

Towards a Canonical View of Peer Assessment

David E. Millard, Karen Fill, Lester Gilbert, Yvonne Howard,
Patrick Sinclair, Damilola O. Senbanjo, Gary B. Wills
*Learning Societies Lab, School of Electronics and Computer Science,
University of Southampton, Southampton, UK*
{dem, kf, lg3, ymh, pass, dos103, gbw}@ecs.soton.ac.uk

Abstract

Peer Assessment (or Peer Review) is a popular form of reciprocal assessment where students produce feedback, or grades, for each others work. Peer Assessment activities can be extremely varied with participants taking different roles at different stages of the process and materials passing between roles in sophisticated patterns. This variety makes designing Peer Assessment systems very challenging. In this paper we present a number of Peer Assessment case studies and show how a simple review cycle can be used as a building block to achieve the more complex cases. We then propose a Canonical Use Case for Peer Assessment, in which a Review Plan is used to describe how review cycles can be combined to achieve the required complexity.

1. Introduction

Peer Review, “is assessment of students by other students, both formative reviews to provide feedback and summative grading” [2]. Whilst there is a long history of formative peer review in universities, particularly in English and the Arts, in the last ten years increasingly innovative approaches, often involving group production and / or review of learning outputs, have been adopted in a wide range of disciplines [5, 12]. Bostock summarised the benefits of peer review as improving motivation; encouraging students to take responsibility for their own learning; treating assessment as part of learning, so that mistakes are opportunities rather than failures; practising transferable skills; providing a model for internal self-assessment of own learning; and encouraging deep rather than surface learning [2].

In this paper we develop a number of case studies of Peer Review in order to construct a canonical view. We outline a common peer review cycle that serves as a building block to support all of them, and show how

it can be used to construct one of the more complex case studies. We suggest that a system that can support this cycle with a varying number of participants at each stage could potentially support all peer review processes, and present a use case that encapsulates the complexity of the cycles in a Peer Review Plan.

2. Background

Since the early 1990s, several computer systems have been developed for performing peer assessment exercises. An early project was MUCH (Many Using and Creating Hypermedia) [10]. Other systems include Peers [9], OASYS [1] and Self and Peer Assessment Resource Kit [4].

More generic Peer Review systems have also been developed. Peer Grader [5] is a web-based peer assessment system that allows students to grade the assignments of other students. Several assignment styles are supported, such as reviewing research papers, researching material on the web, and annotating lecture notes. The system allows author-reviewer mapping patterns to be generated automatically or manually by the lecturer. Students are able to resubmit their work once they have received feedback from their peers.

Some systems have formalized these patterns. For example, Computer Supported Collaborative/Cooperative Learning (CSCL) collaboration scripts [3] have inspired the design of peer assessment systems. OPAS (Online Peer Assessment System) uses a collaboration script to describe the different stages of collaborative peer assessment. Fast (Flexible Assignment System) [11] allows different scripted collaborative learning, including peer assessment, to be planned and executed through a web-based interface. Miao and Koper [8] present an approach for a peer assessment system based on open e-learning standards in order to script peer assessment processes.

In this paper, we are less concerned with the form of particular scripting languages, than a canonical view of

peer assessment itself (which may later underpin such a language). In the following case studies and analysis we suggest a common peer review cycle that can be used to build up more complex scenarios.

3. Case Studies

In order to analyse Peer Review we have looked at a number of case studies. This led us to four factors which might describe a Peer Review process, the number of authors, the number of artefacts that those authors create, the number of reviewers, and the number of reviews that those reviewers return. We mapped our case studies to reflect a range of possible values for these factors, shown in Table 1.

Table 1: Case studies mapped to factors

	Single Author	Multi Author
Single Artefact	Simple (A), Round Robin (B)	Group Activity (C) Group Review (D)
Multi Artefact	Multiplicity (F)	Committee (E) Multiplicity (F)

	Single Reviewer	Multi Reviewer
Single Review	Simple (A) Group Activity (C)	Committee (E) Group Review (D)
Multi Review	Multiplicity (F)	Round Robin (B) Multiplicity (F)

Case Study A: Simple Peer Review

In the simplest form of peer review, authors and reviewers are paired together [7]:

- There are n student *authors* who are also *reviewers*
- authors create one *artefact* each
- reviewers review one artefact each
- reviewers send *feedback* to author and to *tutor*
- the tutor awards a mark to the author.

Case Study B: Round Robin

In Round Robin peer review, participants are grouped, and each participant reviews the work of every other participant in their group. An example, is review reported by Bostock [2]:

- there are n student *authors*
- there are n/y review groups ($y=$ four in Bostock)
- authors create one artefact each
- each reviewer in the group sends feedback and a mark (%) to the author
- authors then revise their artefacts
- each reviewer in the group reviews the artefact again and sends a final mark to the tutor

- tutor compiles all the marks, re-marks the artefacts themselves and allocates a final mark to the author based on all the marks.

Case Study C: Group Activity

Here, a group of authors work together to produce an artefact that is then reviewed by a third party. One example, from the authors' own experience, is an MSc Supervised Work Session:

- x groups of y students
- each group produces an artefact
- tutor reviews the artefact and awards a grade.

Case Study D: Group Review

In Group Review, a group of authors work together to produce an artefact, and then individually review the efforts of their group. An example of this is reported by Gregory *et al* [6]:

- there are n students
- there are n/y groups ($y=$ four in the real case)
- the group creates one *artefact*
- tutor gives preliminary feedback
- the group submits final artefact
- each student submits a self-assessment and an assessment of the group effort
- group meets with tutor to discuss efforts and suggested marks
- tutor allocates final marks
- students receive marks and detailed feedback.

Case Study E: Committee Review

Here, a group of reviewers consider several different artefacts in order to produce one review. One example, from the authors' own experience, is a Multimedia course based around a conference:

- n students
- each student creates an artefact
- each student reviews 4 artefacts
- each student summarises the reviews for a fifth artefact and produces a summary
- the committee passes or fails the original artefact based on the artefact, the reviews and the summary.

Case Study F: Multiplicity

Multiplicity involves multiple authors who create multiple artefacts which are then independently reviewed by multiple reviewers. For example, Wheeler [12] where students give a presentation and answer questions and are assessed by their classmates on both:

- n students, m tutors
- each student delivers a presentation (artefact 1) and are grouped to answer questions (artefact 2)

- students and tutors review/mark the presentations
- only tutors review/mark the answers .

4. A Canonical View

The main structure of a peer review activity is defined by the way in which resources are generated and flow between participants. In addition to the number of authors, artefacts, reviewers and reviews, there are several other factors governing this flow. For example, whatever the numbers of authors and reviewers there must be a process for matching them up (typically so that each reviewer has the same number of non-identical artefacts to review).

Another factor is mark allocation. Marks could be allocated to artefacts regardless of reviews (i.e. the reviews exist purely to give feedback to authors), or in order to help mark the artefacts (either by asking reviewers to assign the mark, or by taking their comments into consideration), or allocated to reviewers for the quality of their reviews.

If we assume that matching and marking occurs outside of the peer review system (for example, by a grouping system, or via a marks spreadsheet) then we can say that the core of Peer Review is concerned with moving resources from those that generate them to those that receive them.

4.1. The Peer Review Cycle

We believe that all Peer Review processes are constructed from combinations of the same basic cycle of Generate, Submit and Distribute, shown in Figure 1.

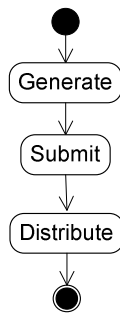


Figure 1: The basic review cycle

For example the Simple Case (A above) can be described as two iterations of the cycle. The author *Generates* an Artefact which is then *Submitted* and *Distributed* to the reviewers. The reviewers then *Generate* a review which they then *Submit* and is *Distributed* back to the artefact's author.

The complexity of Peer Review is accounted for in three ways:

1. The cycle can be started in any one of its three states. For example, to begin an activity the student may be asked to *Generate* an artefact, to *Submit* an existing artefact, or the tutor may provide it, in which case the first task is to *Distribute* it.
2. The cycles can be interleaved, and can occur in parallel as well as in sequence.
3. Each stage within the process may involve 1...n participants (authors/tutors/reviewers), producing 1...m resources (artefacts/reviews/marks).

4.2. Analysing Multiplicity as Cycles

Multiplicity is one of our more complex peer review examples, as it involves multiple artefacts, and multiple reviewers producing multiple reviews. Although in Wheeler [12] it is run as a real world exercise it is possible to imagine a digital version. Figure 2 shows a UML Activity Diagram for the Wheeler case study (F) broken into the different cycles. There are six cycles in total, the first three occur in sequence and the last three in parallel. Note that cycles five and six overlap because they share the same Distribute activity.

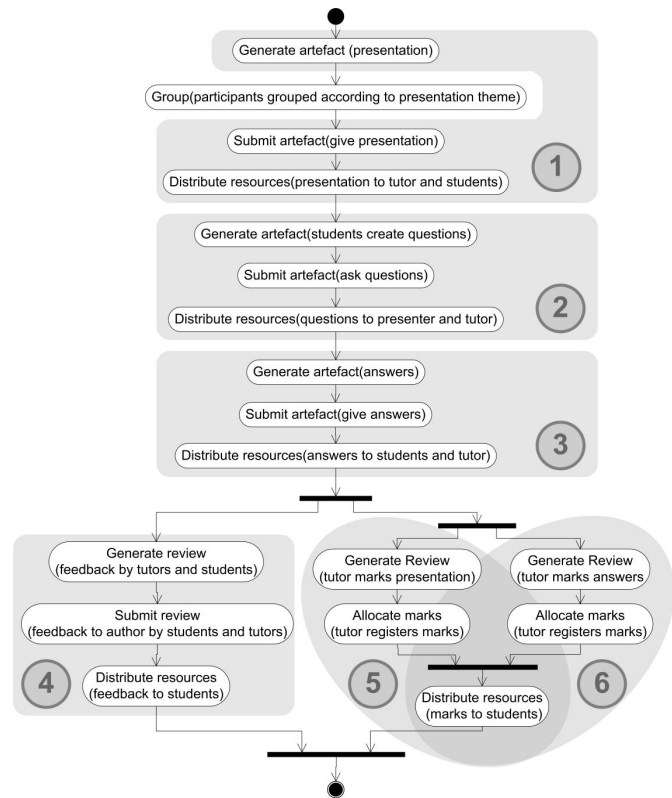


Figure 2: Activity diagram of case study F broken into six overlapping cycles

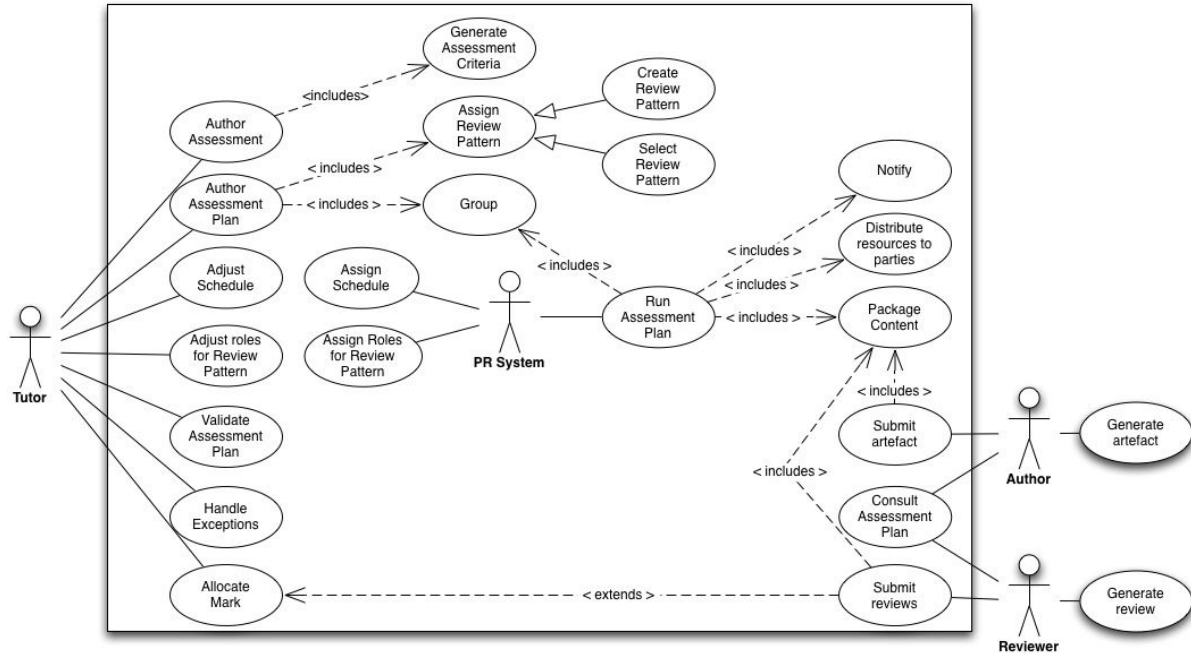


Figure 3: A canonical use case for peer assessment

We originally classified our case studies using the number of authors, artefacts, reviewers and reviews. However, we can now see this is because we were identifying two review cycles in sequence; the first follows the authors' artefacts to reviewers, the second the reviewers' reviews to authors. In a single cycle the characteristics are simply the number of people who *create* resources (authors or reviewers), the number of *resources* they create (artefacts or reviews) and the number of people who *receive* those resources (reviewers or authors). Table 2 shows the value of these characteristics for each cycle in Figure 2 where n is the number of classmate reviewers, and m is the number of questions.

Table 2: Number of creators, resources and receivers in case study F cycles

Cycle	Creators (Authors/ Reviewers)	Resources (Artefacts/ Reviews)	Receivers (Reviewers/ Authors)
1	1	1	$n+1$
2	n	m	2
3	1	m	$n+1$
4	$n+1$	$n+1$	1
5	1	1	1
6	1	1	1

5. Use Case for Generic Peer Review

A generic peer review system capable of complex configurations of review cycles, with $1..n$ Creators, Resources and Receivers in each cycle, should be able to support all the forms of peer review; although, as discussed in Section 4, there are some algorithms concerning creating groups and matching roles that are not fully generic and may need modular support.

Figure 3 shows the Use Case for a generic peer review system. The complexity of the system is encapsulated in a *Peer Review Plan* with three elements:

- A *Peer Review Pattern* (an ordered description of the cycles of peer review and the roles of the participants in each cycle).
- A number of actual *Participants* (possibly arranged into *Groups*) that populate the roles in the plan.
- A *Schedule* of upcoming dates and times, that ties the pattern to a real timescale.

It is the tutor's responsibility to construct, adjust and validate this plan. The tutor must also handle exceptions in the peer review process (for example, a student missing a deadline).

A proactive Peer Review System is at the heart of the process. This executes the plan according to the schedule, notifies the participants as to their tasks, allows them to submit artefacts, reviews and marks, and distributes these according to the pattern. The Peer

Review System can also help the tutor author the plan (for example, by suggesting a schedule based on a start date and the pattern, and roles for participants based on the pattern). Students (Authors and Reviewers) receive notifications from the system, and submit their artefacts and reviews via it. They can also consult the plan, to see what is expected of them.

The Group use case is primarily part of the Author Assessment Plan use case, describing the process of allocating people roles, and/or placing them into groups. However, not all grouping can be done *a priori*; for example in Wheater [12] authors are grouped according to the theme of their presentations, but this is not known until after their presentations have been submitted. For this reason the Group use case is also part of the Running an Assessment Plan use case.

It is also important to note that the Allocate Marks use case is an extension of the Submit Review use case. This is because we regard a mark as a type of review; the difference is that in Allocate Mark we assume that as well as being submitted (for processing according to the pattern) it is also formally recorded in some way (such as being entered into a University marks system).

6. Conclusions and Future Work

In this paper we have presented a number of case studies of Peer Review and explained how they can all be described using different configurations of a simple cycle of Generate/Submit/Distribute, as long as each stage can deal with 1...n Creators, Resources and Receivers. We have suggested that a system could support all these combinations of cycles if it encapsulated the complexity in a review plan comprised of a review pattern, schedule and list of participants and groups. We then presented a use case diagram for such a system.

We intend to develop the use case as the basis for a web-service system called *PeerPigeon* to deal with the submission and distribution aspects of the use case. PeerPigeon will have a web-based interface, and will allow participants to register for email notifications. The biggest challenge in realising the system will be to choose a representation format for the review pattern. LMS Learning Design (LD) and Question Test Interoperability (QTI) are two e-learning standards that might be relevant. CSCL scripts might also be useful, as well as principles from process languages such as BEPL or WS-Configuration. Initially, we will allow tutors to choose from pre-loaded patterns, but hope to develop an authoring tool in the future.

Peer Review is a valuable tool for teachers, that not only has intrinsic pedagogical value, but can also

support teachers who are have limited time and resources. We hope that in the future peer review systems will allow tutors to set up peer review activities more easily, and thus enable them to use peer review in a new and lightweight manner

7. References

- [1] Bhalerao, A. and Ward, A. "Towards Electronically Assisted Peer Assessment: A Case Study", *ALT-J volume 9 no 1* (2001), pp. 26-37
- [2] Bostock S., "Student peer assessment", *Higher Education Academy Article*, 16 Mar 2001
- [3] Dillenbourg, P., "Over-scripting CSCL: The risks of blending collaborative learning with instructional design.", *Three worlds of CSCL. Can we support CSCL*, P.A. Kirschner, eds., 2002, pp.61-91.
- [4] Freeman, M. and J. McKenzie (2002). "Implementing and evaluating SPARK, a confidential web-based template for self and peer assessment of student teamwork: benefits of evaluating across different subjects". *British Journal of Educational Technology* 33(5): 553-572.
- [5] Gehringer, E.F. "Strategies and Mechanisms for Electronic Peer Review". *Frontiers in Education Conference FIE'2000 30 Annual, Kansas City, MO, October 18–21, 2000*, IEEE Education/Computer Society, 1, F1B/2–F1B/7
- [6] Gregory, A., Yeomans, E. and Powell, J. "Peer Assessment and Enhancing Students' Learning", in *Learning and Teaching for Business: Case Studies in Successful Innovation* (eds. Kaye, R. and Hawkrige, D), London, Kogan Page, 2003
- [7] Mangelsdorf K. "Peer reviews in the ESL composition classroom: What do the students think?", *ELT Journal* 1992 46(3):274-284; doi:10.1093/elt/46.3.274, 1992
- [8] Miao Y., Tattersall C., Schoonenboom J., Stevanov K. and Aleksieva-Petrova A., Using open technical e-learning standards and service-orientation to support new forms of e-assessment, *TENCompetence OPen Workshop*, Manchester, 2007.
- [9] Ngu, A. H. H., J. Shepherd, et al. "Engineering the 'Peers' system: the development of a computer-assisted approach to peer assessment." *Research and Development in Higher Education* 18: 582-587. 1995
- [10] Rushton, C., Ramsey, P. and Rada, R., "Peer assessment in a collaborative hypermedia environment: a case study", *Journal of Computer-Based Instruction*, vol 20(3), 75. 1993
- [11] Topcuoglu, H., "FAST - Flexible Assignment System", Proceedings of E-Learn 2006, Honolulu, Hawaii .
- [12] Wheater C.P, Langan A.M. and Dunleavy P.J., "Students assessing student: case studies on peer assessment", *Planet*, 15, 2005, pp. 13-15.