PeerStudio: Rapid Peer Feedback Emphasizes Revision and Improves Performance

Chinmay Kulkarni[1], **Michael S. Bernstein**[1], **Scott Klemmer**[2]
[1] Stanford University, [2] UC San Diego
{chinmay,msb}@cs.stanford.edu, srk@ucsd.edu

ABSTRACT
Rapid feedback is a core component of mastery learning, but feedback on open-ended work requires days or weeks in most classes today. This paper introduces *PeerStudio*, an assessment platform that leverages the large number of students' peers in online classes to enable rapid feedback on in-progress work. Students submit their draft, give rubric-based feedback on two peers' drafts, and then receive peer feedback. Students can integrate the feedback and repeat this process as often as they desire. In MOOC deployments, the median student received feedback in just twenty minutes. Rapid feedback on in-progress work improves course outcomes: in a controlled experiment, students' final grades improved when feedback was delivered quickly, but not if delayed by 24 hours. More than 3,600 students have used PeerStudio in eight classes, both massive and in-person. This research demonstrates how large classes can leverage their scale to encourage mastery through rapid feedback and revision.

AUTHOR KEYWORDS
MOOC; peer assessment; peer learning; mastery learning; deliberate practice

INTRODUCTION

Online learning need not be a loop of watching video lectures and then submitting assignments. To most effectively develop mastery, students must repeatedly revise based on *immediate, focused feedback* [12]. Revision is central to the method of deliberate practice as well as to mastery learning, and depends crucially on rapid formative assessment and applying corrective feedback [16]. In domains as diverse as writing, programming, and art, immediate feedback reliably improves learning; delaying feedback reduces its benefit [23].

Unfortunately, many learning experiences cannot offer tight feedback-revision loops. When courses assign open-ended work such as essays or projects, it can easily take a week after submission to receive feedback from peers or overworked instructors. Feedback is also often coupled with an unchangeable grade, and classes move to new topics faster than feedback arrives. The result is that many opportunities to develop mastery and expertise are lost, as students have few opportunities to revise work and no incentive to do so.

L@S 2015, March 14–15, 2014, Vancouver, Canada.
ACM 978-1-4503-3411-2/15/03.
http://dx.doi.org/10.1145/2724660.2724670

Could software systems enable peers in massive classes to provide rapid feedback on in-progress work? In massive classes, peer assessment already provides summative grades and critiques on final work [25], but this process takes days, and is often as slow as in-person classes. This paper instead introduces a peer learning design tailored for near-immediate peer feedback. It capitalizes on the scale of massive classes to connect students to trade structured feedback on drafts. This process can provide feedback to students within minutes of submission, and can be repeated as often as desired.

We present the *PeerStudio* system for fast feedback on in-progress open-ended work. Students submit an assignment draft whenever they want feedback and then provide rubric-based feedback on two others' drafts in order to unlock their own results. PeerStudio explicitly encourages mastery by allowing students to revise their work multiple times.

Even with the scale of massive classes, there are not always enough students online to guarantee fast feedback. Therefore, PeerStudio recruits students who are online already, and also those who have recently submitted drafts for review but are no longer online. PeerStudio uses a progressive recruitment algorithm to minimize the number of students emailed. It reaches out to more and more students, emailing a small fraction of those who recently submitted drafts each time, and stops recruiting immediately when enough (e.g., two) reviewers have been recruited.

This paper reports on PeerStudio's use in two massive online classes and two in-person classes. In a MOOC where 472 students used PeerStudio, reviewers were recruited within minutes (median wait time: seven minutes), and the first feedback was completed soon after (median wait time: 20 minutes). Students in the two, smaller, in-person classes received feedback in about an hour on average. Students took advantage of PeerStudio to submit full drafts ahead of the deadline, and paid particular attention to free-text feedback beyond the explicit rubric.

A controlled experiment measured the benefits of rapid feedback. This between-subjects experiment assigned participants in a MOOC to one of three groups. One control group saw no feedback on in-progress work. A second group received feedback on in-progress work 24 hours after submission. A final group received feedback as soon as it was available. Students who received fast in-progress feedback had higher final grades than the control group ($t(98)=2.1$, $p<0.05$). The speed of the feedback was critical: receiving slow feedback was statistically indistinguishable from receiving no feedback at all ($t(98)=1.07$, $p=0.28$).

PeerStudio demonstrates how massive online classes can be designed to provide feedback an order of magnitude faster than many in-person classes. It also shows how MOOC-inspired learning techniques can *scale down* to in-person classes. In this case, designing and testing systems iteratively in massive online classes led to techniques that worked well in offline classrooms as well; Wizard of Oz prototyping and experiments in small classes led to designs that work well at scale. Finally, parallel deployments at different scales help us refocus our efforts on creating systems that produce pedagogical benefits at any scale.

RELATED WORK

PeerStudio relies on peers to provide feedback. Prior work shows peer-based critique is effective both for in-person [8,29] and online classes [25], and can provide students accurate numeric grades and comments [13,25].

PeerStudio bases its design of peer feedback on prior work about how feedback affects learning. By *feedback*, we mean task-related information that helps students improve their performance. Feedback improves performance by changing students' locus of attention, focusing them on productive aspects of their work [21]. It can do so by making the difference between current and desired performance more salient [17], by explaining the cause of poor performance [5], or by encouraging students to use a different or higher standard to compare their work against [26].

Fast feedback improves performance by making the difference between the desired and current performance more salient [23]. When students receive feedback quickly (e.g., in an hour), they apply the concepts they learn more successfully [23]. In domains like mathematics, computers can generate feedback instantly, and combining such formative feedback with revision improves grades [18]. PeerStudio extends fast feedback to domains such as design and writing where automated feedback is limited and human judgment is necessary.

Feedback merely changes what students attend to, so not all feedback is useful, and some feedback degrades performance [21]. For instance, praise is frequently ineffective because it shifts attention *away* from the task and onto the self [1].

Therefore, feedback systems and curricular designers must match feedback to instructional goals. Large-scale meta-analyses suggest that the most effective feedback helps students set goals for future attempts, provides information about the quality of their current work, and helps them gauge whether they are moving towards a good answer [21]. Therefore, PeerStudio provides a low-cost way of specifying goals when students revise, uses a standardized rubric and free-form comments for correctness feedback, and a way to browse feedback on previous revisions for velocity.

How can peers provide the most accurate feedback? Disaggregation can be an important tool: summing individual scores for components of good writing (e.g. grammar and argumentation) can capture the overall quality of an essay more accurately than asking for a single writing score [9,24]. Therefore, PeerStudio asks for individual judgments with yes/no or scale questions, and not aggregate scores.

PeerStudio uses the large scale of the online classroom in order to quickly recruit reviewers after students submit in-progress work. In contrast, most prior work has capitalized on scale only after all assignments are submitted. For instance, DeduceIt uses the semantic similarity between student solutions to provide automatic hinting and to check solution correctness [14], while other systems cluster solutions to help teachers provide feedback quickly [6].

ENABLING FAST PEER FEEDBACK WITH PEERSTUDIO

Students can use PeerStudio to create and receive feedback on any number of drafts for every open-ended assignment. Because grades shift students' attention away from the task to the self [21], grades are withheld until the final version.

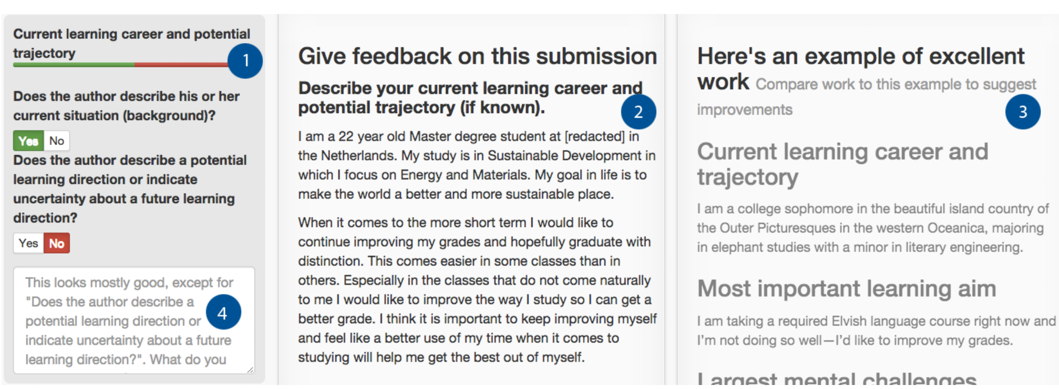

Figure 1: PeerStudio is a peer learning platform for rapid, rubric-based feedback on drafts. The reviewing interface above shows (1) the rubric, (2) the student draft, (3) an example of excellent work to compare student work against. PeerStudio scaffolds reviewers with automatically generating tips for commenting (4).

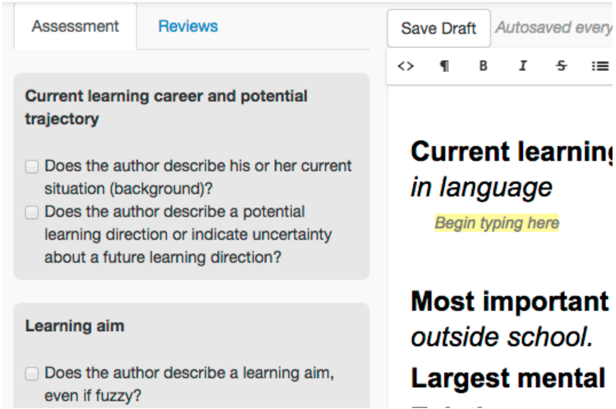

Figure 2: The drafting interface shows the assignment rubric, and a starter template. Reviews on previous versions are also available (tab, top-left).

Creating a draft, and seeking feedback

PeerStudio encourages students to seek feedback on an initial draft as early as possible. When students create their first draft for an assignment, PeerStudio shows them a minimal, instructor-provided starter template that students can modify or overwrite (Figure 2). Using a template provides a natural hint for when to seek feedback—when the template is filled out. It also provides structure to students that need it, without constraining those who don't. To keep students focused on the most important aspects of their work, students always see the instructor-provided assignment rubric in the drafting interface (Figure 2, left). Rubrics in PeerStudio comprise a number of criteria for quality along multiple dimensions.

Students can seek feedback on their current draft at any time. They can focus their reviewers' attention by leaving a note about the kind of feedback they want. When students submit their draft, PeerStudio starts finding peer reviewers. Simultaneously, it invites the student to review others' work.

Reviewing peer work

PeerStudio uses the temporal overlap between students to provide fast feedback. When a student submits their draft, PeerStudio asks them to review their peers' submissions in order to unlock their own feedback [4]. Since their own work remains strongly activated, reviewing peer work immediately encourages students to reflect [27].

Students need to review two drafts before they see feedback on their work. Reviewing is double blind. Reviewers see their peer's work, student's review request notes, the instructor-created feedback rubric, and an example of excellent work to compare against. Reviewers' primary task is to work their way down the feedback rubric, answering each question. Rubric items are all yes/no or scale responses. Each group of rubric items also contains a free-text comment box, and reviewers are encouraged to write textual comments. To help reviewers write useful comments, PeerStudio prompts them with dynamically generated suggestions.

Reading reviews and revising

PeerStudio encourages rapid revision by notifying students via email immediately after a classmate reviews their work. To enable feedback comparison, PeerStudio displays the number of reviewers agreeing on each rubric question, as well as reviewers' comments. Recall that to emphasize iterative improvement, PeerStudio does not display grades, except for final work.

After students read reviews, PeerStudio invites them to revise their draft. Since reflection and goal setting are an important part of deliberate practice, PeerStudio asks students to first explicitly write down what they learned from their reviews and what they plan to do next.

PeerStudio also uses peer assessment for final grading. Students can revise their draft any number of times before they submit a final version to be graded. The final reviewing process for graded submissions is identical to early drafts, and reviewers see the same rubric items. For the final draft, PeerStudio calculates a grade as a weighted sum of rubric items from reviews for that draft.

PeerStudio integrates with MOOC platforms through LTI, which allows students to login using MOOC credentials, and automatically returns grades to class management software. It can be also used as a stand-alone tool.

PEERSTUDIO DESIGN

PeerStudio's feedback design relies on rubrics, textual comments, and the ability to recruit reviewers quickly. We outline the design of each.

Rubrics

Rubrics effectively provide students feedback on the current state of their work for many open-ended assignments, such as writing [2,3], design [25], and art [29]. Rubrics comprise multiple dimensions, with cells describing increasing quality along each. For each dimension, reviewers select the cell that most closely describes the submission; in between values and gradations within cells are often possible. Comparing and matching descriptions encourages raters to build a mental model of each dimension that makes rating faster and cognitively more efficient [15].

When rubric cell descriptions are complex, novice raters can develop mental models that stray significantly from the rubric standard, even if it is shown prominently [24]. To mitigate the challenges of multi-attribute matching, PeerStudio asks instructors to list multiple distinct criteria of quality along each dimension (Figure 4). Raters then explicitly choose which criteria are present. Criteria can be binary *e.g.*, "did the student choose a relevant quote that logically supports their opinion?" or scales, *e.g.*, "How many people did the student interview?"

Our initial experiments and prior work suggest that given a set of criteria, raters satisfice by marking some but not all matching criteria [22]. To address this, PeerStudio displays binary questions as dichotomous choices, so students must choose either yes/no (Figure 4); and ensures that students answer scale questions by explicitly setting a value.

To calculate final grades, PeerStudio awards credit to yes/no criteria if a majority of reviewers marked it as present. To reduce the effect of outlying ratings, scale questions are given the median score of reviewers. The total assignment grade is the sum of grades across all rubric questions.

Scaffolding comments

Rubrics help students understand the current quality of their work; free-text comments from peers help them improve it. Reviews with accurate rubric scores, but without comments may provide students too little information.

To scaffold reviewers, PeerStudio shows short tips for writing comments just below the comment box. For instance, if the comment merely praises the submission and has no constructive feedback, it may remind students "Quick check: Is your feedback actionable? Are you expressing yourself succinctly?" Or it may ask reviewers to "Say more…" when they write "great job!"

To generate such feedback, PeerStudio compiles a list of relevant words from the student draft and the assignment description. For example, for a critique on a research paper, words like "contribution", "argument", "author" are relevant. PeerStudio then counts the number of relevant words a comment contains. Using this count, and the comment's total length, it suggests improvements. This simple heuristic catches a large number of low-quality comments. Similar systems have been used to judge the quality of product reviews online [19].

PeerStudio also helps students provide feedback that's most relevant to the current state of the draft, by internally calculating the reviewer's score for the submission. For a low-quality draft, it asks the reviewer, "What's the first thing you'd suggest to get started?" For middling drafts, reviewers are asked, "This looks mostly good, except for [question with a low score]. What do you suggest they try?" Together, these commenting guides result in reviewers leaving substantive comments.

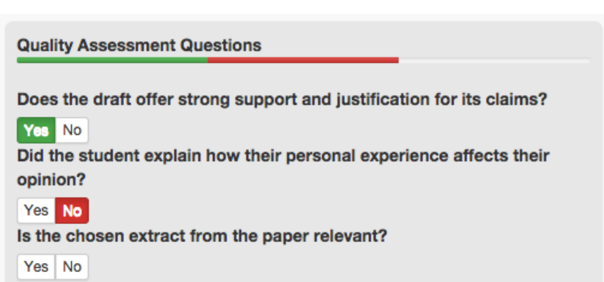

Figure 4: Example dichotomous questions in PeerStudio. The last question is not yet answered. Students must choose yes/no before they can submit the review.

Recruiting reviewers

Because students review immediately after submitting, reviewers are found quickly when there are many students submitting one after another, *e.g.*, in a popular time zone. However, students who submit at an unpopular time still need feedback quickly.

When enough reviewers are not online, PeerStudio progressively emails and enlists help from more and more students who have yet to complete their required two reviews, and enthusiastic students who have reviewed even before submitting a draft. PeerStudio emails a random selection of five such students every half hour, making sure the same student is not picked twice in a 24-hour period. PeerStudio stops emailing students when all submissions have at least one review. This enables students to quickly receive feedback from one reviewer and begin revising.

To decide which submissions to show reviewers, PeerStudio uses a priority queue. This queue prioritizes student submissions by the number of reviews (submissions with the fewest, or no, reviews have highest priority), and by the time the submission has been in the review queue. The latest submissions have the highest priority. PeerStudio seeks two reviewers per draft.

FIELD DEPLOYMENT: IN-PERSON AND AT SCALE

This paper describes PeerStudio deployments in two open online classes: *Learning How to Learn* (603 students submitting assignments), *Medical Education in the New Millennium* (103 students) on the Coursera and OpenEdX platforms respectively. We also describe deployments in two in-person classes: a senior-level class at the University of Illinois at Urbana-Champaign on *Social Visualization* (125 students), and a graduate-level class in education at Stanford University, on *Technology for Learners* (51 students).

All four classes used PeerStudio for open-ended writing assignments. In *Learning how to Learn*, for their first

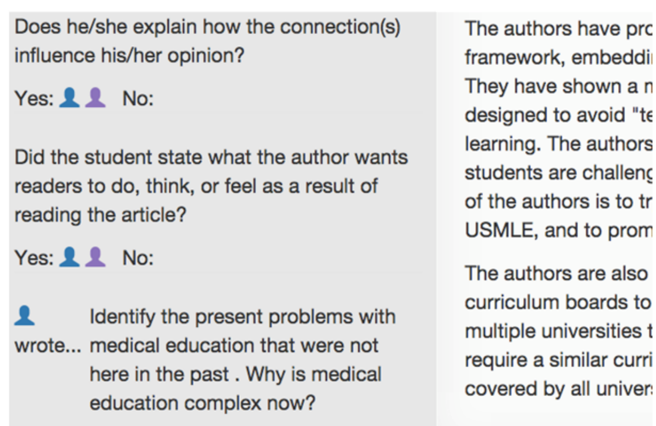

Figure 3: Students see reviews in the context of their draft (right, clipped). PeerStudio displays the number of reviewers (two here) agreeing on each rubric question and comments from each.

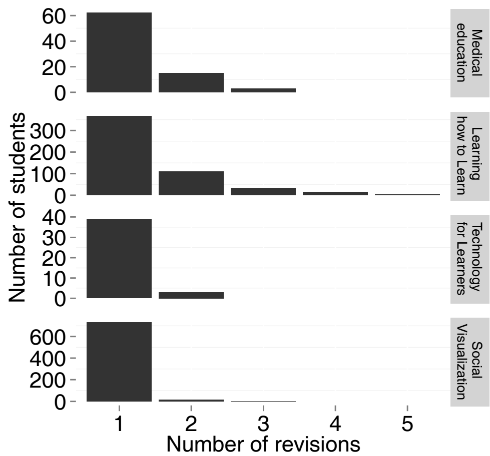

Figure 5: Most students created a single revision. Students in MOOCs revised more than students in in-person classes.

assignment students wrote an essay about a learning goal and how they planned to accomplish it using what they learned in class (*e.g.*, one student wrote about being "an older student in Northern Virginia retooling for a career in GIS after being laid off"). In the second assignment, they created a portfolio, blog or website to explain what they learned to others (*e.g.*, one wrote: "I am a professor of English as a Second Language at a community college. I have created a PowerPoint presentation for my colleagues [about spaced repetition and frequent testing]").

The Social Visualization and Medical Education classes asked students to critique research papers in the area. In Social Visualization, students also used PeerStudio for an open-ended design project on data visualization (*e.g.*, one student team designed a visualization system that used data from Twitter to show crisis needs around the US). Finally, the Technology for Learners class used PeerStudio as a way to critique a learning tool (*e.g.*, ClassDojo, a classroom discipline tool). This class requested its reviewers to sign reviews, so students could follow-up with each other for lingering questions.

Deployment observations

Throughout these deployments, we read students' drafts, feedback, and revisions. We regularly surveyed students about their experiences, and spoke to instructors about their perspectives. Several themes emerged.

Students requested feedback on full rough drafts

Rather than submit sections of drafts, students submitted full rough drafts. Drafts were often missing details (*e.g.,* lacking examples). In the Medical Education critique, one question was "did you find yourself mostly agreeing or mostly disagreeing with the content of the research paper? Why?" In initial drafts, students often pointed out only one area of disagreement, later drafts added the rest. Other drafts were poorly explained (*e.g.,* lacking justification for claims) or too rambling.

Students typically asked for four kinds of feedback: 1) On a specific aspect of their work, *e.g.*, "I guess I need help with my writing, vocabulary and grammar, since I'm not an English native-speaker"; 2) On a specific component of the assignment: *e.g.*, "Can you let me know if part 4 and 5 make sense—I feel like I am trying to say too much all in one go!" 3) As a final check before they turned in their work: *e.g.*, "This draft is actually a 'release candidate'. I would like to know if I addressed the points or if I missed something." 4) As a way to connect with classmates: *e.g.*, "I just want to hear about your opinions :)".

When students revised their draft, we asked, "Overall, did you get useful feedback on your draft?" as a binary question—80% answered 'yes'.

Students revise rarely, especially in in-person classes

Most students did not create multiple drafts (Figure 5). Students in the two MOOCs were more likely to revise than students in in-person classes ($t(1404) = 12.84$, $p < 0.001$). Overall, 30.1% of online students created multiple revisions, but only 7% of those in in-person classes did.

When we asked TAs in the in-person classes why so few students revised, they told us they did not emphasize this feature of PeerStudio in class. Furthermore, student responses in surveys indicated that many felt their schedule was too busy to revise. One wrote it was unfair to "expect us to read some forty page essays, then write the critiques and then review two other people, and then make changes on our work... twice a week." These comments underscore that merely creating software systems for iterative feedback is not enough — an iterative approach must be reflected in the pedagogy as well.

Students see comments as more useful than rubric feedback

Students could optionally rate reviews after reading them and leave comments to staff. Students rated 758 of 3,963 reviews. We looked at a random subset of 50 such comments. In their responses, students wrote that freeform comments were useful (21 responses) more often than rubric-based feedback (5 responses). Students also disagreed more with reviewers' comments (7 responses) than with their reviewers' marked rubric (3 responses). This is possibly because comments can capture useful interpretive feedback, but differences in interpretation lead to disagreement.

An undergraduate TA looked at a random subset of 150 student submissions, and rated reviewer comments on a 7-point Likert scale on how concretely they helped students revise. For example, here is a comment that was rated "very concrete (7)" on an essay about planning for learning goals:

"What do you mean by 'good schedule'? There's obviously more than one answer to that question, but the goal should be to really focus and narrow it down. Break a larger goal like "getting a good schedule" into concrete steps such as: 1) get eight hours of sleep, 2)…

We found 45% of comments were "somewhat concrete" (a rating of 5 on the scale) or better, and contained pointers to resources or specific suggestions on how to improve; the rest of the comments were praise or encouragement. Interestingly, using the same 7-point Likert scale, students rated reviews as concrete more often than the TA (55% of the time).

Students reported relying on comments for revising. For instance, the student who received the above comment wrote, "I somehow knew I wasn't being specific… The reviewer's ideas really helped there!" The lack of comments was lamented upon, "The reviewer did not comment any feedback, so I don't know what to do."

One exception to the general trend of comments being more important was students who submitted 'release candidate' drafts for a final check. Such students relied heavily on rubric feedback: "I have corrected every item that needed attention to. I now have received all yes to each question. Thanks guys. :-)"

Comments encourage students to revise

The odds of students revising their drafts increase by 1.10 if they receive any reviews with free-form comments ($z=4.6$, $p < 0.001$). Since fewer than half the comments contained specific improvement suggestions, this suggests that, in addition to being informational, reviewer comments also play an important motivational role.

Revisions locally add information, improve understandability

We looked at the 100 reflections that students wrote while starting the revision to understand what changes they wanted to make. A majority of students (51%) intended to add information based on their comments, *e.g.*, "The math teacher [one of the reviewers] helped me look for other sources relating to how math can be fun and creative instead of it being dull!" A smaller number (16%) wanted to change how they had expressed ideas to make them easier to understand, *e.g.*, "I did not explain clearly the three first parts… I shall be clearer in my re-submission" and, "I do need to avoid repetition. Bullets are always good." Other changes included formatting, grammar, and occasionally wanting a fresh start. The large fraction of students who wanted to add information to drafts they previously thought were complete suggests that peer feedback helps students see flaws in their work, and provides new perspectives.

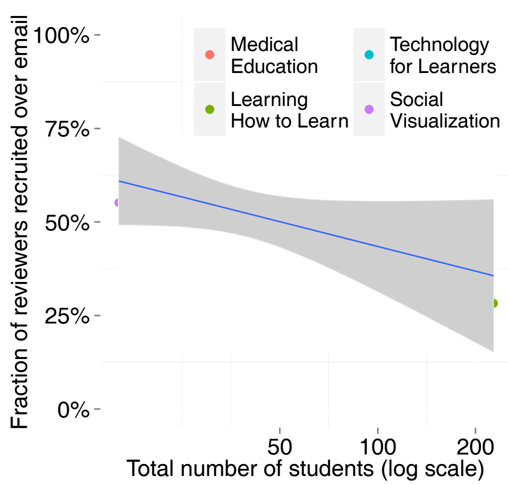

Figure 7: More students in large classes are likely to be online at the same time, so fewer reviewers were recruited by email.

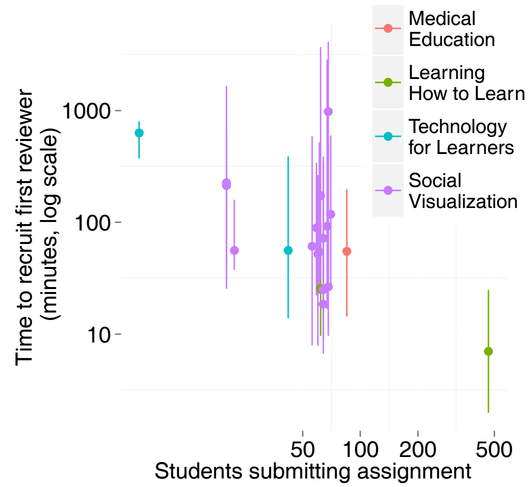

Figure 6: Reviewers are recruited faster in larger classes.

Most students reworked their drafts as planned: 44% of students made substantive changes based on feedback, 10% made substantive changes not based on the comments received, and the rest only changed spelling and formatting. Most students added information to or otherwise revised one section, while leaving the rest unchanged.

PeerStudio recruits reviewers rapidly

We looked at the PeerStudio logs to understand the platform's feedback latency. Reviewers were recruited rapidly for both in in-person and online classes (see Figure 6), but the scale of online classes has a dramatic effect. With just 472 students using the system for the first assignment in Learning How to Learn, the median recruitment-time was 7 minutes and the 75^{th} quartile was 24 minutes.

Few students have long wait times

PeerStudio uses a priority queue to seek reviews; it prioritizes newer submissions given two submissions with the same number of reviews. This reduces the wait time for the *average* student, but unlucky students have to wait longer (e.g. when they submit just before a popular time, and others keep submitting newer drafts). Still, significant delays are rare: 4.4% had no reviews in the first 8 hours; 1.8% had no reviews in 24 hours. To help students revise, staff reviewed submissions with no reviews after 24 hours.

Feedback latency is consistent even early in the assignment

Even though fewer students use the website farther from the deadline, peer review means that the workload and review labor automatically scale together. We found no statistical difference in recruitment time (t(1191) = 0.52, p = 0.6) between the first two and last two days of the assignment, perhaps because PeerStudio uses email to recruit reviewers.

Fewer reviewers recruited over email with larger class size. PeerStudio emails students to recruit reviewers only when enough students aren't already on the website. In the smallest class with 46 students submitting, 21% of reviews came from Web solicitation and 79% of reviews were written in response to an emailed request. In the largest, with 472 students submitting, 72% of reviews came from Web solicitation and only 28% from email (Figure 7). Overall, students responded to email requests approximately 17% of the time, independent of class size.

These results suggest that PeerStudio achieves quick reviewing in small, in-person classes by actively bringing students online via email, and that this becomes less important with increasing class size, as students have a naturally overlapping presence on site.

Reviewers spend about ten minutes per draft
PeerStudio records the time between when reviewers start a review and when they submit it. In all classes except the graduate level Technology for Learners, students spent around 10 minutes reviewing each draft (Figure 9). The median reviewer in the graduate Technology for Learners class spent 22 minutes per draft. Because all students in that class started reviewing in-class but finished later, its variance in reviewing times is also much larger.

Are reviewers accurate?

There is very strong agreement between individual raters while using the rubric. In online classes, the median pairwise agreement between reviewers on a rubric question is 74%, while for in-person classes it is 93%. However, because most drafts completed a majority of the rubric items successfully, baseline agreement is high, so Fleiss' κ is low. The median κ=0.19 for in-person classes, and 0.33 for online classes, conventionally considered "Fair agreement". In in-person classes, on average staff and students agreed on rubric questions 96% of the time.

Staff and peers write comments of similar length

Both in-person and online, the median comment was 30 words long (Figure 8). This length compares well with staff comments in the *Social Visualization* class, which had a median of 35 words. Most reviews (88%) had at least some textual comments, in addition to rubric-based feedback.

Students trade-off reviewing and revising

23% of students reviewed more than the required two drafts. Survey results indicated that many such students used reviewing as an inexpensive way to make progress on their own draft. One student wrote that in comparison to revising their own work, "being able to see what others have written by reviewing their work is a better way to get feedback." Other students reviewed peers simply because they found their work interesting. When told she had reviewed 29 more drafts than required, one student wrote, "I wouldn't have suspected that. I kept reading and reviewing because people's stories are so interesting."

Students appreciate reading others' work more than early feedback and revision

A post-class survey in *Technology For Learners* asked students what they liked most about PeerStudio (30 responses). Students most commonly mentioned (in 13 responses) interface elements such as being able to see examples and rubrics. Reading each other's work was also popular (8 responses), but the ability to revise was rarely mentioned (3 responses). This is not surprising, since few students revised work in in-person classes.

Apart from specific usability concerns, students' most frequent complaint was that PeerStudio sent them too much

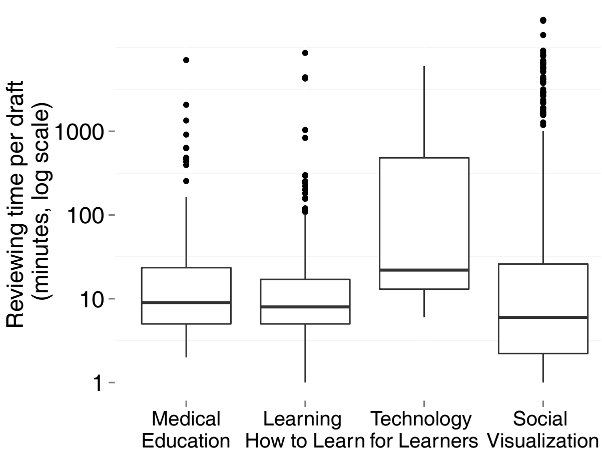

Figure 9: Reviewers spend roughly 10 minutes reviewing each draft. The graduate-level Technology for Learners class spends longer. (The larger variation is because students start reviewing in class, and finish later.)

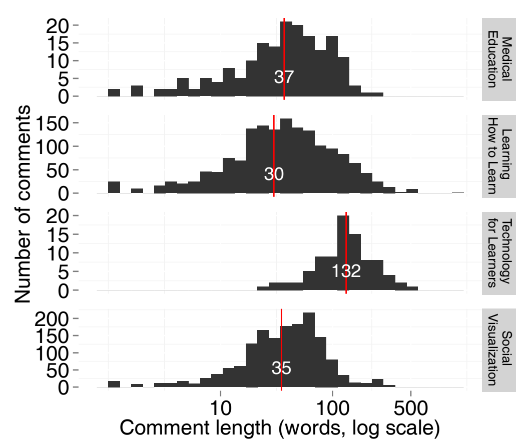

Figure 8: Students write substantive comments, both in-person and online. The graduate level Technology for Learners has longer comments, possibly because reviews were signed.

email. One wrote, "My understanding was that students would receive about three, but over the last few days, I've gotten more." Currently, PeerStudio limits how frequently it emails students; future work could also limit the total number of emails a student receives.

FIELD EXPERIMENT: DOES FAST FEEDBACK ON IN-PROGRESS WORK IMPROVE FINAL WORK?

The prior study demonstrated how students solicited feedback and revised work, and how quickly they can obtain feedback. Next, we describe a field experiment that asks two research questions: First, does feedback on in-progress work improve student performance? Second, does the speed of feedback matter? Do students perform better if they receive rapid feedback? We conducted this controlled experiment in *ANES 204: Medical Education in the New Millennium*, a MOOC on the OpenEdX platform.

Students in this class had working experience in healthcare professions, such as medical residents, nurses and doctors. In the open-ended assignment, students read and critiqued a recent research paper based on their experience in the healthcare field. For example, one critique prompt was "As you read, did you find yourself mostly agreeing or mostly disagreeing with the content? Write about three points from the article that justify your support or dissent." The class used PeerStudio to provide students both in-progress feedback and final grades.

Method

A between-subjects manipulation randomly assigned students to one of three conditions. In the *No Early Feedback* condition, students could only submit one final draft of their critique. This condition generally mimics the status quo in many classes, where students have no opportunities to revise drafts with feedback. In the *Slow Feedback* condition, students could submit any number of in-progress drafts, in addition to their final draft. Students received peer feedback on all drafts, but this feedback wasn't available until 24 hours after submission. Additionally, students were only emailed about their feedback at that time. This condition mimics a scenario where a class offers students the chance to revise, but is limited in its turnaround time due to limited staff time or office hours. Finally, in the *Fast Feedback* condition, students could submit drafts as in the slow feedback condition, but were shown reviews as soon as available, mirroring the standard PeerStudio setup.

Students in all conditions rated their peers' work anonymously; reviewers saw drafts from all conditions and rated them blind to condition. Our server introduced all delays for the Slow Feedback condition after submission. Rubrics and the interface students used for reviewing and editing were identical across conditions.

Measures

To measure performance, we used the grade on the final assignment submission as calculated by PeerStudio. Since rubrics only used dichotomous questions, each rubric question was given credit if a majority of raters marked "yes". The grade of each draft was the sum of credit across all rubric questions for that draft.

Participants

In all, 104 students participated. Of these, three students only submitted a blank essay; their results were discarded from analysis. To analyze results, we built an ordinary-least-squares regression model with the experimental condition as the predictor variable, using *No Early Feedback* as the baseline (R^2=0.02).

Manipulation check

While PeerStudio can provide students feedback quickly, this feedback is only useful if students actually read it. Therefore, we recorded the time students first read their feedback. The median participant in the *Fast Feedback* condition read their reviews 592 minutes (9.8 hours) after submission; the median for the Slow Feedback condition was 1528 minutes (26.6 hours). This suggests that the manipulation effectively delayed feedback, but the difference between conditions was more modest than planned.

Results: fast early feedback improves final grades

Students in the *Fast Feedback* condition did significantly better than those in *No Early Feedback* condition ($t(98)=2.1$, $p<0.05$). On average, students scored higher by 4.4% of the assignment's total grade: i.e., enough to boost a score from a B+ to an A-.

Slow early feedback yields no significant improvement

Surprisingly, we found that students in the *Slow Feedback* condition did not do significantly better than those in the *No Early Feedback* condition ($t(98)=1.07$, $p=0.28$). These results suggest that for early feedback to improve student performance, it must be delivered quickly.

Because of the limited sample size, it is also possible this experiment was unable to detect the (smaller) benefits of delayed early feedback.

Students with fast feedback don't revise more often

There was no significant difference between the number of revisions students created in the *Fast* and *Slow feedback* conditions ($t(77)=0.2$, $p=0.83$): students created on average 1.33 drafts; only 22% of students created multiple revisions. On average, they added 83 words to their revision, and there was no significant difference in the quantity of words changed between conditions ($t(23)=1.04$, $p=0.30$).

However, students with *Fast feedback* referred to their reviews marginally more frequently when they entered reflections and planned changes in revision ($\chi^2(1)=2.92$, $p=0.08$). This is consistent with prior findings that speed improves performance by making feedback more salient.

Even with only a small number of students revising, the overall benefits of early feedback seem sizeable. Future work that better encourages students to revise may further increase these benefits.

DISCUSSION

The field deployment and subsequent experiment demonstrate the value of helping students revise work with fast feedback. Even with a small fraction of students creating multiple revisions, the benefits of fast feedback are apparent. How could we design pedagogy to amplify these benefits?

Redesigning pedagogy to support revision and mastery

In-person classes are already using PeerStudio to change their pedagogy. These classes did not use PeerStudio as a way to reduce grading burden: both classes still had TAs grade every submission. Instead, they used PeerStudio to expose students to each other's work and to provide them feedback faster than staff could manage.

Fully exploiting this opportunity will require changes. Teachers will need to teach students about when and how to seek feedback. Currently, PeerStudio encourages students to fill out the starter template before they seek feedback. For some domains, it may be better to get feedback using an outline or sketch, so reviewers aren't distracted by superficial details [28]. In domains like design, it might be useful to get feedback on multiple alternative designs [10]. PeerStudio might explicitly allow these different kinds of submissions.

PeerStudio reduces the time to get feedback, but students still need time to work on revisions. Assignments must factor this revision time into their schedule. We find it heartening that 7% of in-person students actually revised their drafts, even when their assignment schedules were not designed to allow it. That 30% of online students revised assignments may partly be because schedules were designed around the assumption that learners with full-time jobs have limited time: consequently, online schedules often provide more time between assignment deadlines.

Finally, current practice rewards students for the final quality of their work. PeerStudio's revision process may allow other reward schemes. For instance, in domains like design where rapid iteration is prized [7,11], classes may reward students for sustained improvement.

Plagiarism

Plagiarism is a potential risk of sharing in-progress work. While plagiarism is a concern with all peer assessment, it is especially important in PeerStudio because the system shares work before assignments are due. In classes that have used PeerStudio so far, we found one instance of plagiarism: a student reviewed another's essay and then submitted it as their own. While PeerStudio does not detect plagiarism currently, it does record what work a student reviewed, as well as every revision. This record can help instructors check that the work has a supporting paper trail. Future work could automate this.

Another risk is that student reviewers may attempt to fool PeerStudio by giving the same feedback to every assignment they review (to get past the reviewing hurdle quickly so they can see feedback on their work). We observed three such instances. However, 'shortcut reviewing' is often easy to catch with techniques such as inter-rater agreement scores [20].

Bridging the in-person and at-scale worlds

While it was designed for massive classes, PeerStudio "scales down" and brings affordances such as fast feedback to smaller in-person classes. PeerStudio primarily relies on the natural overlap between student schedules at larger scales, but this overlap still exists at smaller scale and can be augmented via email recruitment.

PeerStudio also demonstrates the benefits of experimenting in different settings in parallel. Large-scale between-subjects experiments often work better online than in-person because in-person, students are more likely to contaminate manipulations by communicating outside the system. In contrast, in-person experiments can often be run earlier in software development using lower-fidelity approaches and/or greater support. Also, it can be easier to gather rich qualitative and observational data in person, or modify pilot protocols on the fly. Finally, consonant results in in-person and online deployments lend more support for the fundamentals of the manipulation (as opposed to an accidental artifact of a deployment).

Future work

Some instructors we spoke to worried about the overhead that peer assessment entails (and chose not to use PeerStudio for this reason). If reviewers spend about 10 minutes reviewing work as in our deployment, peer assessment arguably incurs a 20-minute overhead per revision. On the other hand, student survey responses indicate that they found looking at other students' work to be the most valuable part of the assessment process. Future work could quantify the benefits of assessing peer work, including inspiration, and how it affects student revisions. Future work could also reduce the reviewing burden by using early reviewer agreement to hide some rubric items from later reviewers [24].

Matching reviewers and drafts

PeerStudio enables students to receive feedback from peers at any time, but their peers may be far earlier or more advanced in their completion of the assignment. Instead, it may be helpful to have drafts reviewed by students who are similarly advanced or just starting. Furthermore, students learn best from examples (peer work) if they are approachable in quality. In future work, the system could ask or learn the rough state of the assignment, and recruit reviewers who are similar.

CONCLUSION

This paper suggests that the scale of massive online classes enables systems that drastically and reliably reduce the time to obtain feedback and creates a path to iteration, mastery and expertise. These advantages can also be scaled-down to in-person classrooms. In contrast to today's learn-and-submit model of online education, we believe that the continuous presence of peers holds the promise of a far more dynamic and iterative learning process.

ACKNOWLEDGMENTS

We are grateful to Courtney Noh for contributing to software development; Coursera and OpenEdX for platform integration and encouraging instructors to try PeerStudio; instructors of MOOCs who use PeerStudio in their classes; and the many students who taught and learned with their peers and shared their stories with us. This research was funded in part through NSF grants #1351131 and #1444865, the Hasso Plattner Design Thinking Program, and the Siebel Scholars Program. This research was conducted under Stanford IRB protocol #30324.

REFERENCES

1. Anderson, S. and Rodin, J. Is Bad News Always Bad?: Cue and Feedback Effects on Intrinsic Motivation. *Journal of Applied Social Psychology 19*, 6 (1989), 449–467.
2. Andrade, H.G. The Effects of Instructional Rubrics on Learning to Write. *Current Issues in Education 4*, 4 (2001).
3. Andrade, H.G. Teaching with rubrics: The good, the bad, and the ugly. *College Teaching 53*, 1 (2005), 27–31.
4. André, P., Bernstein, M., and Luther, K. Who gives a tweet? *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12*, ACM Press (2012), 471.
5. Balcazar, F.E., Hopkins, B.L., and Suarez, Y. A critical, objective review of performance feedback. *Journal of Organizational Behavior Management 7*, 2 (1986).
6. Brooks, M., Basu, S., Jacobs, C., and Vanderwende, L. Divide and Correct: Using Clusters to Grade Short Answers at Scale. *Learning at Scale*, (2014).
7. Buxton, B. *Sketching user experiences: getting the design right and the right design.* Morgan Kaufmann, 2007.
8. Carlson, P.A. and Berry, F.C. Calibrated Peer Review and assessing learning outcomes. *Frontiers in Education Conference*, (2003).
9. Dawes, R.M. The robust beauty of improper linear models in decision making. *American psychologist 34*, 7 (1979), 571.
10. Dow, S., Fortuna, J., Schwartz, D., Altringer, B., and Klemmer, S. Prototyping dynamics: sharing multiple designs improves exploration, group rapport, and results. *Proceedings of the 2011 annual conference on Human factors in computing systems*, (2011), 2807–2816.
11. Dow, S.P., Heddleston, K., and Klemmer, S.R. The efficacy of prototyping under time constraints. *Proceeding of the ACM conference on Creativity and cognition*, ACM Press (2009), 165.
12. Ericsson, K.A., Krampe, R.T., and Tesch-Römer, C. The role of deliberate practice in the acquisition of expert performance. *Psychological review 100*, 3 (1993).
13. Falchikov, N. and Goldfinch, J. Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of educational research 70*, 3 (2000), 287–322.
14. Fast, E., Lee, C., Aiken, A., Bernstein, M.S., Koller, D., and Smith, E. Crowd-scale interactive formal reasoning and analytics. *Proceedings of the 26th annual ACM symposium on User interface software and technology*, (2013).
15. Gray, W.D. and Boehm-Davis, D.A. Milliseconds matter: an introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of experimental psychology. Applied 6*, 4 (2000), 322–35.
16. Guskey, T.R. Closing Achievement Gaps: Revisiting Benjamin S. Bloom's "Learning for Mastery." *Journal of Advanced Academics 19*, 1 (2007), 8–31.
17. Hattie, J. and Timperley, H. The Power of Feedback. *Review of Educational Research 77*, 1 (2007), 81–112.
18. Heffernan, N., Heffernan, C., Dietz, K., Soffer, D., Pellegrino, J. W. Goldman, S.R., and Dailey, M. Improving Mathematical Learning Outcomes Through Automatic Reassessment and Relearning. *AERA*, (2012).
19. Kim, S.-M., Pantel, P., Chklovski, T., and Pennacchiotti, M. Automatically assessing review helpfulness. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics (2006), 423–430.
20. Kittur, A., Chi, E.H., and Suh, B. Crowdsourcing user studies with Mechanical Turk. *Proc. of CHI*, ACM Press (2008), 453.
21. Kluger, A.N. and DeNisi, A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin 119*, 2 (1996), 254–284.
22. Krosnick, J.A. Survey research. *Annual review of psychology 50*, 1 (1999), 537–567.
23. Kulik, J.A. and Kulik, C.-L.C. Timing of Feedback and Verbal Learning. *Review of Educational Research 58*, 1 (1987), 79–97.
24. Kulkarni, C., Socher, R., Bernstein, M.S., and Klemmer, S.R. Scaling Short-answer Grading by Combining Peer Assessment with Algorithmic Scoring. *ACM Conf on Learning@Scale*, (2014).
25. Kulkarni, C., Wei, K.P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., Koller, D., and Klemmer, S.R. Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction (TOCHI) 20*, 6 (2013), 33.
26. Latham, G.P. and Locke, E.A. Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes 50*, 2 (1991), 212–247.
27. Marsh, R.L., Landau, J.D., and Hicks, J.L. How examples may (and may not) constrain creativity. *Memory & cognition 24*, 5 (1996), 669–80.
28. Sommers, N. Responding to Student Writing. *College Composition and Communication 33*, 2 (1982), 148–156.
29. Tinapple, D., Olson, L., and Sadauskas, J. CritViz: Web-Based Software Supporting Peer Critique in Large Creative Classrooms. *Bulletin of the IEEE Technical Committee on Learning Technology 15*, 1 (2013), 29.