*Tiffany Potter, UNIVERSITY OF BRITISH COLUMBIA, tiffany.potter@ubc.ca*
*Letitia Englund, UNIVERSITY OF BRITISH COLUMBIA, letitia.englund@ubc.ca*
*James Charbonneau, UNIVERSITY OF BRITISH COLUMBIA, james.charbonneau@ubc.ca*
*Mark Thompson MacLean, UNIVERSITY OF BRITISH COLUMBIA, maclean@math.ubc.ca*
*Jonathan Newell, UNIVERSITY OF BRITISH COLUMBIA, jon.newell@ubc.ca*
*Ido Roll, UNIVERSITY OF BRITISH COLUMBIA, ido.roll@ubc.ca*

# ComPAIR: A New Online Tool Using Adaptive Comparative Judgement to Support Learning with Peer Feedback

## ABSTRACT

Peer feedback is a useful strategy in teaching and learning, but its effectiveness particularly in introductory courses can be limited by the relative newness of students to both the body of knowledge upon which they are being asked to provide feedback and the skill set involved in providing good feedback. This paper applies a novel approach to facilitating novice feedback: making use of students' inherent ability to compare. The ComPAIR application discussed in this article scaffolds peer feedback through comparisons, asking students to choose the "better" of two answers in a series of pairings offered in an engaging online context. In contrast to other peer-feedback approaches that seek to train novices to be able to provide expert feedback (such as calibrated peer review) or to crowdsource grading, ComPAIR focuses upon the benefits to be gained from the critical process of comparison and ranking. The tool design is based on the longstanding psychological principle of comparative judgement, by which novices who may not yet have the compass to assess others' work confidently can still rank content as "better" with accuracy. Data from 168 students in pilot studies in English, Physics and Math courses at the University of British Columbia are reviewed. Though the use of ComPAIR required little classroom time, students perceived this approach to increase their facility with course content, their ability assess their own work, and their capacity to provide feedback on the work of others in a collaborative learning environment.

## KEYWORDS

peer feedback, answer comparison, adaptive comparative judgement, online teaching tools, collaborative learning

## INTRODUCTION

In this paper we introduce and evaluate ComPAIR, a newly-developed online collaborative learning technology through which students evaluate, comment upon, and rank pairs of their peers' submissions, and having done so, return to reflect back on their own answers. With this approach, ComPAIR builds upon the benefits to learning that come with engaging in self-assessment and peer-assessment (Ambrose, Bridges, DiPietro, Lovett, Norman, & Mayer, 2010).

As Maryellen Weimer (2002) and others have documented, student-centered learning looks quite different in classes of different sizes. Large classes in particular—traditionally assumed to operate on a lecture model—can be challenging sites in which to maintain student engagement and to facilitate collaborative, active learning, especially of nuanced, subjective critical processes like literary analysis, illustrating math concepts, and answering complicated, open-ended physics questions. One common approach to facilitating active learning in large classrooms is using peer feedback (Paulus, 1999; Dochy, Segers & Sluijsmans, 1999).

Using peer feedback has numerous benefits: first, it is a practical and scalable solution for giving students detailed feedback on complex assignments (Topping, 1998). Second, peer feedback has been shown to be an effective means of learning, demonstrating long-term benefits (Jhangiani, 2016); in fact, in some cases, feedback provided by multiple peers led to more significant improvements in revisions of papers, compared with feedback provided by experts (Cho & Schunn, 2007). Third, the benefits of peer feedback extend beyond the domain of learning and also impact motivation and self-regulatory skills such as giving feedback (White & Frederiksen, 1998).

While it clearly offers several pedagogical benefits, the peer feedback process also has one clear limitation: it is contingent upon students' (often limited) ability to give valuable feedback. Giving feedback is a complex process that involves being able to recognize limitations of given answers and to formulate clear explanations about how to improve them. Recognizing the limitations of given answers is especially challenging as the feedback providers themselves are novices. Simply put, this is a variant on Meno's learning paradox: how can one novice give valuable feedback to another novice? A range of approaches aimed at guiding students in giving more meaningful feedback have been demonstrated to be beneficial to some degree, including tutoring on how to give feedback (Zhu, 1995); reviewing examples of annotated expert feedback (the practice in the peer-feedback component in edX.org); providing detailed rubrics (Hafner & Hafner, 2003); and providing students with feedback on their own feedback (Robinson, 2001; Nguyen, Xiong, & Litman, 2014).

This paper suggests another approach to facilitating novice feedback: making use of students' inherent ability to compare. The benefits of asking students to compare as part of their learning processes have long been documented across activities and domains (Schwartz & Bransford, 1998; Star & Rittle-Johnson, 2009; Graham, Namy, Gentner, & Meagher, 2010; Burke & Gilmore, 2015). The idea that people perform better when comparing than in more general evaluations is longstanding, going back at least to Thurstone's "Law of Comparative Judgement" (Thurstone, 1927). In brief, Thurstone argues that human beings can reliably make judgements as to the relative quality of one of a pair of artifacts compared with the other. One common explanation is that contrasts help learners to notice deep features of the domain and offer a baseline for evaluation (Schwartz & Bransford, 1998; Roll, Holmes, Day, & Bonn, 2012; Loibl, Roll, & Rummel, 2016). While it is difficult to evaluate an answer without a baseline (what constitutes "good"?), it is easier to compare two alternatives (which is better?). When applied to peer feedback, using comparisons supports learning of the feedback provider (by learning from comparison and contrasts), as well as the feedback recipients (as they receive feedback on the important elements of their answers). A somewhat similar approach in which learners compared answers to expert solutions has been used successfully in a MOOC setting (Kulkarni, Bernstein & Klemmer, 2015). However, while Kulkarni and colleagues asked students to compare their peers' answers to expert solutions, we have asked them to compare them to each other.

In this paper we describe and evaluate a system—ComPAIR—in which students compare and rank pairs of student answers in an online learning environment. The pool of relevant answer samples for comparison is created by the students themselves in the first phase of the exercise, and then pairs of those answers are provided anonymously to students by the technology. The pairing of answers has the effect of speeding up the peer feedback process, as learners consider two answers at once, rather than


Potter, T., Englund, L., Charbonneau, J., Thompson MacLean, M., Newell, J., & Roll, I. (2017). COMPAIR: A new online tool using adaptive comparative judgement to support learning with peer feedback. *Teaching & Learning Inquiry, 5*(2). http://dx.doi.org/10.20343/teachlearninqu.5.2.8

only one. Furthermore, it may create more meaningful comparisons, closer in their level, and thus may encourage more active processing. The "adaptive" component of the tool is designed for pairing answers: as more comparisons are completed across the class as a whole, answers with more similar scores—as calculated by the application's algorithm based on an answer's win/loss record in previous comparisons—are paired together, making the comparisons carried out later in the process more challenging to assess. While students starting earlier in the comparison window may receive nearly random pairs, students starting later will review cases that are more similar in their quality. In this project, we apply adaptivity in the form of "Adaptive Comparative Judgement" (Pollitt, 2012).

With this new technology, we hope to enhance the learning of both feedback providers (as a result of the critical process involved in assessment informed by specific assignment criteria), and feedback recipients (as a result of higher-quality feedback fostered through the process, as well as reflection on their own answer after having analyzed and compared other students' answers). In terms of students' experience, the target outcomes for the use of ComPAIR included increasing students' sense of active and collaborative learning as part of a large-class community; facilitating the development of capacities related to self-assessment and the assessment of the work of others; providing students with new opportunities to practice answers related to specific skill sets; and helping students to acquire the ability to articulate a clear sense of the qualities of more versus less effective academic work in multiple disciplines.

The system is the result of a participatory design process, in which instructors, researchers, and technical experts designed the system together. At each stage in this one-year iterative process, instructors provided relevant use cases. After discussing potential approaches for implementation, the technology experts would develop a prototype, carry out usability testing with students, and present it at the next meeting. The group would provide feedback and continue to define the system capabilities and interface.

We begin with a description of the ComPAIR application and then review data from three courses that piloted the new application in 2015: one introductory English, and two advanced Math and Physics courses at the University of British Columbia—a top-tier Canadian university. In this article, we assess UBC students' perceptions of the ComPAIR technology and its capacity to facilitate specific learning benefits:

1. Learning from providing peer feedback (in the 'student as tutor' model) through comparisons and associated comments to explain their reasoning.
2. Learning from feedback received from peers on their own work, prompting self-reflection on their work.
3. Learning translated from the application to the classroom (online or physical), when instructors can have access to a class-generated list of submissions ranked by quality, which can be reincorporated into future class sessions in a range of diverse pedagogical and disciplinary approaches.

The pilot described and assessed in this paper sought to evaluate whether ComPAIR facilitates the processes required above. This study investigates whether students perceive benefit from learning through comparison, and whether the application is an effective tool in any or all of the three disciplines in which it was piloted: introductory English literature, and senior-level Physics and Math.

## THE COMPAIR APPLICATION

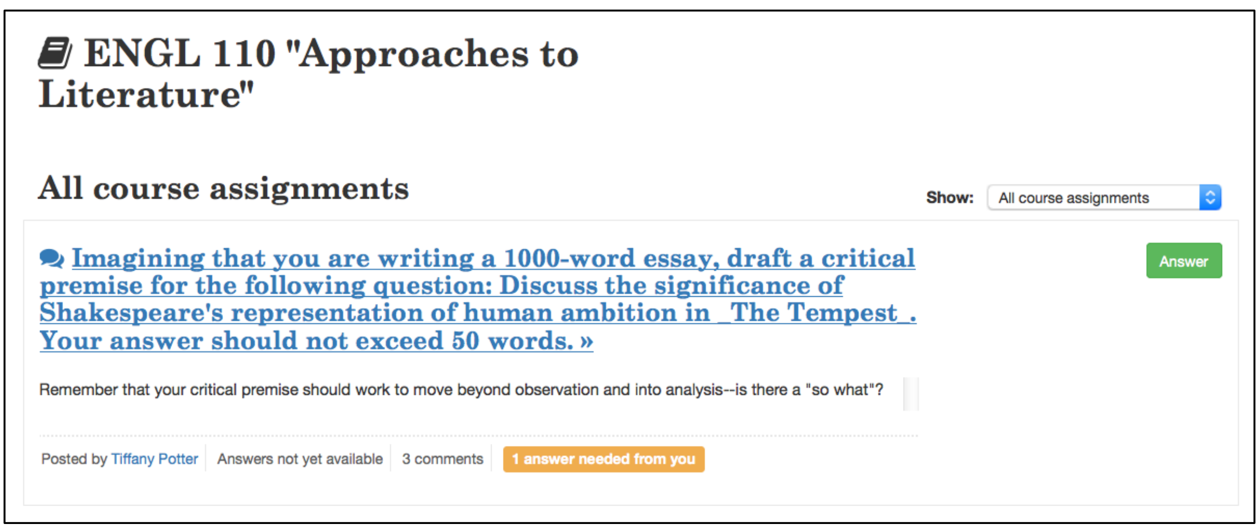
ComPAIR seeks to enhance student learning through a series of connected steps. For each assignment, students:

1. Complete and submit an answer to a question.
2. Read two answers from other students presented anonymously in pairs to discover multiple approaches to the question.
3. Undertake the critical process of determining and articulating what makes a "better" answer through writing individual feedback and choosing preferred answers from the presented pairs.
4. Students repeat steps two and three: the default setting for the application is for students to evaluate, provide feedback, and rank three pairs.
5. Participate in self-reflection, class discussion, or other follow-up emerging from the exercise.

### The student interface

When registered students log in, they can see all active assignments created for a course (Figure 1), and select the assignment they wish to complete. In the first step of the assignment (Figure 2), students answer the selected question (in rich text, with a pdf, or both). In the second step, students read and compare pairs of peer answers. The comparison process starts with giving feedback to the authors of the answers (Figure 3), so that students are encouraged to think critically about each individual answer in the specific context of a comparison. That is, students can use each answer as a source of reference in commenting upon the other, the comparator answer providing a basis for the identification of strengths or weaknesses that might not be evident with just a single isolated answer. The students then select a preferred answer from each pair for each instructor-set assessment axis (criteria) and explain their selection for each (Figure 4).

**Figure 1.Course screen from student view**




 Potter, T., Englund, L., Charbonneau, J., Thompson MacLean, M., Newell, J., & Roll, I. (2017). COMPAIR: A new online tool using adaptive comparative judgement to support learning with peer feedback. *Teaching & Learning Inquiry, 5*(2). http://dx.doi.org/10.20343/teachlearninqu.5.2.8

**Figure 2. Assignment answer screen**

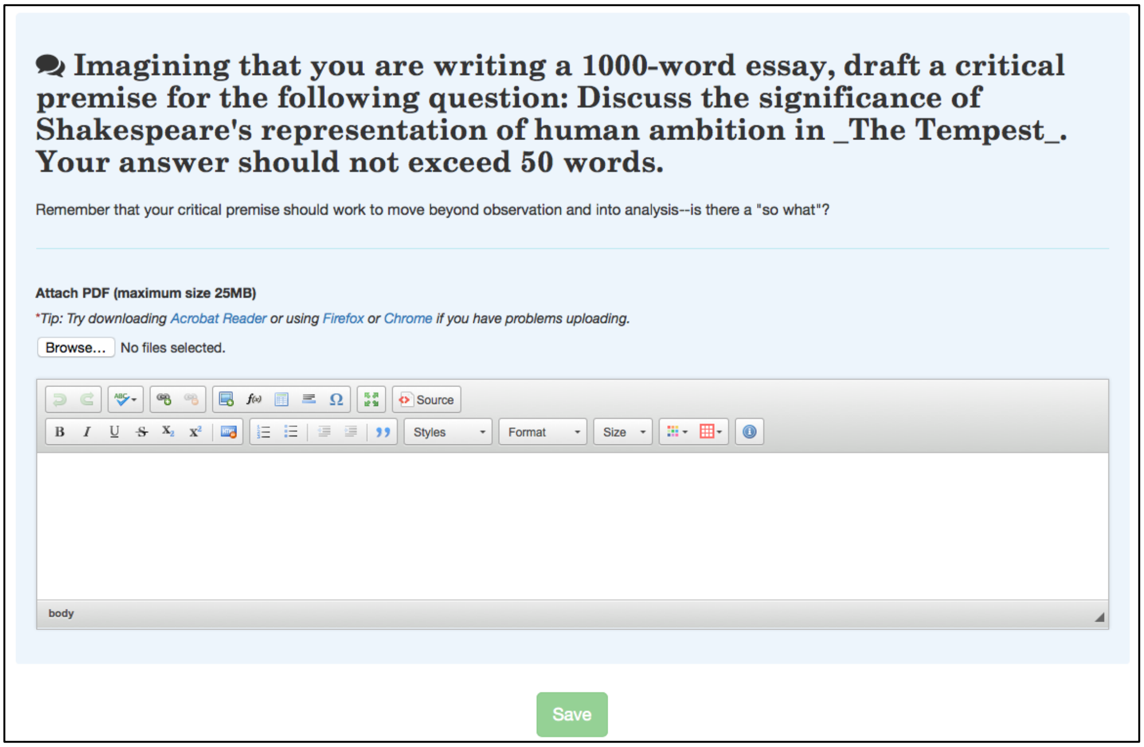



**Figure 3. Comparison screen: giving feedback**

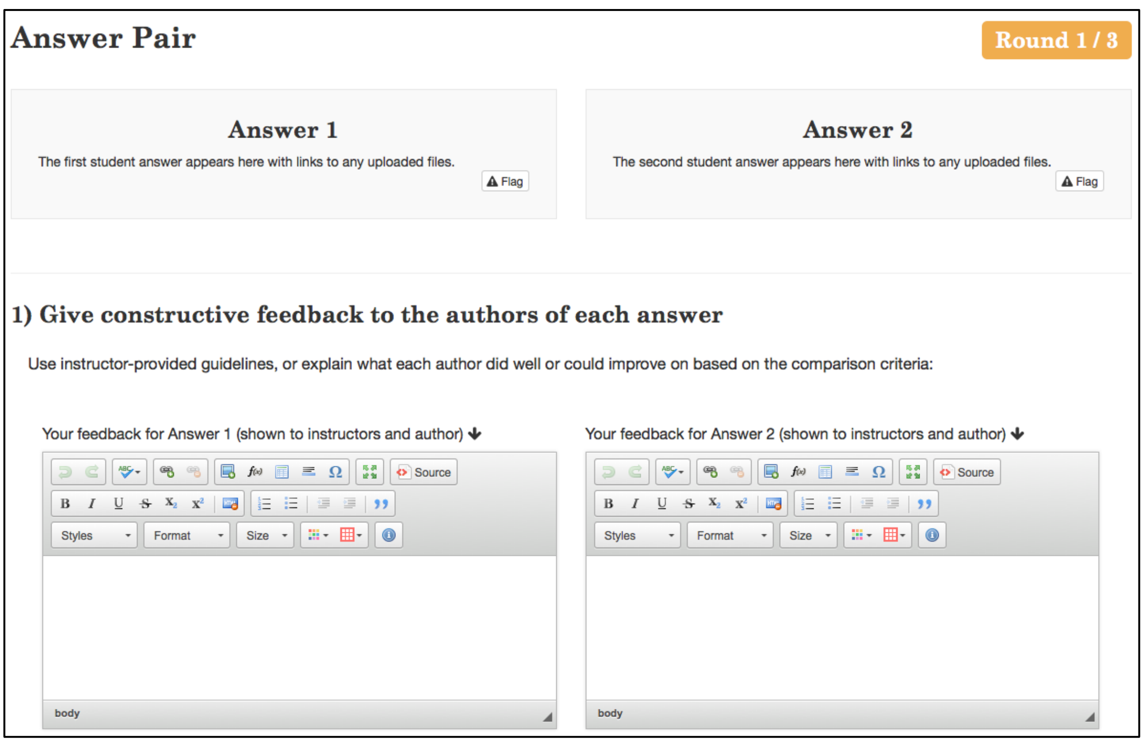
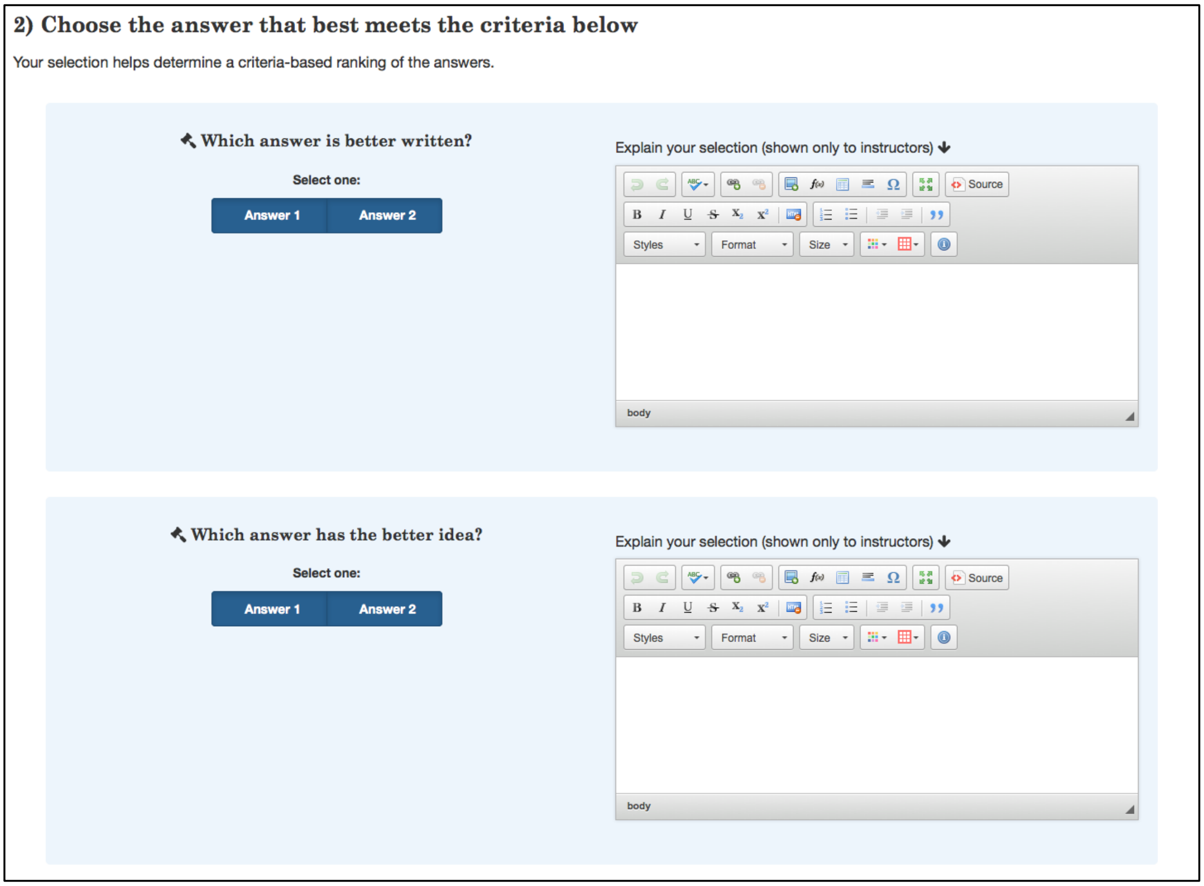


**Figure 4. Comparison screen: choose the better answer**



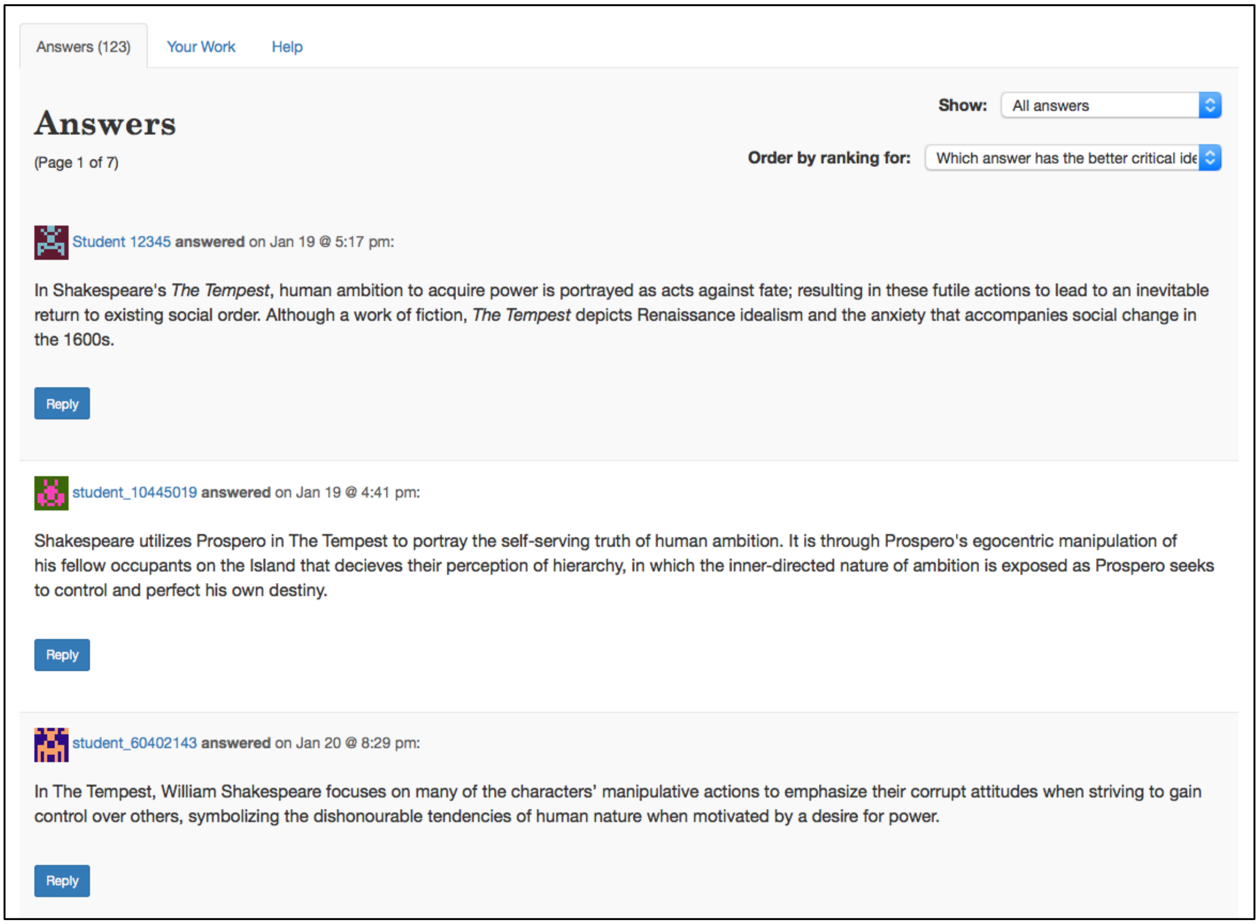

For each pair comparison, students submit their feedback, preferences, and explanations in a single form. After completing one comparison, they are automatically directed to the next pair of answers. Because, according to the theory of Adaptive Comparative Judgement, three comparisons represent the minimum for a reliable ranking scheme (Pollitt 2012), the system's default is to ask students to compare three sets, or six individual submissions (though instructors can change that setting). That the algorithm pairs answers with similar scores means that the comparisons carried out later in the process are expected to be more challenging to assess and rank.

Following the peer-feedback and comparison stages, the instructor may choose to add an additional step wherein students receive a self-evaluation prompt that asks them to reflect and comment on their own original answer, with a critical sense informed by the assessments of peer answers they have just completed. If the instructor forgoes this step, the student sees the completion screen.

Once students finish the required comparisons, they are able to access the peer feedback given to their own original answer, as well as the complete list of answers given by their peers (Figure 5). The list may be sorted by how well answers performed in the pairs according to each instructor-provided criterion, with the most-preferred answers appearing higher on the list. Instructors also have access to percentages revealing more closely how students ranked each answer.

 Potter, T., Englund, L., Charbonneau, J., Thompson MacLean, M., Newell, J., & Roll, I. (2017). COMPAIR: A new online tool using adaptive comparative judgement to support learning with peer feedback. *Teaching & Learning Inquiry, 5*(2). http://dx.doi.org/10.20343/teachlearninqu.5.2.8

**Figure 5. Student view of ranked answers for a selected criterion**



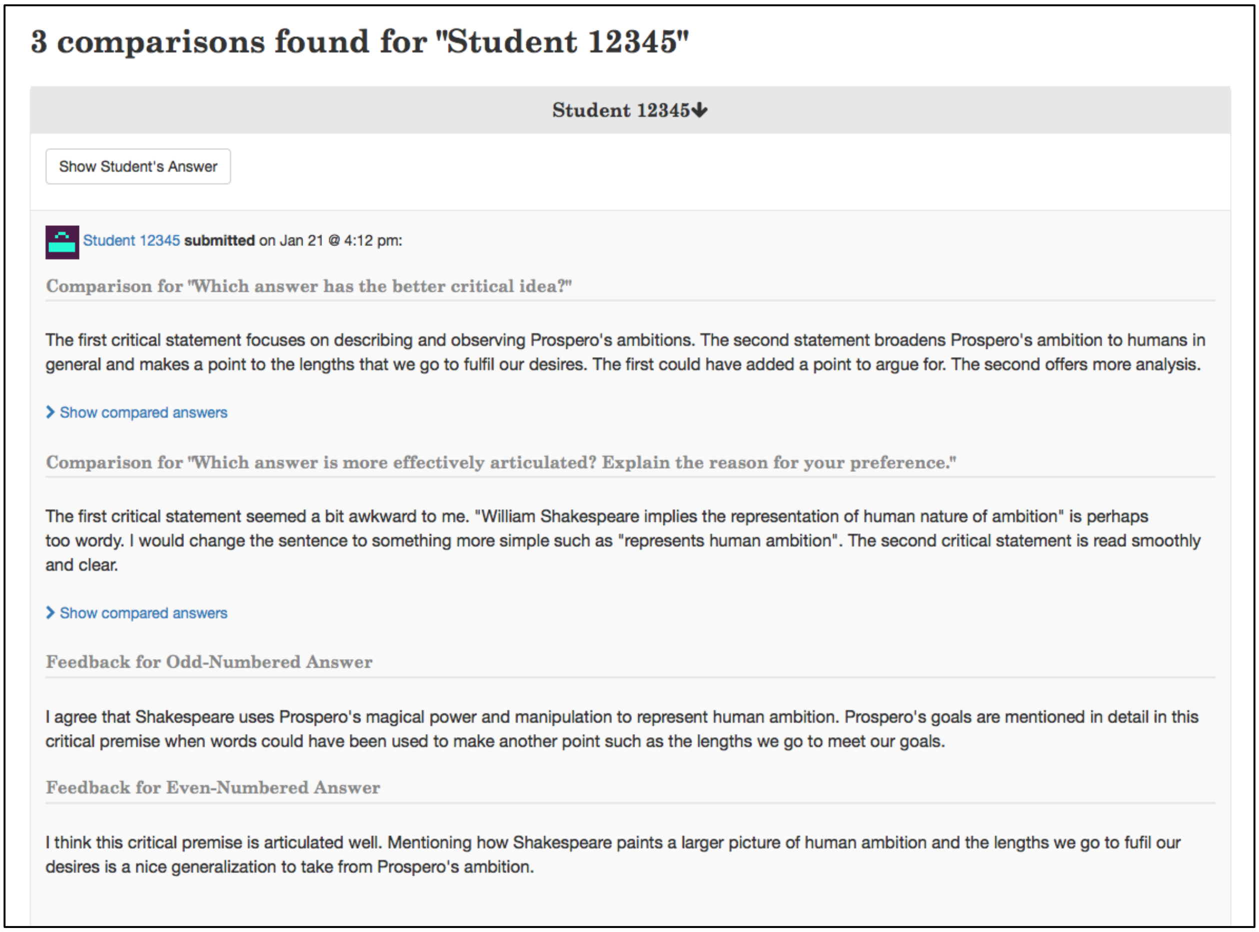

### The instructor interface

On the administrative side, instructors set up five structural elements for each assignment in the application, a process that can be completed in under ten minutes:

1. The assignment itself: the short form of the assignment question and any additional assignment details, which can be provided in an attached PDF or added with rich text.
2. The schedules: the timelines for when students a) answer the assignment and b) compare answers given by their peers. By default, comparing begins directly after answering ends, though these may be designed to overlap.
3. The criteria: the questions (or a single question) students are asked to consider in comparing the answer pairs. The default is "Which is better?"; however, instructors may change this to custom criteria to suit learning outcomes. A detailed explanation may also be given for each criterion.
4. The comparisons: the number of answer pairs students will compare and whether a formal self-evaluation will follow the comparisons. The recommended number for more accurate student-chosen rankings is three or more pairs (Pollitt, 2012).
5. The online discussion: whether students may leave comments and feedback on one another's answers outside of the comparison process.

During and after an assignment, instructors can read and comment (privately or publicly) on student answers and view the feedback and comparisons each student submits for the presented answer pairs (Figure 6).

**Figure 6. Instructor view of one student's completed feedback and comparisons**




Next, we describe an evaluation of ComPAIR in three UBC courses.

## DEVELOPMENT AND ASSESSMENT METHODS

### Ethical research

Evaluating new pedagogies and technologies in the classroom offers students potentially better instruction, while also exposing them to the risk of ineffective teaching. To minimize the risk, we were very attentive to students' reactions throughout the term. We further sought approval of the UBC Behavioural Research Ethics Board. This approval covered collecting and analyzing student data from the courses that used ComPAIR. Students whose data fell under the approval were notified of the study, provided information about the study, and given the chance to opt out of having their anonymized data included (up until the final exam). Students were also given an optional invitation to participate in normal pilot evaluation activities (surveys, focus groups). They were assured that neither these activities nor the data analysis would influence their grade in any way, and instructors did not see any results until after final grades were submitted for the relevant term. Furthermore, instructors did not know which of their students chose to participate in the additional activities, and whether students chose to opt out. De facto, none of the students opted out.

### Participatory design

The design of ComPAIR adhered to the principles of participatory design. In participatory design, the users are not merely the customers; instead, they are partners in the process that creates technologies which they eventually use (Schuler & Namioka, 1993). ComPAIR was developed by a team of faculty members, programmers, researchers, and students from the very beginning of the project, before a single line of code was written. The process started out with the three core faculty members envisioning use cases for how ComPAIR would be piloted in their courses at the same time that the full development team discussed the research questions that they wanted to answer. This informed both how the software would be structured and how the database would be designed. Between meetings, the faculty members would refine their target scenarios, while the technical team sketched interfaces and prototypes. This allowed for rapid iterations in the design.

This participatory environment, where everyone knew who was responsible for each aspect of the project, allowed the team to be extremely responsive, and most importantly, facilitated the involvement of students directly in the design cycle through early usability testing of the interfaces and workflows during scheduled testing sessions. The design process led to a minimally viable version of ComPAIR that was piloted in a live offering of Physics 333. The faculty member could give an assignment, and the students would quickly provide feedback on both the pedagogical elements of the assignment and the usability of ComPAIR. If there was a design issue or a back-end issue, the faculty member knew exactly which member was responsible for that part.

During this live pre-pilot, the program evolved as the needs of the students and the faculty were identified. For instance, the design of gradebooks and assignment views were influenced by the live use of the program. Perhaps even more interestingly, as the software evolved, the pedagogy and design of the assignments evolved. Seeing what could and couldn't be done with the software influenced the ways in which instructors used the software in assignments. Data in the rest of this paper focuses on the three pilot studies that followed this early pre-pilot.

### Courses in the pilot study

ComPAIR was piloted in English, Physics, and Math courses in 2015. As noted above, instructors representing each discipline formed a key part of the working group that developed ComPAIR and received customized technical support in using the beta version of the application throughout the term.

UBC's ENGL 110 (*Approaches to Literature*) typically is offered to classes of 150 first-year students, with 'lecture' for two hours a week and TA-led 30-student tutorials for one hour. The teaching team behind every section strives to create opportunities to invite active engagement and to challenge the disheartening assumption that many students make about such a large class: that they are anonymous and that no one notices whether or not they are there (Boyd, 2010; Buckley, Bain, Luginbuhl, & Dyer, 2004). ComPAIR was piloted in one English 110 section (final enrolment 129) between January and April 2015.

Students in English 110 are expected to acquire skill in literary analysis and close reading, and the ability to analyze a group of documents to develop a cohesive and original critical position that engages (and does not merely parrot) existing bodies of scholarship. The students are expected to write original pieces of critical literary scholarship, and some of them find that an intimidating prospect. One important learning outcome in English 110 is the ability for students to translate their initial perceptions into verbal and then on-page articulations that communicate both ideas and analytical validity clearly and effectively. This seems straightforward enough, but actually requires intense integration of several high-level critical faculties, and it is a process with which many students struggle: the familiar sense that "I know what I think; I just don't know how to say it." The two-phase assignment in the ComPAIR pilot in ENGL 110 focused specifically on this foundational learning outcome.

UBC's PHYS 333 (*Climate and Energy*) is a third-year online course available to students outside the physics major degree program. The course typically runs with about 25 students who participate in a series of online lessons, constructed response questions, take-home experiments, and class-wide 'Big Picture' problems addressed collaboratively. ComPAIR was piloted in the September 2014 and September 2015 course offering, with data collected in the latter pilot.

An important learning outcome of Physics 333 is that students be able to participate in discussions about climate, and debates about similarly socially complex scientific topics, using numbers, not adjectives. This outcome requires students to learn to take complicated, and sometimes ill-defined, questions that are generally intractable and make informed assumptions that allow them to perform relatively simple, accurate calculations to arrive at conclusions. These are often called Fermi, or back-of-the-envelope calculations, and are the foundation of what it means to 'think like a physicist'. To practice these skills, Physics 333 students are given three 'Big Picture' questions throughout the semester. These are complicated, vaguely-stated questions that students work on together to solve under the guidance of the instructor. ComPAIR is used as part of the assessment of these assignments.

UBC's MATH 317 (*Calculus IV*) is a standard vector calculus course typically offered to third-year mathematics and physics students, though students from many programs take it. This course runs with about 275 students each year and uses a lecture-based format with assessment through online homework and take-home assignments, as well as midterms and a final exam. This is a highly technical course and students generally struggle to understand the big-picture conceptual ideas behind the mathematical machinery they are learning because of the heavy emphasis on computational skills. ComPAIR was piloted in one section of MATH 317 in January 2015.

### Students in the pilot study

UBC is a large research-focused university with a strong international ranking and reputation. Admission is highly competitive, and students typically enter with very high high-school averages and very strong records of academic achievement. In 2015-16, 21.6% of UBC's students were non-Canadian "international students."

Because Vancouver has a very high cost of living, and because student populations have become less homogenous over the last few decades, a substantial proportion of UBC students have responsibilities beyond the academic (childcare, eldercare, part- or full-time work), and many have commutes of 1-2 hours each way. These circumstances mean that the traditional group projects valued in collaborative learning may place an unreasonable burden on students who are not able to arrange meetings with their peers outside of class times.

Up to 65% of the students in a typical English 110 class are from outside of the Faculty of Arts, some of whom are genuinely excited about literature, with others dutifully meeting requirements, and a handful somewhat resentful of being forced by their Faculty to take a literature course.

The students in Physics 333 are upper-level Science students who have a genuine interest in climate issues. Students typically use the course to fulfill an upper level science credit and, even though

 Potter, T., Englund, L., Charbonneau, J., Thompson MacLean, M., Newell, J., & Roll, I. (2017). COMPAIR: A new online tool using adaptive comparative judgement to support learning with peer feedback. *Teaching & Learning Inquiry, 5*(2). http://dx.doi.org/10.20343/teachlearninqu.5.2.8

they may be on campus daily, prefer the flexibility of a course with no scheduled class time. Many take the course in an attempt to cobble together an interdisciplinary degree in sustainability. Students in this course are typically a strong cohort, generally well-versed in the social, economic, and scientific aspects of sustainability.

The students in Math 317 are upper-level Science and Engineering students who have a strong interest in mathematics. The material in this course provides students with the mathematical background to take advanced courses involving electricity and magnetism, fluid dynamics, and differential geometry. Most of the students have done well in their previous mathematics courses and have a strong interest in mathematics. Students generally consider MATH 317 to be a difficult course because they are working with unfamiliar concepts and there is a high degree of computation required.

### The ComPAIR assignments

In English 110, each student completed two scaffolded ComPAIR assignments. In each, they created a 50- to 100-word critical premise in answer to a specific question assigned in class.[1] The first assignment asked students to integrate several distinct ideas that had been addressed in class discussion on Shakespeare's *Tempest* into a 50-word critical premise to frame the argument as if they had been assigned a short essay.[2] The second, more complex task followed three weeks of work on the skills involved in the critical reading of poetry; the question required reading a short modern poem and developing a critical premise to frame a single extended argument using different tools of poetic analysis.[3]

Students were assigned questions Mondays at noon, and had until Tuesday at 9:00am to enter an answer. Between 9:00am Tuesday and 6:00pm Thursday, they were asked to evaluate three pairs of answers. Using specific criteria given within the application, students then carried out the evaluation and comparison process, filling in separate comment boxes for each of two axes (quality of the idea and effectiveness of its expression). After comparing three pairs, students reviewed peer assessments of their answer, and could choose to see where their answer placed in the class-wide rankings. In the final step, the TAs then downloaded the students' answers, printed them, and returned them to students in Friday's tutorial. Each student received individual assessment and comments, but the exercise was intentionally kept low-stakes, with ComPAIR counting only toward class contribution marks. In tutorial, TAs reviewed both excellent answers and common pain points, and then used samples from the assignment as teaching tools in follow-up collaborative exercises on the design of an essay in literary criticism.

ComPAIR was used three times in Physics 333 as part of the "Big Picture" assignments. Completing each 'Big Picture' assignment involves participating in a series of tasks that mirror Polya's Method for problem solving (Polya, 1945). The steps in Polya's method are 1) understand the question 2) make a plan 3) execute the plan and 4) reflect on your answer.

For the first two steps, students collaborate in a forum (in this online course, with Piazza) to determine whether or not the question is well posed, find what they need to answer the question, and start to formulate a plan to actually perform the calculation. Each of the first steps takes a week to complete and is heavily facilitated by the instructor. The third and fourth steps take place on ComPAIR. The students execute their plan by submitting a written solution to ComPAIR. Solutions are three to six pages long, a combination of calculation and exposition. Students then complete three comparisons in ComPAIR. Finally, they complete a self-reflection on their work.

Two rubrics provided to the students are the key to making this process work in Physics. The first tells them how to assess each other's work during the comparisons, providing clear expectations for students' written solutions, as well as a clear way to determine which assignment is better. The second rubric is used by the instructor or TA to determine students' grades on the overall assignment. This second rubric grades participation in the forum in Steps 1 and 2, and whether or not students completed the submission. Most importantly, it grades the quality of the feedback students have given to other students during the comparison. This means that at no point in the process is their grade on the project directly linked to the content of their submission; grading is based upon contributions to the process.

In Math 317, students were given two optional assignments at the end of two major units in the course. Each unit developed over a one-month period a major mathematical tool and the idea was to give students an assignment that asked them to work with each tool to demonstrate they understood the conceptual relationship between the tool and the geometry it was intended to explore. The first assignment asked students to draw a series of figures to illustrate how to show whether or not a given space curve lies in a plane. The second assignment asked students to draw a series of figures to illustrate how computing the line integral of a particular kind of vector field around a simple closed plane curve generates the area of the region bounded by that curve. The students could download a template of panels on which they could draw their figures and then they uploaded a scan or photograph of their work to ComPAIR. Both of these assignments force students to make the links between the analysis they are learning and the geometry connected to it. Students were given one week to do each assignment and one week to do the comparisons.

### Instruments for measuring the student experience

For all courses, students were asked to communicate their experience with the application near the end of term. The research team collected this data, keeping the evaluation process separate from the teaching teams until after final grades were submitted. The research team administered paper surveys during class time (with an online survey for the online course in Physics); they also held hour-long focus group sessions outside of class. In total, 168 students completed the survey (98 of 129 total registered students in English, 16 of 19 in Physics, and 54 of 85 in Math), and four students from the English and Math courses participated in one of two different in-person focus groups (students in the online Physics courses did not attend focus groups).

For the English pilot, the UBC Student Evaluations of Teaching (SEoT) completed at the end of term provided additional data. These course evaluations were done online on the student's time, without an instructor, TA, or researcher present. Though students were not prompted to address the application specifically, a significant number named it in response to more open-ended questions.

## RESULTS

Along with identifying any technical or usability issues, our data collection and analysis sought student input in three key areas.

1. Perceived impact of specific learning activities: What parts of the assignment did students feel contributed to their learning and which detracted?
2. Perceived learning benefits: Where did students feel they did or did not improve their skills?
3. Overall ComPAIR experience: How did students feel about using the application?

One would expect significant discrepancies among the three pilots, given variability in contexts such as discipline, course, course level, class size, and student characteristics, as well as in implementation (the assignments themselves, wrapping activities, and grading). Other variations lie in


Potter, T., Englund, L., Charbonneau, J., Thompson MacLean, M., Newell, J., & Roll, I. (2017). COMPAIR: A new online tool using adaptive comparative judgement to support learning with peer feedback. *Teaching & Learning Inquiry, 5*(2). http://dx.doi.org/10.20343/teachlearninqu.5.2.8

online versus face-to-face course delivery and natural variations in instructional styles. Thus, it is particularly interesting to note the very consistent results that were observed in our surveys and focus groups. The results highlight consistent student perception and experiences within these varied implementations and contexts. Student comments in the survey and focus groups shared common elements in the larger scope, though they were grounded in the specific assignment for their classes. Our discussion considers these commonalities and addresses the observed differences.

### Perceived impact of specific learning activities

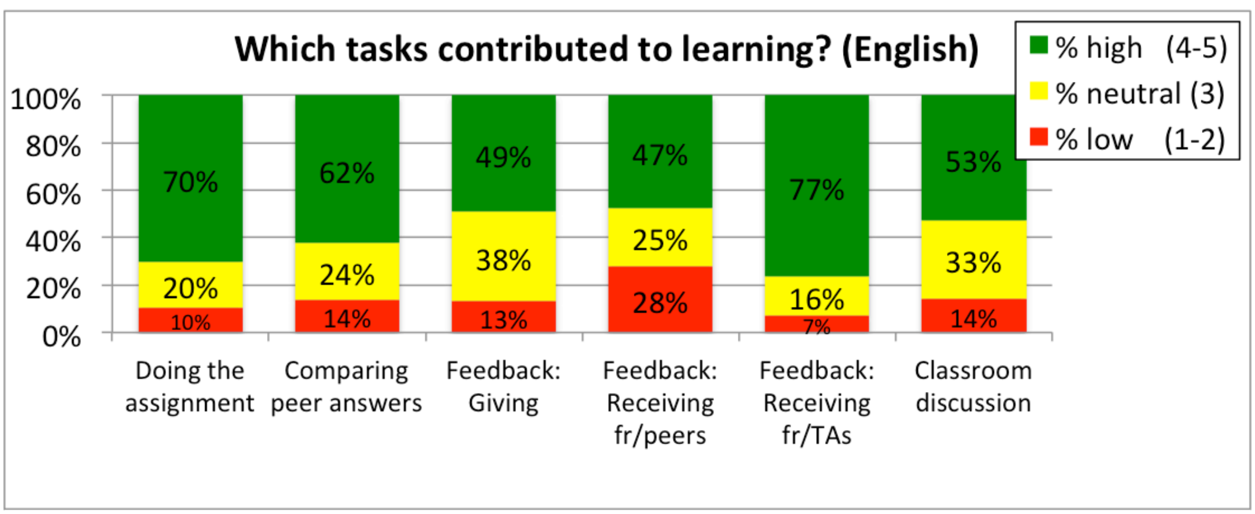
In the survey, students rated their impressions of the relative benefit of six learning activities: doing the assignment; comparing peer answers; giving feedback; receiving peer feedback; receiving TA/instructor feedback; and assignment-related follow-up discussion following the completion of each exercise. These are reflected in the charts below.

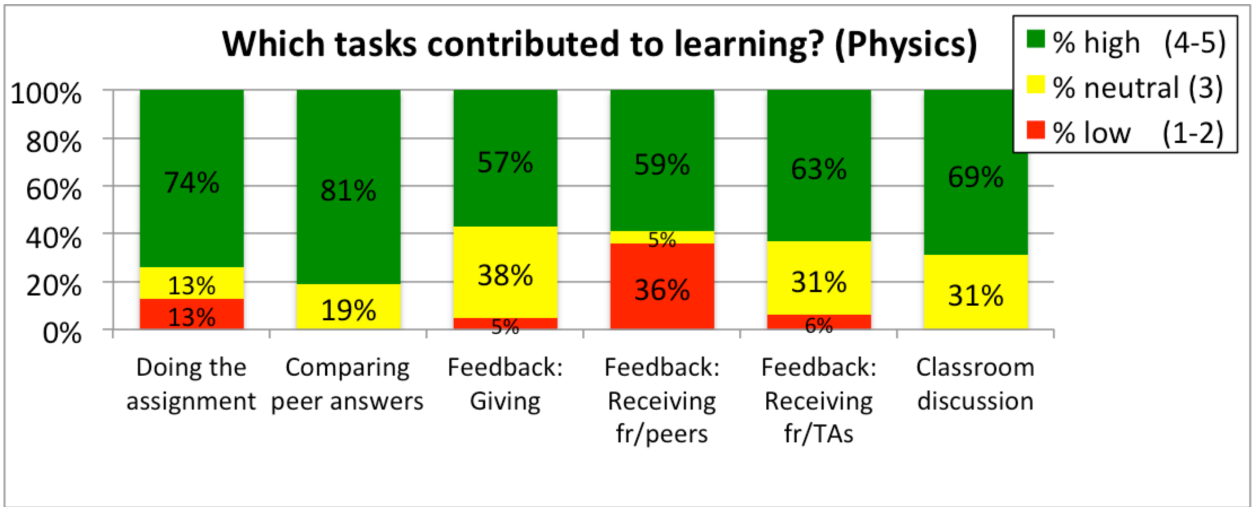
As seen in Figure 7, students in all three cases valued doing the assignment itself, receiving expert feedback, and comparing examples. The peer feedback itself, be it providing or receiving, was rated as somewhat less beneficial. English students reported that they learned most from receiving the TA feedback and doing the assignment, then from comparing peer answers and assignment-related follow-up discussion. English students said they learned least from giving and receiving peer feedback, with slightly more than 50% rating the effectiveness of these learning activities as neutral or low.

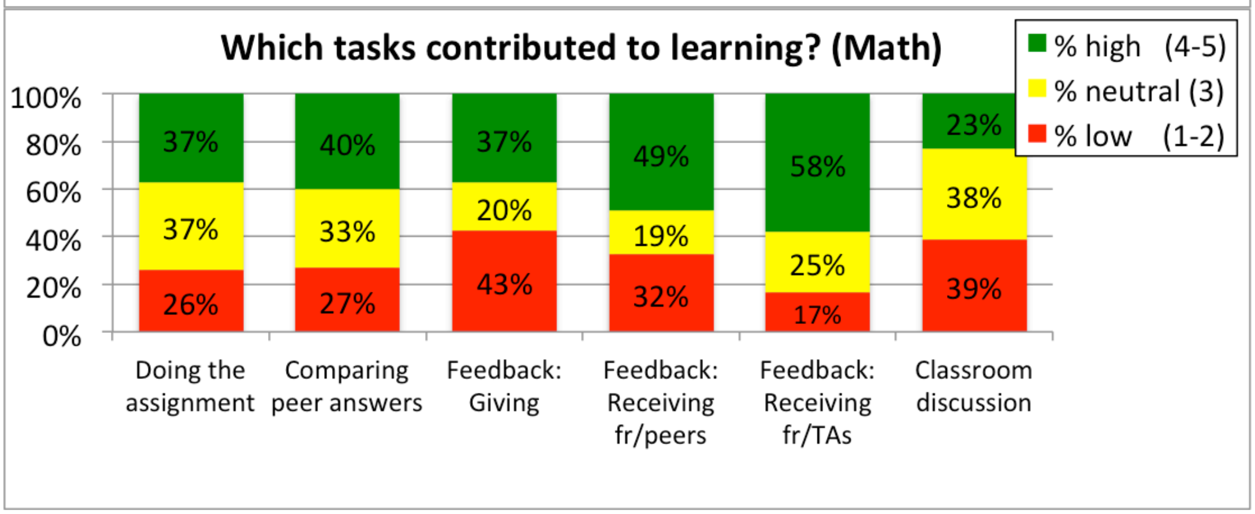
Physics students reported that they learned most from comparing peer answers, then from doing the assignment, assignment-related follow-up discussion, and receiving instructor feedback. Like the English students, Physics students said they learned least from giving and receiving peer feedback, with about 40% rating this activity neutral or low.

Math students reported that they learned most from receiving the TA feedback and peer feedback, then from comparing peer answers, doing the assignment, and giving peer feedback. Math students said they learned least from assignment-related follow-up discussion, with 77% rating this learning activity neutral or low. Notably, Math students gave lower assessments across the board; we elaborate on this pattern in the discussion below.

**Figure 7. Student perceptions of learning activities**






The data was substantiated and further illuminated in each of these areas by student comments in the surveys, course evaluations, and focus groups.

*Doing the assignment*

English students viewed the assignment as training for their in-class essay exams and valued the assignments because they clearly supported the students' learning outcomes and course grade. Focus group participants agreed that assignments were "*highly relevant*" to the course, and numerous course evaluation comments referred to the assignments as effective "*practice*" or "*preparation*" for the exams.

Like the English students, the Physics students also valued the assignments and also experienced them as a highly-integrated part of the course. Of all the pilots, the Physics one also weighted the assignments most heavily, with ComPAIR-related activities accounting for approximately 15% of the overall course grade (vs. 5% or less in the other two courses).

In contrast, Math students did not see as clear a relation between the assignment and their course success, and valued the assignments less as standalone exercises. One focus group participant thought the assignments "*weren't central to the course*" and that the optional nature of the second assignment further drove home this perception (although based on comments, this participant and others did like the general concept of the assignments).

 Potter, T., Englund, L., Charbonneau, J., Thompson MacLean, M., Newell, J., & Roll, I. (2017). COMPAIR: A new online tool using adaptive comparative judgement to support learning with peer feedback. *Teaching & Learning Inquiry, 5*(2). http://dx.doi.org/10.20343/teachlearninqu.5.2.8

*Comparing peer answers*

The English and Math students rated comparing as the third most-helpful learning activity, and placed it lower overall than the Physics group. In part, students' lower ratings for comparing may be tied to the uncertainty they expressed on how to choose a preferred answer, particularly when the paired answers seemed closer in quality (an expected feature of the application in later comparisons). Students in the focus groups expressed a desire for "*more guidance on how to choose which one is better*" in both courses, and admitted they sometimes refreshed the answer pairs until they got an easier pair to distinguish.

Despite uncertainty around comparing correctly, students from English and Math said they enjoyed seeing their peers' answers. Many positive comments in the focus groups and surveys reflected how students thought it was good *"to see what other people were thinking about the material*" and "*to look at other people's work and try to evaluate it."* Most of all, students appreciated how the comparison process informally "*allowed you to compare your work to others,*" giving them a sense of where they stood in relation to their peers.

Physics students, in contrast, rated comparing as their top learning activity, with 81% rating it highly. These students received the most detailed rubrics for doing the comparison, with a points system for evaluating the answers. This likely contributed to students feeling they benefitted from the comparison process. As one student reported, "*At first, I thought that doing 3 comparisons was too much. But [...] I learned how to skim through people's assignments quickly and could quickly tell what was a 'good' assignment and what was a 'bad' one*" using the rubrics provided.

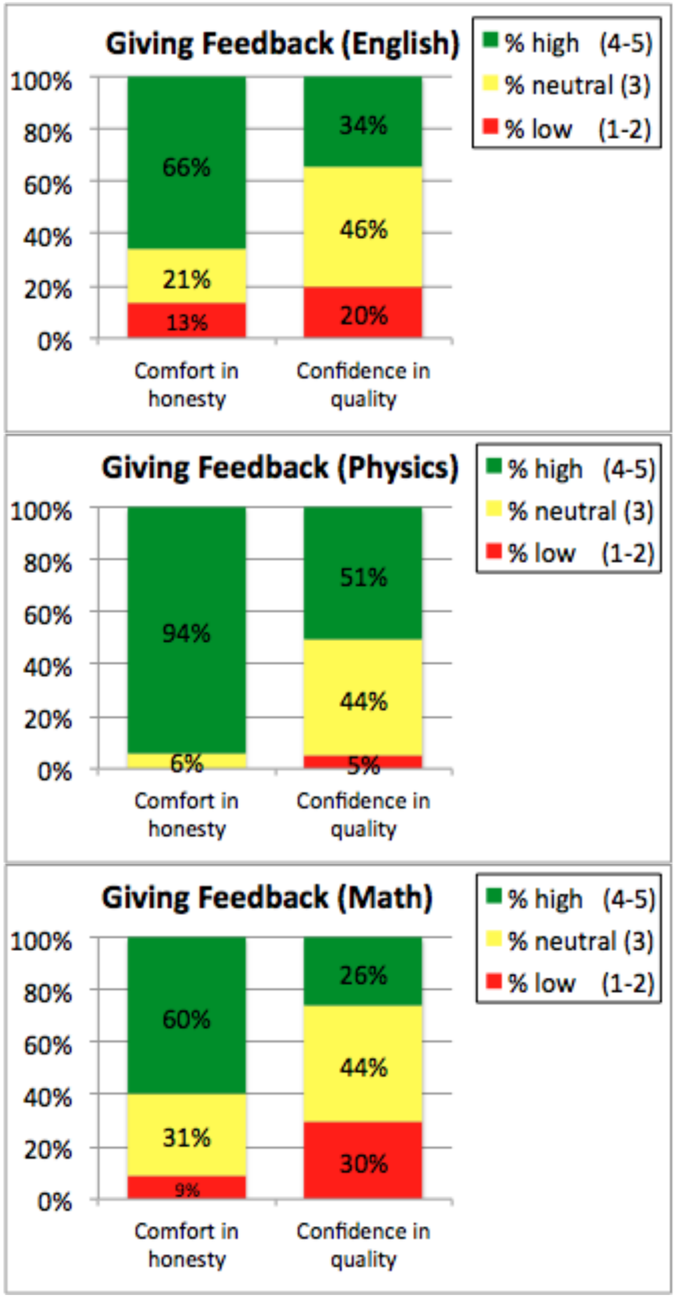
*Giving and receiving feedback*

All groups valued TA feedback highly—especially the English students, where 77% rated TA feedback high. The relatively more subjective nature of literary analysis and its evaluation generates an important effect here: English students received a greater range of peer feedback that addressed potential implications and nuances in the critical reading of the assigned texts. In the focus groups and surveys, several students reported this as "*contradictory*" or "*mixed*" feedback and wanted their TAs to weigh in and clarify what was actionable peer input. One student suggested modifying the application to include "*feedback from the instructor on...the critiques my peers gave me*" to clear up the confusion directly.

Math and Physics students did not express similar confusion over the content of their peer feedback. One Math focus group participant said of the feedback, "*it is just [to] explain your understanding*" of the drawings in the pair, so either peers understood what you drew or they did not. Math and Physics students still valued TA/instructor input but did not depend on it for clarification, though both of these student groups also expressed a desire for higher-quality peer feedback. One Physics student proposed using a more structured multiple-choice form for this activity to "*allow for way more qualitative and quantitative feedback*" during the comparison process.

Figure 8 shows students' perceptions of their ability to provide feedback. The majority in all disciplines reported feeling comfortable being honest in peer feedback, but only 34% in English, 51% in Physics, and 26% in Math were highly confident in that they gave quality feedback, as noted in the graphs below.

**Figure 8. Student perceptions of giving feedback**




Student comments clarified part of the reason for these perceptions: as was the case in comparing answers, students wanted more guidance in how to undertake this task. In focus groups, English and Math students requested more "*specific questions about peer work rather than general feedback*" questions or having feedback response *"divided into two sections—improvements/criticisms and things that were good*" to better direct the critical thinking process. A Physics student further suggested that "*an example showing how detailed the feedback should be would help*" in addition to the rubrics provided for that course.

*Assignment-related follow-up discussion*

The disciplinary contrast is clearly present in student response to the follow-up discussion activity, where English and Physics students responded more positively than did Math students. English follow-up discussion happened in the classroom, in the small tutorial groups of 26-28 students, led by graduate-student TAs; follow-up was organized around assignment-related activities and group work that built upon the specific question and directed it toward what would be the next phase of a writing process. Physics discussion, due to the nature of the course, was fully online—in forums and over email—and involved the instructor outlining the successes and failures of the problem solving strategy, specifically focusing on what should be kept the same and could be improved during the next iteration of problem solving. In Math, classroom discussion happened in the large lecture with the instructor, which included talking about the problem after completion and going over the instructor's solution but did not include hands-on activity.

The effect, then, of the students' preferences for more authoritative responses on grade-related, subjective work seems magnified by the distinction in teaching practice, wherein the English and Physics assignments were part of a larger, scaffolded learning process and the Math assignment was summative and did not call for further work or discussion. One Math student commented that the assignment itself revisited "*a topic that had already been thoroughly covered*" in the course.

**Perceived learning benefits**

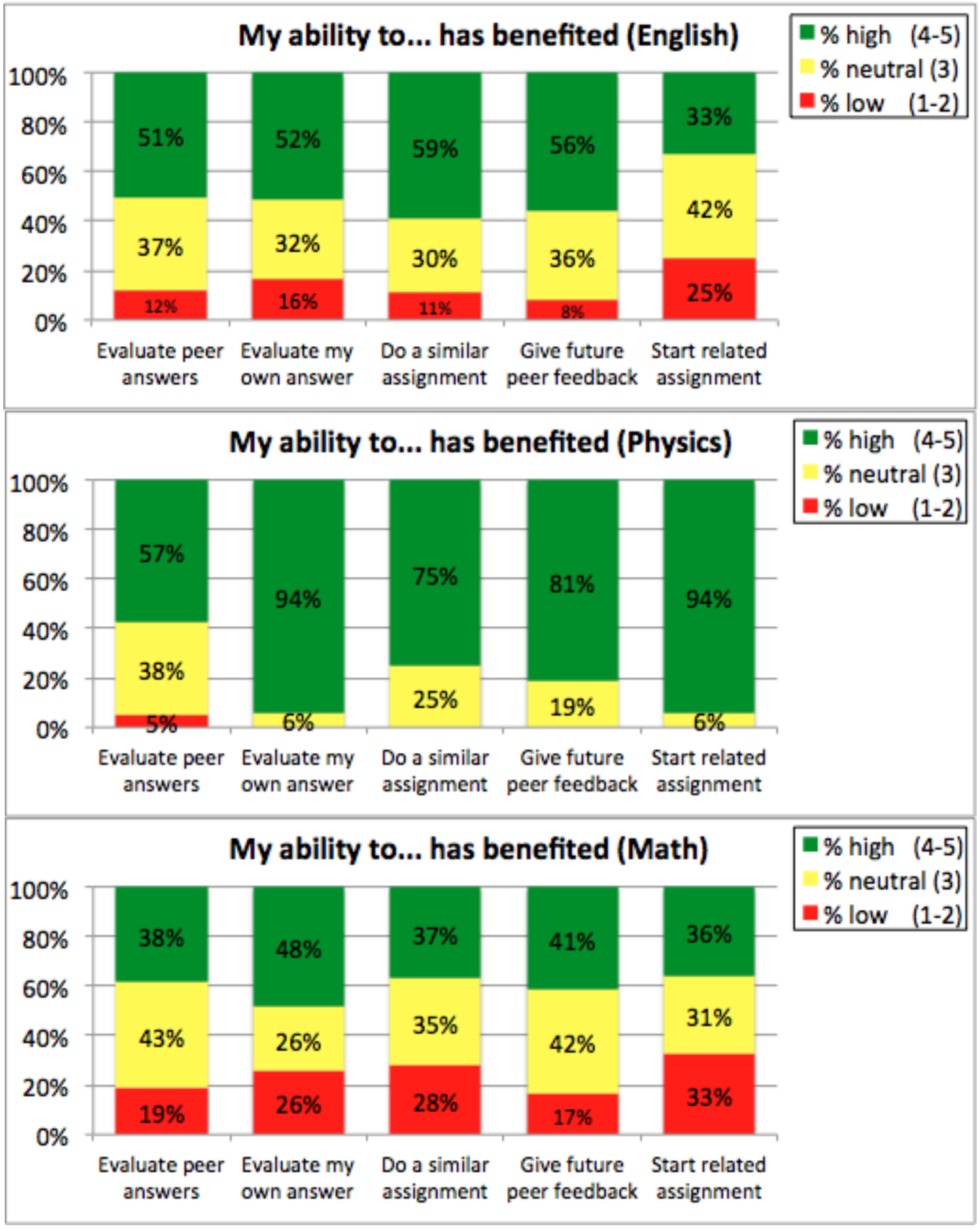
In the survey, students were asked if they had benefited from the ComPAIR assignments in their ability to evaluate peer answers; evaluate their own answer; complete a similar assignment; give peer feedback; or confidently use the assessed skill in a related course assignment. There were substantial differences in the answers here, with Physics students being the most positive about the impact of the assignments (see Figure 9). One possible explanation is the tight integration of the assignments into the course and the emphasis on feedback. Another possibility emerges from the fact that a core focus of this particular course is approaches to problem solving; the course explicitly encourages more reflective and self-directed learning, and—particularly as an online course—may attract a more reflective and self-directed group of students.

The lack of consistent ranking on perceived learning benefits among the three pilots is of interest. English students reported the greatest benefits in their ability to do a similar assignment and give peer feedback, then in their ability to evaluate their own and their peers' answers. Confidence in starting the term paper (the related assignment identified for English) benefited least, rated neutral or low by 67% of respondents.

Physics students reported the most benefit in their ability to evaluate their own answer and answer future "Big Picture" questions (the related assignment identified for Physics), then in their ability to give peer feedback and do a similar assignment. Evaluating peer answers benefited least, rated neutral or low by 43% of students.

Math students reported top benefits in their ability to evaluate their own answer, then in their ability to give peer feedback, evaluate peer answers, and do a similar assignment. Confidence in explaining complex Math answers (the related assignment identified for Math) benefited least, rated neutral or low by 64% of students.

**Figure 9. Student perceptions of learning benefits**




We can again spot disciplinary differences in these results that are further brought to light through student comments. English students saw the most benefit in transferring skills to a similar assignment. This perception may be related to the previously noted finding that many English students understood well what larger skill they were working on (clearly articulated in the focus groups as "*getting practice writing thesis statements*").

Physics and Math students saw the most benefit in evaluating their own specific answer, possibly driven in part by being able to see other students' solutions (a less common occurrence in these types of courses) and the self-reflection required at the end of the comparison process for both courses. As one

Math focus group participant explained, reading peer answers helped to clarify "*what the answer was supposed to be*" and revealed where the participant didn't "*get it*" with the initial submitted answer. A Physics student further commented that critically looking at other student answer pairs "*lets you think about your own work as well as how you could improve.*"

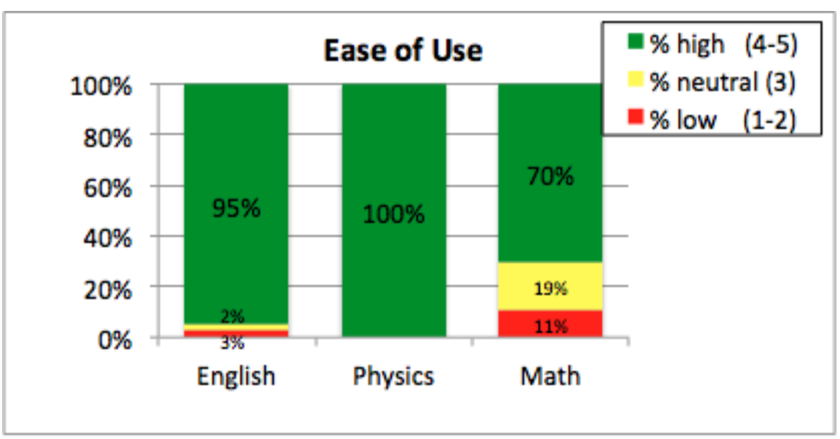
English and Physics students also seemed to find more benefit in evaluating other answers than did Math students. English students received two criteria to work with when comparing pairs and Physics students had detailed rubrics, while Math students had one criterion and no rubrics. The difference may have affected perception of learning. Interestingly, though, several English students noted that the evaluation felt "*redundant*" in requiring explanations for multiple criteria (especially with the number of required text boxes to fill in); yet some Math students did not feel they had enough to work from, with one survey participant wondering, "*what should we be focused on?*" (in response to the "Which is better?" default criterion used).

A notable similarity was that all groups reported high benefit in the process of giving peer feedback. Giving feedback was rated relatively high despite the lack of certainty noted earlier about the quality of the feedback students felt they gave (Figure 8 above). So although many students did not feel entirely confident while giving feedback, the act of practicing the skill helped them learn and acquire confidence moving forward. Particularly given the normally face-to-face contexts of the English and Math courses, the fact that students gave feedback through an application instead of in-person may have influenced the learning here, with one focus group participant describing the environment as a "*safe and anonymous*" one to practice this type of critical thinking in.

### Overall ComPAIR experience

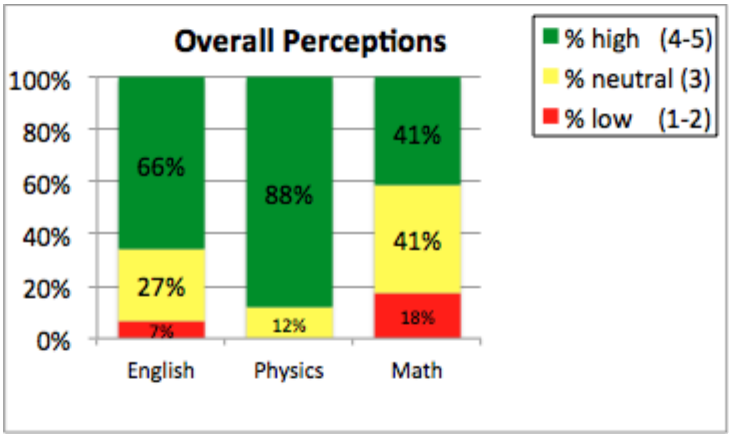
The majority of students in all courses rated ease of use positively in the survey, with 95% of English, 100% of Physics, and 70% of Math rating the usability high. Multiple survey and focus group participants also called the application "*straightforward*" and "*easy to use.*" Students noted some technical issues in the courses—moreso in Math (where PDFs had to be uploaded and viewed) than English or Physics—and this likely influenced the lower usability rating from the Math students visible in Figure 10.

**Figure 10. Student perceptions of application usability**




In the survey, students also rated their overall experience using ComPAIR on a scale ranging from positive to negative. As Figure 11 shows, for the English course, 66% of students rated their experience as positive, 27% as neutral, and 7% as negative. For Physics students in the fully online course, 88% reported a positive experience, 12% reported neutral, and 0% reported a negative experience. The Math students were divided in their rating of the experience between positive and neutral (at 41% each), with 18% rating the experience as negative.

**Figure 11. Student perceptions of overall ComPAIR experience**



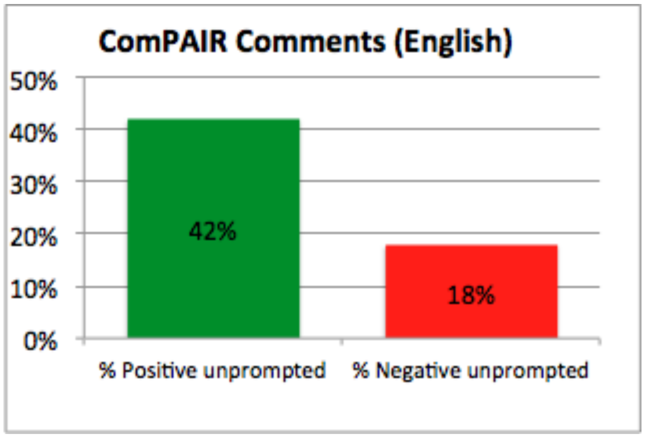

While this pilot did not have a control group and cannot quantify effects on student learning, both instructors and teaching assistants perceived ComPAIR as having a positive impact on student learning. Across all three disciplines, instructors and TAs reported a good quality of work in the assignments, as well as an apparent effect of increased levels of student preparation and participation, particularly in English tutorial sessions. Some English TAs found that the assignments "*ensured that students came prepared*" to sessions in the ComPAIR weeks and that ComPAIR "*made a difference in thesis generation*." The Physics instructor noted that students "*clearly put more effort into*" their answers, and the Math instructor thought some students "t*ook the opportunity to produce something better*" than they would typically hand in (for similar past assignments).

In the English pilot, the official course evaluations provided an important final data set for interpreting the students' overall experience (see Figure 12). In these evaluations for the Faculty of Arts, one question asks "What assignments were most valuable in helping you learn about this subject?" followed by a blank comment box.[4] Of 83 students who wrote narrative comments in response, 41 identified ComPAIR as the most valuable assignment (other answers included term papers, in-class essays, group discussions, lecture material, and doing the readings).

In the next evaluation question, 75 students responded to the reversed prompt "What assignments were the least valuable in helping you to learn about this subject?" Of these 75, 17 students mentioned ComPAIR. One of the comments noted technical problems as the source of unhelpfulness, and two mediated their naming ComPAIR with explanations of relative value among several useful assignments.

With 97 students submitting course evaluations, this means that 42% of English students gave unprompted positive feedback of ComPAIR and 18% gave unprompted negative feedback of ComPAIR, as seen in the graph below. The positive feedback also included more detailed commentary, with assignments frequently described as "*effective,*" "*helpful,*" and "*useful.*"

**Figure 12. Unprompted comments in english course evaluations**




## CONCLUSIONS

The goal in developing ComPAIR was to increase the pedagogical usefulness and effectiveness of peer assessment and feedback by adding the element of comparison in an online tool with the capacity to support student learning across Arts, Humanities, and STEM disciplines. In addition to the expectation of increased student engagement in comparison and ranking (rather than just commenting), the core innovation is the use of an adaptive algorithm such that the process of comparing other students' work is rendered progressively more challenging as the tool uses data from earlier students' comparisons to generate pairs that are increasingly similar. Because the type of assignment and follow-up varied by discipline in the pilot, students' impressions of their learning did too, but overall, 66% of students in the English pilot, 88% in the Physics pilot, and 41% in the Math pilot found the experience positive, and a surprisingly substantial plurality of students in the English pilot chose in the final course evaluation to name ComPAIR as the most helpful assignment of the course. These numbers roughly align to the percentages of each group that rated doing the assignment as contributing to their learning (70% in English, 74% in Physics, 37% in Math).

These rankings highlight the first lesson learned from this work. Not surprisingly, the way that students felt about the specific assignments had a strong impact on their overall experience with the application. Specifically, students saw less value in the Math assignments: it was less clear to students how these assignments aligned with course goals, and, being optional, the assignments could be understood to be less essential for the course. There is nothing inherently good or bad about a technology: it is a matter of how it is used. Students who saw the basic assignment as beneficial appreciated the opportunity to provide and receive feedback on it, using ComPAIR. Students who did not perceive the basic task as beneficial were less likely to appreciate having to spend more time on this using the application.

One particularly significant result is that, regardless of their level of confidence in writing quality peer feedback (Figure 8), many students perceived a learning benefit to their ability to give feedback from the act of practicing online using paired answers (Figure 9). Improving students' confidence in their ability to provide feedback is a significant contribution of ComPAIR, similar to other forms of training (White & Frederiksen, 1998). Requiring the recommended minimum of three comparisons created a valuable opportunity for students to practice peer feedback and to strengthen their ability and confidence in this area through repetition. In teaching environments seeking to emphasize collaborative learning, this is an important outcome gained from an intervention requiring little classroom time. As noted in the introduction, the ability of students to give high-quality feedback is a major challenge in peer-feedback systems (Zhu, 1995). Training on this task takes time that is not dedicated to learning the course subject matter. Using comparisons to support this process allows students to stay on task and reduces the need to dedicate much-needed time (and attention) to feedback training. Learning by providing feedback is a valuable experience (Lundstrom & Baker, 2009). Being able to support that using ComPAIR is a major benefit of the tool.

Many students saw ComPAIR as having benefits for both immediate assignments and for their learning more generally, and they were more likely to see value and learning in a comparison assignment when they could identify a clear connection between the assignment tasks and their overall course success. As part of their training to use comparative judgement generally or ComPAIR specifically, students need to understand how this teaching strategy is expected to help them improve a core course skill and not simply how the assignment works (i.e., it is comparing answers and ranking based on wins/losses). A known challenge for online learning is that students often do not understand well what is expected of them (Song, Singleton, Hill, & Koh, 2004). Our pilot demonstrates that an integrated approach to assignments can positively impact learning with a comparison-based application like ComPAIR: using answer comparison and feedback as a formative starting point or near-final step of a larger process—rather than as a standalone summative exercise—appears to increase the overall value of the assignments in students' minds.

Students very much liked being able to read peer responses, and they were interested both in how others answered the question and what peers said about their work. However, our study demonstrated that students need clear guidance and motivation when comparing answers and giving feedback in order to turn this natural curiosity into a concrete learning activity. Otherwise, we risk students' ending up frustrated by the process (not knowing what to look for, especially with well-matched answer pairs) or by the quality of what they receive from peers. The value of using rubrics to support peer feedback has been shown previously (cf. Suen 2014). Also in our case, a combination of multiple criteria or rubrics along with significantly weighting the feedback were demonstrated to help circumvent this frustration.

In addition to considering best practices for assignment integration and instruction, the team also took away ideas for improving the tool itself. To focus the comparison process more on writing quality peer feedback, the next version of the tool will contain only two text boxes on the comparison screen (in contrast to Figure 3 and Figure 4)—one responding to each answer author. Students will also select an answer for each criterion in the first step instead of the second, to better guide the thinking process around comparing the answers as a basis for providing feedback (which becomes the second step). Instructors will have the option to add an instructor-generated practice answer pair that all students compare first as a warm-up round, and instructors can choose to hide the final answer rankings to remove any sense of competition or confusion that some students expressed over their answer placement.

At the process level, we believe that much of the early success of the tool is to be credited to the participatory design process. Participatory design of learning environments is yet to become a common practice (Könings, Seidel, Jeroen, & van Merriënboer 2014). In this work, it had two primary forms: the joint development team, composed of instructors, developers, and researchers; and the early feedback cycles with students. In the future, we will attempt to include students as members of the design group.

Overall, the qualitative and quantitative data presented above suggests that ComPAIR offers a worthwhile learning experience, especially when integrated strategically into the course. This analysis focuses on the learning process within the online and classroom environments, as well as students' and instructors' perceptions of the value of the tool and the assignment. Future research should look at other aspects of learning with ComPAIR. First and foremost, it is of interest to evaluate how students use the feedback they receive. Are students investing the needed time to reflect on their initial responses, make sense of the feedback, and learn? Subsequently, do we see improvement in learning outcomes? Improvement in outcomes can be evaluated at the disciplinary level: that is, whether students can demonstrate a fuller understanding of the relevant content or skill. Improvement in outcomes can also be evaluated at the feedback level: that is, whether students learn to rank more insightfully and to provide better feedback. Evaluating students' rankings will also allow us to evaluate the use of ComPAIR for peer assessment purposes (adding to the current design for pedagogical use of peer feedback). Future analysis may indicate whether the overall ranked list of submissions can be used to grade complex assignments at scale, while keeping instructor load manageable.

This pilot suggests that ComPAIR is a potentially highly useful teaching tool for both students and instructors in courses across academic disciplines, particularly in its capacity to strengthen students' abilities to self-assess the quality of their own answers to an assigned question and in fostering the ability to evaluate and provide feedback on others' work. Using only a small amount of classroom time, assignments with ComPAIR had positive impact on this most essential of learning outcomes: the ability to recognize strengths and weaknesses and initiate self-reflection of academic work. We invite readers to experiment with UBC's open-source ComPAIR here: https://github.com/ubc/compare. ComPAIR's code can also be downloaded from this site.

## ACKNOWLEDGEMENTS

This project was supported by a grant from the UBC Teaching and Learning Enhancement Fund and with support in kind from the UBC Centre for Teaching, Learning and Technology. The ComPAIR application and related analysis would not have been possible without the support of numerous colleagues at UBC. We would like to thank our contributors for their roles in the project's success: Simon Bates, Professor of Teaching in Physics and Academic Director of the Centre for Teaching, Learning and Technology at UBC (UBC CTLT); Marianne Schroeder, Associate Director, Teaching and Learning Technologies at UBC CTLT; Pan Luo, John Hsu, and Michael Tang, Programmer Analysts, UBC CTLT; Philip Lee and Kolja Schwenghagen, co-op students at UBC CTLT; Dallas Hunt, Garth McClure, Morag McGreevey, and Michael Taylor, ACJ project assistants and Teaching Assistants for the pilot section of ENGL 110.

*Tiffany Potter is Associate Head, Curriculum, in the Department of English at the University of British Columbia (UBC) in Vancouver, BC, Canada.*

*Letitia Englund is a User Experience and Interface Analyst at the Centre for Teaching, Learning and Technology at UBC, Canada.*

*James Charbonneau is a Physics Instructor and Associate Director of the integrated Science One and Coordinated Science Programs at UBC, Canada.*

*Jonathan Newell teaches in the Department of English at UBC, Canada.*

*Mark Thomson MacLean is Professor of Teaching in the Department of Mathematics at UBC, Canada.*

*Ido Roll is Director, Institute for Scholarship of Teaching and Learning, and Senior Manager, Research and Evaluation, at the Centre for Teaching, Learning and Technology at UBC, Canada.*

## NOTES

1. A critical premise is a one-or-two-sentence statement that establishes a position that could be argued in an essay of the assigned length.
2. ComPAIR question for January: "Imagining that you are writing a 1000-word essay, draft a critical premise for the following question: Discuss the significance of Shakespeare's representation of human ambition in *The Tempest*. Remember that your critical premise should work to move beyond observation and into analysis--is there a "so what"? Your answer should not exceed 50 words."
3. ComPAIR question for February: "Read the poem below, and make notes on syntax, figurative language, patterns of diction and imagery, and form. Then, imagining that you are writing a 1200-word essay, write a critical premise as we have discussed in class. For your answer, write ONLY the premise of 20-100 words. [followed by text of Lorna Crozier's poem, "Transplanted"]."
4. The UBC Student Evaluation of Teaching is comprised of a University module, with six questions asked for every course in the university, then a Faculty module, in which each Faculty adds questions related to its own teaching goals. The Faculty of Science does not include a question on the relative value of different assignments.

## REFERENCES

Ambrose, S., Bridges, M.W., DiPietro, M., Lovett, M.C., Norman, M.K., & Mayer, R.E. (2010). *How learning works: Seven research-based principles for smart teaching*. San Francisco, CA: Jossey-Bass.

Boyd, J. (2010). The best of both worlds: The large lecture, writing-intensive course. *Communication Teacher*, *24*(4), 229-237. DOI: 10.1080/17404622.2010.513992

Buckley, G.L, Bain, N.R., Luginbuhl, A.M., & Dyer, M.L. (2004). Adding an 'active learning' component to a large lecture course. *Journal of Geography*, *103*(6), 231-237. DOI: 10.1080/00221340408978607

Burke, J and Gilmore, B. (2015). *Academic Moves for College and Career Readiness, Grades 6-12.* Thousand Oaks, CA: Sage.

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education*, *48*(3), 409-426. DOI: 10.1016/j.compedu.2005.02.004

Dochy, F. J. R. C., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, *24*(3), 331-350. DOI: 10.1080/03075079912331379935

Graham, S. A., Namy, L. L., Gentner, D., & Meagher, K. (2010). The role of comparison in preschoolers' novel object categorization. *Journal of Experimental Child Psychology*, *107*(3), 280-290. DOI: 10.1016/j.jecp.2010.04.017

Hafner, J. and Hafner, P. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education 25*(12), 1509-1528. DOI: 10.1080/0950069022000038268 DOI: 10.1007/s11251-013-9300-7

Jhangiani, R. (2016). The impact of participating in a peer assessment activity on subsequent academic performance. *Teaching of Psychology* 43(3), 180-186. DOI: 10.1177/0098628316649312

Könings, K. D., Seidel, T., Jeroen, J., & van Merriënboer, G. (2014). Participatory design of learning environments: integrating perspectives of students, teachers, and designers. *Instructional Science*, *42*(1), DOI: 10.1007/s11251-013-9305-2

Kulkarni, C. E., Bernstein, M. S., & Klemmer, S. R. (2015). PeerStudio: Rapid peer feedback emphasizes revision and improves performance. *L@S '15: Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, 75-84. New York, NY: ACM. DOI: 10.1145/2724660.2724670

Loibl, K., Roll, I., & Rummel, N. (2016). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review.* (2016). *DOI:10.1007/s10648-016-9379-x* https://link.springer.com/article/10.1007/s10648-016-9379-x

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing, 18*(1), 30-43. DOI: 10.1016/j.jslw.2008.06.002

Nguyen, H., Xiong, W., & Litman, D. (2014). Classroom evaluation of a scaffolding intervention for improving peer review localization. *International Conference on Intelligent Tutoring Systems*, 272-282. New York: Springer International Publishing. DOI: 10.1007/978-3-319-07221-0_34

Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*, *8*(3), 265-289. DOI: 10.1016/S1060-3743(99)80117-9

Pollitt, Alistair (2012). The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice*, *19*(3), 281-300. DOI: 10.1080/0969594X.2012.665354

Polya, G. (1945). *How to Solve it: A New Aspect of Mathematical Method*. Princeton, NJ: Princeton University Press.

Robinson, R. (2001). Calibrated Peer Review™: An application to increase student reading & writing skills. *The American Biology Teacher*, *63*(7), 474-480. DOI: 10.1662/0002-7685(2001)063%5B0474:CPR%5D2.0.CO;2

Roll, I., Holmes, N. G., Day, J., & Bonn, D. (2012). Evaluating metacognitive scaffolding in guided invention activities. *Instructional Science*, *40*, 691-710. DOI: 10.1007/s11251-012-9208-7

Schuler, D. & Namioka, A. (Eds.). (1993). Participatory design: Principles and practices. Boca Ratan, FL: CRC Press.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, *16*(4), 475-522. DOI: 10.1207/s1532690xci1604_4

Song, L., Singleton, E. S., Hill, J. R., & Koh, M. H. (2004). Improving online learning: Student perceptions of useful and challenging characteristics. *The Internet and Higher Education*, *7*(1), 59-70. DOI: 10.1016/j.iheduc.2003.11.003 http://dx.doi.org/10.1207/s1532690xci1604_4

Star, J. R., & Rittle-Johnson, B. (2009). It pays to compare: An experimental study on computational estimation. *Journal of Experimental Child Psychology*, *102*(4), 408-26. DOI: 10.1016/j.jecp.2008.11.004

Suen, H. K. (2014). Peer assessment for massive open online courses (MOOCs). *The International Review of Research in Open and Distributed Learning*, *15*(3). DOI: 10.19173/irrodl.v15i3.1680

Thurstone, L. L. (1927). A Law of Comparative Judgment. *Psychological Review, 34*(4), 273-286. DOI: 10.1037/h0070288

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249-276. DOI: 10.3102/00346543068003249

Weimer, M. (2002). *Learner-Centered Teaching: Five Key Changes to Practice*. San Francisco, CA: Jossey-Bass.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, *16*(1), 3-118. DOI: 10.1207/s1532690xci1601_2

Zhu, W. (1995). Effects of training for peer response on students' comments and interaction. *Written Communication, 14*(2), 492-528. DOI: 10.1177/0741088395012004004