CHAPTER 11.
THE POTENTIAL OF PEER REVIEW SOFTWARE THAT FOCUSES ON THE REVIEW, NOT THE DRAFT

Nick Carbone
Aptara

New software designed specifically for teaching students to give, to receive, to use, and to reflect on peer feedback offers writing instructors powerful new ways for making workshop pedagogies central in any course across the curriculum where writing is taught. The software can invigorate student-to-student workshops and help move peer review from an infrequently used pedagogy on the margins of a course to a regularly used pedagogy closer to the center of the course. When I say “margins of a course,” I have in mind those courses that might do one peer review assignment per essay assigned. I consider that relatively infrequent. By “regularly used pedagogy,” I mean peer review happens once or twice a week, at a minimum, becoming a central and constant course activity. Or put another way, writing workshops happen once or twice a week.

New peer review software makes it possible to shift peer review to a more central role in a course by making the work of peer review more visible—and thus more teachable—than prior technologies. The software treats the work of peer review—the writing of comments, the reading of comments, the decisions about which comments to apply during revision—as essential, even more essential to learning to write than the final draft of the paper under review.

The software, by making the work of review more visible, helps both teachers and students. For teachers, it makes it easier to see peer review as it is happening, thus making it easier to coach reviewers and writers. For students, as writers, they have tools for choosing comments to use, ranking their usefulness, and making choices about applying the feedback. As reviewers, students can see how their feedback is used and how it compares to feedback given by other reviewers. By being designed for teaching peer review first and foremost, the new software instantiates a belief in writing as process, in writing classrooms as sites of writers and reviewers workshopping their writing.

This essay will touch on four examples of this software. These examples were chosen for two reasons. First, they are currently available as I draft. And second,
I’ve had experience either working in them or reviewing their features in my work developing educational technology and as writing across the curriculum consultant. The four products used in this essay to illustrate the benefits of using peer review software are Eli Review, Peerceptiv, My Reviewer, and Calibrated Peer Review.

Before looking at this software, however, I want to take a brief look at past peer review technology. The past will help show the promise of the present. Early electronic tools allowed students and teachers to more easily access documents and freed them from the space and time restraints of the classroom. More importantly, though, they sowed the seeds that have made more recent technologies so effective—they began the process of aggregating peer reviews apart from the document under review and thus uncovering insights into what kind of comments are most likely to lead to revision.

THE TECHNOLOGY OF HARDCOPY AND WRITING WORKSHOPS

Student to student feedback on writing has always required technology. In the work of Murray, Elbow and Belanoff, Bruffee, Gere and even much of Ede and Lunsford, an assumed and underlying technology was simple proximity—writers in the same room, working in pairs or small groups.

In its classic form, after proximity, the second central peer review technology was, and may still be in many classrooms, paper and pen: writers come to class with one or more hard copies of their drafts. They sit in groups and read (or hear being read by the author) each other’s drafts. If reviewers are not writing on their peers’ drafts, then often the writer makes notes on feedback received. This is a workshop model.

As writers and their readers work through the peer review assignment, the instructor moves around the room, checking in with each group one a time. This helps keep writers on task as the instructor listens in, advises, and helps reviewers give better comments, and helps writers learn how to weigh the feedback.

With the advent of photocopiers and then printers, peer review was better able to become homework. Instead of discussing drafts during a class workshop, writers would come to imitate more the kind of solo review scholars do when they peer review an article for an academic journal. At the next class meeting, writers would receive their reviews, written by reviewers who worked in isolation. Though the review isn’t blind, let’s call this the academic journal model.

First uses of electronic technology did not change the workshop nor academic journal models. In workshops, students might work from laptops or shared files instead of hard copy. In the academic journal model, they might get the file via email to work on at home. Still, electronic tools were boons. In hard copy and oral-driven
review technology, the work of peer review was no longer available to the reviewer once the review was given. With electronic review, both the writer and the reviewer would have a copy of the review work. That simple change allowed students, should instructors call for it, to reflect on their work as reviewers without relying on memory alone. They could call up the files and read through their comments.

I don’t know of research that shows the prevalence of workshop peer review compared to academic journal peer review. The point here is that prior to the new technologies explored below, in both the workshop and academic journal models, the draft remains at the center. Comments reside in the draft margins.

In these kinds of technologies, where an instructor wants to coach students on giving and using feedback, the work becomes prodigious. A single writer may have to work with two copies of his or her draft, but the instructor with a class of 24 students, would need to look through 48 copies of drafts.

An instructor can choose to try to gather all those drafts and all the comments and then to coach the reviewers and writers on how to give better comments and make better use of the feedback. But that takes time most instructors do not have. And very often because it is hard to teach these two aspects of peer review, both students and instructors become disenchanted with the value of peer review. Writers don’t find the feedback useful. Reviewers don’t believe they can give feedback writers will find useful.

The criticism of peer review is well known. Before offering strategies for addressing them, Linda Nilson (2009) summarizes the complaints her research uncovered about the quality of student peer review work:

- too lenient or uncritical; focused on whether the evaluator likes or agrees with a work rather than its quality; overly critical and harsh; inaccurate; superficial; focused on trivial problems and mechanical errors; focused too much on content alone; unrelated to the assignment’s requirements; and not referenced to specific instances in the work. (2)

Confirming Nilson’s analysis, other studies show peer review can be effective when well designed and well taught (Cho and MacArthur; Cho, Schunn, and Charney; Min; Zundert, Sluijsmans, and Merriënboer; Strasma; Cahill; Brammer and Rees; and Shih).

Students giving poor feedback is not a technology problem; it is a learning challenge. No technology on its own will make peer review more effective. However, for those committed to making peer review work, new technologies can help in powerful ways that were unavailable before. As I discuss below, new technologies provide methods for making peer review central and teachable, but instructors must be willing and able to reimagine their pedagogy in order to do this.
MULTIPLE REVIEWERS IN ONE DOCUMENT

Some of the work of seeing peer review in context has been simplified. It is possible for a class of 24 students do peer review in groups of three (two reviewers per document) and for instructors to see the comments from reviewers in just 24 documents instead of 48. The most prominent example of this is Google Docs, a platform that allows multiple students to make marginal or in-text comments which are viewable by writers, reviewers, and the instructors.

But even with this, coaching peer review remains a challenge. For faculty to get a sense of what is happening in peer review, they can collect the Google Doc URLs and visit each Doc one at a time and look at the comments. However, that is still 24 URLs to collect and visit.

Discerning trends in the review work, finding and sharing good comments to serve as models, coaching a writer on which advice to follow, can still take a long time. The only step saved has been in reducing the number of documents to open from 48 to 24.

Too often, peer review falters because it is only after the review cycle has been completed, after new drafts have been submitted, that instructors discover peer review comments have not been used by writers. By then, asking reviewers to improve their comments or writers to rethink their choices about using comments is too late.

THE POWER OF AGGREGATION

New peer review software aggregates comments. Instead of an instructor going from draft to draft to read comments, the software collects the comments. The new software also allows for comment types that use Likert Scales, writing criteria or feature identification, and other approaches that create data. The new software also allows writers to rate the usefulness of review comments, creating another form of data.

This aggregation of comments and data provides both qualitative and quantitative insights. Patterns can be revealed more quickly. Writer’s judgments about the useful of reviews can be summarized and probed while the revision cycle in process. Instructors can better intervene in the review process to address the quality of the review work before the review cycle is complete.

Being able to see more fully the peer review dynamic as it occurs opens the door to making peer review more teachable. Aggregating comments from across drafts creates a corpus of writing that can be given the time and attention it needs to be reviewed and revised. Figure 11.1 represents a student document, “Student work under review,” surrounded by feedback. The feedback is in the document.
SOFTWARE THAT AGGREGATES PEER REVIEW COMMENTS

In 2002, when I was working for Bedford/St. Martin’s, a college textbook company that is now an imprint of Macmillan Learning, I worked with a professor and first year writing teacher from the University of Hawai’i named Walter Creed. Creed had written his own code to collect drafts and encourage peer review, and Bedford/St. Martin’s helped bring it to market. We called the product *Comment*. It was the first software I saw that used aggregation (see Figure 11.2) to elevate the peer review comment from the margins of essays under review to a collection of work that could be read on its own.

*Figure 11.1. The essay at the center, review comments on the margin.*

*Figure 11.2. Feedback aggregated.*
With feedback aggregated, writers, reviewers, and the instructor can make feedback their focus. Instead of using the writing under review to find and read the feedback, one can explore and read the feedback directly, referencing back to the work under review as needed. In formal academic peer review, very often an editor writes a cover letter or prepares a review collation as a way to aggregate review feedback for the writer. By doing this kind of aggregation automatically, it is easier to see patterns in feedback, to compare differing advice from among reviewers.

Comment had a feature that collected all the comments a student had written for classmates from all the different drafts upon which those comments were made. By aggregating a student’s comments, it made it easier to make commenting more central to the economy and ecology of the course. Students could easily collect and see all the comments they wrote, either by assignment or across assignments.

This made reflecting on their ability to write comments easier to assign in the course. And this ability to have students reflect from an already assembled collection of their comments, helped me elevate the importance of peer review in the course. As I used the feature more, and increased student focus on the comments they wrote and the skill of writing good comments, I came to see that I was treating the feedback comment as a genre worthy of a writing course.

My evocation of genre is fairly informal. But what clicked for me then and now was a phrase from Carolyn Miller’s “Genre as Social Action” where she writes “that a rhetorically sound definition of genre must be centered not on the substance or the form of discourse but on the action it is used to accomplish” (151). And for me, the path from that quote to seeing peer review writing as a genre was this: the comments have “action[s] . . . to accomplish.”

Peer review comments have an audience, a context, and a purpose. The intellectual work of writing a good comment is as rich as the work of writing a good essay. We collect student essays and drafts for assessment of student growth as writers. Comment made it possible to do the same with student peer review writing: it collected it and made it more visible.

For me as a writing teacher, the ability to see a student’s collected comments was a new window into how that student was evolving as a reader of writing, a thinker of writing, and—because the comments are written—a writer on writing. The same collection became a window for each student into those same processes and abilities. Their comments persisted as a collected body of work they could review and reflect upon.

FROM COMMENT TO TODAY: AGGREGATING VALID DATA

Despite being a wonderful little bit of software, Comment never found a foothold in the college composition classroom, and Bedford/St. Martin’s stopped
offering it about twelve years ago. Since then, a new wave of tools has emerged that foregrounds peer review. These new programs go beyond what Comment attempted. In addition to aggregating review comments, these newest programs have been designed to use quantitative and qualitative data to help instructors and students see patterns in the reading, thinking, and writing of both the work under review and the work of review.

The data comes not just from counting—such as the number of comments received in a document as you might find in Google Docs, or the number of edits made in a document as you might find on a Wiki page—but also from capturing judgments and choices. But for now, the thing to keep in mind is that newer software looks for and reports patterns. And in a writing course, that’s a powerful thing.

The programs mentioned earlier—Eli Review, Peerceptiv, My Reviewer, and Calibrated Peer Review—emphasize student-to-student feedback and offer tools for helping teachers and students see the work of review.

INSIGHTS ON WRITING DERIVED FROM WORK OF REVIEW

Figure 11.2 shows the concept of aggregating review comments. What these four programs do that breaks new ground is not only aggregating comments, but they also aggregate data—qualitative and quantitative data—generated both from peer review comments themselves and, in three of the programs, from writers rating the usefulness of the feedback they received. What I hope to show in the following pages is why that ability to gather data and to express insights—patterns in the data about both the writing under review and the writing of reviews—via analytics offers such potential.

But first a reminder: we should not be put off by the term “data.” Writing teachers have always relied on data. By data I mean simply information and knowledge of student performance collected and stored for the purposes of analysis. The result of the analysis becomes analytics—reports, patterns, insights—that lead to teaching and learning decisions.

Gradebooks, for example, contain data on student performance. Many programs require faculty to calculate mid-semester grades. Faculty must notify the administration about students whose grades are low so that academic advisors can reach out to the students and offer support. Or consider another example. Instructors might require assignment page counts or word counts; they may return work that is too short, keeping track of how many students need to resubmit their work. From an analysis of that data, they may create an analytic in the form of a list of students who consistently struggle to meet assignment requirements. Instructors might then meet with those students to find out why they struggle.
While data and analytics in writing courses are not new, what is new, in the context of this article, is how peer review software uses data and analytics to offer new insights into student drafting and revising processes. The value of data, thus insights from its analytics, will depend on good writing and peer review assignment design.

Let’s look at an example from Eli Review. Eli offers a review question type called “Trait Identification.” That tool simply asks a reviewer to check a radio button if a trait occurs. The instructor using Eli can define what a trait is and can create the directions and selections as required by their course, their assignment, and their teaching.

The traits in Figure 11.3 are tied to a writing assignment where students analyze a data set and make recommendations. To draft the assignment, students have to reference themes found in the data (first trait), explain how the themes are related (second trait), explain the implication of the relationship (third trait), and suggest a response or outcome or action (fourth trait).

For students to be able to draft writing that has these traits, and to review writing for these traits, the instructor will have had to taught students about these traits and their importance. This screenshot also signals that students, then, are learning about these traits in at least three ways: first, from the teaching about these traits prior to writing; second, from drafting writing that seeks to fulfill the traits; and third, from reviewing writing to see if the traits have been accomplished.

With that pedagogy and practice as context, what you see in this example is that the instructor knows where students are struggling. The report in Figure 11.3 shows fewer than half of the students reviewed have “a passage that suggests an appropriate outcome, response, or action to be taken.” That’s a significant insight. Normally to gain such an insight, an instructor would collect the drafts and take several days to read through them.

![Figure 11.3. Trait identification.](image)
But with the tool shown in Figure 11.3, the instructor learns about the struggle of students to meet the final trait in one peer review session. If this is the first question asked and is done during class, the instructor will learn this in 15 minutes. Imagine that. In 15 minutes instead of week an instructor knows that an element of the assignment causes half the class to struggle. So now the instructor can address the struggle and investigate with the class why a key trait is missing in half the drafts under review.

The key to making review data and reports valuable comes in crafting good questions to guide peer reviewers (Liu and Carless). That is, as the example above illustrates, one should only ask students to look for things they've been taught. The software brings the work of good instruction and of student writing and reviewing into view. It makes the learning from peer review visible in ways that were not before possible.

REVIEWING THE REVIEWERS—MAKING PEER REVIEW WORK ASSESSABLE

All the software under discussion also takes to heart, far more than software that came before it, is the value of writers giving feedback (Lundstrum and Baker). Eli, Peerceptiv, and CPR, in their user guides for students, explicitly discuss the benefits of giving reviews has for students who are learning to write.

For example, in advice to reviewers, Peerceptiv reminds students as they review to think of “aspects of your own work you want to improve” by asking “What can you take away from each review [you give] that allows you to become a better writer?”

Calibrated Peer Review points out that “students not only learn their discipline by writing, they also learn and practice critical thinking by evaluating . . . submissions from their peers” (n.p.).

Writing for Eli Review, Melissa Graham Meeks explains:

. . . students’ mastery of giving helpful comments depends on their inclusion of signals related to “describe-evaluate-suggest” pattern. The absence of these three moves results in bad feedback, which writers can’t use to revise. But, bad feedback has a larger consequence: it doesn’t lead to givers’ gain. Givers’ gain is the benefit reviewers get when they apply to their own work what they see other writers doing or not doing. (np).

A key feature of Eli, Peerceptiv and My Reviewers is for writers to rate the usefulness of reviews received. Each program provides each reviewer a helpfulness rating, and in each there is some comparison of a reviewer’s helpfulness—whether
on a particular review or overall—to class helpfulness averages. We’ll use Peerceptiv screen shots to explore the value of this feature.

Peerceptiv’s “Back Evaluate” allows a writer to give feedback on reviews received. Notice from the blue box labeled “Summary,” that in Peerceptiv instructors can create prompts and guidelines for how review comments should be rated. The feedback asks (see notations 2 and 3) students to provide both quantitative and qualitative feedback on reviews.

Figure 11.5 shows the peer reviewers their “Back-Evaluation” helpfulness ratings, the comments the ratings address, and the qualitative feedback. Figures 11.4 and 11.5 are from peerceptiv.zendesk.com.

*Figure 11.4. Peerceptiv’s “back-evaluate”.*

*Figure 11.5. Helpfulness ratings.*
How each program uses helpfulness measures varies, but their insights can be important to learning and learning motivation.

First, learning to write a good review takes practice. It’s a close reading and writing skill students can and should get better at. Writing a good review helps not only the writer but also the reviewer because it gives reviewer more practice thinking carefully about writing.

Second, one of the reasons peer review often fails as a pedagogy is because reviewers don’t know how or whether a writer uses the advice given. When reviewers get feedback from writers on how their reviews are being considered and might be applied, then those reviewers come to see that their review work isn’t just busy work. In other words, it’s hard to care about peer review if you think the writer or instructor aren’t going to care about the quality of the feedback.

Now, one of the concerns I have heard from instructors about having students rate the feedback received from classmates is that they may just give all feedback good ratings. But that concern, really, is no different than that of instructors who avoid doing peer review at all because they don’t believe students can or will give constructive feedback on writing.

However, we know that students can learn to give feedback on writing. And so it follows that students can also learn to rate and comment well on the feedback they received. Teaching students to give good feedback, as we see above in the brief look at the trait identification example from Eli Review, is matter of making good use of the software to create peer review guidelines and assignments that match what students are being taught. Similarly, in our look at Peerceptiv’s “back-evaluation” tool, we see an example of guidelines being used to help students give meaningful feedback on comments received from reviewers.

Both examples show evidence of pedagogical thought and planning. The software builds on an instructor’s pedagogical designs by aggregating feedback and revealing trends or allowing an instructor to focus on a particular student’s performance. Thus the software makes it more possible to help teach students how to be better reviewers. Which helps them also to be become better writers if only because they are getting more practice working with writing through the reviewing writing, considering feedback received, and evaluating that feedback.

Christian D. Schunn, one of the creators of Peerceptiv, has been doing research on peer review for close to 20 years. Many of the tools and features from Peerceptiv grew out of that research. As Peerceptiv has evolved, his research on it and how it shapes peer review in classrooms has continued. A 2017 study co-authored with Melissa M. Patchan and Russell J. Clark found that students who believed their grade for doing peer review would be influenced by “back-evaluation” ratings of their feedback’s helpfulness gave more feedback using comments on the texts and with a greater focus on critiques and solutions for writers to try.
These same students also did a better job at using the back-evaluation tool more consistently and richly.

That is, when students believe the work of review and rating reviews is important enough in the course, they do the work of rating reviews well. One of the benefits of this new crop of tools for teaching peer review as central to drafting and revising is this: their ability to collect and analyze student work, and their ability to build in evaluation by writers on reviews received, makes it aspects of the writing process more visible than ever before. And because this work can be created and seen, it can contribute to a course’s assessment and grading practices.

A side note is in order here, Patchan, Schunn, and Clark’s study focused on student perceptions that their helpfulness scores would be a factor in grading. Eli Review recommends against basing review grades on helpfulness ratings because that could lead students to inflate ratings. The goal is to teach students to give honest and constructive ratings in the same way they are being taught to give honest and constructive feedback. So while helpfulness ratings can have a place in the course grade, that place should be found through practices such as reflecting on what makes an effective comment, or being able to describe how writers used feedback given, and on how to rate feedback in ways that are honest and that will help classmates become better at giving feedback.

To further help assure that helpfulness ratings aren’t inflated, both Eli and My Reviewers tie the ranking of feedback directly to revision planning. For example, at the University of South Florida where My Reviewers was developed, a common revision plan assignment requires students to “Summarize this feedback, analyze which comments you find most helpful, and then determine how you will revise your draft.”

By tying analysis of a comments to how they will be used in revision, My Reviewers sets up a pedagogy that asks students to focus on the revisions they will make and to analyze and assess reviews with those needs in mind. In this way, revision planning influences how writers will rate peer feedback, increasing the likelihood that ratings will reflect analysis.

In Eli, there’s a “Revision Plan” tool. If teachers assign revision plans, students follow three steps after a review.

In step 1, writers see all their feedback. Writers can rate the helpfulness of written comments using a five-star scale, with five being most helpful.

In step 2, writers create revision plans moving comments received into the revision plan. In Eli, the software records which comments go into a revision plan. For a reviewer, having a comment added to a classmate’s revision plan becomes another factor in that reviewer’s helpfulness rating.

In step 3, writers annotate the comments they’ve added to their revision plans, outlining why the comment is included and how the writer will work with it.
These revision plans are then available for the instructor to review. The revision plans show the kinds of decisions writers make about how they will use the feedback received. They are both a metacognitive document and a practical revision tool that gives a writing instructor documented insight into how writers are using feedback.

Figure 11.6 shows an instructor view of a student’s revision plan. In this plan, the writer, Katherine, has chosen a comment from a classmate that will guide her revision. Katherine added that comment to her revision plan. She also made a note about the comment and has indicated by a star ranking how helpful it is to her. Her revision plan concludes, in “Revision Notes,” with her broader thinking about her next draft. The instructor has a box for giving the writer feedback on her plan.

![Figure 11.6. Revision plan. Image from https://elireview.com/support/guides/instructor/tasks/revision/](https://elireview.com/support/guides/instructor/tasks/revision/)
If revision plans are shared with the reviewers whose comments make up the plan, those reviewers can also see the impact their feedback has in helping a writer revise. Reviewers learn not only what a writer feels about a comment based on the number of stars it may earn, but also whether the feedback is actually going to help a writer revise.

**FINAL THOUGHTS**

The more I explore and think about this new peer review software, the more I see it, in so many ways, as an ideal tool for enacting writing and the teaching of writing as a form of cognitive apprenticeship, a term introduced in a 1991 essay by Allan Collins, John Seely Brown, and Ann Holum titled, “Cognitive Apprenticeship: Making Thinking Visible.” I quote their article title in full because so much of what the newest peer review software does is exactly that: it attempts to make thinking visible.

Collins et. al. noted that in a traditional apprenticeship, people learn by observing or being shown how a process or skill works, and then by being given more and more responsibility for the task over time, building up complexity and nuance as the apprenticeship advances.

The authors write, “Apprenticeship involves learning a physical, tangible activity. . . . In schooling, the processes of thinking are often invisible to both the students and the teacher. Cognitive apprenticeship is a model of instruction that works to make thinking visible” (38).

In writing—as Donald Murray most famously noted—writing evolves and writers grow not just from receiving feedback, but also from giving it. It’s why we have workshop pedagogies. In professional contexts—including academic contexts especially—peer review is a professional activity and skill. One grows as an academic thinker and writer by being both able to write and to review. The same holds for most of the careers students will go into—they will be called upon to write and to review.

What I like about these new tools for peer review is that they make very visible the cognitive work students are doing as writers, as reviewers, and as emerging thinkers. Peer comments are visible thinking on writing. Writer ratings are of visible thinking on comments. Revision plans are visible thinking on decisions writers are making.

Most importantly this thinking is visible to students. They can see their review comments as a corpus of writing worthy of reflection. They can see the choices as on which comments to use. They can see as reviewers the kinds of comments writers use to guide revision.
In the end, the more we accurately see, the better we can teach, and the more students can learn. Used well, these new tools for peer review let us see more than we ever could before.

**WORKS CITED**


