking (Davies, 2005). Throughout the various stages of development of this system the importance of feedback and quality of comments (Davies, 2004 & 2006) has been emphasised as being of great value to the owner of the essay. The rewarding of students for performing the marking and commenting in a qualitative manner has become one that has necessitated the introduction of a compensation process that automatically adjusts the marker's marks prior to the production of a compensated peer mark that acts as the final grade for a particular essay. Students have commented in the past that they find they have two major concerns in performing the peer-marking process:

a) that they maintain consistency throughout the peer-marking process
b) they are able to perform the 'task' well compared with other students in the group.

Following a successful internal grant application made to the University of Glamorgan's Teaching and Learning Committee, the opportunity arose to further develop the functionality of the CAP system to permit the students to amend their marks and/or comments for a particular essay having been permitted to review their previous marking of the essay. During this process they were also permitted to view the comments of their peers who had also marked this particular essay.

The new functionality of the system was then included in a trial study undertaken with a postgraduate cohort on a module teaching E-Learning within the academic year 2006-7. This paper describes the assessment process undertaken by these students and highlights the effect that this new functionality has had upon the peer-marking process.

Statistics are presented that show the increased time scales required for this aspect of the peer-assessment and whether the introduction of this review stage has had any subsequent effect on the quality and consistency of the peer-marking prior to the owner of an essay viewing their grades.

Discussion is also included that highlights the difficulty in providing an automatic reward for the peer-marking process undertaken by the students that maps to their quality of grading and commenting.

Assessment Description

As part of their coursework assessment within a module teaching E-Learning, a postgraduate cohort of 13 students were requested to produce an essay in the form of a fully referenced RTF document that explained how to develop 'a distance learning Powerpoint presentation to teach 10 year olds something of a technical nature (in this particular assessment they'd been previously introduced to the Golden Ratio Phi as an example) but they were advised that this aspect of the assignment should not be subject specific. This report was to be addressed at the level of their peers and it was suggested that it was to be a maximum of three pages plus references. It was also requested that the main source of referencing be off the web (however some books & journals were to be expected). The reason for this being that in the peer-marking timescale permitted it would be difficult for a marker to be able to find book and journal references but as the CAP system supports an embedded web browser it would be easy for them the judge the relevant research undertaken by the essay developer. The students were given two weeks to research, develop and submit this essay.

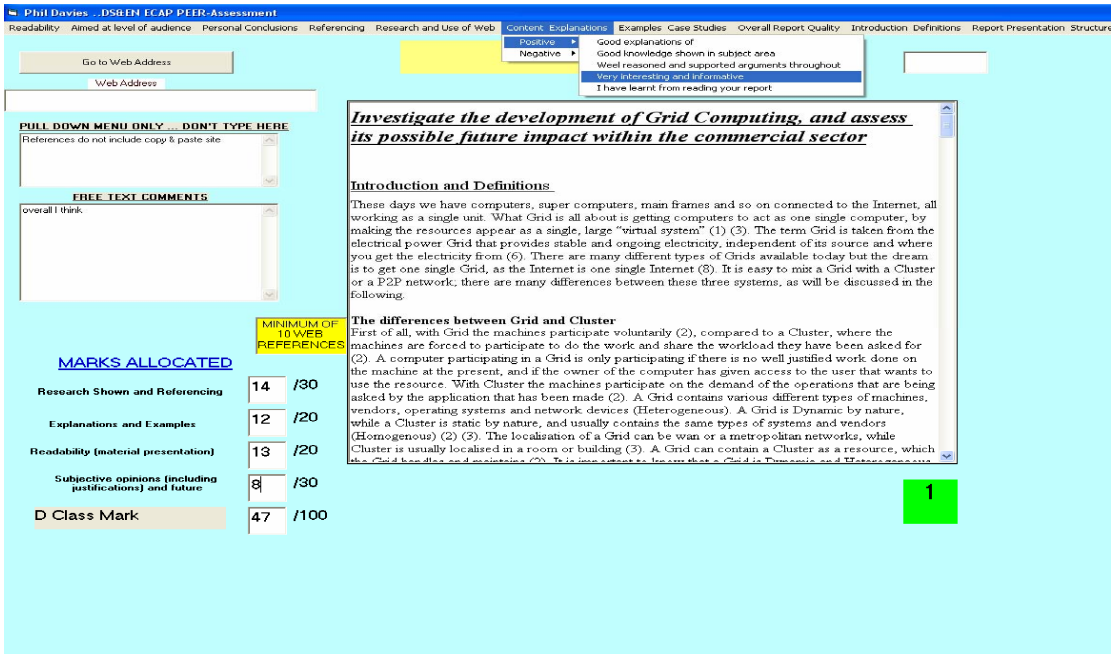
Having performed this aspect of the assignment they were then expected to peer-mark and -comment at least six of their peers' work making use of the CAP system. The comments bank and criteria they used to assess the essays

will be explained later in this paper. Prior to the students undertaking the peer-assessment aspect of this assignment they were asked to use the marking system to self-assess their own work. This is an aspect of assessment that students in the past have found to be extremely difficult. The mark generated by this self-assessment process is not necessarily of great importance with regard to the outcomes of this assignment, however by performing this aspect of assessment it has been reported in the past that it has provided a means of the students

a) getting used to the computerized assessment system
b) having a way of creating a standard for themselves that they can use throughout the peer-marking process

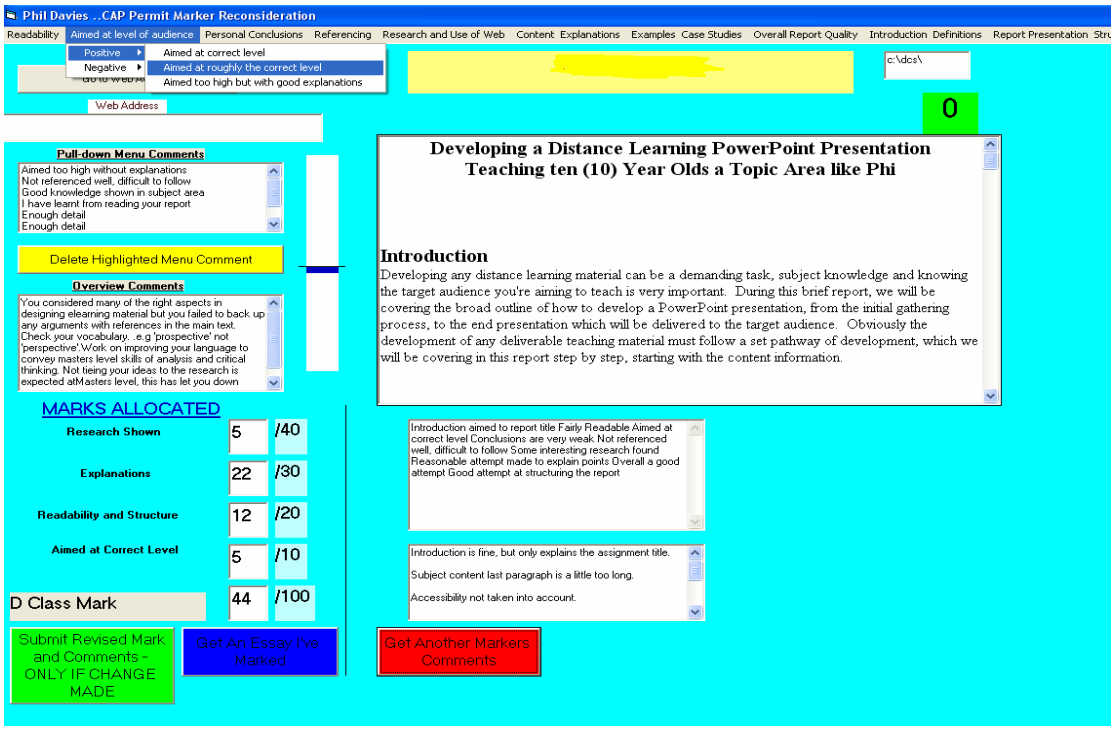
The students were then given a week to perform the peer-marking process making use of the CAP marking system (Figure 1).

Figure One



Having completed the peer-marking process the students were then given a week to make use of the new review functionality added to the CAP system (Figure 2) which permitted them to view the comments of their peers concerning essays that they had previously marked. This paper reports upon the effect that this new review aspect has had upon the peer-marking process.

Figure Two



Subsequently the students were permitted to view the marks and comments of their peers with regard to their own submitted essays. They were allowed to view the median derived peer mark for their essay not the compensated peer-mark that would represent the final grade they were to be awarded for their essay.

In addition to this grade for their essay they were allocated a mark for the consistency shown in the peer-marking process that they had performed.

On completion of the assignment they were provided with a questionnaire requesting them to comment on how they had found the overall assessment process.

CAP Application – Setting the Weighted Comments Bank

Prior to the self- and peer-marking of the assignment, the students were requested to develop an appropriate bank of comments that they could use within the ten categories used within the CAP menu driven marking system. Prior to the assessment being undertaken the students were offered the opportunity of replacing some of these ten categories and also to suggest suitable marking criteria for this particular assignment. Through discussion it was decided to leave the commenting categories as in the past, namely:

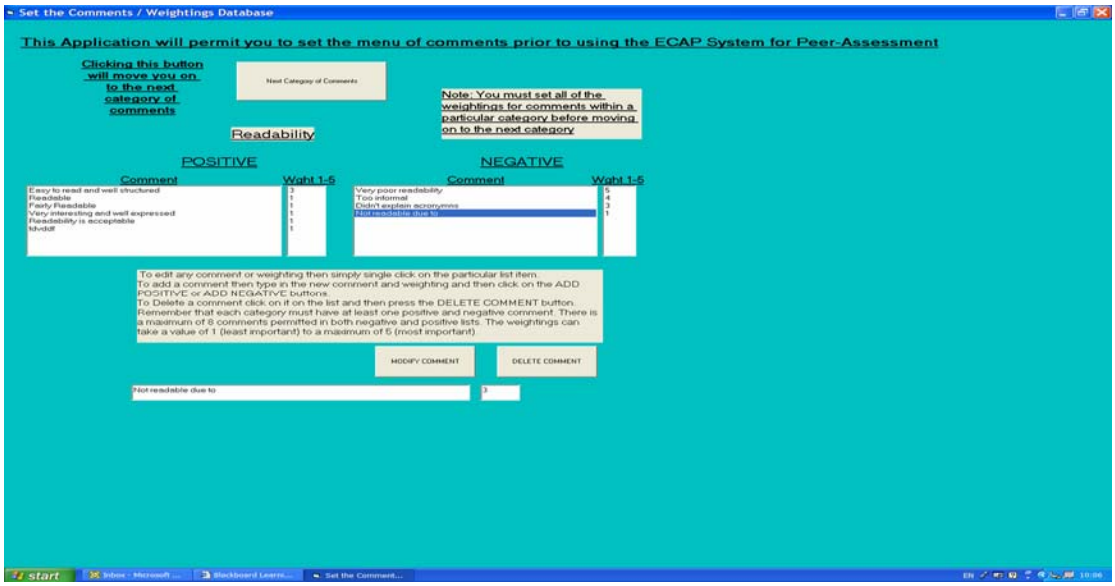
Readability, Aimed at correct level, Personal conclusions, Referencing, Research & use of web, Content & explanations, Examples & case studies, Overall report quality, Introduction & definitions and Report presentation & structure.

The marking criteria categories were:

Research Shown 40%, Explanations 30%, Readability & Structure 20% and Aimed at Correct Level 10%.

The students then made use of the Comments and Weightings setting application (Figure 3) to set comments that they felt suitable for their marking and including weightings per comment to include subjective importance for their commenting. This is described in more detail in Davies, 2005.

Figure Three



Results

13 students undertook the assessment process however one of these did not complete the peer-marking process as requested. The result of this student has been included as the essay was peer-marked.

The overall compensated peer-mark generated for the essays was 60% with a standard deviation of 11.59. In order to generate this compensated average peer-mark for an essay, the possibility that a marking student is a ‘hard’ or ‘easy’ marker (often mapping to personal expectations) has to be taken into account. It would be unfair (unfortunate) from a student’s perspective were they to be peer-marked by six hard markers compared to another student who was marked by six easy markers. In order to provide some form of compensation process, each marker has to be judged with regard to their average over- or under-marking methods. Each essay therefore needs to have a provisional average grade produced for it (the median is deemed to be a fairer reflection than the mean). Having created this, each marker’s mark is compared against the average mark for the essay they’ve marked and an over- or under- average ‘mark difference’ is created. The essays now marked by this student are amended by this mark difference and a compensated peer mark is generated for each essay. Therefore the final peer-mark produced for

an essay is compensated taking into account the 'bias' shown by a marker. In the past uses of the CAP system the use of this compensation process has not really had a major influence upon the final grade produced, but certainly does allay the fears of students with regard to them being 'fairly' graded for their essays and not being disaffected by particular markers.

In past uses of the CAP system particular emphasis has been placed on the required quality of the comments produced mapping to the actual marks presented. Table 1 below shows the correlation between the compensated peer marks and the average feedback indexes for these essays (the quantified value taken into account the menu driven positive and negative comments):

Table One

+7	**+6**	**+5**	**+4**	**+3**	**+2**	**+1**	**+0**	**-0**	**-1**	**-2**	**-3**	**-4**
81	68	61	72		53		60	52		43		42
72		65						51				
								51				

As in past uses of the CAP system the above results on the whole indicate a very positive mapping of the comments received for an essay to the actual mark attained.

What this study has tried to ascertain is whether the students once offered an opportunity to review/modify their initial mark actually will do so. The preliminary analysis of this work indicates that out of a total number of 76 markings that took place there 41're-markings' where either the menu driven comments and/or marks were changed. At this early stage of the analysis it is difficult to make any assertions as to in which way these re-markings have affected the overall peer-assessment results. It is possible that some students clicked on the 'submit a modified marking' button without actually performing a change?

The average time that a student took to mark an essay was 42 minutes (however this is not an exact timing that correlates to actual effort). It is interesting to note that the range of times included within this process was from 3-72 minutes. The students made good use of their menu comments with on average 16 comments being provided per marking.

Within the 41 're-markings' 26 of these actually resulted in a change of the original mark produced. The actual mark changes are detailed below:

+1, +9, +1, +2, +8, +6, +18 (71->89), +7, +6

-1, -2, -2, -8, -3, -5, -2, -1, -4, -6, -5, -7, -6, -3, -7, -7, -2, -5

Out of the 13 students involved in the study, 8 of these students made some form of amendments to their original markings, with 2 of these students actually 'modifying' all of their markings.

The conference presentation will provide a more detailed analysis of the findings of this study with regard to the re-marking process and whether by including this 'review' stage it has had any effect upon the actual final peer-marks that were produced.

A further aspect of this study is to attempt to automatically reward the students for the quality of their marking and commenting. A mark consistency figure has been generated that indicates the consistency shown by the marker (further explanation of how this is derived can be found in Davies, 2004).

It was decided in order to compare 'like with like', to map the consistency marks against the actual final compensated peer generated marks produced for the essays. In this way the 'range of abilities' of the students was used as a boundary to the percentage grade awarded.

For this group the average essay grade produced was 60% with a range of 81% to 42%. Thus the percentage points above the average being 21 and below being 18.

With regard to the mark consistencies produced the average being 4.87, with a range of 2.31 to 10.78 (keeping in mind that a low score is good whilst a high score is poor with regard to mark consistency). The resultant point range of a 'good' student below the average being 2.56 and that of a 'poor' student above the average being 5.91.

Therefore, mapping a good student's marking consistency to a good essay results in 21/2.56 = 8.2% for every mark consistency point below the average to be added to the essay average mark of 60%.

Similarly a poor student's marking consistency to a poor essay results in 18/5.91 = 3.05% for every mark consistency point above the average to be taken from the average essay mark of 60%

e.g. suppose a student has a mark consistency grade of 5.9. This is above the average mark consistency grade of 4.87, therefore it indicates a below average marking performance. To work out the percentage grade for this marking would result from an essay average (60%) – difference between the student's mark consistency (5.9) minus the average mark consistency figure (4.87) i.e. 1.03. This figure is then multiplied by the weighting for a poor result (i.e. 3.05%). Therefore the mark awarded to this student being 60% - (1.03*3.05)% = 60 – 3.14 = 56.86%.

This method is obviously 'raw' and illustrates the difficulty in mapping an actual percentage grade to 'reward' the marking process in a qualitative manner.

Further analysis will take place and be reported upon concerning the feedback consistency and the effect that the re-marking has upon these consistencies.

From initial student comments this form of assessment has been met with general approval. A full analysis of the questionnaire results will be included at the presentation.

Conclusions

At this early stage of the data analysis no major conclusions can be made as to the effect that the review stage has had upon the peer-marking process. The presentation will attempt to identify any significant trends, however these will be limited due to the small sample used within this study. Initially the results appear to indicate that the review stage does not have a major effect upon the peer-marks produced, thus the need for the compensation process remains.

At the onset of this study the author had mixed feelings concerning the possible outcomes of the introduction of the review stage. In past uses of the CAP marking system students have requested that they would have liked to have had the opportunity to re-assess their original markings, however the inclusion of this extra stage has been avoided in the past as it was felt that this may result in the students not setting their criteria for peer-marking clearly prior to performing marking due to the fact that they'd have a 'second chance'. The preliminary results appear to indicate that the students even though they knew that this second chance would be available took every care in their original marking (mainly due to the fact that they noted that they would be allocated a grade for performing this marking in a qualitative manner). The mark changes were relatively minor and appear to have little bearing on the overall results produced.

This addition to the functionality of the system has again met with the general approval of the students in that it has provided then with an opportunity to get a realistic appraisal of what their peer-assessment of the essays was in comparison with others within their group. Again it must be noted that this addition to the peer-marking system has resulted in an increase in the assessment time scales, and as such great care has to be taken in mapping an appropriate reward for the additional effort expected from the students in performing the peer-assessment process. This as in the past uses of the CAP system has to be mapped to the quality of the process not map just the time taken.

References

Davies, P. (2000), Computerized Peer-Assessment, *Innovations in Education and Teaching International (IETI),* 37, 4, pp 346-355.

Davies, P. (2003), Closing the Communications Loop on the Computerized Peer Assessment of Essays, *Association of Learning Technology Journal (ALT-J)*, 11, 1, pp 41-54.

Davies, P. (2004), 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, *Association of Learning Technology Journal (ALT-J)*, 12, 3, pp 263-279.

Davies, P. (2005), Weighting for Computerized Peer-Assessment to be Accepted, *in Danson, M (Ed) Proceedings of the 9th Annual International CAA Conference*, Loughborough University, pp 179-192, ISBN 0-9539572-4-1.

Davies, P. (2006), Peer-Assessment: Judging the quality of student work by the comments not the marks?, *Innovations in Education and Teaching International (IETI)*, 43, 1, pp 69-82.