

**SELF, PEER AND TUTOR
ASSESSMENT OF TEXT ONLINE:
DESIGN, DELIVERY AND ANALYSIS**

Richard Parsons

Self, Peer and Tutor Assessment of Text Online: Design, Delivery and Analysis

Dr Richard Parsons
r.parsons@dundee.ac.uk
Centre for Learning and Teaching
University of Dundee
Dundee DD1 4HN, UK

Abstract

A complete system has been developed that allows a lecturer to create an online self and peer-based assessment exercise. Software and supporting material are now available for use by lecturers to create an assessment for their students, consisting of a question and marking criteria. The software utilises the internet, which ensures it can be operated at any time or place and it is supported by a website which provides access to the software, examples of its use, instructions and guidance materials.

The software was designed to permit a lecturer to create a text-based exercise to be marked by criterion-based referencing by both the student author and their peers. Students can be provided with guidance for marking and giving feedback. The program interface has been designed to be straightforward to use for both the lecturers designing activities and the students completing the exercises.

The program can be delivered from within Virtual Learning Environments, or as a stand-alone activity linked from web pages or email messages. A number of examples are available to demonstrate the program, but it is recognised that successful academic deployment of the exercises requires careful thought and innovation by the tutors.

Analysis of initial use of the program illustrates that tutors and students appreciate the opportunity to get involved in the assessment process. Students, in particular, learn exactly how marks are gained and they recognise that their performance can be improved. Although substantial variation occurred, the marks from the self, peer and tutor based assessment using criterion-based referencing correlated positively with norm-based referencing by an expert assessor. Self and peer assessment will continue to be developed as a key activity enhancing student education, providing them with reflective and assessment skills that are crucial to their personal development. Completing these exercises online ensures that efficiency gains in staff time can be realised.

Keywords

self assessment, peer assessment, online, automatic, evaluation, criteria reference.

Creating a Reusable Learning Object

Computer aided assessment can take many forms, and undoubtedly one of its key roles is in supporting formative assessment to drive learning. Enabling students to carry out self assessment of their text, and taking this further by enabling them to carry out peer assessment of their colleagues' text, represents a "reusable learning object" that teaching staff can develop and deploy to enable and encourage student learning. Like many aspects of CAA, self and peer assessment can be used as an "innovation" with the students, and they embrace the technology enthusiastically while progressing their education. This paper describes an online resource that can be used to efficiently create a self and peer assessment exercise for students.

Design

Self and peer evaluation of student work is valuable as a teaching tool to encourage students to become responsible and reflective (Dochy *et al.* 1999). Examples of both the successful use of both self (Boud, 1995) and peer (Topping and Ehly, 1998) assessment are available from hundreds, if not thousands of years ago, making our present consideration most relevant when focussed on the learning methods that are in use today. Students develop their own assessment skills and the marks obtained can be a fair indication of performance (Stefani, 1994). The web provides an ideal environment for developing abilities to answer and assess questions with written text answers. The challenge to produce such systems lies in developing both the academic environment of the exercise and the programming to enable appropriate presentation, file storage and processing.

Using computers to permit self or peer assessment of text has an innovative history. As this project has developed, others have been publicised including a peer assessment systems by Stephen Bostock of Keele University, Phil Davis of Glamorgan University, Ashley Ward and Abhir Bhalerao of Warwick University. These systems, and the one described here, are now making use of the important combination of internet-based convenience and carefully designed, challenging exercises. This combination produces a powerful learning tool that can enhance student learning.

One common difficulty cited in the use of self and peer assessment systems is that they are difficult to manage and maintain. The constraints of preparing for and administering self and peer assessment, without the use of specialist software, is recognised as being a higher load than conventional marking for small classes, and lower for large classes (Boud 1995). With specialist software, this concern evaporates and both small and large classes can be involved in self and peer assessment as appropriate to support their learning. There is clearly a requirement for such software, although it is undoubtedly a complex matter to create a system which is both easy to use, relatively secure and with sufficient options to be useful. At present students benefit from the innovative aspects of the programs, and they are unlikely to have completed similar exercises online. Should this mode of assessment and learning become more widespread, then clearly this pioneering spirit will be lost.

The level of tutor moderation within a self and peer assessment exercise is generally a matter that has to be judged by the exercise instructors. For a group of skilled and mature students involved in an exercise contributing at only a relatively low level to overall assessment, little or no tutor moderation may be expected. Alternatively, with a relatively inexperienced, not entirely trustworthy group completing an assessment with a moderate stake in overall assessment, then careful moderation and even full tutor marking should be carried out. Again the teacher planning and leading the exercise is usually well placed to make this decision.

Delivery

Throughout the remainder of this paper, the self and peer based system that has been developed at the University of Dundee will be described and evaluated. In a similar way to many systems, it first began as a prototype program operating in a limited fashion to meet a specific purpose (permitting electronic collection of a text-based exercise on cell size and the self assessment of this exercise against a set of defined criteria), and it has been developed into a much more comprehensive and adaptable system.

The present system is designed for use by academic authors to create a text-based exercise for their students to complete. Both the tutor creation of an exercise and the student use of the subsequent exercise are completed by a web interface. The system is now fully operational and is presently available for use without registration or charge. It can be accessed at the URL presented below, which provides access to both the program and supporting material to provide assistance to staff and student users.

<http://www.dundee.ac.uk/learning/leu/ilt/selfpeer.htm>

Design of the program

To ensure successful use of the program by staff who are experts in specific disciplines, but not necessarily in self and peer assessment or internet technologies, it was clear that the program should be straightforward to use and clear in operation. The system also had a requirement for considerable functionality. To enable a “reusable learning object” to be created, there had to be a mechanism to save the work of tutors creating and deploying exercises, and a certain amount of flexibility is required to support tutors own innovative approaches, which frequently vary across different disciplines. The system is entirely web-based and operates from a PC, Mac or other hardware, and has been successfully tested with both Internet Explorer and Netscape Navigator.

Central to the development of the system has been feedback obtained from staff and student users, and also through the informative activity of observing people using the system. Some early problems arose, particularly through the use of the browser “Back” button, which because of the technical operation of the program potentially let one student submit multiple pieces of text, or permitted a single student to assess more than their share of their peers’ work. This issue has been dealt with.

The system has been developed from an original online self-assessment exercise that was created to cope with a lecture clash. It has taken about two years to develop and in the few months the system has been available it has been used in the teaching of history, surgery, staff development, biology and environmental sciences. Technically the program consists of a single master PERL program that receives instruction from the webform completed by the academic instructor. This central program then creates a set of HTML pages and additional PERL programs to process the specific exercise that has been created. Student text and assessment details are stored in flat files available for analysis and processing. This approach has proved useful as it allows gradual development and enhancement of the program, while ensuring that exercises in use by staff and students are not influenced by these developments. Feedback to the students uses email messages that are generated automatically using SendMail and short text messages by sending a specialist email message to a commercial text messaging service (textforce.co.uk).

The program is compatible with modern VLE systems such as Blackboard or WebCT, and it can also be used in a standalone mode. Initial student access to the system is best managed using a specially constructed link from within a VLE module where learning material to support the activity may be present. Instructions are provided on the subsequent pages presented to the students as they complete the exercise, but ideally they should be introduced to the exercise with a demonstration, process diagram and guidance text.

Operation of the Program

Use of the software by a lecturer and their students is outlined here and is summarised visually in Figure 1. Although both the construction and operation of the exercise should be kept as straightforward as possible, the multi-stage process required means that there is potential for confusion and lecturers should consider including an overall diagram of the process to their students at the outset of the exercise.

Lecturer and student activity stages

1. Lecturer designs the assessment question, any specific instructions, marking criteria (maximum of 12), mark distribution and a model answer.
2. Lecturer uses online system to create a Self and Peer Assessment Exercise.
3. Lecturer makes exercise available to students, simply using a VLE link, email or a web link.
4. Students complete the exercise
 - a. Students research the exercise and prepare their text answer.
 - b. Students individually submit their answer online.
 - c. Student completes their self assessment using criteria defined by the lecturer.
 - d. Student continues to mark two of their anonymous peer's work. Results are emailed anonymously to the original author. A text message can also be sent to the original author notifying them of their result.

- e. Students are presented with summary marks and offered a URL or additional instructions as required.
5. Lecturer may choose to mark all or a sample of the work.
6. A complete summary of the answers, marks awarded, comments made and mean overall marks is available to the lecturer for moderation or publication as required.

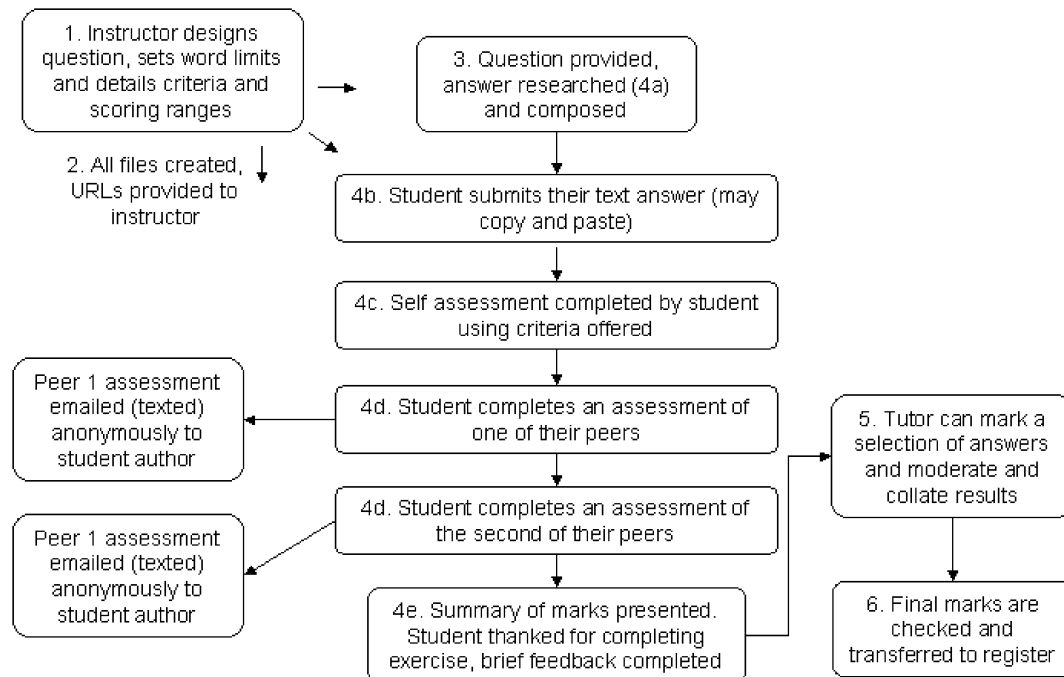


Figure 1. Outline of the operation of the online self and peer assessment program. The lecturer creates the text-based exercise, and students complete a series of web-based submissions to complete the exercise.

Guidance Provided to Students

It is normally appropriate to provide some guidance to students who are about to complete a self and peer assessment exercise. This guidance will include technical instructions for how to complete the exercise, but is likely to also include some guidance about how to carry out assessments and communicate the results of this assessment to their peers. The website supporting the program contains some guidance (700 words) that can be edited as required by instructors and provided to students to assist them in their learning about assessment. It specifically deals with describing criteria-based and norm-based reference assessment, which may be new concepts for students, and also provides advice on how to make constructive comments in feedback.

Innovative Features of the Software

This software is straightforward for instructors to use, although users are advised to test the system initially in a trial environment with their colleagues as test students. A second exercise can be created by reloading the stored question data, and this can be altered and fine tuned to create the final exercise. The instructor must submit a model answer which is presented to the students to enable them to see an excellent answer and to provide a reference as they complete their assessments. To provide the first two students completing the exercise with material to mark, the instructor must also enter two “student” answers.

When students start the exercise they enter their email address which provides a unique key for the program and the route for getting peer assessment results to the students. If a student attempts a second submission, they receive an error message informing them that their email address has already been used. To avoid use of the “Back” button, pages of the program are presented in a popup browser window with no menu and with right-click disabled. Students can thus move sequentially through the exercise as they should. If instructors wish, the students can be offered the facility to enter their mobile phone number (UK only at present) which will automatically receive a text communicating a summary of each peer and tutor mark awarded to the student (Figure 2). The email that is automatically sent to the student author following each peer or tutor assessment of their text contains a detailed description of how marks were awarded (according to the marking criteria provided) and also includes an open text response allowing assessors to provide justification and advice.



Figure 2. A novel feature of the software is the ability to have a text message sent to the student with a summary of each peer and tutor assessment.

Boud (1995) has described how staff moderation of self and peer assessment can be made efficient if it is focused on results which show significant discrepancy. Within this program it is possible to use the reporting features of the software to focus on results with significant variations between the different assessments. It is also possible to tutor mark a random selection of papers with the software, thereby encouraging accurate and careful assessment by the

student participants.

Further Developments

There will be further development of the self and peer assessment system in the future. Ideas enhancing the program's use include; linking images and multimedia; developing automatic plagiarism detection; automatically highlighting assessments with significant discrepancies, and using automated free-text marking. Some of the most significant developments and usage of the system is likely to come from new collaborations and the author welcomes requests for modifications or developments. It is possible to make the marking criteria available to the students before their final text is submitted, enabling an openness that may enhance quality in the text submissions. The operation of the program could be split into two periods so that all the text submissions are made first, and self and peer assessment is completed after the deadline for submission is complete. This method would ensure that the marking criteria are not released until all submissions are made, but suffers from the need to get the students back online to complete their assessments.

Students recognise the potential and significance of the approach. They are initially motivated to produce a quality answer and research their topic well. Before completing the assessments, they can be instructed in methods of assessment, and as they complete their self assessment they recognise how their work has deviated from the model answer and the marking criteria. Through peer assessment they recognise that others also face challenges in improving their own work, and they can assist with this process. They may be presented with examples of very good work, and will begin to understand how their approach to assessment can be improved. There is the risk that poorer or lazy students do not apply themselves to the peer assessment part of the exercise and it may be appropriate to include a facility within the software to reward careful and correct assessment. This could involve bonus marks, if peer assessors score similar to tutor assessors, or withholding assessment results until peer assessments are completed correctly.

Analysis of Use

The program is probably best understood by using an example to illustrate the setup, activity and analysis of the system. The specific example presented here is the first of two self and peer assessment exercises carried out with a first year Environmental Science class of 120 students. The students were timetabled to receive a two hour computer-based class and during this period, they were introduced to the exercise, encouraged to research the topic via the WWW and their lecture notes, able to compose their text-based response, submit their work and carry out the self and peer assessment exercise. They were given oral and written guidance to the exercise and the lecturer was on hand to deal with minor issues and assist with their research.

The following specific learning outcomes were presented to the students:

1. Understand how assessment of work may be completed by criterion-based referencing or norm-based referencing
2. Recognise that defined criteria used for marking work, can be predicted and fulfilled
3. State the typical assessment criteria that a piece of text may be judged on
4. Demonstrate skills of appraisal and assessment
5. Develop a deeper understanding of two important Environmental Science topics

The first exercise question was:

Write a short piece of text that describes the major scientific reasons why global warming is almost certainly occurring at present. The total length of the text must be between 200 to 300 words.

As the students carried out the exercise they were given guidance in the use of the internet for this form of research and provided with help as they composed their test answers. Some prepared carefully structured word-processed submissions, although it was pointed out that most formatting was going to be lost when the text was submitted by a webform. Students were encouraged to make their submissions and leave time for the self and peer assessment (30 minutes). They were generally observed to complete the assessment carefully, although for this initial exercise the depth and approach of assessment was clearly a new experience for many. Recognising that assessment is a learning activity, this was the first of two self and peer assessment exercises, and it was noted that their approach and outlook improved for the second exercise.

To illustrate how self and peer assessment is carried out the specific marking criteria for this exercise are reproduced here:

- Was the answer proofread carefully? Was it free of significant grammatical, spelling or typographical errors? (max. 10 marks)
- Did the text start with an introductory sentence and finish with a concluding sentence, and was it informative and easy to read? (max. 10 marks)
- Did the answer mention the greenhouse effect or glasshouse effect specifically? (10 marks)
- Did the answer mention the hypothetical nature of the theories, that nothing was absolutely certain? (5 marks)
- Did the answer explain the greenhouse effect and the involvement on long and short wave radiation? (10 marks)
- Did the answer specifically mention the key gases carbon dioxide, methane, nitrous oxide and fluorocarbons? (max. 10 marks)
- Did the answer mention that the greenhouse gases differ in their activity and impact on global warming? (10 marks)
- Did the answer mention that it is human activity that has largely lead to an increase in the concentration of greenhouse gases? (5 marks)

- ❑ Was an example gas included with specific data presented? (5 marks)
- ❑ Did the answer mention that natural events also influence climate change? (5 marks)
- ❑ Did the answer mention climate models that predict a specific (and appropriately correct) increase in global surface temperatures? (5 marks)
- ❑ Were some brief references included to give authority to the answer and also to permit further research to be carried out? (5 marks)
- ❑ Answer within word limits (200 to 300 words, checked automatically) (10 marks)

In addition to collecting self and peer assessment results for the exercise, each piece of student text was also assessed by a tutor, using an identical set of marking criteria. To provide a comparison with norm referenced assessment, the first 24 pieces of text and the model answer were provided anonymously to an expert marker, who was provided with the question, but not the specific marking criteria used. They were asked to complete assessment using norm-referencing of their own standard for the level of a first year Environmental Science student.

Table 1. Illustration of mark distribution obtained for three student answers that were assessed by the online criteria-based reference scheme (self, peer 1, peer 2 and tutor) and by an expert assessor using a norm-based scheme.

	Self (%)	Peer 1 (%)	Peer 2 (%)	Tutor (%)	Mean (%)	Expert (%)
Lowest	40	25	44	35	36	10
Moderate	58	60	61	47	56.5	50
Highest	73	81	79	85	79.7	70

The results of this exercise, and in particular the 24 students whose work was also marked by an expert using norm referencing, indicated that there was some variation in marking standards. In general, across the marking techniques used (self, peer, tutor and expert) poor work received a low mark, and very good work received a very good mark. Specific examples of marks awarded for the lowest, a moderate and the highest scores are presented in Table 1. The average deviation from the mean for all the criteria-based marking (self, peer 1, peer 2, and tutor) was 8.5 marks, which could be judged as acceptable, given the predominately formative nature of this particular exercise. The evidence presented here generally supports the conclusion that students can carry out accurate assessment when provided with specific marking criteria (Miller *et al.*, 1998).

There was significant variation between the criteria-based marks and the expert norm-based marks (Figure 3). In general, the expert marker applied a harsher system for awarding marks and was using the full marking scale to effectively discriminate answers. The criteria-based online scheme awarded marks for being within the word limits (10 marks), quality of grammar and spelling (maximum of 10 marks) and the form and readability of the text (maximum 10 marks). In this way, an answer that was scientifically poor may have gained a moderate mark. Again, given the formative and innovative nature of the exercise, it was probably correct that students who attempted the exercise were rewarded with getting some of the marks relatively easily. The model answer which scored 95 to 100% under the criteria-based marking scheme was also submitted to the expert for norm-based marking along with the student answers, and it was gratifying to see this answer receive the highest awarded expert mark of 95%.

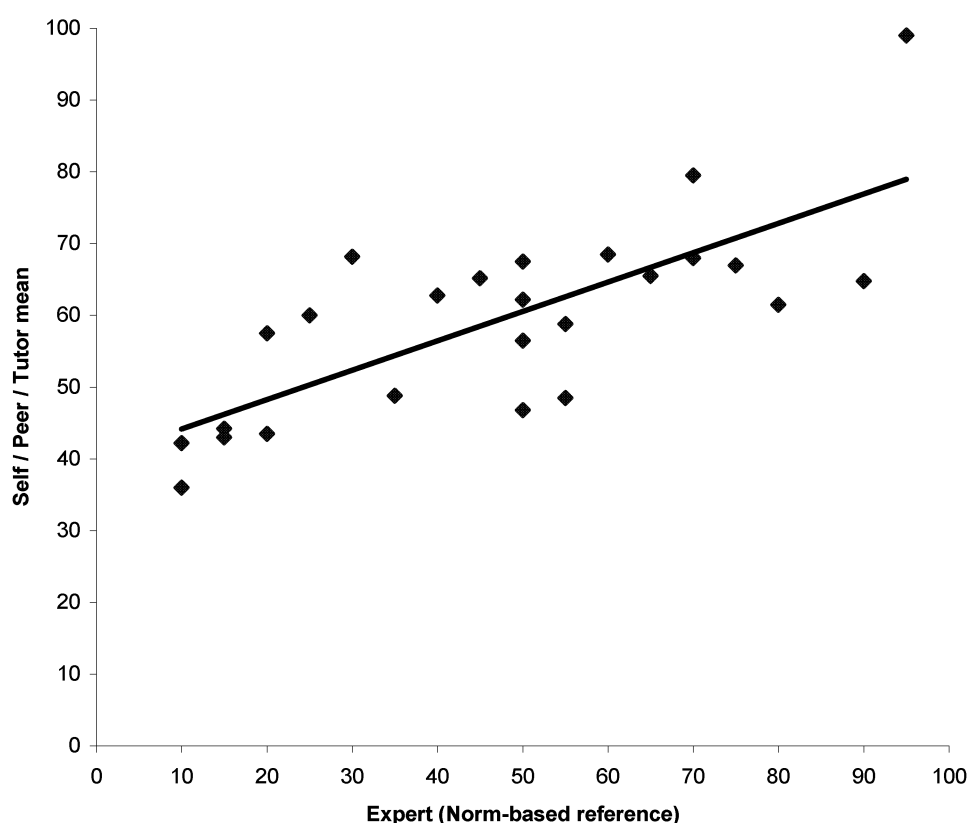


Figure 3. Plot of the mean mark obtained by the online criteria-based marking scheme against a single mark awarded by an expert using a norm-based reference scheme for 25 pieces of text work by Environmental Science students.

Student and Staff Evaluation

Students and staff have been asked to evaluate the system and many responses have been obtained by both the author and other lecturers who have used the system. Overall, users of the system are very positive towards it. Examples of student comments on the system are presented here.

- "A good computer program that more classes should use"
- "A great learning tool for the students. Enjoyable too, better than the labs"
- "Useful for understanding what markers look for in a piece of work"
- "It was interesting to see how the marking process takes place, and how you can have an essay that in its own right is well written but does not contain the points needed by the marker"
- "Assessing others gives me comparison to my own"
- "Good to see how we can improve our work by our mistakes"
- "Useful for the ability to receive results via text message"
- "It was good to be able to compare your own work with others and offered an insight into the basic processes staff must use"
- "Makes you realise that you need more depth in your answer and to read the question"
- "It is quite funky"

The Environmental Science students who had completed the exercise detailed above were surveyed for their opinions on the self and peer exercises they had completed. In response to specific questions, 70% agreed that the exercises were educationally valuable and 80% thought that the exercises were useful for learning about what criteria are used to mark work. Two other highly rated items were that it helped with understanding the overall assessment process (60%) and was useful for learning how to self assess work (61%). These results generally matched the learning outcomes for the exercise and represent strong educational support for the system.

Conclusions

Learning is about doing, and the self and peer assessment system described here represents a useful tool for teachers to encourage students in their learning, by actively involving them in their assessment. The role of the teacher is changed from one of setting an exercise, marking and giving feedback, to the more challenging activity of setting an exercise appropriate for self and peer assessment, defining the criteria for marking, teaching the students how to carry out fair assessment and managing and moderating the process of self and peer assessment.

A balance has to be reached between the summative marks awarded for the exercise and the extent of trust and moderation that is required. Fortunately the software assists with this as all marking is anonymous, but instructors have full access to the student submissions, marks and feedback. This system is not designed as a high security, high stakes exercise, it is predominately a student learning activity. For instructors, the significant

challenge remains both designing a text-based exercise appropriate for self and peer assessment and committing to the exercise with enthusiasm to ensure the students are focussed and empowered.

Acknowledgements

The author would like to thank his colleagues, Margaret Adamson, Hamlyn Jones and Jonathan Weyers, for comments on both the development of the program and the text of this paper. The LTSN Bioscience Centre is thanked for supporting development of the program.

References

Boud D. (1995). *Enhancing Learning through Self Assessment*. Kogan Page Ltd, London.

Dochy F, Segers M & Sluijsmans D. (1999). *The use of self-, peer and co-assessment in higher education: a review*. *Studies in Higher Ed.*, 24, 331-350.

Miller AH, Imrie BW & Cox K. (1998). *Student Assessment in Higher Education*. Kogan Page Ltd, London.

Topping K & Ehly S. (1998). *Peer-Assisted Learning*. Lawrence Erlbaum Associates Inc. Mahwah, NJ.

Stephani LAJ. (1994). *Peer, self and tutor assessment: relative reliabilities*. *Studies in Higher Ed.*, 19, 69-75.