

Research Article

## Computer-Assisted and Peer Assessment: A Combined Approach to Assessing First Year Laboratory Practical Classes for Large Numbers of Students

Gayle Ferguson, Elizabeth Sheader and Ruth Grady

Faculty of Life Sciences, The University of Manchester, UK

Date received: 05/11/2007

Date accepted: 26/02/2008

---

### Abstract

*Providing fair assessment with timely feedback for students is a difficult task with science laboratory classes containing large numbers of students. Throughout our Faculty, such classes are assessed by short-answer questions (SAQs) centred on principles encountered in the laboratory. We have shown recently that computer-assisted assessment (CAA) has several advantages and is well received by students. However, student evaluation has shown that this system does not provide suitable feedback. We thus introduced peer assessment (PA) as a complementary procedure. In October 2006, 457 students registered for a first-year practical unit in the Faculty of Life Sciences, University of Manchester. This unit consists of ten compulsory biology practical classes. The first four practicals were assessed using PA; the remaining six practicals were assessed by CAA and marked by staff or postgraduate student demonstrators. The reliability and validity of PA were determined by comparing duplicate scripts and by staff moderation of selected scripts. Student opinions were sought via questionnaires.*

*We show that both assessments are valid, reliable, easy to administer and are accepted by students. PA increases direct feedback to students, although the initial concerns of student groups such as mature and EU/International students need to be addressed using pre-PA training.*

**Keywords:** Computer-assisted assessment; peer assessment; laboratory practical class assessment; feedback; large class assessment

---

### Introduction

Assessment of student learning serves many purposes (Anon, 2000; Race, 2001a), one of which is the opportunity for students to demonstrate that they have achieved the intended learning outcomes (ILOs) for a particular course. Some ILOs can be difficult to assess within large class environments, where students generally work in pairs or in groups; assessing practical skills under exam conditions can be expensive, time-consuming and may not provide the best conditions for students to demonstrate their ability. For these reasons, in laboratory practical classes in the Faculty of Life Sciences, University of Manchester, UK, the ILOs (Table 1) are assessed either formatively in the laboratory by giving immediate feedback on performance, or summatively, by the use of data-handling exercises and short-answer questions (SAQs) based on principles encountered in the laboratory. The questions are accessed and marked via computer (computer-assisted assessment, CAA) using dedicated software (Assess by Computer, ABC) developed by the School of Computer Sciences, The University of Manchester (Sargeant *et al.*, 2004). The software allows questions to be posed, answers entered and marks allocated by staff via a computer interface. CAA for SAQs has been used with some success (Bull and Collins, 2002; Pain and Le Heron, 2003;

Wang *et al.*, 2004) and has been found to compare favourably with traditional paper-based assessment in our own Faculty (Sheader *et al.*, 2006), especially with regards to ease of implementation, ability to spot plagiarism and reducing variability between markers, by having one member of staff mark one question for all students.

**Table 1** *Intended learning outcomes (ILOs) for the compulsory level 1, 10-credit unit “Introduction to Laboratory Science”*

Overall intended learning outcomes for first year practical units	Method of assessment
<i>By the end of their first year, students are expected to:</i>	Formative (F)
Summative using SAQs (S)	
be competent in a range of practical techniques and skills appropriate to the biosciences	F
conduct experiments taking into consideration health and safety requirements	F
make detailed experimental observations, and to record, analyse and evaluate experimental and other scientific data	F
analyse experimental data using appropriate statistical methods	S
be able to modify or design related experiments	S
relate knowledge acquired in the laboratory to theoretical material covered in the lecture units	S
work both independently or as part of a team as required	Not assessed
be able to make critical evaluation of both their own work and that of their peers	F
reflect upon their skills development during their first year	Not assessed
behave according to laboratory professional standards. These include being on time, arriving prepared for the practical with all the necessary resources and equipment, respecting the lab equipment and following the health and safety procedures and standards	F

However, our evaluation has shown that CAA scores poorly on student satisfaction surveys, particularly with regard to providing timely and satisfactory feedback. Recent student evaluation also revealed that many students were unfamiliar with these types of SAQ and that they were uncomfortable with their lack of certainty in knowing what was expected of them in completing the assessment; especially as all other assessments in the first year were by multiple choice questions alone.

PA is a distinct method of assessment that actively engages students in the marking process (Ballantyne *et al.*, 2002; Ellery and Sutherland, 2004; Orsmond, 2004). In doing so, PA gives students an insight into how assessment ‘works’ at University and a chance to see directly the attainment level of another member of their year group. This gives students the valuable opportunity to judge how well they are performing compared with their peers and prepares students to tackle assessments in the future: important criteria defining effective feedback to students (Falchikov, 1995; Gibbs, 1999; Race, 2001b; Rust, 2002). For these reasons, PA was introduced in 2006 in our Faculty as a complementary assessment to CAA for a compulsory first year practical unit.

As PA was introduced as part of the summative assessment for the unit, we wished to establish whether PA is as reliable and valid as other currently used methods of assessment. In particular we were interested to see if weaker students (*i.e* those that gained low marks

themselves) were able to recognise and reward good scripts, and to ensure that no student 'lost out' because of this assessment approach.

This study describes a combined assessment approach for large student numbers. We provide evidence that both assessment methods are valid, reliable, easy to administer and are accepted by students. We also provide evidence that PA has added benefits for students, including increasing feedback and enabling students to feel more confident for future assessments.

## Methods

### Assessment of the unit

For the academic year 2006–2007, 457 students registered for the Introduction to Laboratory Science 10-credit unit in the Faculty of Life Sciences, University of Manchester, UK. This first year unit comprises ten practicals covering a range of biological subjects (including biochemistry, human physiology, molecular biology and biodiversity), which in previous years had been assessed by SAQs. The SAQs were marked by staff or postgraduate student demonstrators via computer using specialised software (Sargeant *et al.*, 2004). For the year reported herein, the first four practicals (taken in weeks 2–5 of the semester; see *Table 2*) were assessed by SAQs completed on paper, but marked by fellow students (PA). The remaining six practicals (taken in weeks 7–11 or in weeks 2–11; see *Table 2*) were assessed by SAQs, which were submitted online by the students and marked via computer by academic staff or postgraduate students. Questions were assigned different marks ranging from 1, for questions requiring one-word answers, up to 10 marks for questions requiring longer, more detailed responses; the total number of marks for PA practicals 1–4 was 61 and for CAA practicals 5–10 was 131. All students attempted the SAQs for all ten practicals. The summative mark for the unit comprised 20% for attending and completing the practicals and 80% for the SAQs.

**Table 2** Timeline of events for the compulsory level 1, 10-credit unit "Introduction to Laboratory Science"

	Week	Activity
<b>Semester 1</b> (September to January)	1	Pre-assessment questionnaire
	2–5	Practicals 1–4 completed
	6	'Reading Week' (no classes)
	7	Monday: deadline for SAQs for practicals 1–4; Wednesday: PA of SAQs
	8	PA marks revealed
	7–11	Practicals 5–7 completed
	2–11	Practicals 8–10 completed
	12	Deadline for SAQs for practicals 5–10 (CAA) General feedback given for practicals 5–10
<b>Christmas Break and examination period</b>		
<b>Semester 2</b> (February to May)	1	CAA marks released
	2	Post-assessment questionnaire

### **Completion of SAQs and administration of PA**

SAQs covering material from practicals 1–4 were distributed and subsequently completed on paper in the students' own time. Scripts were submitted to the Undergraduate Office by the deadline date (Monday of Week 7). Administrative staff recorded the submission from each student and then covered the student names to allow anonymous marking.

### **PA of practicals 1–4**

Two days after the submission deadline, the students attended the peer-marking sessions which were held in a lecture theatre (session length 50 minutes). Attendance at the marking session was compulsory; non-attendance was penalised with the student losing 50% of their own marks for practicals 1–4. Half of the student cohort attended either one of two sessions. Each student was given a script to mark and signed the coversheet to confirm that they had marked the work to the best of their ability. One member of staff revealed the expected answers to the SAQs using a PowerPoint presentation, allowing enough time for students to allocate marks for each question. Questions from students were not encouraged so that students had to settle on marks independently. Scripts were collected and the marks collated by administrative staff. Students that could not complete the marking in the time allocated were allowed to finish the marking in a seminar room after the peer marking session. A selection of scripts (n = 66) was moderated by staff, with particular emphasis given to high-scoring scripts (arbitrarily judged to be more than 47 out of 61 marks) and low-scoring scripts (less than 25 marks out of 61).

Marks were released 10 days after the PA sessions. Students were given 1 week to challenge their mark, and were given the opportunity to have their script remarked by staff if they were unhappy with how another student had marked it.

### **Completion of SAQs and administration of CAA**

The ABC software allows inputted questions to be accessed via an external website, with a separate URL for each practical. All students were given a unique identifier (student registration number) and password so that they could access the URLs from any internet-linked computer. The students could enter the websites and attempt the questions in their own time by entering text-based answers. Answers could be saved and edited prior to submission by the deadline. Only one submission was allowed and submissions were not possible after the deadline had lapsed. An email receipt was sent to students to confirm answer submission. Submitted answers were saved on file and could be accessed by the administrator as required. Questions were set by several staff, although only one member of staff was required to enter the questions into the ABC system. One computer technician was then required to oversee the process. This involved registering the students' registration numbers to use the system and monitoring any problems that the students had in accessing the websites.

### **CAA of practicals 5–10**

For CAA, answers saved on file were passed to tutors for marking, with scripts made anonymous automatically. To mark the assessment questions via computer, tutors could view the student answers on-screen, and depending on length of answer, could visualise several scripts at a time. The student responses could be compared with a pre-inputted 'model answer' and marks allocated manually depending on the quality of the answer, as in traditional assessment. Student answers could be sorted and viewed according to answer length or similarity to model answer. Key-words in the answer could be selected and highlighted to aid marking if required. Final marks were totalled automatically and presented in a spreadsheet for analysis. All answers to a particular question were marked by the same member of staff or postgraduate student demonstrator. In week 12, after the deadline for practicals 5–10 had passed, students were

encouraged to attend a feedback session in a lecture theatre. During the session, students were given general guidance over what was expected from the SAQs for practicals 5–10. Individual marks were not made available until after the Christmas break.

## Evaluation of the assessment methods

### Perceptions of the assessment methods

All students were surveyed via an anonymous questionnaire (*Appendix 1*) in week 1, prior to starting the unit (*Table 2*). Several questions required students to rate statements on a 5–point Likert-scale ('strongly disagree/disagree/neutral/agree/strongly agree'). Room was provided to allow students to write free-comments regarding assessment. Completed questionnaires were received from 249 (55%) students. After both types of assessment were completed, and marks revealed, a second questionnaire (*Appendix 2*) was distributed to all students at the start of the second semester, of which 283 (62%) were returned.

Opinion regarding feedback were acquired via the University-wide student satisfaction survey that requires students to rate feedback for a unit using a scale from -2 to +2. This survey is administered centrally and is not controlled by Faculty.

### Validity and reliability of PA

The robustness of PA as a summative assessment procedure was determined in several ways. Firstly the means and distributions of marks awarded by PA and CAA in 2006 were compared both with each other and with the previous year's data using the Kolmogorov-Smirnov test for differing probability distributions and the Mann-Whitney U tests for differing means (SPSS).

A selection of scripts obtaining low (< 25/61), average (36 or 37/61) and high marks (> 47/61) was moderated by a staff member. The scripts marked by students who scored poorly or scored well were also moderated.

Five scripts were duplicated, to allow different undergraduate students (for each script  $n = 4$ ; a different UG student marked each script) and postgraduate student demonstrators (for each script  $n = 4$ ; one PG demonstrator marked five different scripts) to mark the same script. Mean marks were compared using the Mann-Whitney U Test.

## Results

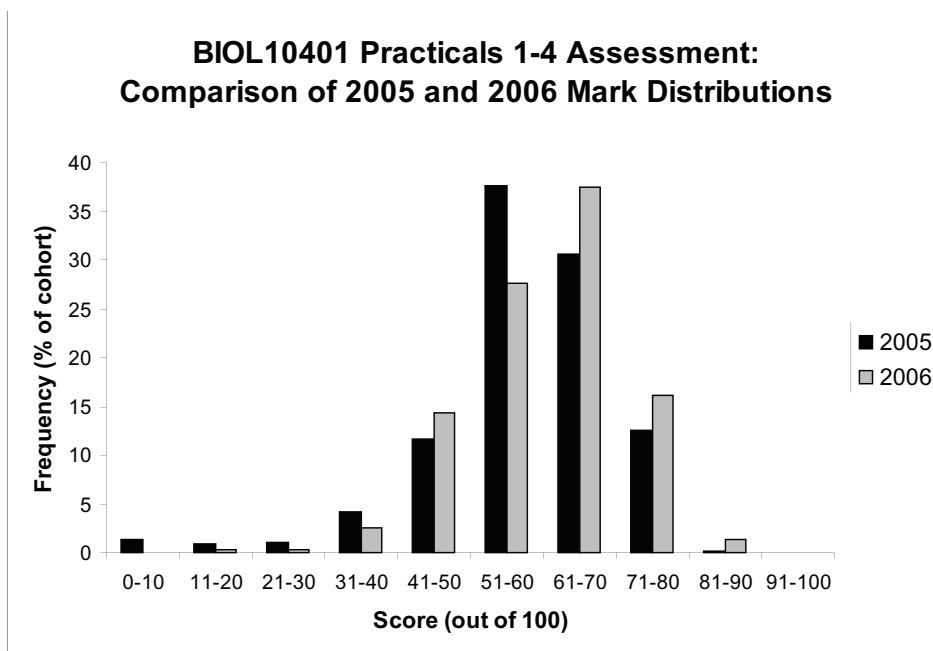
Of the 457 students who were due to attend the peer marking sessions, 6 were absent without reason (1%) and 2 presented with sicknotes. Ten students challenged their mark following publication of results, with 6 (1%) subsequently requesting a remark. These re-marked scripts were all subsequently increased by 2–13%.

### Validity of PA

The practicals in 2005 (the first year in which this unit ran) were all assessed via CAA. The same learning outcomes were assessed in 2005 and in 2006; however, the actual questions used were not identical. The distribution of marks for practicals 1–4 assessed by PA in 2006 was significantly different from the distribution of marks awarded for the same practicals in 2005 (K-S  $p < 0.001$ ; Fig. 1). Neither was normally distributed (K-S,  $p = 0.05$  for both years). The mean in 2006 ( $60 \pm \text{SD } 10$ ) was significantly higher than the mean in 2005 ( $57 \pm \text{SD } 13$ ; Mann-Whitney U test  $p = 0.002$ ).

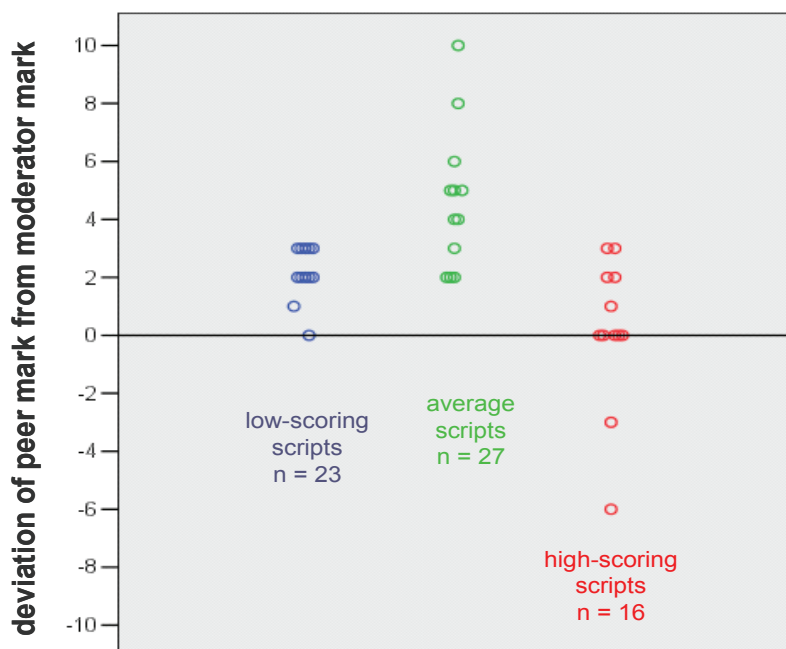
Practicals 5–10 were the same in 2005 and in 2006 and both were assessed by CAA. The assessment questions were not identical but were of an equivalent standard. The distributions

of marks were not significantly different to each other and neither is normally distributed (data not shown). Means were identical for both years ( $58 \pm \text{SD } 13$ ).



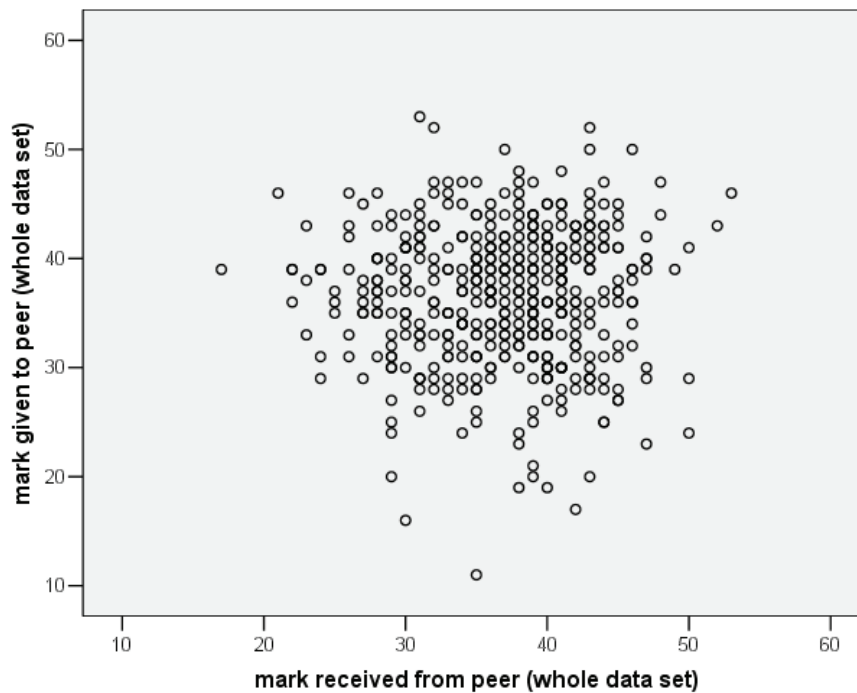
**Figure 1** Distribution of marks for Practicals 1–4 assessed by CAA in 2005 and by PA in 2006. The practicals were the same although the assessment questions were not identical. Marks are not normally distributed and are significantly different to each other (Kolmogorov-Smirnov tests,  $p = 0.05$  for both years. Mean in 2006 =  $60 \pm 10$ ; mean in 2005 =  $57 \pm 13$ ).

A selection of scripts with low, average or high marks was moderated by a staff member. Marks were generally raised in all categories (Fig. 2).



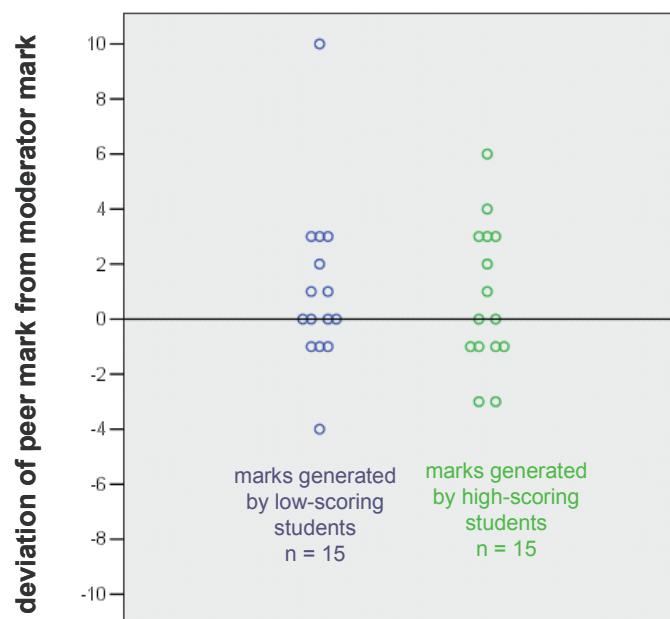
**Figure 2** Scripts from different mark categories were moderated and the deviations shown above. Mark deviations ranged from -6 to + 10 (– 10% to + 16%). The assessment was marked out of 61. Mark categories were defined as low (< 25/61), average (36 or 37/61) and high marks (> 47/61)

Moderated marks for low-scoring scripts increased from 0 to 5%; average-scoring scripts were increased by 3 – 16%; high-scoring scripts varied from a 10% decrease up to a 5% increase. There was no correlation between the mark a student received and the mark a student gave (Fig. 3).



**Figure 3** The mark a student received compared to the mark the student gave. The data show no correlation (Pearson correlation test, two-tailed,  $p = 0.85$ ,  $n = 436$ ). The assessment was marked out of 61.

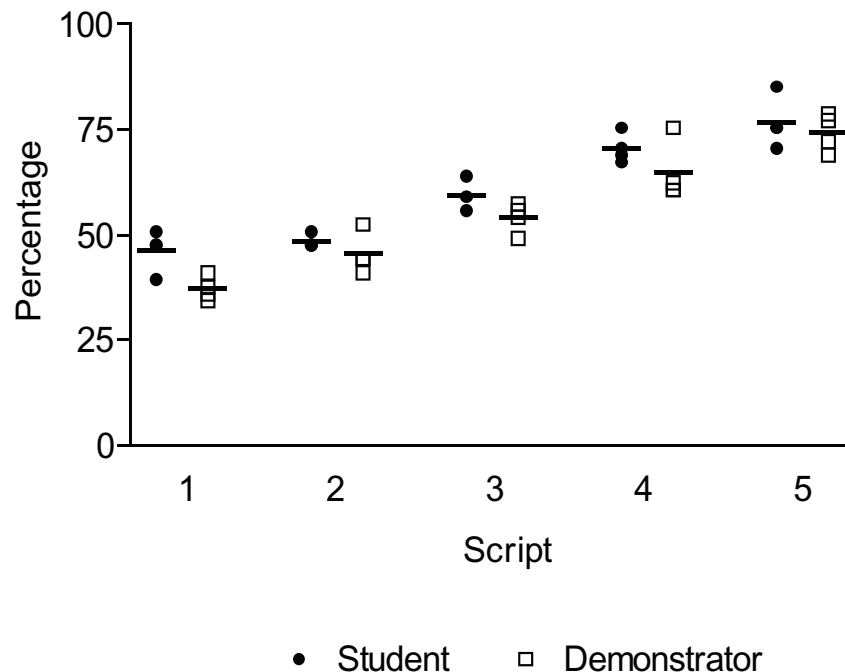
Students scoring low marks were no more or less likely to mark high or low than students who performed well (Fig. 4) with both just as likely to be moderated up or down.



**Figure 4** The scripts marked by low-scoring and high-scoring students were moderated by a staff member. Mann-Whitney U-test analyses show no significant difference in variability of marking between low-scoring & high-scoring students ( $p = 0.428$ ). The assessment was marked out of 61. Mark categories were defined as low ( $< 25/61$ ) or high ( $> 47/61$ ).

## Reliability of PA

Duplicate scripts marked by different students did not always return identical marks (Fig. 5). Generally UG students marked more generously than the PG student demonstrators; although there was no significant difference in the mean marks for any of the scripts (Mann-Whitney U test p-values range from 0.057 to 0.886).



**Figure 5** The percentage marks for each script are shown for the five duplicated scripts marked by either a PG student demonstrator (open squares) or by UG students (filled circles). The horizontal bars represent the mean marks for each student group for each script. Mean marks are not significantly different for the same script using the Mann-Whitney test. Script 1  $p=0.057$ ; Script 2  $p=0.434$ ; Script 3  $p=0.057$ ; Script 4  $p=0.200$ ; Script 5  $p=0.886$ .

## Student perceptions of PA and CAA

### Pre-assessment questionnaire data

Over half the student cohort (55%) returned a completed questionnaire to determine attitudes prior to starting the unit. The respondents reflected the demographics of the 2006–2007 cohort overall. Respondents were predominantly aged 20 or under (92%), Home (UK) students (84%) and had entered the University following 'A' level study (91%). The small number of mature students (*i.e.* aged over 21;  $n = 20$ ) consisted of 57% EU/International students ( $n = 11$ ) and 43% UK students ( $n = 9$ ). The percentage of female respondents (63%) was slightly higher than the cohort overall (56% female intake in 2006).

Most students (70%) had not undertaken peer assessment before; this was the same for all ages and for both UK and EU/International students. Interestingly, 11% did not know whether they had done PA before.

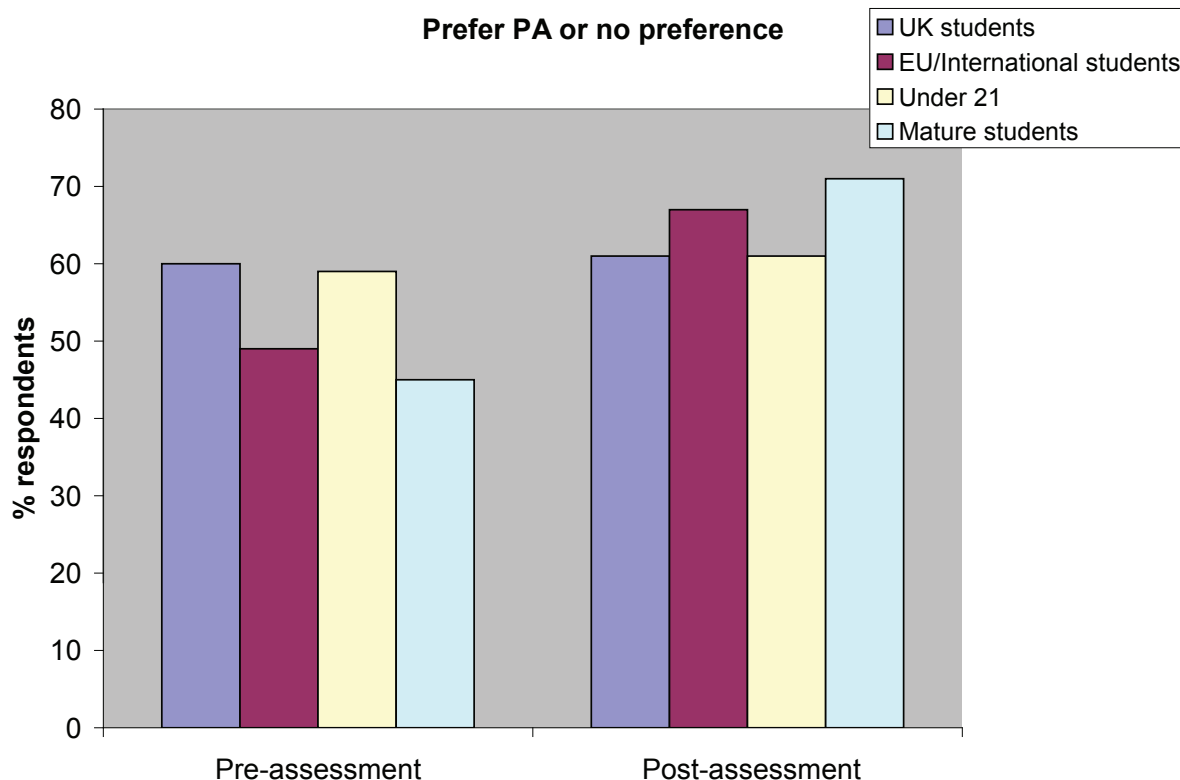
### Post-assessment questionnaire data

In Semester 2, following release of the unit marks, a post-assessment questionnaire was returned by 62% of students. The demographics of the respondents reflected the demographics

of the student cohort overall (see above). Most attitudes did not change significantly after partaking in PA, although there were some exceptions.

Prior to starting the unit, students were asked if they thought that PA was a fair method of assessment. Approximately one-third of respondents agreed with this statement; this was independent of gender, the students' origin or whether the students had undertaken PA before or not. Students not comfortable with PA commented "*I do not feel my peer-marked questions were marked fairly. I feel more comfortable having academic staff mark my work*" and "*Work marked by peers was rushed and therefore unfair/inaccurate*". There was a small difference for mature students, of whom only 20% agreed that PA was a fair assessment. Post-assessment, the agreement figure rose to approximately 40% of respondents, although this figure remained lower for mature (21%) and EU/International students (22%).

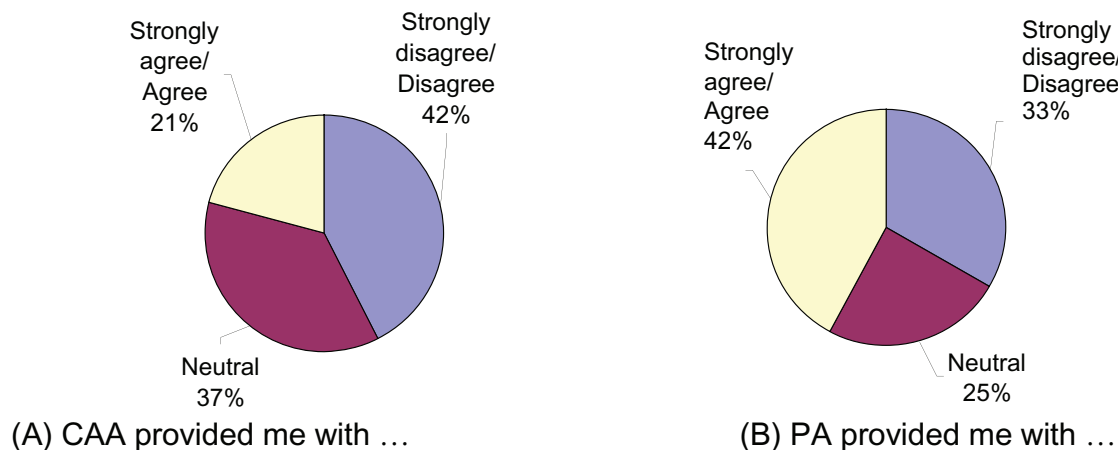
Attitudes to CAA as a method of assessment were more positive, with 60% of the cohort believing CAA to be fair. When asked to choose between assessment by peers or by staff only, approximately 60% of the cohort opted for 'no preference' or 'some marking by peers'. This proportion did not change between the pre- and post-assessment questionnaires. However, if the student cohort is broken down into different categories, differing trends emerge. Figure 6 shows the different attitudes of UK students and EU/International students: here, the option for PA rose from 49% to 67% in the latter group following assessment. If separated into age groups, the response of mature students increased from 45% to 71%. One student noted "*I don't care who marks it, but I would like to see where I went wrong*".



**Figure 6** Percentage of respondents that opted for 'marking by peers' or 'no preference' both pre- and post-assessment.

Prior to PA, over half of the student cohort (51%) thought that they would learn more if staff marked their work for practicals 1–4. Post-assessment, this figure had dropped to just 19%,

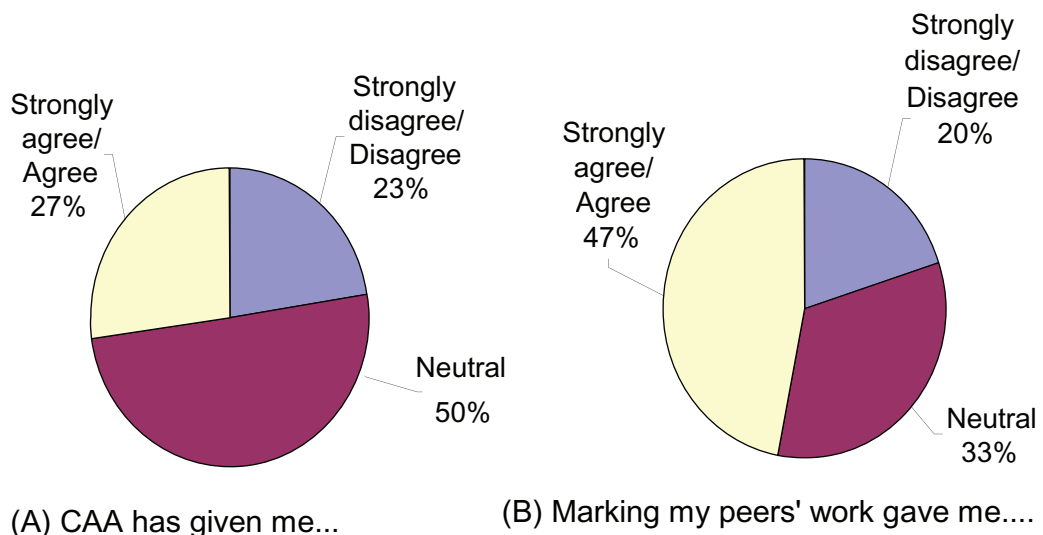
although 29% of mature and 27% of EU/International students still thought this would be the case.



### ... adequate feedback about my own performance

**Figure 7** Students were asked whether (A) CAA or (B) PA provided adequate feedback regarding their own performance

Most students found that they could mark their peer's work adequately. Overall, 19% thought that they could NOT mark adequately, although this response was recorded by a higher percentage of mature students (29%) or EU/International students (29%). PA was perceived to offer greater feedback than CAA. For PA, over 40% of students agreed with the statement 'PA provided me with adequate feedback about my own performance', whereas the figure was approximately 21% when polled about CAA (Fig. 7).



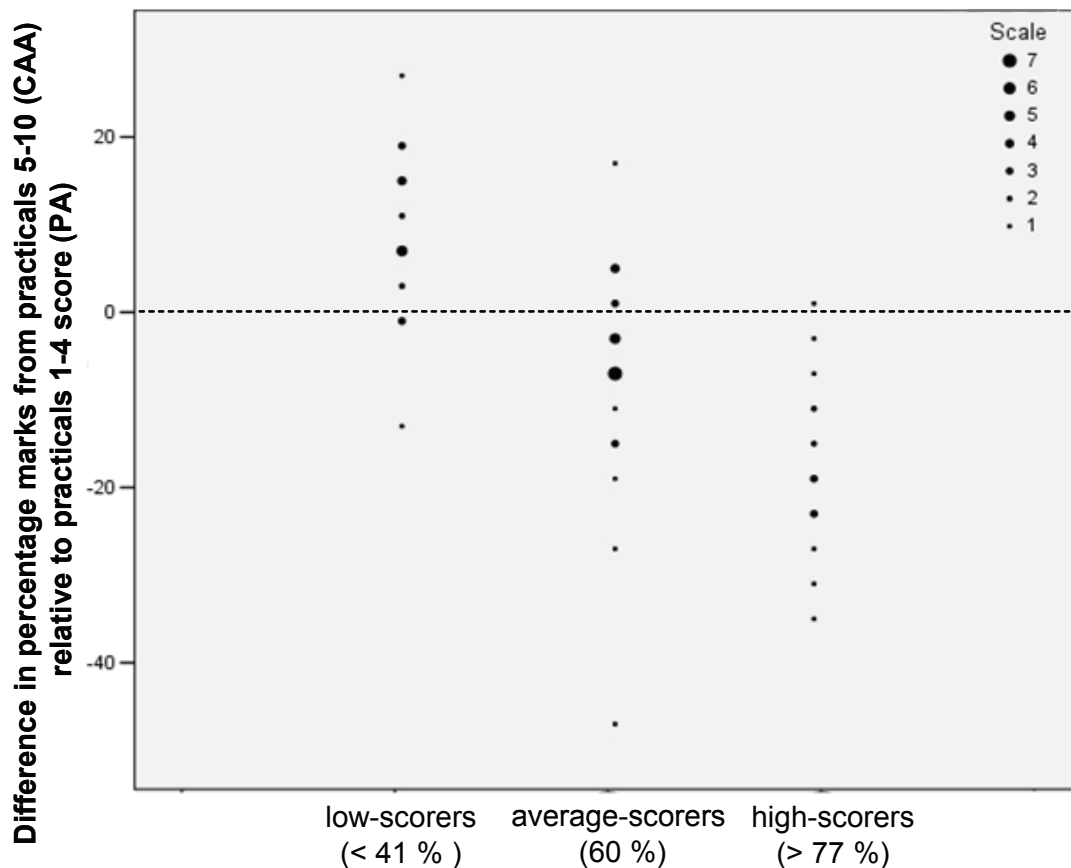
### ... confidence in answering further

**Figure 8** Students were asked whether PA or CAA increased their confidence in answering further assessment questions.

In the student satisfaction survey (scale -2 to +2), the rating for feedback for the unit as a whole increased from 0.35 ( $\pm 1.03$ ) in 2005, to 0.97 ( $\pm 0.87$ ) in 2006. Indeed, students that did not

vote in favour of PA could also see the benefit of it for learning, as one such student noted "... however, the direct feedback from the peers marking was very useful".

PA was perceived to help the students answer further assessment questions. Nearly half of students answered positively regarding PA, compared with approximately 27% for CAA (Fig. 8). The figure for mature students regarding PA was lower with only 29% responding positively. One student noted "It was productive to see how the system worked before staff marking", whilst another volunteered "I think marking someone else's work helps me to better understand the level of work expected from me".



**Figure 9** Differences between percentage marks obtained for practicals 5–10 and practicals 1–4. The x-axis shows results from practicals 1–4. The size of the circles represents the number of students showing a particular mark difference.

The majority of students who performed poorly on practicals 1–4 assessment (< 41%) improved their mark significantly on practicals 5–10 (median improvement of 8%; Fig. 9) whereas there was a marked tendency for students who obtained average marks (60%) or high (> 77%) on practicals 1–4 to perform less well on the practicals 5–10 (median improvement of -6% and -19%, respectively). These differences are statistically significant (Mann-Whitney U test;  $p < 0.001$ ).

## Discussion

It has been demonstrated previously in our Faculty that CAA saves staff time, is easy to administer, can detect plagiarism and allows anonymity of marking (Sheader *et al.*, 2006). However, reservations regarding timely feedback and a lack of students' previous experience of answering SAQs prompted us to supplement the assessment of a first year practical unit with PA. As examined in Falchikov and Goldfinch (2000) and Sivan (2000), PA has been used for many different types of exercises, in different disciplines, in different countries and has been used both formatively and as a summative assessment. Often PA exercises are combined with group or self-assessments for small to medium sized classes; here we have introduced PA to over 400 students in one compulsory first-year practical unit, with only 1% of students opting out of the process (unexplained absences) and 1% challenging the assessment mark given (requiring a re-mark). It is perhaps worth mentioning that all students who challenged their mark were justified in doing so, as all of these papers were awarded a higher mark after challenge. A similar number of students (2% or approximately 5 students) requesting a re-mark was noted by Hughes (2004); however, this number is lower than the 9% reported elsewhere (Ballantyne *et al.*, 2002).

Studies addressing the reliability (comparing peer marks with each other) and validity (comparing peer marks with the 'correct' mark) of PA have been reviewed extensively (Topping, 1998; Dochy *et al.* 1999, Falchikov and Goldfinch, 2000). The reliability of marking varies, depending on the context of application, but is generally favourable. Reliability of examination marking by professional teachers is itself often inconsistent (Newstead and Dennis, 1994), although scientific disciplines tend to fare better than essay-style examinations in humanities or social sciences (Falchikov and Boud, 1989). It should be noted that only a few of the SAQs in this assessment were calculations with numerical answers; the majority of SAQs required explanations of biological phenomena or descriptions of experimental procedure that were open to interpretation by the marker. In this study, the reliability of the marks awarded was determined by comparing duplicate scripts marked by different peers. The spread of marks was comparable with the spread of marks returned by postgraduate student demonstrators. The mean marks for both cohorts were not significantly different, showing the validity of the PA. This was true for scripts generating low, average or high marks. Marks were compared with PG demonstrators and not staff to mimic the situation for the unit if PA had not been introduced. It is known that inexperienced markers often mark conservatively (that is, do not mark too low or high even if the script deserves it; Stefani, 1994; Mowl and Pain, 1995) so it is pleasing to see that although the PA mark distribution is slightly skewed to the right, marks are returned from the entire scale for the whole cohort.

The validity of marks was also confirmed by staff moderation of the marks, from scripts drawn at random and from selected lower and higher-scoring scripts. If the PA marks given by students were not valid marks, we would expect to see low-scoring scripts moderated to much higher marks, high-scoring scripts all moderated lower and average-scoring scripts to contain a mixture of both. This was not the case: moderated scripts tended to be under-graded by students regardless of script quality, indicating that low scores cannot be attributed to overly harsh student marking or high scores to overly generous student marking. Other studies have confirmed the validity of peer marks against a tutor standard (Stefani, 1994; Sivan *et al.*, 1995; Falchikov and Goldfinch, 2000) whereas some show poor agreement (Mockford, 1994; Mowl and Pain, 1995).

The ability of the student did not seem to influence the mark awarded as there was no correlation between the mark a student got and the mark a student gave. We found that there was no difference in the amount of moderation (up or down) of scripts marked by students

who themselves had obtained low or high marks. This shows that the variability of both groups was equivalent. Hughes and Large (1993) also found this to be the case for PA of oral communications, but as noted by Topping (1998), there are few studies looking at the relative abilities of assessors and assessee. In this study, the marks for PA for all students were significantly higher than for CAA practicals, although in absolute terms this was not of any consequence for this unit.

After undertaking PA, students' perceptions of the assessment process were generally favourable, albeit tinged with caution as most other studies have also shown (Searby and Ewers 1997; Brindley and Scoffield, 1998; Ballantyne *et al.*, 2002). CAA was perceived to be 'fairer' than PA, although it is unclear if students realised fully that the CAA work was mostly marked by postgraduate student demonstrators and not all by staff, as this may well have altered their responses to this question. Two thirds of the cohort would either opt for PA for future assignments or had no preference in assessment procedure. Most students found that they could mark their peers' work adequately and realised that they would not have learnt any more even if staff had marked their work. Feedback was perceived to be enhanced for PA rather than CAA and the unit as a whole was rated better for feedback than in previous years. PA was also thought by students to help with answering future questions, although the marks obtained for practicals 5–10 did not seem to support this optimism. It is interesting to note that the students who performed poorly on practicals 1–4 generally improved their marks significantly on the remaining practicals, whereas the opposite was true for students who obtained average or high marks. It is not clear whether this was an effect of PA, or was the result of student effort in light of summative feedback from their first assessed assignment or a reflection of the contents of the individual practicals suiting differing students; this merits further study. Nevertheless, the process of demystifying assessment at University is an important one and has been noted as one of the strengths of PA (Brindley and Scoffield, 1998; O'Moore and Baldock, 2007). This was also borne out by comments offered by the students themselves, as even students who did not show a preference for PA in future assessments were able to see the added benefits of partaking in this type of assessment.

Attitudes to PA were not correlated with gender or, perhaps surprisingly, with prior experience of PA. In our Faculty, students are predominantly from the UK and are aged under 20 on entry, although these figures may change in future intakes. The students who were more negative in their attitudes to PA were mature and EU/International students. This was apparent in opinions regarding 'fairness' of PA, whether the students would learn more by staff marking the assessments, were able to mark adequately and if PA would help in future assessments. Interestingly, despite their reservations regarding specific aspects of PA, both of these groups were more likely to opt for PA or have no preference in future when compared with UK or students under 20. Sivan (2000) found that mature students were more inclined to accept the added benefits of PA, possibly due to the likelihood of being exposed to 'peer review-type' activities in the workplace prior to starting university. In future years, the worries of minority groups such as mature and International students will be addressed during week 1, when the students are introduced to the concept of PA, in order to reassure the students that the process will work for all. Practice or training in taking part in PA effectively is recommended by several authors (Mowl and Pain 1995; Hanrahan and Isaacs 2001; Ballantyne *et al.* 2002). This was also suggested by our students and so practice questions will be made available in future.

### Concluding remarks

PA has been reported as being stressful (Brindley and Scoffield, 1998), time-consuming (Falchikov, 1986; Searby and Ewers 1997) and difficult to deliver for large classes (Ballantyne *et al.*, 2002). However, here we have found that the combination of PA with CAA worked well in our Faculty for the summative assessment of a large undergraduate practical class for over

400 students. Students took the PA task seriously and marks were not prejudiced as a result of partaking in PA. Any student who did feel unfairly treated had the option of requesting a re-mark by staff, although very few students chose to do so. Mature students and EU/International students were more guarded regarding PA, although these students were more likely to opt for this type of assessment process after completion of the unit. Pre-PA training may alleviate the concerns of these students and this will be addressed in future.

We think that in addition to providing an overall mark for the unit, the numerous benefits of PA harmonised with the previously reported benefits of CAA (Sheader *et al.*, 2006). With PA, the students could compare their level of achievement directly with a peer; learn how and why marks are awarded for a particular question; begin to understand the complexities involved with awarding marks and appreciate at first-hand the importance of presenting work in a legible and logical manner. Hopefully, all of these benefits will help the students in completing similar assessment tasks in the future. In addition, the ability to appraise and judge another's work is an integral part of the practice of peer-review in research science, thus PA is a useful early introduction to gaining experience of such a process at a formative stage in the students' careers.

### Acknowledgements

This study has been supported by a grant from the HEA Centre for Bioscience Departmental Teaching Enhancement Scheme 2006. The CAA software was provided and supported by Phil Reed, John Sargeant and Mary McGee Wood from the School of Computer Sciences, The University of Manchester. Thank you to Ian Hughes and Ruth Anderson-Beck (The University of Leeds) for valuable advice regarding PA procedure and to Nick Ashton (FLS, The University of Manchester) for helpful suggestions for the manuscript.

### Communicating Author

Ruth Grady, 1.800 Stopford Building, Oxford Road, Manchester M13 9PT  
 Phone 0161 2754680 Fax 0161 2754565 email [ruth.grady@manchester.ac.uk](mailto:ruth.grady@manchester.ac.uk)

### References

- Anon. (2000) Code of practice for the assurance of academic quality and standards in higher education. The Quality Assurance Agency for Higher Education available at [http://www.qaa.ac.uk/academicinfrastructure/codeOfPractice/section6/COP\\_AOS.pdf](http://www.qaa.ac.uk/academicinfrastructure/codeOfPractice/section6/COP_AOS.pdf) (Accessed February 2008)
- Ballantyne, R., Hughes, K. and Mylonas, A. (2002) Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education* **27**, 427-441
- Brindley, C. and Scoffield, S. (1998) Peer assessment in Undergraduate Programmes. *Teaching in Higher Education*, **3**, 79-89
- Bull, J. and Collins, C. (2002) The use of computer-assisted assessment in engineering: some results from the CAA national survey conducted in 1999. *International Journal of Electrical Engineering Education*, **39**, 91-99
- Dochy, F., Segers, M. and Sluijsman, D. (1999) The use of self-, peer- and co-assessment in Higher Education: a review. *Studies in Higher Education*, **24**, 331-350
- Ellery, K. and Sutherland, L. (2004) Involving students in the assessment process. *Perspectives in Education*, **22**, 99-110.
- Falchikov, N. (1986) Product comparisons and process benefits of collaborative peer and self-assessments. *Assessments and Evaluation in Higher Education*, **11**, 146-166
- Falchikov, N. (1995) Improving feedback to and from students. In *Assessment for Learning in Higher Education*, pp. 157-166. P. Knight, ed. London, Kogan Page

- Falchikov, N. and Boud, D. (1989) Student self-assessment in Higher Education: a meta-analysis. *Review of Educational Research*, **59**, 395-430
- Falchikov, N. and Goldfinch, J. (2000) Student peer assessment in Higher Education: a meta-analysis comparing peer and teacher marks. *Review of Educational Research*, **70**, 287-322
- Gibbs, G. (1999) Using assessment strategically to change the way students learn. In *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches*, eds S. Brown, and A. Glasner, pp. 41-53, Buckingham, Society for Research into Higher Education & Open University Press
- Hanrahan, S. and Isaacs, G. (2001) Assessing self- and peer-assessment: the students' views. *Higher Education Research and Development*, **20**, 53-70
- Hughes, I (2004) Peer-assessment of practical write-ups using an explicit marking schedule In: *Self- and peer-assessment: guidance on practice in the biosciences. Teaching Bioscience Enhancing Learning Series*, P. Orsmond, S. Maw, J. Wilson, and H. Sears, eds. pp. 1-47 Leeds, The Higher Education Academy Centre for Bioscience
- Hughes, I. and Large, B. (1993) Staff and peer-group assessment of oral communication skills. *Studies in Higher Education*, **18**, 379-385
- Mockford, C. (1994) The use of peer group review in the assessment of project work in higher education. *Metoring and Tutoring*, **2**, 45-52
- Mowl, G. and Pain, R. (1995) Using self and peer assessment to improve students' essay writing: a case study from geography. *Innovations in Education and Teaching International*, **32**, 324-335
- Newstead, S. and Dennis, I. (1994) Examiners examined: the reliability of exam marking in psychology. *The Psychologist*, **7**, 216-219
- O'Moore, L. and Baldock, T. (2007) Peer assessment learning sessions (PALS): an innovative feedback technique for large engineering classes. *European Journal of Engineering Education*, **32**, 43-55
- Orsmond, P. (2004) Self- and peer-assessment: guidance on practice in the biosciences. In *Teaching Bioscience Enhancing Learning Series*, eds S. Maw, J. Wilson, and H. Sears, pp. 1-47 Leeds, The Higher Education Academy Centre for Bioscience
- Pain, D. and Le Heron, J. (2003) WebCT and online assessment: the best thing since SOAP? *Educational Technology and Society*, **6**, 62-71
- Race, P. (2001b) A briefing on self, peer and group assessment. In *LTSN Generic Centre Assessment Series* pp. 1-25 York, Learning and Teaching Support Unit
- Race, P. (2001a) Designing Assessment and Feedback to Enhance Learning. In: *The Lecturer's Toolkit*, 2<sup>nd</sup> edition. Kogan Page Ltd., London, UK, p. 31-103
- Rust, C. (2002) The impact of assessment on student learning: how can the research literature practically help to inform the development of departmental assessment strategies and learner-centred assessment practices? *Active Learning in Higher Education*, **3**, 145-158
- Sargeant, J., Wood, M. and Anderson, A. (2004) A human-computer collaborative approach to the marking of free-text answers. Eighth International CAA Conference, Loughborough University, Loughborough, UK. pp 361-370.
- Searby, M. and Ewers, T. (1997) An evaluation of the use of peer assessment in Higher Education: a case study in the School of Music, Kingston University. *Assessment & Evaluation in Higher Education*, **22**, 371-383.
- Sheader, E., Gouldsbrough, I. and Grady, R. (2006) Staff and student perceptions of computer-assisted assessment for physiology practical classes. *Advances in Physiology Education*, **30**, 174-180
- Sivan, A., Yan, L. and Kember, D. (1995) Peer assessment in hospitality and tourism management. *Hospitality and Tourism Educator*, **7**, 17-20
- Sivan, A. (2000) The implementation of peer assessment: an action research approach. *Assessment in Education*, **7**, 193-213

- Stefani, L. (1994) Peer, self, and tutor assessment: relative reliabilities. *Studies in Higher Education*, **19**, 69-75
- Topping, K. (1998) Peer assessment between students in colleges and universities. *Review of Educational Research*, **68**, 249-276
- Wang, T., Wang, K., Wang, W., Huang, S. and Chen, Y. (2004) Web-based assessment and test analyses (WATA) system: development and evaluation. *Journal of Computer Assisted Learning*, **20**, 59-71