



Purchase, H. and Hamer, J. (2018) Perspectives on peer-review: eight years of Aropä. *Assessment and Evaluation in Higher Education*, 43(3), pp. 473-487. (doi:[10.1080/02602938.2017.1359819](https://doi.org/10.1080/02602938.2017.1359819))

This is the author's final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/144627/>

Deposited on: 25 July 2017

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

Perspectives on Peer-Review: Eight Years of Aropä

Helen C. Purchase^{a*} and John Hamer^a

^aSchool of Computing Science, University of Glasgow, Glasgow, UK

Correspondence details:

School of Computing Science

University of Glasgow

18 Lilybank Gardens

Glasgow G12 8QQ

UK

Email for corresponding author: helen.purchase@glasgow.ac.uk

Biographical notes:

Dr Helen Purchase started her interest in e-learning with a PhD in Intelligent Tutoring Systems, and has several publications relating to the use of technology in education. As well as co-creator of Aropä, she leads the GUSTTO project – a social media system for practising academics to share ‘Teaching Tips’ with their colleagues.

Dr John Hamer received his PhD in Computer Science from the University of Auckland in 1990, where he served as Senior Lecturer until 2010. He initially created Aropä to support his own teaching. He currently works as a commercial software developer, and has an honorary affiliation with the University of Glasgow.

Word count: 8845 (including references and appendix)

Perspectives on Peer-Review: Eight Years of Aropä

Drawing on eight years of observation and correspondence from the Aropä project, we report on the issues important to academics who conduct on-line student peer-review activities, and the features they request to support their own instructional designs. The Aropä project is unusually broad, having so far supported over one hundred instructors at twenty institutions across nine countries, almost a thousand activities, and more than 36 thousand individual students. As the designers, developers, maintainers and advisors of an evolving and widely used tool, we use our unique position to report on the perspectives and priorities of instructors and students in this important and developing field.

Keywords: online peer-review, Aropä

Introduction

Aropä is an online system that has provided free support for peer-review activities continuously since 2009, worldwide. Aropä is wholly designed, developed, and maintained by the authors of this paper, both academic computing science staff members at a Scottish university. It has significant global reach: from Australia and New Zealand to the US and Canada, including Malaysia, Spain and Slovakia in between. Its use covers a wide range of subjects: anthropology and accounting, classics and commercial law, human biology and health policy, palaeontology and pharmacology. To date, the system has supported the development of critical thinking and reflective skills in over 36 thousand students worldwide.

Up to 31st December 2016, 961 successful peer-review assignments devised by 106 instructors have been supported at twenty institutions in nine countries. In the last two academic years, nearly 13 thousand individual students used the system to write reviews on their peers' work. Large classes are easily supported; for example, in August 2015, 948 students each submitted a report for a Commercial Law course at The

University of Auckland, and together wrote a total of 2,716 reviews of their peers' work within the period of a week.

The authors of this paper are the sole designers, developers and maintainers of the Aropä system. We have complete control, can observe all activity, and are in direct communication with the instructors who use it. We are therefore in a unique position to report on the range and scope of the peer-review activities, including trends, emerging issues and individualisation.

During the past eight years, we have had hundreds of conversations about peer-review with over 70 instructors in at least 26 different subject areas. These conversations reveal the wide range of issues that instructors and their students consider important for successful peer-review: some of them relate to the peer-review activity itself, some focus on the requirements for an online peer-review system. In many cases, instructors have requested enhancements to the system so as to support their own assessment design. In this paper, we report on the issues arising from our interactions with instructors, classifying them with respect to operations, pedagogy and administration. Based on this data and our own observations, we conclude by discussing the dominant perspectives of both instructors and students when conducting successful peer-review activities.

Peer-review and peer-review systems

Peer-review

In its basic form, the process of peer-review entails an author submitting an artefact (such as an essay, a research article, a computer program, presentation, etc.), and their peers stating their opinion of it. This process is common practise both in academia (especially with respect to research publication) and in industry. In the case of student

peer-review, students are typically first provided with an assignment specification, they produce the artefact as required by the specification, and then submit it (as would be the case with typical assessment activities). The set of student submissions are distributed amongst all students, and each student provides feedback on one or more of their peers' submissions, with reference to a given set of criteria. At the end of this process, students reflect on the range of feedback that has been provided on their own submission.

Students therefore take on a variety of roles:

- as author – in submitting an assignment;
- as reader – in examining submissions written by peers;
- as critic – in judging submissions against a set of criteria; and
- as the criticised – in reflecting on feedback from peers.

The benefits of peer-review

Taylor et al. (2015) summarise the benefits of peer-review as being four-fold:

- the social context of learning as a process of creating consensus creates an active learning environment where students and instructors build knowledge together as a Community of Practice (Wenger, 1998);
- breaking down barriers between students from diverse backgrounds fosters development of peer-learning networks, with peers acting as role-models;
- the provision of extensive formative feedback (without increasing staff marking loads); and
- the noticeable improvement in students' performance as a result of the peer-review activity as demonstrated by several research studies.

Several articles demonstrate performance benefits through experience reports in different subjects: including, for example, nursing (Ramm et al, 2015), science writing (Walker & Sampson, 2013), biology, calculus and physiology (Reinholz, 2016), physics (Price et al., 2016) and instructional design (Brill, 2016).

Performance benefits of student peer-review activities include the development of evaluation, critical thinking, written expression and other transferrable skills. Weaver et al (2016) demonstrate an increase in student performance in the areas of science literacy and critical thinking as a result of peer-review activities in a biology research course. In their analysis of peer-review activities in a critical skills course, Tighe-Mooney et al (2016, p14) identified the development of ‘critical skills including, analysis and evaluation, making judgments and providing evidence and support for decisions made’ as well as improvements in the ability to ‘communicate, both in writing and verbally, to actively give and receive feedback, as well as reflect on the process and their input into it.’

Li et al (2010), writing in relation to student projects in teacher education, demonstrate that it is the providing of peer feedback (rather than the receiving) that has greatest effect on final performance; that is, the quality of the reviews written by a student were significantly correlated with their final project score, with no relationship found between the quality of the reviews received by a student and final project score. This observation is reinforced by Nicol et al (2014), who researched peer-review activities of undergraduate engineering students, and identified the wide range of cognitive activities involved in creating feedback.

The scholarly literature describing peer-review activities is mostly positive. A few authors mention drawbacks or problems that they have experienced during their own peer-review activities. Mulder et al (2014) discuss the problem of students giving

poor-quality reviews, and emphasise the importance of training students to write reviews and of providing an appropriately structured marking rubric. Crowe et al (2015) note that in-class peer-review does not improve students' performance, and suggests that it wastes valuable contact time. Cassidy (2006) suggests that students might object to peer-review on the basis that other students might 'steal their ideas' – although he does not find concrete evidence to support this view.

Peer-review tools

Several online tools for peer-review have been developed in the past couple of decades; in many cases, these have been developed in-house by computing science academics for use in their own classes. Useful reviews of these tools can be found in Hains-Wesson (2014) and Luxton-Reilly (2009).

Four systems initially developed and used in-house have over the years been distributed and used more widely. SWoRD (Cho & Schunn, 2007) and PeerScholar (Paré & Joordens, 2008) are now licenced pay-for-use products (Panther Learning and Pearson respectively). PRAZE (Mulder & Pearce, 2007), and Aropä (Hamer, Purchase et al, 2014) are offered through their university connections (Melbourne, Auckland/Glasgow respectively), with facilities to extend their use to other universities worldwide.

PRAZE is administered and managed by the University of Melbourne Learning Environments Unit, and offers extensive online documentation; use at other universities was discontinued in 2015. Aropä is hosted by the universities of Glasgow and Auckland, and administered, managed and supported by the authors; although it is open source, it has proven to be more effective if hosted locally.

The core features provided by all of these systems are:

- Anonymity. Authors should not know who their reviewers are; allowing reviewers to know who the authors are is occasionally useful.
- Allocations. The instructor should be able to specify the number of reviews each student should complete.
- Reviewing criteria. Instructors should be able to specify a rubric for reviewers to follow.
- Mark calculation. Where the rubric includes numeric marks, an overall mark for every student can be calculated based on a (possibly weighted) average of the marks given by student reviewers.
- Submission methods. It should be possible to specify one or more types of file required for submission.

The four tools listed above that have been extended for use after publication as research articles all largely satisfy these core criteria. A recent (2017) introduction to the online peer-review landscape is PeerGrade (<https://intercom.help/peergrade/>) administered by David Kofoed Wind from The Technical University of Denmark; there is evidence of it having been used at several universities so far. Unlike the other systems mentioned above, PeerGrade did not initially arise as the outcome of a situated academic research project.

The Aropä system

Aropä, created in 2007, was originally intended for use by computing science academics. In 2009, its use was extended to other subjects, and in 2010 provision was made for multi-institutional access. The steps comprising an Aropä assignment are:

- (1) The instructor specifies the

- assignment name
 - submission and review deadlines
 - a rubric to guide students in their reviewing
 - author/reviewer allocation method (random, groups, tags).
- (2) Students work on their submissions, and submit them to Aropä.
 - (3) At the submission deadline, reviewers are automatically allocated to submissions. A pairing between a submission and a reviewer is known as an ‘allocation’.
 - (4) Students use the system to read their peers’ submissions, and write reviews based on the rubric.
 - (5) After the review deadline, students read their peers’ feedback.

Aropä is hosted on two servers, at the Universities of Glasgow and Auckland, and has run continuously since early 2009. The system is entirely web-based, and there is nothing for instructors or students to install. There is no requirement to integrate with existing institutional authentication systems or course management systems: Aropä can be run as a self-contained system. However, support for several common institutional authentication protocols (e.g., LTI (IMS Global, 2012) and LDAP (UnboundID, 2015)) is provided. Class lists are entered by the instructor, using a simple cut-and-paste from an existing class roll.

Perspectives on peer-review

We are fortunate to have direct communication with all system users and have gathered extensive data on issues that matter to academics when conducting peer-review exercises, and the features they deem important to a successful peer-review activity. In particular, we have frequently implemented specific feature requests for individual

users; requests for new features are suggestive of the particular manner in which instructors view the purpose and risks of peer-review. Requests vary from straightforward infrastructure requests (e.g. increasing the maximum file upload size), through administrative issues (e.g. being able to download all submissions) to features specifically intended to enhance the pedagogy (e.g. allowing students to respond to their reviewers).

The instructors who use our system are typically peer-review novices who want to try out a new pedagogy in their class: in three cases we are aware that an Aropä assignment had previously been conducted manually or using an alternative technology. The level of experience of these instructors varies from newly appointed academics to established staff.

Data Analysis

We analysed eight years of email correspondence with Aropä users (from the start of the project until 31st December 2016) – a total of 2,258 emails – the aim being to identify:

- system features deemed important to instructors for successful peer-review;
- discussions about the pedagogical motivation for the features; and
- whether the features discussed already existed in the system, were implemented on request, or were justifiably not implemented.

A total of 203 emails contained relevant information – these were emails from 71 people: 60 instructors, 1 student and 10 members of staff from university Centres for Learning and Teaching (or institutional equivalent). These emails provide naturally occurring data; that is, the data was not collected systematically (e.g. by questionnaire) or collected at regular intervals. We are therefore unable to make any comparative

observations (relating, for example, to differences between disciplines, institutions or trends).

69 different features were discovered in these emails, and classified according to:

- features that existed in the basic (pre-2009) version: emails where instructors were seeking confirmation that a feature was available in the system (19);
- new features requested and implemented by us (29);
- new features requested, but for which we have a good reason for not implementing (8);
- new features requested, and which are currently on our wish list (13).

We note that these emails do not (and cannot) represent all the communication we have had with the instructors over the years, since we regularly meet with users face-to-face, or talk on the phone. Additional categories are therefore based on our recollections of verbal communication with individuals. We acknowledge the anecdotal nature of this informal, undocumented data source. However, since this article presents our personal experiences of actively supporting peer-review instructors over many years, it does not make sense to ignore it. These two additional categories are:

- features that existing in the basic (pre-2009) version, that had been included based on our own pre-2009 experiences and those of computing science colleagues (6);
- new features requested verbally since 2009, and implemented by us (6).

The 81 features are described in detail in the Appendix.

Thematic classification

The 81 features were placed into three themes, based on their motivation: *operational expediency*, *enhanced pedagogy*, and *administrative ease*. The first two of these themes are discussed below by highlighting the features identified in the data, together with brief case studies of interesting cases. The 26 administrative features are less pedagogically interesting, and so are omitted from the discussion; they are, however, included in the table in the Appendix.

Operational expediency

Flexible assessment design. The basic model, in which all students in a class submit a single piece of work by a specified deadline and then all complete reviews by a further deadline, is actually quite rare: instructors all have their own activity design, and all require flexibility in specifying the parameters. Aside from the basic parameters of submission and reviewing deadlines, submission file type(s), and the number of submissions that each student should review, different assignment practices include allowing students to submit on behalf of a group, having students review as a group, students associating a ‘tag’ with their submission, and having allocations created only between students in specified teams.

Instructors want complete control over the form of the rubric (including both open and closed responses), and over the reviewing allocations – while the automatic creation of allocations is useful, instructors like to be able to make manual changes if necessary.

An early feature was a facility for giving extensions, so that students can submit or review after official deadlines. Aropä automatically adjusts the allocations to take extensions into account (including the ‘tag’ of the late submission if necessary). This

facility is regularly used; however, Aropä is non-forgiving for students who submit just after a deadline – a couple of seconds will mean the difference between being able to submit or not. Many instructors are happy to allow students to submit ‘just a little bit’ late: allowing instructors to specify a brief ‘implicit’ extension period is one of our planned enhancements.

Two particular features were implemented on request to improve flexibility:

- An engineering mathematics instructor wished reviewers to annotate submissions. A document can be uploaded into a review allowing reviewing students to print their allocated submissions, hand-write comments onto the documents, and then scan and upload these into the Aropä review form. We understand this feature is no longer used as the instructor had difficulty persuading his class of the benefits of peer-review - the additional administrative load for the students in printing and scanning documents was seen as unreasonable. Tablet and PDF annotation technology may lead to greater future use of this feature.
- An urban studies instructor wanted her students to write their assignment under exam conditions, in a computer laboratory with no internet access. The students’ files (each named by their student number) can then be uploaded in into Aropä in a zip file – removing the requirement that each student submit their document separately. The review process then continues as usual.

Two flexibility requests we have not implemented: specifying different rubrics for peer-reviewers and tutor-markers (since Aropä is primarily a peer-review system, not an online marking tool), and allowing instructors to specify in which order students review their allocated submissions (since it is not clear why this would be necessary, and this

would remove control from the student).

Most instructors have very specific requirements for their own assignments, and do not like being squeezed into an alternative model simply because it is the only one that the technology supports: flexibility in assessment design is crucial.

Monitoring. Instructors like to be able to monitor student activity during the assignment so that they can see which students have uploaded submissions, which reviewers have written reviews, whether students with extensions have met revised deadlines, and which students have read their reviews. These tasks are particularly important for first-time instructors (who are typically anxious as to whether the students will engage in the activity) and if the instructor wishes to send reminders to the class if progress is slow, or to particular tardy students.

We implemented on request a facility for all details about review and read-review activity to be downloaded into a spreadsheet, for the case where an instructor was awarding summative marks for review activity; as a result of this, the system also now indicates when a review has only been partially completed.

Instructors particularly value the ability to ‘impersonate’ a student in Aropä – that is, see the system as a student sees it – particularly useful when a student says that things have gone wrong with their submission or reviews, and for testing purposes. The impersonate feature also allows an instructor to act on behalf of a student.

Encouraging good student behaviour. Much of the success of a peer-review activity depends on students doing what is expected of them: at the simplest level, Aropä warns students if they have uploaded a file of the incorrect type.

More serious, however, is the case where students do not complete all reviews allocated to them. This happens particularly when the peer-review activity is optional –

indeed, we can typically easily identify when an Aropä peer-review assignment is voluntary, since usually around half of the class submit their work, and less than a third of those who have submitted complete reviews. But even in non-voluntary summative assignments, a small percentage of students do not complete their reviews. Many instructors consider it unfair that such students receive feedback on their work, having not provided any for their peers. We implemented a feature whereby students are not able to access the reviews of their own work unless they have completed at least one (or all) of the reviews allocated to them.

Even when reviews are provided, they are sometimes sparse. However, we declined a request from a biology instructor for the system to prevent a student saving their review unless the comment boxes contained a minimum number of words, on the grounds that if student is prevented from submitting a review because a comment is not long enough, we would not want to encourage them to simply fill a word quota with irrelevant text. We want students to think carefully about what they write, and setting a minimum comment length suggests we do not care what they write (simply how much they write).

Fairness to students. Instructors are very aware that any assessment deemed ‘unfair’ by students will result in complaints – the basis of assessment is that all students are treated equally. Students complain if their workload is greater than their peers: Aropä’s allocation algorithm ensures that all students have the same number of reviews to write; in the rare cases where this is not the case (caused by, for example, highly unbalanced group size membership or several late submission extensions), the instructor can see this – we intend to include a feature where imbalanced allocations are explicitly highlighted. If the instructor knows the allocations are unbalanced, they can be adjusted manually, or the affected students can be warned in advance of the inequality.

Some features ensure, as much as possible, that reviewers provide ‘fair’ assessments of their peers’ work. An instructor can indicate that only those students who have submitted may take part in the review process – ensuring that all reviewers understand the nature and scope of the submission. If there is a range of essay topics, submissions can be ‘tagged’ so that students only write reviews on the topic of their own essay.

Reviewers are required to download the submission before they are permitted to review it – we cannot be sure that they have read the submission after download, of course, but this does prevent a student from completing a nonsense review without accessing the submission at all.

The process depends on anonymity: authors never know the identity of their reviewers, allowing reviewers to express their opinion honestly. The system cannot, however, ensure complete author anonymity if a student puts their name at the top of their document, or labels the file with their own name – although we plan to strip meta-data from submitted files where possible.

Enhanced pedagogy

Focus on quantitative feedback . Rubrics can contain comment items (reviewers enter free text) as well as radio buttons (reviewers choose one option from a given set) – providing both qualitative and quantitative feedback. If there is more than one set of radio buttons in a rubric, they are distinguished by different colours. There is increasing use by instructors of quantitative responses in the reviews; as a result, instructors can now associate marks with each radio button option; thus, for example, in the case of a set of three radio buttons, 10 marks might be given for the ‘excellent’ option, 8 marks for ‘very good’, and 2 marks for ‘fair’.

Associating marks with radio buttons increases the flexibility of the rubric, especially if the quantitative marks given by reviewers downloaded are used in any formative or summative manner. Aropä will not only aggregate marks for each student – for each review, and over all reviews – it also highlights large discrepancies between reviewers’ quantitative assessments. Outlying marks can be identified and removed from aggregation, and, indeed, the entire set of marks from a ‘rogue reviewer’ can be ignored. This process reassures instructors of fairness when reporting aggregated results.

We have resisted, however, allowing rubric items that solely represent marks. An engineering instructor wished to have a rubric feature where the student would specify a mark (e.g., a value from 1 to 10) using a drop-down list. We resisted this on the basis that giving marks out of ten (for example) suggests that there are ten distinct levels of achievement, and a drop-down list of numbers does not give space for describing these different achievement levels. Thus, there is more chance of variability in the marks given by students, since it is not clear what each mark (from one to ten) actually means. The existing radio-button rubric elements require that a descriptive label be associated with each option.

Graduate attribute skills . There is increasing emphasis on facilitating the development of ‘Graduate Attributes’ or ‘Transferrable Skills’ as part of an undergraduate degree programme (Hager and Holland, 2007). Peer-review is an excellent vehicle for encouraging such skill development, and many of the features highlighted by instructors demonstrate their commitment to developing critical analysis skills.

Assisting students in the practice of critical analysis are features that allow students to review their own submission at the same time as reviewing peers’ submission (self-review), and, after the review deadline, to see how other students have

assessed the submissions that they themselves reviewed – allowing them to calibrate their own opinion with those of their peers.

An early (pre-2009) feature allows reviews to be marked – a second-level peer-review exercise takes as input the reviews from the primary assignment, and allows students to critique the quality of a randomly allocated set of their peers' reviews. This exercise encourages students not only to engage in critical analysis, but to think about what good critical analysis is: useful for training students to become constructive reviewers. This exercise can also be done by tutors (see 'The right feedback' below) so students receive feedback from tutors as to the appropriateness of the reviews they wrote.

This feature was extended on the request of an anatomy instructor to allow students to provide an anonymous response to their own reviewers. Her request was primarily as a result of students complaining that they thought that the criticism they had received was 'unfair', and she wanted to give them an opportunity to think about why they thought it unfair, and to express this view in writing.

Students are sometimes expected to use their review feedback to enhance a draft, thus focussing on the skills of analysis (interpreting the reviews), critical judgement (selecting which comments should be addressed), reflection (reflecting on the quality of the draft) and revision and integration (making appropriate changes without complete re-writing). An instructor in biomedical sciences wished to assess the extent to which students had used their reviews in producing their final document submitted for formal assessment. Submissions can now be downloaded together with their reviews, facilitating easy access to all the information that she needs when assessing each student.

Subsequent to the ‘response to reviewers’ feature implemented for the anatomy assignments described above, an accounting lecturer used this feature as a means by which students could describe what specific changes they made to their initial draft in response to the reviews.

Students are often only interested in the ‘mark’ that they get for an assignment. An option enables the instructor to specify that radio buttons choices are suppressed when reviews are presented to the author, so as to encourage students to focus on analysing the qualitative comments.

Only one instructor has highlighted that students might benefit from some training in critical analysis before reviewing peers’ work. We implemented a ‘non-student author’ feature whereby an instructor in English language can submit a document written by himself, and then ask students to review it. His own critique of this document (or one written by a tutor) is made available to the students as soon as they have written their own review (using a ‘locking’ facility that allows students to lock-in their own review, removing the requirement to wait until the review deadline) so that they can compare it with their own review. A similar calibration process is described by Price et al. (2016) whose students critique sample solutions and get feedback on their evaluations before evaluating their peers.

We have had requests to include a discussion forum that will allow students to engage in ongoing online discussion with their critics – this would be a significant endeavour which would introduce a very different focus to the system, and we have not implemented this feature.

Pedagogic innovation. Many instructors who use peer-review are keen on promoting educational innovation. Indeed, many go on to publish their experiences with Aropä in the educational scholarship literature (e.g. Patterson, 2009), often without our

knowledge. Many of the administrative features listed in the Appendix are intended to support instructors with their scholarly pedagogic activities by providing a means for repeating assignments year on year.

As well as creating assignments for each class, instructors can create a survey administered through Aropä, as a way of eliciting student feedback on their peer-review experience. On request, we have defined a standard Aropä feedback survey that can be uploaded and used by any instructor.

Plagiarism. Instructors are increasingly concerned about the possibility of peer-review facilitating plagiarism, especially if the peer-review process relates to a draft to be revised in response to feedback, and then submitted for summative assessment. Indeed, anecdotally, this is one of the most common reasons instructors give us for not wishing to engage in peer-review activity. We also hear plagiarism concerns expressed by students, especially those in the later years of their degree study.

In response to this concern from a classics lecturer, we extended the original ‘tagged’ allocations feature. When students ‘tag’ their submission according to one of a set of tags specified by the instructor, the original ‘within tag’ allocation method (see ‘Flexible assessment design’ above) ensures students only review other submissions with the same tag as their own. The new ‘between-tag’ extension ensures that, as far as possible, students only review essays on a different topic from their own. The between-tag algorithm does the best allocation that is possible: depending on the distribution of topics amongst the tags, some students may necessarily review on their own topic, but the number of students who do this is minimised by the allocation algorithm.

The 'right' feedback. Instructors and students alike are sceptical that student reviewers will not give the ‘right’ feedback; together with plagiarism, this is one of the most common objections to peer-review. While it is expected that a learning outcome of a

peer-review exercise is that students are able assess critical feedback on their work, filter it, and decide which (potentially contradictory) feedback to take on board, both students and instructors are anxious that students might get feedback that is in some respects ‘wrong’.

We implemented a feature whereby official ‘markers’ could review students’ submissions in addition to the peer-review activity. Official (‘right’) reviews are written by tutors or instructors who are given a ‘marker’ role in the system, and can be allocated either all or a proportion of the submissions to review. These markers are directed to the student interface when they log in to complete reviews. The fact that any type of submission can be marked in this way – individual, tagged, or submitted by group – means that some instructors now use Aropä simply as a convenient way for doing online marking.

Since tutors can use the second-level ‘review marking’ feature (see ‘Graduate attributes’ above) to mark students’ reviews, in some cases students are awarded marks for summative assessment based on the quality of their reviews. When this feature was used in a first-year classics assignment, students were so concerned that they might take on board peer-reviews of their work that tutors later marked as ‘wrong’ that they requested a feature whereby they could see how the reviews of their work had been assessed by tutors – thus enabling them to see whether each review was consider ‘good’ or not (and therefore whether it should be taken seriously). Our own view is that an important part of the peer-review process is that students should sift through potentially contradictory reviews and decide themselves what to take on board and what to discard. We discussed our perspective with the instructor, but after three consecutive years of her getting this request from students, we implemented this feature. She subsequently decided not to use it after all, having, through her experience with the system,

eventually come around to our own view. Interestingly, the more recent cohorts of students have not requested this feature.

One way some instructors hope the students might get ‘correct’ feedback is by including correct answers (or a model solution) within the rubric itself, thus ensuring that all students know what the ‘right’ answer to the question is before they provide feedback. Effectively, in this case, the peer-review process is then used as ‘peer-marking’ process (with respect to a ‘correct’ answer). We had not anticipated that a model solution would be used in rubrics; indeed, most of our prior peer-review examples comprised submissions of essays or other artefacts for which a single ‘correct’ model solution would be inappropriate. To facilitate this process for a molecular biology assignment which included a model solution in the rubric, we implemented a feature whereby images could be included in rubrics.

Discussion

The priorities and concerns of instructors interested in peer-review are multi-faceted and ever changing. In gathering information about instructors’ perspectives on online peer-review, there is a balance between pedagogically-motivated and administratively-motivated requests for online support. We find that instructors:

- want to make students happy – persuading students of the benefit of peer-review is not always easy, and fighting against resistance is hard (especially in the first year of introduction);
- have varying views on whether ‘the right feedback’ is necessary – some will go to enormous lengths to ensure that students are not exposed to poor advice, while others are more willing to allow the students to make up their own minds;

- have varying views on the potential for plagiarism – independent of whether the submission is a draft for a final summative assessment document or not;
- are innovative, and value flexibility that will allow them to use the system in unusual ways – they are reluctant to compromise the peer-review model they want to use so as to fit with system constraints;
- want their job to be made easy – if the system can do internally what they currently have to do externally, so much the better;
- value high usability – many instructors often use clumsy and difficult-to-use institutional software on a regular basis, and are pleased to find that Aropä is easy and intuitive to use.

Over the many years that we have been monitoring the use of Aropä over a wide range of subjects, we have noticed that problems with the success of the peer-review activity are often caused ‘outside’ the system: students forget to submit their work or write their reviews, some students take the review activity less seriously than others, students violate anonymity by including their own name in their submission, the instructor devises a rubric that only includes closed questions so does not permit useful constructive feedback, students forget their passwords, the class list provided by administrators at the start of the semester is incomplete, etc. While we can give advice to both instructors and students, we cannot dictate how they use the system.

Much of the success of any peer-review activity depends on the way in which it is introduced to the students. In the early days of Aropä, peer-review was often simply ‘tacked on’ to the end of a regular assignment, as a way of students getting additional feedback. In recent years, the trend has been more towards making the peer-review exercise an integral part of the assessment regime.

Conclusion

The benefits of an in-house peer-review system created, developed, and nurtured over time by practising academics are clear. Our approach to the inclusion and availability of system features and to the responsive service we provide to our instructors is led first and foremost by a firm understanding of the needs of the academic process – from both instructor and student perspectives.

But the Aropä project is more than simply an academic service – our close collaboration with the instructors and our knowledge of their contexts of use means that we consider the project to be a scholarly endeavour in itself, allowing us to gather data and comment on approaches to peer-review activities worldwide, subject-wide and over time.

Based on our experiences and data, our main conclusions are:

- An increasing numbers of instructors are willing to try peer-review. We speculate that this is due to increasing student numbers and decrease in staff:student ratios, and the increased emphasis on ‘assessment and feedback’ in response to poor student evaluations (Pitt and Norton, 2017). More positively, successful peer-review experiences are increasingly being shared amongst instructors, encouraging others to give it a try, and the increased use of teaching-only contracts in universities means that more instructors are being encouraged to engage in academic innovation activities.
- Instructors either like using peer-review in their classes (and continue using it) or don’t. From our post-activity communications with the instructors, we believe this is more to do with the nature of the students in the class or the relationship between instructor and students than with the nature of the peer-review activity

itself. Students are often resistant at first, and the instructor's perseverance in the face of dissatisfaction and/or their willingness to change the nature of the activity (for example, by offering incentives) are important factors affecting continuance.

- Students mostly dislike it, for a variety of reasons. They want 'expert' feedback, they consider writing reviews unnecessary extra work, and they consider that their own reviewing efforts are not adequately rewarded. They are often resentful of reviews that are critical of their work, and sometimes their confusion over contradictory feedback leads to their being dismissive of the whole process. They are, however, glad to have the opportunity to see other students' work, and are genuinely grateful for constructive comments that allow them to improve their work. Key to this dissatisfaction is, we believe, a tendency for students to underestimate the importance of generic transferrable skills: they seldom recognise that there is personal benefit to themselves (in terms of skill development) in critiquing someone else's work, and in analysing a range of (possibly contradictory) opinions of their own work.
- It is hard to persuade students to take part in a peer-review activity if there is no extrinsic incentive. Those assignments where no incentive has been offered are easy for us to identify in the system – typically less than a third of the class provide reviews. This lack of engagement is self-perpetuating: students who have voluntarily provided reviews in an assignment where participation has been low are unlikely to have received comments on their own work (and so are less likely to voluntarily provide reviews the next time they are asked to do so). There needs to be a critical mass of participation for students to think that the

reviewing process has been worth their while: few voluntary peer-review activities reach this critical mass.

- Technology can only take you so far, and cannot make the important academic decisions. Aropä will support the mechanics of peer-review, but this is only a small part of the activity – the rubric, the allocation method, incentives, and the instructions given to students are all key to success. Instructors who have believed that conducting a peer-review activity simply means using Aropä have quickly discovered the flaw in their approach.

Innovative pedagogical trends come and go, but peer-review is a fundamental educational activity that persists – what can be more natural than asking the person sitting next to you what they think of your work? With ever increasing class sizes, and students' desire to improve their work based on feedback, online peer-review will continue to play an essential role in higher education in years to come. This paper has demonstrated that even the basic notion of a 'peer-review activity' is shifting, developing and extending, and encompasses a wide range of perspectives and priorities. Only flexible online tools that can be easily and quickly adapted in response to instructors' requests can keep abreast of the pedagogical innovations of enthusiastic instructors.

Acknowledgements. We are grateful for the support of the School of Computing Science at the University of Glasgow and the Department of Computer Science at the University of Auckland for providing hosting services. Many instructors over the years have been willing to discuss features, advise on enhancements, and to try out our new features; they also gave their approval for their email correspondence to be analysed for this paper. Ethical approval for this study was given by the Ethics Committee of the College of Science and Engineering, University of Glasgow (ref: 300160176).

Indirectly, we are also grateful to the over 36 thousand students worldwide who have exercised their critical faculties by using Aropä to write reviews and assess their own.

References

- Brill, J. 2016. "Investigating peer review as a systemic pedagogy for developing the design knowledge, skills, and dispositions of novice instructional design students." *Education Tech Research Development* (64): 681–705.
- Cassidy, S. 2006. "Developing employability skills: peer assessment in higher education." *Education and Training* 48(7): 508 – 517.
- Cho, K. and Schunn, C.D. 2007. "Scaffolded writing and rewriting in the discipline: A web based reciprocal peer review system." *Computers and Education* 48(3): 409–426.
- Crowe, J.A., Silva, T. and Ceresola, R. 2015. "The effect of peer review on student learning outcomes in a research methods course." *Teaching Sociology* 43(3): 201-213.
- Hager, P., and Holland, S., eds. 2007. *Graduate attributes, learning and employability*. (6): 1-15. Springer Science & Business Media.
- Hains-Wesson, R. 2014. "An evaluation of peer and self assessment tools." *Learning Innovations*, Deakin University, Snippets and Sides (3).
- Hamer, J., Purchase, H., Luxton-Reilly, A., and Denny, P. 2015. "A comparison of peer and tutor feedback." *Assessment and Evaluation in Higher Education*, 40(1): 151-164.
- IMS Global. 2012. *Learning Tools Interoperability® as a SSO Mechanism*.
<http://www.imsglobal.org/learning-tools-interoperability-ss0-mechanism>.
- Luxton-Reilly, A. 2009. "A systematic review of tools that support peer assessment." *Computer Science Education* 19(4).

- Li, L., Liu, X. and Steckelberg, A. 2010. "Assessor or assessee: how student learning improves by giving and receiving peer feedback." *British Journal of Educational Technology*, 41: 523-536.
- Mulder, R. A., & Pearce, J. M. 2007. "PRAZE: Innovating teaching through online peer review." *24th Conference of the Australasian Society for Computers in Learning in Tertiary Education* 727-736.
- Mulder, R.A., Pearce, J.M. and Baik, C. 2014. "Peer review in higher education: Student perceptions before and after participation." *Active Learning in Higher Education* 15(2): 157-171.
- Nicol, D. , Thomson, A. and Breslin, C. 2014. "Rethinking feedback practices in higher education." *Assessment and Evaluation in Higher Education* 39(1): 102-122.
- Paré, D. and Joordens, S. 2008. "Peering into large lectures: examining peer and expert mark agreement using peerScholar, an online peer assessment tool." *Journal of Computer Assisted Learning*, 24(6): 526–540.
- Patterson, D. 2009. "Two peer review systems for summative assessment in a Green Processing engineering course." *Engineering Our Future* 2281-2295. Engineers Australia.
- Pitt, E. and Norton, L. 2017. "Now that's the feedback I want!' Students' reactions to feedback on graded work and what they do with it." *Assessment and Evaluation in Higher Education*, 42(4): 499-516.
- Price, E., Goldberg, F., Robinson, S., and McKean, M. 2016. "Validity of peer grading using Calibrated Peer Review in a guided-inquiry, conceptual physics course." *Physical Review Physics Education Research* (12) 020145.
- Ramm, D., Thomson, A., and Jackson, A. 2015. "Learning clinical skills in the simulation suite." *Nurse Education Today* 36(6): 823–827.

- Reinholz, D. 2016. "The assessment cycle: a model for learning through peer assessment." *Assessment and Evaluation in Higher Education* 41(2): 301-315.
- Taylor, S., Ryan, M. and Pearce, J. 2015. "Enhanced student learning in accounting utilising web-based technology, peer-review feedback and reflective practices." *Higher Education Research and Development* (34:6): 1251-1269.
- Tighe-Mooney, S., Bracken, M. and Dignam, B. 2016. "Peer assessment as a teaching and learning process." *All Ireland Journal of Teaching and Learning in Higher Education* 8(2).
- UnboundID. 2015. *LDAP: Open, Vendor-Neutral. Industry Standard.*
<https://www.ldap.com>.
- Walker, J. and Sampson, V. 2013. "Argument-driven inquiry." *Journal of Chemical Education* 90(10): 1269-1274.
- Weaver, K., Morales, V., Nelson M., Weaver, P., Toledo, A. and Godde, K. 2016 "The benefits of peer review and a multise­mester capstone writing series on inquiry and analysis skills in an undergraduate thesis." *CBE - Life Sciences Education* (15): 15.
- Wenger, E. 1998. *Communities of Practice*. Cambridge University Press.

Appendix Eighty-one peer-review features deemed important by instructors and students.

Feature	Description
Operational: flexible assessment design	
flexible deadline setting	instructors can set their own submission and reviewing deadlines
flexible submission and reviewing	authors can be everyone in the class or one member from each group; reviewers can be individuals or specified groups
submissions can be in a range of different forms	instructors can specify what file type students should upload; multiple files can be uploaded
equal number of reviews	instructors can specify how many submissions each student should review
group submissions	instructors can specify group membership, and individual students can submit on behalf of a group
flexible rubric definition	instructors can define their own rubric of any length, using radio buttons and/or comment boxes
manual review allocations	instructors can manually change the allocations at any time
extensions	the deadline for submission and review can be extended for specified students
within tag extensions	the deadline for within-tag submission and review can be extended for specified students, while specifying the expected tag
upload submissions	instructors can upload the set of all submissions for a class
upload file in rubric	reviewers can upload a file into their rubric as part of their review
within team reviewing	students can be placed in teams so that each student only reviews submissions from within their own team
between tag extensions	the deadline for between-tag submission and review can be extended for specified students, while specifying the expected tag
Implicit extensions	instructors can specify a 'grace period' after submission and review deadlines, during which time the system will accept submissions and reviews
ordering of viewing	instructors can specify which order students review their peers' submissions
different rubrics for tutor markers	separate rubrics are defined for students and for tutor markers
Operational: monitoring	
identifying errant students	information about students who have not yet submitted or who have missed

	extension deadlines is highlighted to the instructor
monitoring empty or incomplete reviews	instructors can see which students have submitted incomplete reviews
monitor reviewing	throughout the course of the assignment, instructors can see who has submitted, and who has completed their reviews
monitor reading of reviews	after the review deadline, instructors can see who has read their reviews
export the 'monitor review' page information	instructors can export the information on who has completed reviews and who has read their own reviews
impersonating student	instructors can impersonate a student so that they can see exactly what a student sees (and can act on behalf of that student if necessary)
Operational: encouraging good student behaviour	
submitting correct file type	if a student tries to upload a file of the wrong type, the system will highlight this
completing reviews	if a student tries to submit an incomplete review, the system will highlight this
restricting access to reviews	Instructors can specify that only those students who have done some or all of their reviews are able to see the feedback on their own submission
minimum review comment length	students are unable to save their review unless the comment boxes contain a minimum number of words
Operational: fairness to students	
read submission before reviewing	students are unable to review a submission until they have accessed it - either by downloading it, or accessing an on-screen version
reviewer anonymity	students never know who their reviewers are
within-tag allocations	students can tag their submissions by topic, and will only review submissions on the same topic
restricted reviewing	reviewers can be everyone in the class, or only those who uploaded a submission
allocation priorities	students are always allocated to the same number of submissions to review as their peers
meta data	information that identifies the author of a submission is (as much as is possible) removed

imbalanced allocations	instructors are informed if the allocations created are such that some students are required to review more submissions than others; in the particular case of between-tag allocations, instructors are informed if some students will be required to review submissions on their own topic
Pedagogy: Quantitative feedback	
radio button sets in rubric	different sets of radio buttons in a rubric are demarcated by colour when separated by a horizontal line
marks allocated to radio buttons	mark values can be allocated to each option in a radio button list in the rubric
aggregate marks	marks given by students using radio buttons are aggregated, and made available to instructors
download marks	marks given by students using radio buttons are aggregated, and can be downloaded in a spreadsheet
highlight discrepancies	marks given by students using radio buttons are aggregated, and reviewers whose marks are out of line with others are highlighted
pull down list in rubrics	the rubric allows drop-down lists of numeric marks
Pedagogy: graduate attributes	
response to reviewers	students can write a response to their peer-reviewers
self-review	students can use the reviewing rubric to review their own work
review marking (by students)	students can provide feedback on the quality of reviews
review marking (by tutors)	tutor markers can provide feedback on the quality of reviews
download submissions and reviews	instructors can access submissions together with all their reviews
only comments are revealed	if an instructor wishes students to focus on qualitative comments, quantitative responses are suppressed
calibrate opinion with other reviews	students see the reviews from other students on the submissions that they themselves reviewed
non-student authors	students can review artefacts submitted by an instructor, so as to gain experience and training in the review process
review locking	students can 'lock' their reviews when complete, allowing immediate access to reviews of their own work
dialogue between students	a discussion forum allows students to have dialogue with each other
Pedagogy: innovation	

standard survey	the Examples Class that all new instructors are given guest access to includes a standard Aropā survey that elicits feedback on the students' experiences of peer-review
Pedagogy: plagiarism	
between-tag allocations	students can tag their submissions by topic, and will only review submissions on topics other than their own
Pedagogy: the 'right' feedback	
tutor review of submissions	some users are given special 'tutor marker' status, which allows them to login and review individual student submissions, using the student reviewing interface.
tutor review of group submissions	some users are given special 'tutor marker' status, which allows them to login and review group submissions, using the student reviewing interface.
show tutor marking of reviews of own work	students can see how tutors have marked the reviews that they have received, and so see which ones tutors have considered 'valid'
images in rubrics	rich text and images can be used in the rubric, typically used to specify a model solution
Administration: effort reduction	
WYSIWYG rubric editor	the rubric editor shows exactly what the reviewers will see when writing their reviews
automatic allocations	allocations between authors and reviewers are created automatically after the submission deadline: no intervention from the instructor is necessary
institutional authentication	users at the two main universities, Auckland and Glasgow, can login using the same credentials as for other institutional systems
copying a class	classes can be copied in their entirety for use in the following year
group upload	information about groups can be uploaded from text files, or from other assignments
integration with Moodle	links to Aropā from Moodle do not require additional authentication (Glasgow University only)
reusing rubrics	instructors can upload and use previously defined rubrics
email reminders	students are emailed to remind them of upcoming submission and review dates

email confirmation	students are emailed to confirm receipt of their submission and reviews
integration with local learning management system (LMS)	student data can be accessed from, and passed to, an institutional LMS
class list upload	class lists can be accessed directly from an institutional LMS
mark reviews of same essay	the allocation of review marking ensures that tutor markers read the minimum number of student submissions
integration with Turnitin	submissions are automatically passed to Turnitin for plagiarism detection
integration with other software	Application Program Interfaces allow integration with other custom software
Administration: easy administration	
class instructor ownership	instructors can give other people access to their Aropä classes (as instructors or guests)
persistent data	instructors always have access to their classes, even those from previous years
deleting class	entire classes (including all the assignments) can be deleted
deleting assignment	instructors can delete assignments, together with submissions and reviews
file size limit	the file size limit for submissions is increased approximately every two years on request
images in submission instructions	rich text and images can be used in the submission instructions, typically used to specify assignment requirements
Administration: high usability	
access code	instructors can set an 'access code' for a class, as a one-time password for students' initial login
setting password	instructors can reset students' passwords, and can set passwords for people with tutor, guest or instructor role within their class
time zones	the times specified on the user interface are always shown in local time
multi-lingualisation	the interface supports character sets other than Roman
visually impaired users	the interface complies with web standards for screen readers, so to support users with visual impairments
student password	students can reset their own password