Designing Scalable and Sustainable Peer Interactions Online

Chinmay Kulkarni, Yasmine Kotturi, Michael S. Bernstein, and Scott Klemmer

Abstract When students work with peers, they learn more actively, build richer knowledge structures, and connect material to their lives. However, not every peer learning experience online sees successful adoption. This chapter first 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. PeerStudio demonstrates how rapid feedback on in-progress work improves course outcomes. We then articulate and address three adoption and implementation challenges for peer learning platforms such as PeerStudio. First, peer interactions struggle to bootstrap critical mass. However, class incentives can signal importance and spur initial usage. Second, online classes have limited peer visibility and awareness, so students often feel alone even when surrounded by peers. We find that highlighting interdependence and strengthening norms can mitigate this issue. Third, teachers can readily access "big" aggregate data but not "thick" contextual data that helps build intuitions, so software should guide teachers' scaffolding of peer interactions. We illustrate these challenges through studying 8500 students' usage of PeerStudio and another peer learning platform: Talkabout. Efficacy is measured through sign-up and participation rates and the structure and duration of student interactions. This research demonstrates how large classes can leverage their scale to encourage mastery through rapid feedback and revision, and suggests secret ingredients to make such peer interactions sustainable at scale.

C. Kulkarni (✉) • M.S. Bernstein
Stanford University, Stanford, CA, USA
e-mail: chinmay@cs.stanford.edu

Y. Kotturi • S. Klemmer
University of California, San Diego, CA, USA

1 Introduction

Many online classes use video lectures and individual student exercises to instruct and assess students. While vast numbers of students log on to these classes individually, many of the educationally valuable social interactions of brick-and-mortar classes are lost: online learners are "alone together" (Turkle 2011).

Social interactions amongst peers improve conceptual understanding and engagement, in turn increasing course performance and completion rates (Porter et al. 2013; Konstan et al. 2014; Kulkarni et al. 2015; Crouch and Mazur 2001; Smith et al. 2009). Benefits aren't limited to the present: when peers construct knowledge together, they acquire critical-thinking skills crucial for life after school (Bransford and Schwartz 1999). Common social learning strategies include discussing course materials, asking each other questions, and reviewing each other's work (Bransford et al. 2000).

However, most peer learning techniques are designed for small classes with an instructor co-present to facilitate, coordinate, and troubleshoot the activity. These peer activities rely on instructors to enforce learning scripts that enable students to learn from the interaction, thus imposing challenges to implementation of peer learning platforms online, at scale (O'Donnell and Dansereau 1995).

How might software enable peer benefits in online environments, where massive scale prevents instructors from personally structuring and guiding peer interactions. Recent work has introduced peer interactions for summative assessment (Kulkarni et al. 2013). How might peer interactions power more pedagogical processes online? In particular, how might software facilitate social coordination?

1.1 Two Peer Learning Platforms

Over the last 2 years, we have developed and deployed two large-scale peer-learning platforms. The first, Talkabout (Fig. 1), brings students in MOOCs together to discuss course materials in small groups of four to six students over Google Hangouts (Cambre et al. 2014). Currently, over 4500 students from 134 countries have used Talkabout in 18 different online classes through the Coursera and Open edX platforms. These classes covered diverse topics: Women's Rights, Social Psychology, Philanthropy, Organizational Analysis, and Behavioral Economics. Students join a discussion timeslot based on their availability, and upon arriving to the discussion, are placed in a discussion group; on average there are four countries represented per discussion group.

We have seen that students in discussions with peers from diverse regions outperformed students in discussions with more homogenous peers, in terms of retention and exam score (Kulkarni et al. 2015). We hypothesize that geographically diverse discussions catalyze more active thinking and reflection. A detailed description in our previous year's chapter in *Design Thinking Research*, discusses
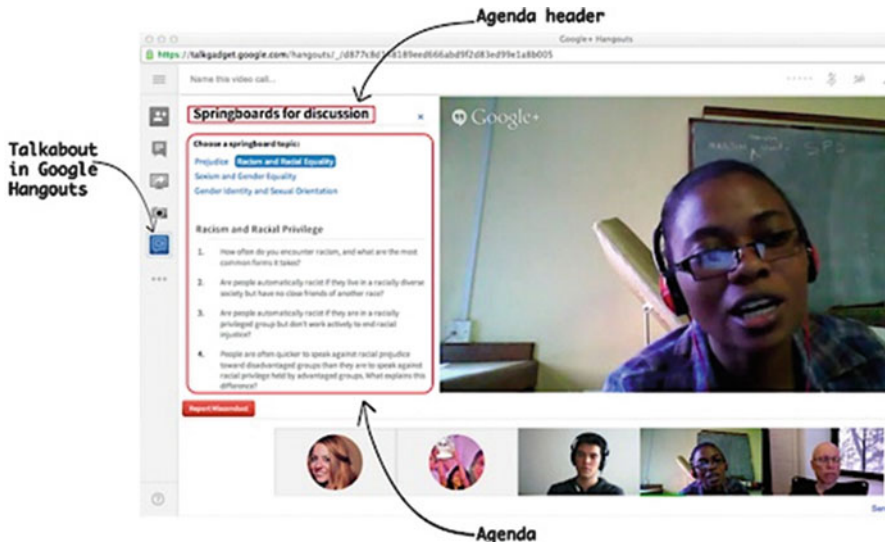
Fig. 1 Our experiences developing and deploying Talkabout, a tool that brings students in massive courses together, helps guide our discussion in the latter part of this chapter

Talkabout's design and pedagogical implications in depth (Plattner et al. 2015). The second platform PeerStudio (Fig. 2) provides fast feedback on in-progress open-ended work, such as essays (Kulkarni et al. 2015). For the implementation and adoption data on PeerStudio, we refer to its use by 4000 students in two courses on Coursera and Open edX. However, we begin with a deep dive into the motives and design for PeerStudio.

1.2 An In-Depth Look: PeerStudio

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* (Ericsson et al. 1993). 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 (Guskey 2007). In domains as diverse as writing, programming, and art, immediate feedback reliably improves learning; delaying feedback reduces its benefit (Kulik and Kulik 1987).

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

Fig. 2 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)

opportunities to develop mastery and expertise are lost, as students have few opportunities to revise work and no incentive to do so.

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 (Kulkarni et al. 2013), but this process takes days, and is often as slow as in-person classes. This chapter 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 chapter first 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: 7 min), and the first feedback was completed soon after (median wait time: 20 min). 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 h 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

2 Peerstudio: Related Work

PeerStudio relies on peers to provide feedback. Prior work shows peer-based critique is effective both for in-person (Carlson and Berry 2003; Tinapple et al. 2013) and online classes (Kulkarni et al. 2013), and can provide students accurate numeric grades and comments (Falchikov and Goldfinch 2000; Kulkarni et al. 2013).

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 (Kluger and DeNisi 1996). It can do so by making the difference between current and desired performance more salient (Hattie and Timperley 2007), by explaining the cause of poor performance (Balcazar et al. 1986), or by encouraging students to use a different or higher standard to compare their work against (Latham and Locke 1991).

Fast feedback improves performance by making the difference between the desired and current performance more salient (Kulik and Kulik 1987). When students receive feedback quickly (e.g., in an hour), they apply the concepts they learn more successfully (Kulik and Kulik 1987). In domains like mathematics, computers can generate feedback instantly, and combining such formative feedback with revision improves grades (Heffernan et al. 2012). 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 (Kluger and DeNisi 1996). For instance,

praise is frequently ineffective because it shifts attention *away* from the task and onto the self (Anderson and Rodin 1989).

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 (Kluger and DeNisi 1996). 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 (Dawes 1979; Kulkarni et al. 2014). 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, Deduce It uses the semantic similarity between student solutions to provide automatic hinting and to check solution correctness (Fast et al. 2013), while other systems cluster solutions to help teachers provide feedback quickly (Brooks et al. 2014).

3   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 (Kluger and DeNisi 1996), grades are withheld until the final version.

3.1   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 (Fig. 3). 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 (Fig. 3, left). Rubrics in PeerStudio comprise a number of criteria for quality along multiple dimensions.

Fig. 3 The drafting interface shows the assignment rubric, and a starter template. Reviews on previous versions are also available (tab, *top-left*)
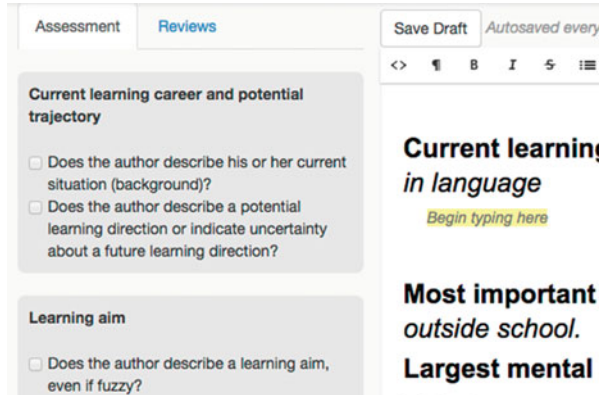
Students can seek feedback on their current draft at any time. They can focus their reviewers' attention by leaving a note about the kind of feedback they want. When students submit their draft, PeerStudio starts finding peer reviewers. Simultaneously, it invites the student to review others' work.

3.2 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 (André et al. 2012). Since their own work remains strongly activated, reviewing peer work immediately encourages students to reflect (Marsh et al. 1996).

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.

3.3 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.

4   Peerstudio Design

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

4.1   Rubrics

Rubrics effectively provide students feedback on the current state of their work for many open-ended assignments, such as writing (Andrade 2001; Andrade 2005), design (Kulkarni et al. 2013), and art (Tinapple et al. 2013). 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 (Gray and Boehm-Davis 2000) (Fig. 4).

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 (Kulkarni et al. 2014). To mitigate the challenges of multi-attribute matching, PeerStudio asks instructors to list multiple distinct criteria of quality along each dimension (Fig. 5). 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 (Krosnick 1999). To address

Fig. 4 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
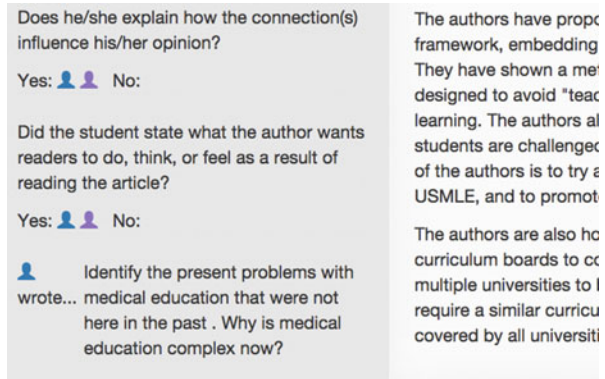
Fig. 5 Example dichotomous questions in PeerStudio. The last question is not yet answered. Students must choose yes/no before they can submit the review
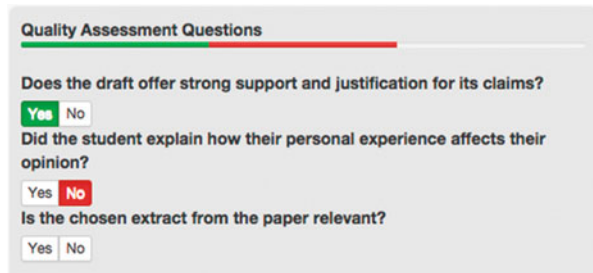
this, PeerStudio displays binary questions as dichotomous choices, so students must choose either yes/no (Fig. 5); and ensures that students answer scale questions by explicitly setting a value.

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.

4.2 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 (Kim et al. 2006).

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.

4.3   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-h 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.

5   Field Deployment: In-Person and at Scale

This chapter 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 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.

5.1 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.

5.1.1 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".

Fig. 6 Most students created a single revision. Students in MOOCs revised more than students in in-person classes
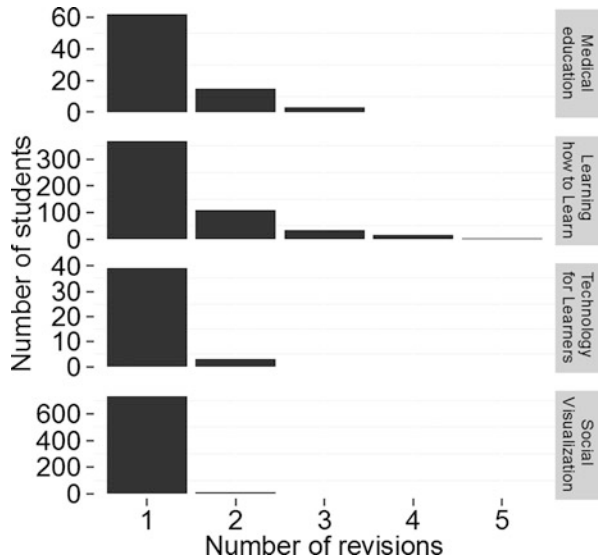
When students revised their draft, we asked, "Overall, did you get useful feedback on your draft?" as a binary question—80 % answered 'yes'.

5.1.2 Students Revise Rarely, Especially in In-Person Classes

Most students did not create multiple drafts (Fig. 6). 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 40 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.

5.1.3 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 3963 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 8 h 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. :-)"

5.1.4 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.

5.1.5 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.

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.

5.2   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 Fig. 7), 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 min and the 75th quartile was 24 min.

5.2.1   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

Fig. 7 Reviewers are recruited faster in larger classes
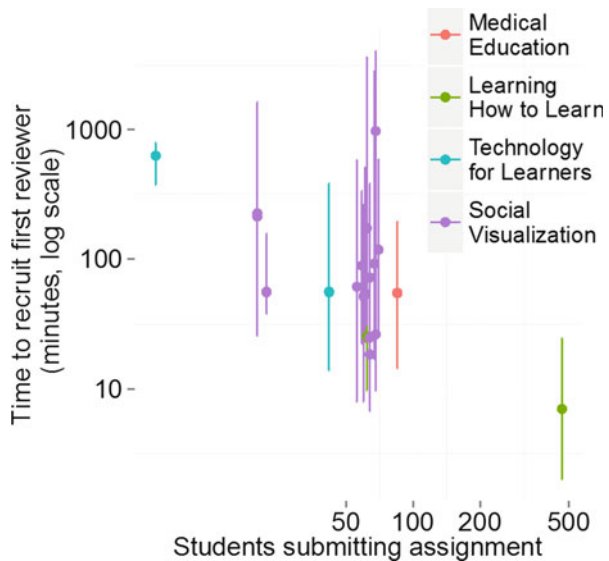
for the *average* student, but unlucky students have to wait longer (e.g. when they submit just before a popular time, and others keep submitting newer drafts). Still, significant delays are rare: 4.4 % had no reviews in the first 8 h; 1.8 % had no reviews in 24 h. To help students revise, staff reviewed submissions with no reviews after 24 h.

5.2.2 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.

5.2.3 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 (Fig. 8). 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.

5.2.4 Reviewers Spend About 10 min 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 min reviewing each draft (Fig. 9). The median reviewer in the graduate Technology for Learners class spent 22 min per draft. Because all students in that class started reviewing in-class but finished later, its variance in reviewing times is also much larger.

Fig. 8 More students in
large classes are likely to be
online at the same time, so
fewer reviewers were
recruited by email
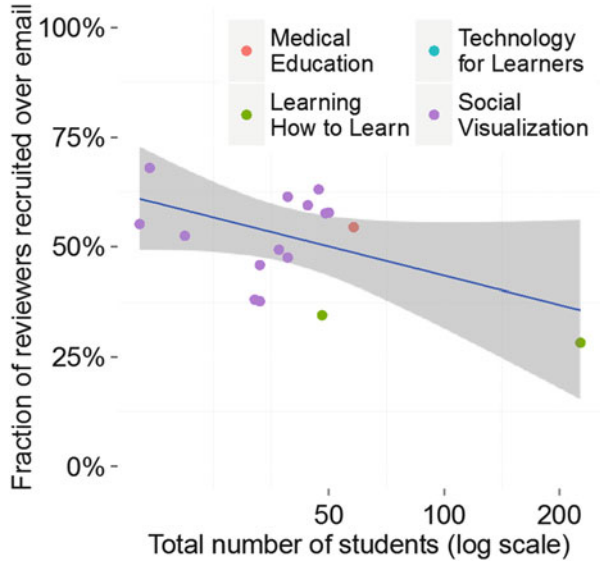
Fig. 9 Reviewers spend
roughly 10 min reviewing
each draft. The graduate-
level Technology for
Learners class spends
longer. (The larger variation
is because students start
reviewing in class, and
finish later.)
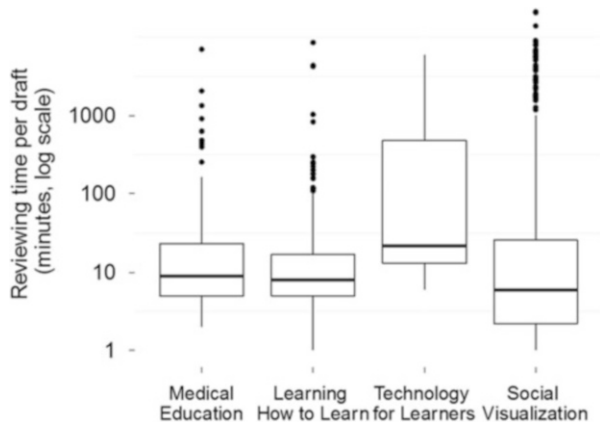
5.3 Are Reviewers Accurate?

There is very strong agreement between individual raters while using the rubric. In
online classes, the median pair-wise 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' k is low. The median $k = 0.19$ 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.

Fig. 10 Students write substantive comments, both in-person and online. The graduate level Technology for Learners has longer comments, possibly because reviews were signed
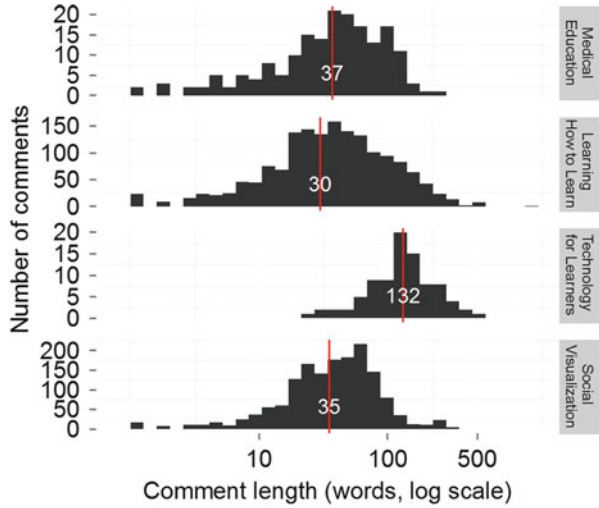
5.4   Staff and Peers Write Comments of Similar Length

Both in-person and online, the median comment was 30 words long (Fig. 10). 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.

5.5   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."

5.6   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 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.

6 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.

6.1 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 h 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.

6.2   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.

6.3   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 ($R^2 = 0.02$).

6.4   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 min (9.8 h) after submission; the median for the Slow Feedback condition was 1528 min (26.6 h). This suggests that the manipulation effectively delayed feedback, but the difference between conditions was more modest than planned.

6.5   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−.

6.5.1 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.

6.5.2 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.

7 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?

7.1 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 (Sommers 1982). In domains like design, it might be useful to get feedback on multiple alternative designs (Dow et al. 2011). 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 (Buxton 2007; Dow et al. 2009), classes may reward students for sustained improvement.

7.2 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 (Kittur et al. 2008).

7.3 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).

7.4 Future Work for PeerStudio

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 min reviewing work as in our deployment, peer assessment arguably incurs a 20-min 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 (Kulkarni et al. 2014).

7.4.1 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.

8 Sustainable Peer Interactions: Three Adoption Challenges and Solutions

With an in depth look into PeerStudio's motives and design, we can better understand the implementation and adoption challenges surrounding peer learning platforms. In this second portion of the chapter, we discuss three such challenges, that have consistently recurred as we have introduced peer learning (PeerStudio and Talkabout) into massive online classes.

First, many courses falsely assume that students will naturally populate the peer learning systems in their classes: "build it and they will come". This assumption often seems natural; after all, students naturally engage with social networks such as Facebook and Twitter. However, students don't yet know why or how they should take advantage of peer learning opportunities. Peer learning platforms sit not in a social setting, but in an educational setting, which has its own logic of incentives: both carrots and sticks are required to keep the commons vibrant. Participation in educational settings has a different incentive structure than a socialization setting. In particular, the benefits of participation are not immediately apparent. For instance, many American college graduates retrospectively credit their dorms as having played a key role in their social development (Dourish and Bell 2007). Yet, universities often have to require that freshmen live in the dorms to ensure the joint experience. We encourage instructors to take a similar reinforcing approach online: integrating peer-learning systems into the core curriculum and making them a required or extra-credit granting part of the course, rather than optional "hang-out" rooms.

The second challenge is that students in online classes lack the ambient social encouragement that brick and mortar settings provide (Erickson and Kellogg 2000). The physical and social configurations of in-person schools (especially residential ones) offer many opportunities for social encouragement (Crouch and Mazur 2001; Dourish and Bell 2007). For example, during finals week, everyone else is studying too. However, other students' activity is typically invisible online, so students do not receive the tacit encouragement of seeing others attend classes and study (Greenberg 2009; Dourish and Bellotti 1992). We hypothesize that in the minimal social context online, software and courses must work especially hard to keep students engaged through highlighting codependence and strengthening positive norms.

The third challenge we have encountered is that instructors can, at best, observe peer interactions through a telescope clouded by big data exhaust: there are few visible signals beyond engagement (e.g. course forum posts and dashboards) and demographics. Student information is limited online (Stephens-Martinez et al. 2014), and knowing how to leverage what demographics instructors do know is non-obvious. In-person, instructors use a lot of information about people to structure interactions (Rosenberg et al. 2007). For example, instructors can observe and adapt to student reactions while facilitating peer interactions. The lack of information in online classes creates both pedagogical and design challenges (Kraut et al. 2012). For instance, in an online discussion, do students completely ignore the course-related discussion prompts and, instead, talk about current events or pop culture? To address such questions, teachers must have the tools to enable them to learn how to scaffold peer interactions from behind their computers.

This half of the chapter addresses these three logistical and pedagogical challenges to global-scale peer learning (Fig. 11). We suggest socio-technical remedies that draw on our experience with two social learning platforms—Talkabout and PeerStudio—and with our experience using peer learning in the classroom. We report on these challenges with both quantitative and qualitative data. Quantitative
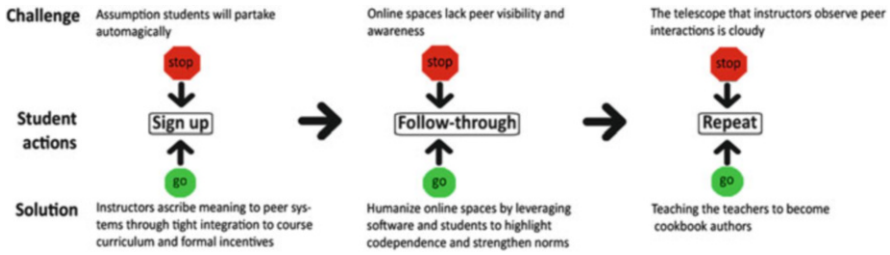
Fig. 11 The challenges and remedies of adoption of peer learning systems presented in this chapter

measures of efficacy include sign-up and follow-through rates, course participation and activity, and participation structure and duration. Qualitative data includes students' and instructors' comments in surveys and interviews. We describe how peer learning behavior varies with changing student practices, teacher practices, and course materials.

9   Social Capabilities Do Not Guarantee Social Use

Peer learning systems share many attributes with collaborative software more generally (Grudin 1994). However, the additional features of the educational setting change users' calculus. Throughout the deployments of our platforms, we've observed different approaches that instructors take when using our peer systems with their material.

Often, instructors dropped a platform into their class, then left it alone and assumed that students would populate it. For example, one course using Talkabout only mentioned it once in course announcements. Across four weeks, the sign-up rate was just 0.4 %, compared to a more successful sign-up rate of 6.6 % in another course; sign-up rate being the number of students who signed up to participate in the peer system out of the number of active students (students who watched a lecture video) in the course. Low percentages represent conservative estimates as the denominator represents students with minimal activity. When this theme recurred in other Talkabout courses, it was accompanied with the same outcome: social interactions languished. Why would instructors who put in significant effort developing discussion prompts introduce a peer learning system, but immediately abandon it?

Through discussions, we noticed that instructors assumed that a peer system would behave like an already-popular social networking service like Facebook where people come en masse at their own will. This point of view resonates with a common assumption that MOOC students are extremely self-motivated, and that such motivation shapes their behavior (Breslow et al. 2013; Kizilcec and Schneider 2015). In particular, instructors were *not* treating the systems like novel learning

technology, but rather as bolted-on social technology. The assumption seemed to be that building a social space will cause students to just populate it and learn from each other.

However, peer learning systems may need more active integration. The value of educational experiences is not immediately apparent to students, and those that are worthwhile need to be signaled as important in order to achieve adoption.

Chat rooms underscored a similar point of the importance of pedagogical integration. Early chat room implementations were easily accessible (embedded in-page near video lectures) but had little pedagogical scaffolding (Coetzee et al. 2014). Later, more successful variants that strongly enforced a pedagogical structure were better received (Coetzee et al. 2015).

9.1   Peer Software as Learning Spaces

Even the best-designed peer learning activities have little value unless students overcome initial reluctance to use them. Course credit helps even students to commit, and those who have committed, to participate. Consider follow-though rates: the fraction of students who attend the discussion out of the students signed up for it. In an international women's rights course, before extra credit was offered, Talkabout follow-through rate was 31 %. After offering extra credit, follow-through rate increased to 52 %. In other classes, we've seen formal incentives raise follow-through rates up to 64 %.

Faculty can signal to students what matters by using scarce resources like grade composition and announcements. We hypothesize that these signals of academic importance and meaning increase student usage. For example, in a course where the instructors just repeatedly announced Talkabout in the beginning, 6.6 % of active students signed up, a large increase from the 0.4 % sign-up rate when there was only one mention of Talkabout.

We saw similar effects with PeerStudio. When participation comprises even a small fraction of a student's grade, usage increases substantially. In one class where PeerStudio was optional, the sign-up rate was 0.8 %. The fraction of users was six times higher in another class where use of PeerStudio contributed to their grade: the sign-up rate was 4.9 %. To maintain consistency with insights from Talkabout, sign-up rates for PeerStudio also represents the number of students who signed up to participate out of the number of active students (students in the course who watched a lecture video).

Students look up to their instructors, creating a unique opportunity to get and keep students involved. One indicator of student interest is if they visited the Talkabout website. Figure 12 shows Talkabout page views after instructors posted on the course site discussing Talkabout, and a decrease in page views when no announcement is made. Talkabout traffic was dwindling towards the end of the course, so the instructor decided to offer extra credit for the last round of Talkabout

Fig. 12 When instructors highlight peer learning software, students use it. Talkabout pageviews of a women's rights course. Instructor announcements are followed by the largest amount of Talkabout pageviews throughout the course. R1 represents Round 1 of Talkabout discussions, and so on, with orange rectangles framing the duration of each round. When instructor does not mention Round 4 and 6, pageviews are at their lowest
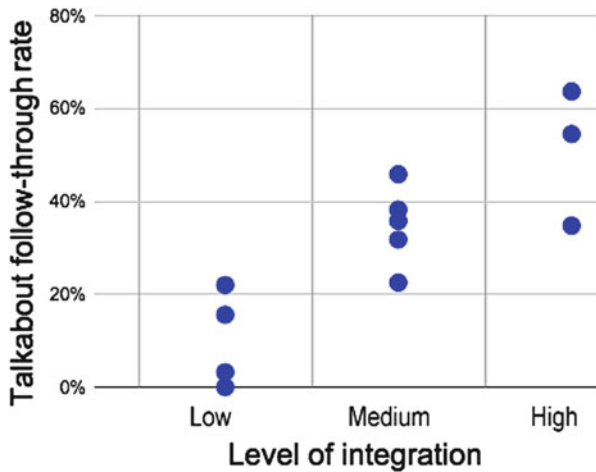

Fig. 13 Follow-through rate from 12 Talkabout courses increases as integration increases

discussions. During the extra-credit granting Talkabout discussions, page views increase around twofold the previous four rounds.

To understand how pedagogical integration and incentives, and follow-through rate interact, we divided 12 Talkabout courses into three categories, based on how well Talkabout was incentivized and integrated pedagogically (see Fig. 13). Courses that never mentioned Talkabout or mentioned it only at the start of the course are labeled "Low integration". Such courses considered Talkabout a primarily social opportunity, similar to a Facebook group. Few students signed up, and even fewer actually participated: the average follow-through rate was 10 %. The next category, "Medium integration," was well integrated but poorly incentivized, classes. These classes referred to Talkabout frequently in announcements, encouraged students to participate, and had well-structured discussion prompts, but they had no formal incentive. Such classes had an average follow-through rate of 35 %.

Well-incentivized and integrated classes, "High integration," offered course extra credit for participation and continuously discussed Talkabout in course announcements, and averaged 50 % follow-through rate. This visualization highlights the pattern that the more integrated the peer learning platform is, the higher the follow-through rate is. We have found that offering even minimal course credit powerfully spurs initial participation, and that many interventions neglect to do this. As one student noted in a post-discussion survey, "I probably wouldn't have done it [a Talkabout session] were it not for the five extra credit points but I found it very interesting and glad I did do it!"

The Talkabout course with the highest follow-through rate not only offered Talkabout for extra credit, but also offered technical support, including a course-specific Talkabout FAQ (Talkabout has an FAQ but it is not course specific). Looking at the forums, the role of the FAQ became apparent: many students posted questions about their technological difficulties and the community TAs and even other students would direct students to this FAQ—loaded with pictures and step by step instructions to help these students understand what Talkabout is and how it's related to them. Moreover, the course support team answered any questions that could not be answered by the FAQ, ensuring that anyone who was interested in using the peer learning platform got the chance to do so.

However, online classes must also accommodate students with differing constraints from around the world. For instance, Talkabout is not available to some students whose country (like Iran) blocks access to Google Hangouts. Other students may simply lack sufficient reliable Internet bandwidth. One course offered small-group discussions for credit that were held either online (with Talkabout) or in-person in order to combat this challenge. When the strongest incentives are impractical, courses can still improve social visibility to encourage participation.

10 Social Transluscence Is Limited Online

Online students are "hungry for social interaction" (Kizilcec and Schneider 2015). Especially in early MOOCs, discussion forums featured self-introductions from around the world, and students banded together for in-person meet-ups. Yet, when peer-learning opportunities are provided, students don't always participate in pro-social ways; they may neglect to review their peers' work, or fail to attend a discussion session that they signed up for.

We asked 100 students who missed a Talkabout why they did so. 18 out of 31 responses said something else came up or they forgot. While many respondents apologized to us as the system designers, none mentioned how they may have let down their classmates who were counting on their participation. This observation suggests that social loafing may be endemic to large-scale social learning systems. If a student doesn't feel responsible to a small set of colleagues and the instructor instead diffuses that responsibility across a massive set of peers, individuals will feel less compunction to follow through on social commitments.

To combat social loafing, we must reverse the diffusion of responsibility by transforming it onto a smaller human scale. Systems that highlight co-dependence may be more successful at encouraging pro-social behavior (Cialdini and Goldstein 2004). In a peer environment, students are dependent on each other to do their part for the system to work. Encouraging commitment and contribution can help students understand the importance of their participation, and create successful peer learning environments (Kraut et al. 2012).

10.1 Norm-Setting in Online Social Interaction

Norms have an enormous impact on people's behavior. In-person, teachers can act as strong role models and have institutional authority, leading to many opportunities to shape behavior and strengthen and set norms. Online, while these opportunities diminish with limited social visibility, other opportunities appear, such as shaping norms through system design. Platform designers, software and teachers can encourage peer empathy and mutually beneficial behavior by fostering pro-social norms.

Software can illuminate social norms online. For instance, when PeerStudio notices that a student has provided scores without written feedback, it reminds them of the reciprocal nature of the peer assessment process (see Fig. 14). As a different example, students that are late to a Talkabout discussion are told they won't be allowed to join the discussion, just as they'd not like to have a discussion interrupted by a late classmate. Instead, the system provides them an option to reschedule. Systems need not wait until things go wrong to set norms. From prior work, we know students are highly motivated when they feel that their contribution matters (Bransford et al. 2000; Ling and Beenen 2005). As an experiment, we emailed students in two separate Talkabout courses before their discussion saying that their peers were counting on them to show up to the discussion (see Fig. 15). Without a reminder email, only 21 % of students who signed up for a discussion slot actually showed up. With a reminder email, this follow-through rate increased to 62 %.

10.2 How Can We Leverage Software and Students to Highlight Codependence and Ascribe Meaning?

PeerStudio recruits reviewers by sending out emails to students. Initially, this email featured a generic request to review. As an experiment, we *humanized* the request by featuring the custom request a student had made. For example, the generic boilerplate request became the personalized request that the student had written
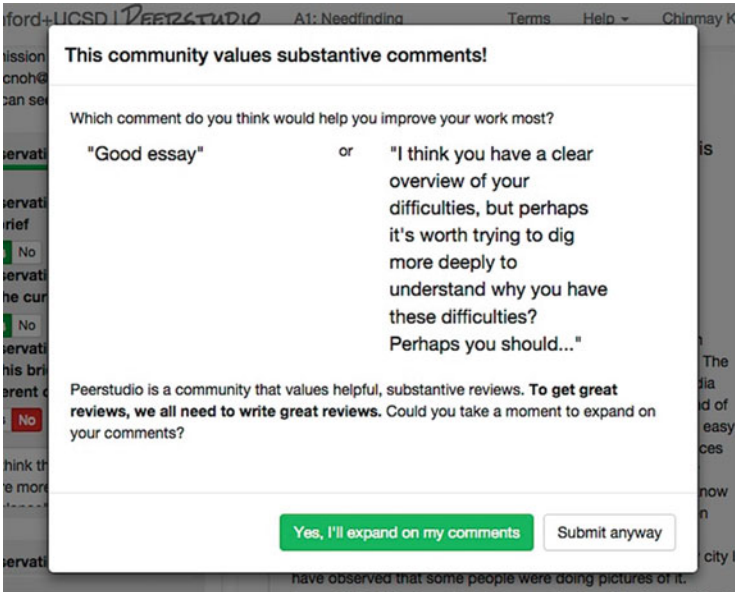
Fig. 14 When PeerStudio detects a review without comments, it asks the reviewer if they would like to go back and improve their review by adding comments

Fig. 15 An email sent to students prior to their discussion, reminding them of the importance of their attendance, increases follow-through rate 41 %

before submitting his draft. Immediately after making this change, review length increased from an average of 17 words to 24 words.

Humanized software is not the only influencer: forum posts from students sharing their peer learning experiences can help validate the system and encourage others to give it a try. For example, one student posted: "I can't say how much I love discussions. . .and that's why I have gone through 11-12 Talkabout sessions just to know, discuss and interact with people from all over the world." Although

unpredictable (Cheng et al. 2014), this word-of-mouth technique can be highly effective for increasing stickiness (Bakshy et al. 2009). When students shared Talkabout experiences in the course discussion forums (2000 posts out of 64,000 mentioned Talkabout, 3 %), the sign-up rate was 6 % (2037 students), and the follow-through rate was 63 %. However, the same course offered a year later, did not see similar student behavior (260 posts out of 80,000 mentioned Talkabout, 0.3 %). The sign-up rate was 5 % (930 students) and follow-through rate was 55 %. Although influenced by external factors, this suggests that social validation of the systems is important.

10.3   Leveraging Students' Desire to Connect Globally

Increasing social translucence has one final benefit: it allows students to act on their desire for persistent connections with their global classmates. For example, incorporating networking opportunities in the discussion agenda allocates times for students to mingle: "Spend 5 min taking turns introducing yourselves and discussing your background." However, we note that this is not a "one-size-fits-all" solution: certain course topics might inspire more socializing than others. For instance, in an international women's rights course, 93 % of students using Talkabout shared their contact information with each other (e.g. LinkedIn profiles, email addresses), but in a course on effective learning, only 18 % did.

11   Designing and Hosting Interaction from Afar

Like a cook watching a stew come to a boil and adjusting the temperature as needed, an instructor guiding peer interactions in-person can modulate her behavior in response to student reactions. Observing how students do in-class exercises and assimilating non-verbal cues (e.g., enthusiasm, boredom, confusion) helps teachers tailor their instruction, often even subconsciously (Klemmer et al. 2006).

By contrast, the indirection of teaching online causes multiple challenges for instructors. First, with rare exceptions (Chen 2001), online teachers can't see much about student behavior interactively. Second, because of the large-scale and asynchronous nature of most online classes, teachers can't directly coach peer interactions. To extend—and possibly butcher—the cooking metaphor, teaching online shifts the instructor from the in-the-kitchen chef to the cookbook author. Their recipes need to be sufficiently stand-alone and clear that students around the globe can cook up a delicious peer interaction themselves. However, most instructors lack the tools to write recipes that can be handed off and reused without any interactive guidance on the instructor's part.

11.1 Guidelines for Writing Recipes: Scaffolding Peer Interactions from Behind a Computer

Most early users of Talkabout provided both too little student motivation and discussion scaffolding. Consequently, usage was minimal (Kulkarni et al. 2015). Unstructured discussion did not increase students' academic achievement or sense of community (Coetzee et al. 2014). To succeed, we needed to specifically target opportunities for self-referencing, highlight viewpoint differences using boundary objects, and leverage students as mediators (Kulkarni et al. 2015). To understand this range of structure, we looked the discussions from 12 different courses and compared agenda word length and discussion duration. We split discussions into two categories: long and short discussion agendas, with 250 words as the threshold, and compared credit-granting and no credit discussions (see Fig. 16). Average discussion duration was 31 min for short agendas. However, only those long agendas that awarded credit successfully incentivized students to discuss longer: the average discussion with credit was 49 min, and without was 30 min. All agendas asked students to discuss for 30 min; students were staying the extra time voluntarily.

We worried that over-structuring an interaction with lengthy and tiresome agendas would leave no space for informal bond-building. However, even with sufficient structure, students can easily veer from the schedule and socialize, exchange social networking information, and offer career advice.

Software systems, platforms, and data-driven suggestions each play a more active role in helping teachers create effective recipes. While most early Talkabout instructors provided too little discussion scaffolding, our data showed instincts led early Talkabout instructors to worry too much about scheduling. For example, time
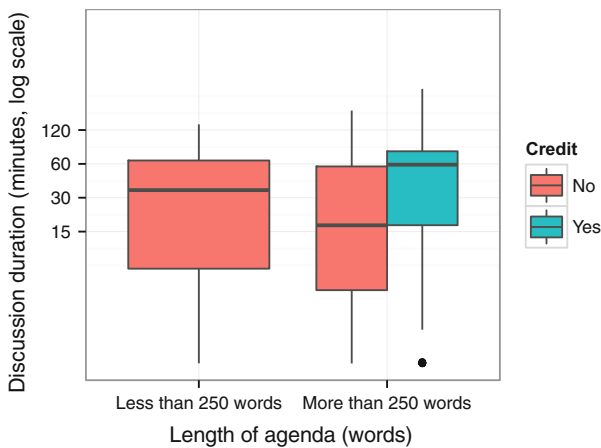
Fig. 16 Longer discussion agendas incentivize students to discuss longer, but only when they are accompanied by course credit for participation

Fig. 17 Data from nine classes and 3400 students shows that most students discuss in the evening, but there are students that will discuss at all 24 h
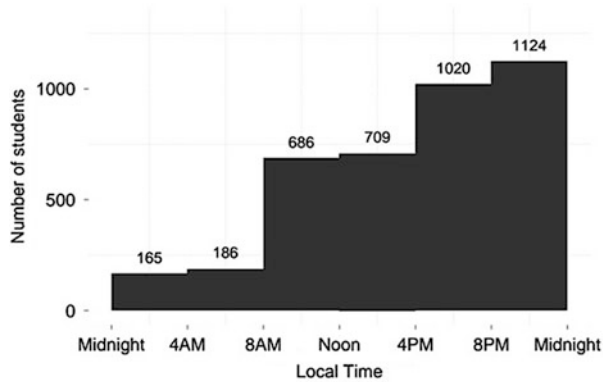
zones are a recurring thorn in the side of many types of global collaboration, and peer learning is no exception. Every Talkabout instructor was concerned about discussion session times and frequency, as this a major issue with in-person sections. Instructors often asked if particular times were good for students around the world. Some debated: would 9 pm Eastern Time be better than 8 pm Eastern Time, as more students would have finished dinner? Or would it be worse for students elsewhere? Other instructors were unsure of how many discussions timeslots to offer. One instructor offered a timeslot every hour for 24 h because she wanted to ensure that there were enough scheduling options. However, an unforeseen consequence of this was that the participants were too spread out over the 24 discussions, and thus some students were left alone.

Analyzing when students participate in discussions taught us that most students prefer evenings for discussions. Yet, different students prefer different times, with every time of day being preferred by someone (Fig. 17). This data suggests that it is unimportant for instructors to find a particular scheduling "sweet spot," and instead their time is better utilized elsewhere: creating the discussion agendas, for example. In summary, these examples illustrated where intuitions can lead teachers and system designers astray. Data-driven suggestions are important to transform expert cooks into cookbook authors.

12   Teaching Teachers by Example

Even fantastic pedagogical innovation can be hamstrung when there is a mismatch between curricular materials and platform functionality. When curricula did not match to the needs of the setting, the learning platforms languished. We emphasize the importance of *teaching by example*: creating designs and introductory experiences that nudge instructors toward the right intuitions. While always true with educational innovation, the online education revolution is a particularly dramatic change of setting, and instructor scaffolding is particularly important.

One of the most robust techniques we have found for guiding instructors is to provide successful examples of how other teachers have used the learning platform. In many domains, from design to writing research papers, a common and effective strategy for creating new work is to template off similar work that has a related goal (Klemmer 2015). During interviews with Talkabout instructors, a common situation recurred: the instructor was having a hard time conceptualizing the student experience. Therefore, to help instructors navigate the interface and create effective discussion prompts, we added an annotated example of a Talkabout discussion (see Fig. 1). Still, we observed that many instructors had difficulty creating effective discussion agendas, e.g. they were very short and did not leverage the geographic diversity Talkabout discussions offer. As an experiment, we walked an instructor through Talkabout—in a Talkabout—and showed an excellent example agenda from another class. This helped onboard the new instructor to working with Talkabout: she was able to use the example as a framework that she could fill in with her own content (see Fig. 18). Next, we showed example course

| Prejudice | Racism and Racial Equality | Sexism and Gender Equality | Gender Identity and Sexual Orientation |

Cross-Cultural Dialogue

1. How much prejudice is there in your country compared with other countries?

2. Which forms of prejudice are declining in your country, and which ones aren't?

3. In your country, which forms of prejudice are the most socially acceptable, and which ones are the least acceptable? Why are some forms more acceptable than others?

4. What are the biggest barriers to social justice in your country, and how can psychology lower them?

5. Are there specific practices that have increased social justice in your country and might work well if adopted in other countries? (Note: Such practices could be at the individual, group, community, or societal level, and could be psychological, educational, legal, political, or economic in nature.)

6. In your opinion, what would the ideal degree of racial and ethnic integration be in your country? Is the same ideal shared by people of all races in your community?

| Learning Theories | Learning Across Cultures | Educational Technologies | Open Conversation |

Constructionism

1. What examples of constructionism, either from your own experience or things you have heard about, strike you as effective ways to teach and learn?

2. What do you see as some of the challenges in using constructionist approaches and tools?

Active Learning

3. What examples of active learning, either from your own experience or things you have heard about, strike you as effective ways to teach and learn?

4. Do you think active learning techniques are commonly used? Would you hope for more educators to adopt these strategies? What might help them do that?

Fig. 18 Two discussions prompts: top used as a template to show a new Talkabout instructor an excellent example. Bottom prompt generated based off the example

announcements that described Talkabout using layman's terms and offering pictures of the Talkabout discussion. Since course announcements are viewed by most online students, it is important to describe peer learning platforms in basic terms to convey a straight forward message.

The next step was to help instructors gain an understanding of what occurs during student discussions. To do this, we showed an instructor a video clip of a Talkabout discussion along with a full discussion summary. In response, the instructor said, "The most interesting point was around the amount of time each student spoke. In this case, one student spoke for more than half of the Talkabout. This informs us to be more explicit with time allocations for questions and that we should emphasize that we want students to more evenly speak." By helping her visualize the interactions, she was able to restructure her discussion prompts in order to achieve her desired discussion goal; in this case, encouraging all students to have equally share their thoughts.

13   Conclusion

This chapter first 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. This chapter then provides evidence for three challenges to global-scale adoption of peer learning, and offered three corresponding socio-technical remedies. We reflect on our experience from developing, designing and deploying our social learning platforms: Talkabout and PeerStudio, as well as our experience as teachers in physical and online classes. We looked at student practices, teacher practices and material design, and assessed the relationship between those and peer learning adoption. When peer systems and curricula are well integrated, the social context is illuminated, and teachers' and system designers' intuitions for scaffolding are guided by software, students do adopt these systems.

References

Anderson S, Rodin J (1989) Is bad news always bad?: cue and feedback effects on intrinsic motivation. J Appl Soc Psychol 19(6):449–467

Andrade HG (2001) The effects of instructional rubrics on learning to write. Current Issues Educ 4:4

Andrade HG (2005) Teaching with rubrics: the good, the bad, and the ugly. Coll Teach 53(1): 27–31

André P, Bernstein M, Luther K (2012) Who gives a tweet? In: Proceedings of the ACM 2012 conference on computer supported cooperative work—CSCW'12, ACM Press, p 471

Bakshy E, Karrer B, Adamic LA (2009) Social influence and the diffusion of user-created content. In: Proceedings of the 10th ACM conference on electronic commerce, pp 325–334

Balcazar FE, Hopkins BL, Suarez Y (1986) A critical, objective review of performance feedback. J Organ Behav Manag 7:2

Bransford J, Brown A, Cocking R (2000) How people learn. The National Academies Press, Washington, DC

Bransford J, Schwartz DL (1999) Rethinking transfer: a simple proposal with multiple implications. Rev Res Educ 24:61–100

Breslow L, Pritchard D, DeBoer J, Stump G, Ho A, Seaton D (2013) Studying learning in the worldwide classroom: Research into edX's first MOOC

Brooks M, Basu S, Jacobs C, Vanderwende L (2014) Divide and correct: using clusters to grade short answers at scale. Learning at Scale

Buxton B (2007) Sketching user experiences: getting the design right and the right design. Morgan Kaufmann, San Francisco

Cambre J, Kulkarni C, Bernstein MS, Klemmer SR (2014) Talkabout: small-group discussions in massive global classes. Learning@Scale

Carlson PA, Berry FC (2003) Calibrated peer review and assessing learning outcomes. In: Frontiers in education conference

Chen M (2001) Design of a virtual auditorium. In: MULTIMEDIA'01 Proceedings of the ninth ACM international conference on multimedia, pp 19–28

Cheng J, Adamic L, Dow P, Kleinberg J, Leskovec J (2014) Can cascades be predicted? In: Proceedings of the 23rd international conference on world wide web, pp 925–936

Cialdini R, Goldstein N (2004) Social influence: compliance and conformity. Annu Rev Psychol 55:591–621

Coetzee D, Fox A, Hearst MA, Hartmann B (2014) Chatrooms in MOOCs: all talk and no action. In: Proceeding of the ACM conference on learning @ scale, ACM Press, pp 127–136

Coetzee D, Lim S, Fox A, Hartmann B, Hearst MA (2015) Structuring interactions for large-scale synchronous peer learning. In: CSCW: ACM Conference on computer supported collaborative work

Crouch CH, Mazur E (2001) Peer instruction: ten years of experience and results. Am J Phys 69(9):970

Dawes RM (1979) The robust beauty of improper linear models in decision making. Am Psychol 34(7):571

Dourish P, Bell G (2007) The infrastructure of experience and the experience of infrastructure: meaning and structure in everyday encounters with space. Environ Plan B Plan Des 34(3):414

Dourish P, Bellotti V (1992) Awareness and coordination in shared workspaces. In: Proceeding CSCW'92 Proceedings of the 1992 ACM conference on computer-supported cooperative work, pp 107–114

Dow SP, Heddleston K, Klemmer SR (2009) The efficacy of prototyping under time constraints. In: Proceeding of the ACM conference on creativity and cognition, ACM Press, p 165

Dow S, Fortuna J, Schwartz D, Altringer B, Klemmer S (2011) Prototyping dynamics: sharing multiple designs improves exploration, group rapport, and results. In: Proceedings of the 2011 annual conference on human factors in computing systems, pp 2807–2816

Erickson T, Kellogg W (2000) Social translucence: an approach to designing systems that support social processes. In: ACM Transactions on Computer-Human Interaction (TOCHI)—Special issue on human-computer interaction in the new millennium, Part 1 7(1), pp 59–83

Ericsson KA, Krampe RT, Tesch-Römer C (1993) The role of deliberate practice in the acquisition of expert performance. Psychol Rev 100:3

Falchikov N, Goldfinch J (2000) Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks. Rev Educ Res 70(3):287–322

Fast E, Lee C, Aiken A, Bernstein MS, Koller D, Smith E (2013) Crowd-scale interactive formal reasoning and analytics. In: Proceedings of the 26th annual ACM symposium on user interface software and technology

Gray WD, Boehm-Davis DA (2000) Milliseconds matter: an introduction to microstrategies and to their use in describing and predicting interactive behavior. J Exp Psychol 6(4):322–335

Greenberg S (2009) Embedding a design studio course in a conventional computer science program. In: Creativity and HCI: From Experience to Design in Education. Springer, pp 23–41

Grudin J (1994) Groupware and social dynamics: eight challenges for developers. Commun ACM 37(1):92–105

Guskey TR (2007) Closing achievement gaps: revisiting Benjamin S. Bloom's "Learning for Mastery". J Adv Acad 19(1):8–31

Hattie J, Timperley H (2007) The power of feedback. Rev Educ Res 77(1):81–112

Heffernan N, Heffernan C, Dietz K, Soffer D, Pellegrino JW, Goldman SR, Dailey M (2012) Improving mathematical learning outcomes through automatic reassessment and relearning. AERA

Kim S-M, Pantel P, Chklovski T, Pennacchiotti M (2006) Automatically assessing review helpfulness. In: Proceedings of the 2006 conference on empirical methods in natural language processing, association for computational linguistics, pp 423–430

Kittur A, Chi EH, Suh B (2008) Crowdsourcing user studies with mechanical turk. In: Proceeding of CHI, ACM Press, p 453

Kizilcec RF, Schneider E (2015) Motivation as a lens to understand online learners: towards data-driven design with the OLEI scale

Klemmer SR (2015) Katayanagi Lecture at CMU: the power of examples

Klemmer SR, Hartmann B, Takayama L (2006) How bodies matter: five themes for interaction design. In: Proceedings of the 6th conference on designing interactive systems, pp 140–149

Kluger AN, DeNisi A (1996) The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychol Bull 119(2): 254–284

Konstan JA, Walker JD, Brooks DC, Brown K, Ekstrand MD (2014) Teaching recommender systems at large scale. In: Proceeding of the ACM conference on learning @ scale conference, ACM Press

Kraut RE, Resnick P, Kiesler S, Burke M, Chen Y, Kittur N, Konstan J, Ren Y, Riedl J (2012) Building successful online communities: evidence-based social design (Google eBook), MIT Press

Krosnick JA (1999) Survey research. Annu Rev Psychol 50(1):537–567

Kulik JA, Kulik C-LC (1987) Timing of feedback and verbal learning. Rev Educ Res 58(1):79–97

Kulkarni C, Wei KP, Le H, Chia D, Papadopoulos K, Cheng J, Koller D, Klemmer SR (2013) Peer and self assessment in massive online classes. ACM Trans Comput-Hum Interact (TOCHI) 20(6):33

Kulkarni C, Socher R, Bernstein MS, Klemmer SR (2014) Scaling short-answer grading by combining peer assessment with algorithmic scoring. In: ACM conference on learning @ scale, ACM

Kulkarni C, Cambre J, Kotturi Y, Bernstein M, Klemmer S (2015) Talkabout: making distance matter with small groups in massive classes. In: CSCW: ACM conference on computer supported collaborative work

Latham GP, Locke EA (1991) Self-regulation through goal setting. Organ Behav Hum Decis Process 50(2):212–247

Ling K, Beenen G (2005) Using social psychology to motivate contributions to online communities. J Comput-Mediat Commun 10(4):00

Marsh RL, Landau JD, Hicks JL (1996) How examples may (and may not) constrain creativity. Mem Cogn 24(5):669–680

O'Donnell AM, Dansereau DF (1995) Scripted cooperation in student dyads: a method for analyzing and enhancing academic learning and performance. In: Interaction in cooperative groups: the theoretical anatomy of group learning. Cambridge University Press, Cambridge, UK, pp 120–141

Plattner H, Meinel C, Leifer L (eds) (2015) Design thinking research: making design thinking foundational. Springer, Heidelberg

Porter L, Lee CB, Simon B (2013) Halving fail rates using peer instruction: a study of four computer science courses. In: Proceeding SIGCSE'13 Proceeding of the 44th ACM technical symposium on computer science education, pp 177–182

Rosenberg J, Lorenzo M, Mazur E (2007) Peer instruction: making science engaging. In: Handbook of college science teaching. NSTA Press, Arlington, VA, pp 77–85

Smith M, Wood W, Adams W, Wieman C, Knight J, Guild N, Su T (2009) Why peer discussion improves student performance on in-class concept questions. Science 323(5910):122–124

Sommers N (1982) Responding to student writing. Coll Compos Commun 33(2):148–156

Stephens-Martinez K, Hearst MA, Fox A (2014) Monitoring MOOCs: which information sources do instructors value? In: Proceedings of the first ACM conference on learning @ scale conference, pp 79–88

Tinapple D, Olson L, Sadauskas J (2013) CritViz: web-based software supporting peer critique in large creative classrooms. Bull IEEE Techn Comm Learn Technol 15(1):29

Turkle S (2011) Alone together: why we expect more from technology and less from each other. Basic Books, New York