International Journal of Teaching and Learning in Higher Education 2015, Volume 27, Number 1, 56-68
http://www.isetl.org/ijtlhe/ ISSN 1812-9129

Effectiveness of Guided Peer Review of Student Essays in a Large Undergraduate Biology Course

Lauren Kelly
Washington State University

Instructors and researchers often consider peer review an integral part of the writing process, providing myriad benefits for both writers and reviewers. Few empirical studies, however, directly address the relationship between specific methodological changes and peer review effectiveness, especially outside the composition classroom. To supplement these studies, this paper compares types of student commentary received between a control and guided rubric in an introductory biology course in order to determine if guided questions augment the amount of "feedforward" responses, questions and suggestions that consider the next draft and are reported to be more beneficial than feedback. Results indicate that guided rubrics significantly increase "feedfoward" observations and reduce less useful categories of feedback, such as problem detection and meanness. Differences between rubrics, however, had limited influence on student attitudes post-peer review. Consequently, potential strategies for further improving student ratings and keeping mean commentary at a minimum are discussed.

Peer review, a widespread procedure in both educational and professional environments, is often lauded as beneficial by both researchers and instructors. Reflecting on the numerous times and contexts in which peer review is performed both formally and informally, Topping (2009) asserts that involvement in the peer review processes allows students to "develop transferable skills for life" (p. 21). Such skills include fostering a sense of student ownership and responsibility for the paper and assessment process, handling mistakes as opportunities to learn rather than failures, and allowing students to practice evaluative skills that can be applied in their careers (Vickerman, 2009). Furthermore, studies also demonstrate that peer review helps the reviewer as well as the student being reviewed. Reviewers may increase the time they spend on task, obtain a greater understanding of the assignment and their own errors, and reflect more on future assignments (Cho & MacArthur, 2011; Topping, 2009). Studies that ask students to evaluate the peer review process also indicate that such work can increase student thoughtfulness and knowledge about what is required in the assignment (Pain & Mowl, 1996).

More research is required, however, to support this optimistic viewpoint, especially because empirical evidence indicates that peer review is not always an effective process, in part due to student perceptions. Nelson and Carson (1998) found that peer review did not successfully support the instructors' goal of developing student papers, attributing the majority of the failure to students viewing the process as an exercise in identifying mistakes and correcting sentence-level error. Though they worked specifically in an ESL classroom, other research corroborates that a focus on evaluation and correction may be the default mode for all students (Crossman & Kite, 2012). In addition, students' attitudes about peer review can also be mixed or negative (Van Zundert, Sluijsmans, & Van, 2010). In a study by Levine, Kelly, Karakoc, and Haidet (2007), students provided negative comments about the peer assessment process instead of explanations for why they gave their peers the marks they did. Pain and Mowl (1996) assessed the effectiveness of peer review in a first-year geography course and found that, even after training, approximately half of the students did not perceive the benefits of peer (or self) assessment.

Taken together, these conflicting results suggest that further studies are necessary for a more comprehensive understanding of peer review methodology and its effect on student opinions, which influence implementation and future peer review interactions. The particular form of peer review, of course, varies based on course type, assignment and objectives. Some studies define peer review, also known as peer assessment, as an evaluation of a final product by peers (Gennip, Segers, & Tillema, 2010). Others refer to peer review as a scaffolded process where formative feedback is available prior to the development of the final product (Odom, Glenn, Sanner, & Cannella, 2009). Given that other work has examined assessment in the non-composition classroom (Harris, 2011; Walvoord, Hoefnagels, Gaffin, Chumchal, & Long, 2008), this study focuses on ratings and commentary on two different rubrics for rough drafts of student essays in an introductory biology course. Such an analysis is critical due to an increase in writing across the curriculum (WAC) initiatives (Beason, 1993) and other writing intensive (WI) departmental requirements, which encourage peer review activities due to pragmatic concerns, such as large class sizes (Covill, 2012; Kelly, 1995). Consequently, peer review may be used frequently across disciplines, perhaps

before experimental studies can assess what factors constitute effective peer review in context. Therefore, in order to benefit WAC and WI programming and their goals, this study contributes to preliminary research analyzing peer review in the science classroom.

By examining student commentary, this study complements work by Cho and MacArthur (2011), whose research categorized peer feedback in an introductory physics lab, Artemeva and Logie (2003) and Dominguez, Cruz, Maia, Pedrosa, and Grams (2012), whose experiments examined categories of peer review commentary for engineering students, and Beason (1993), whose study quantified peer responses in a variety of writing-enriched courses, including dental hygiene. Comparing this study's results to experiments performed outside the humanities will allow for a better understanding of how peer review functions in the context of writing across the curriculum. In analyzing such commentary, this study also considers an understudied category of student response described as inflammatory language (Nelson & Schunn, 2009) or failure/meanness (Rysdam & Johnson-Shull, 2011). This category includes comments that are so harsh that they are no longer constructive (Nelson & Schunn, 2009) or responses that announce failure or emphasize the negative (Rysdam & Johnson-Shull, 2011). Such an examination will facilitate a deeper knowledge about the variables that influence unnecessarily harsh commentary, including anonymity and the use of support materials, such as rubrics.

Rubrics and Guided Peer Review

Rubrics, the framework that guides this research, are defined as guidelines that provide information about what features of student performance matter most. Written by instructors, they often provide criteria and rating scales for final evaluation (Petkov & Petkov, 2006). Covill (2012) indicates that, though rubrics used by instructors and administrators have been extensively considered, few empirical studies have examined an instructional rubric aimed for scaffolded student use and how it influences their "beliefs, practices, and performance" (p. 1). For example, while rubrics are often provided in the appendices of research on peer review, information about their construction and the type of written commentary they procure is often absent. Nelson and Schunn (2009) acknowledge that different instructional prompts result in different forms of commentary, but they go no further in their analysis of rubric construction and its effects. In "Eliciting formative assessment in peer review," Goldin and Ashley (2012) assert, "Rubrics may be used within peer review to support assessment, but few studies examine rubrics *per se*…[though] the choice of rubric influences the experience of both reviewers and authors" (p. 211). Ideally, well-constructed rubrics augment students' self-efficacy, motivation and performance (Covill, 2012).

In response to Goldin and Ashley (2012), this study assists in granting rubrics the critical attention they deserve by examining the effects of a definitive addition, the inclusion of guided questions (see Appendix A), on the types of student commentary present on a problem-specific rubric. This assessment is critical considering the dearth of experiments directly linking outcomes and methodologies in peer review (Van Zundert et al., 2010). Specifically, I hypothesize that guided questions will increase student commentary in the "feedfoward" category, one that has been previously considered in the context of the writing center (e.g., Murtagh and Baker, 2009). In contrast to observations about what occurred in the writer's work (i.e. feedback), feedforward comments include questions and suggestions that focus on what the writer could do in the future. Feedforward is posited as more effective because it results in less defensiveness and an emphasis on revision instead of failure (Goldsmith, 2003). Pragmatically, focusing on specific changes in rubric methodology is also a way for instructors to improve student responses and the success of peer review without spending significantly more time on the process. Previous work suggests that, at least in the short term, peer review may actually require more resources in terms of training, organization and monitoring (Rubin, 2006). Thus, this study aims to examine how even slight changes could advance the process without requiring a significant increase in instructor effort.

Methods

Background

Participants in this study were enrolled in an introductory biology course for non-major students at a large, public, land-grant institution that is one of two research-oriented universities in the state. The approximately 550 participating students were evenly divided between males (48%) and females (52%), and the majority of them were freshman and sophomores who spoke English as their first language. During the semester, students were assigned a writing prompt requiring them to evaluate news articles on a controversial scientific topic. The aim was to provide students with a greater understanding of how science is portrayed in popular media, and assessment was largely focused on the student's ability to effectively complete four tasks: summarize the news articles, identify the articles' key assumptions, assess the articles' validity, and present their own opinion on the topic. Peer review was implemented in lab sections (groups of ~35) run by

teaching assistants (TAs) who were charged with introducing the assignment and helping students revise their rough drafts. Thus, though written instructions and rubrics were standardized, verbal directions and time spent discussing the assignment may have varied between lab sections, and no set tutorials on writing quality or peer review were provided. All peer reviews were done during lab in the same week, and each lab section was randomly assigned the control or guided rubric. The rubrics were identical except for the exclusion (control rubric) or inclusion (guided rubric) of guiding questions (Appendix A). Peer review was worth 5% of the final grade for the assignment. The week following peer review, rough drafts and rubrics with written comments were returned to students, and they were given a questionnaire aimed at examining their attitudes concerning the process. Time dedicated specifically to verbal peer review discussions in lab was not provided.

Rubric Design and Implementation

The rubrics were developed with a consideration of relevant research as well as previous experience with instructional rubrics in the course. I developed a problem-specific rubric, which focuses on content related to the assignment, because research indicates it is more effective than a domain-relevant rubric, which focuses on general comments within domains (e.g. issue, argument), in terms of validity and lower inter-dimension correlation (Goldin & Ashley, 2012). In addition, a problem-specific rubric is particularly useful in a WAC/WI course, where writing assignments are less frequent, because rubrics do not need to be continuously modified to fit the larger context of other projects. Because lengthy and highly-detailed rubrics may be impractical or not positively affect results, the control and guided rubrics both emphasized the four main parts of the assignment (Colvill, 2102; Popham, 1997). Directly relating the rubrics to the assignment prompt aimed to facilitate cognitive gains, such as a reexamination of the assessment criteria and reflection (Colvill, 2012). Portions of the rubrics were also included or modified based on results from a version of the control rubric that was previously used during peer review of a similar assignment. Control and guided rubrics were revised and approved by the TAs and the professor prior to implementation.

Both rubrics asked students to evaluate the author's response to the four main parts of the assignment on a 3-point scale (1 = weak or missing, 2 = good, and 3 = strong). However, the more general follow-up statement on the control rubric ("Explain") was replaced with specific, guiding questions on the guided rubric ("What questions do you have for the author? What steps might the author take to improve…"; see Appendix A). Every student randomly received another student's work to review within the lab group, and both authors and reviewers were identified on the rubric. Following peer review, rubrics were collected with permission from a total of 366 students, with 198 students assigned to the control rubric and 168 students assigned to the guided rubric.

Questionnaire Design and Implementation

After students received their peer review feedback, they were given a questionnaire aimed at examining their attitude about the peer review process. The questionnaire rated students' familiarity with peer review at the university on a 5-point Likert-type scale, and, on a 10 point Likert-type scale, both their attitudes toward peer review in general and peer review in the course. Students were subsequently asked to explain why they provided their rating of the peer review in the course, what reviewer comments were most and least helpful for improving their final draft, and if assessing another student's paper helped them improve their own. Student responses were paired with their corresponding peer reviews whenever possible, so that peer responses and their relationship to perceived utility could be directly assessed. Because some students did not allow their rubrics or responses to be used in the study, this pairing was only possible for 70% of the peer review rubrics (148 control rubrics and 121 guided rubrics).

Coding

Student comments from both rubrics were sorted into one of eight functional categories: problem detection, explanation, praise, guidance, questions, summation, doubt, or reader response. Categories were constructed based on existing research (see, for example, Beason, 1993; Nelson & Shunn, 2003; Rysdam & Johnson-Shull, 2011; Zhu, 2001) and preliminary observations of the types of comments received. Names and definitions of these categories are provided in Table 1, as well as aforementioned WAC/WI studies' corresponding categories for assessing peer review commentary. Comments representing "inflammatory language" or "failure/meanness" were noted and also coded as one of the other 8 categories (predominantly problem detection). Students' explanations of their ratings for the course's peer review were separated into units addressing a single topic, otherwise referred to as idea units (Nelson & Schunn, 2009), and sorted into one of ten categories: useful, lack of time/effort, peer inadequate, depends on peer, vague/confusing, instructor better, already knew, bad rubric, harsh grader, and personal inadequacy (see Table 2 for examples). Thus, a response that indicated that peer review was useful, but that instructor commentary

Table 1
Categories of Commentary, Definitions, and Examples

Categories of Commentary				Definition	Examples
Goldsmith 2003	Dominguez et al. 2012	Cho & MacArthur 2011	Current study	Current Study	Current Study
Feedback	Problems	Problem detection	Problem detection	Points out flaw	"Writing is not clear." "Could flow better."
Feedback	Problems	Problem detection	Explanation	Elaborates on flaw through localization or examples	"You touch on the findings but don't get into arguments, numbers, mistakes or ethics behind the studies." "Thesis statement is not in the first paragraph."
Feedback	Praise	Praise	Praise	Describes strength	"You explain the evidence well."
Feedforward	Solutions	Solution Suggestion	Guidance	Suggestion(s) for improvement	"Develop your point on skepticism if you have [one]." "Maybe use cases as examples in your paper to give evidence of the misuse of BPA and other chemicals."
Feedforward	NA	NA	Question	Asks question	"What is your opinion?"
Feedback	Summarization	NA	Summation	Describes essay without evaluation	"The key assumption in the article is that false la[b] reports are not an accident." "Both sides were indeed brought up in the conclusion."
Feedforward	NA	NA	Reader response	Describes reviewer's opinion	"…this makes me wonder if what we put in our body should really be solely up to us as consumers."
NA	NA	NA	Doubt	Unsure of advice	"Not sure if [you] need more."

would be preferred would be coded as "useful" and "instructor better." Responses were categorized as useful even with qualifiers (e.g. good, but could have been better). Students who stated that peer review helped them with their own work were also reported (Table 2).

Statistical Analysis

To control for the effect of TAs, who might have influenced confounding aspects of the peer review process (e.g., amount of explanation, timing of peer review activity in relation to other lab tasks, etc.), an analysis of covariance (ANCOVA) was used to analyze differences between the control and guided rubric in the type of commentary procured. Correlations between the effectiveness ratings for the course and the number of responses in each category (e.g. problem detection, guidance etc.) were evaluated using Spearman's rank correlation coefficients.

Results and Discussion

Commentary and WAC/WI Courses

Students in the course provided a total of 3,021 comments across 366 rubrics, resulting in an average of approximately 8 comments per rubric. Two students provided 21 comments, the highest number of comments left on a single peer review rubric, and nine students left less than 4 comments, meaning that they did not provide responses for all the scores they gave. Summation was the most common category of review response across treatments, followed by problem detection, guidance and praise. On average, students contributed one positive comment per peer review and only explained one problem that they pointed out through localization or example. Doubt and reader response were rarely noted (Table 3). Overall approximately 48% of students found peer review useful as an author, while approximately 63% of students found it useful as a reviewer. Though the questionnaire did not directly assess why reviewing was useful, several students provided reasons for why being a reviewer was effective in their comments about peer review in general. For example, one student commented, "It helped me with my own paper, [because] the [paper] I peer reviewed was very well written," and another student stated, "I think it's effective to see other people's papers and learn from their accomplishments and mistakes." A third student recognized the importance of reexamining the assignment guidelines: "It's [effective] because it made everybody go back to the grading rubric and confirm if the paper met the grading rubric's expectations." Even a student who was dissatisfied with her reviewer admitted, "The [rubric] helped a little." Thus, as shown in Table 2, students may perceive the benefits of reviewing even if they are frustrated with the comments they receive.

When the results of this study were limited to the four categories of commentary examined by Cho and MacArthur (2011), one of the few studies to assess peer review responses in a science classroom, the number of comments per rubric as well as percentages of problem detection, explanation, praise and guidance were strikingly similar (Table 4). In addition, their study also demonstrated the importance of cognitive gains for the reviewer, showing that reviewers who identified problems and offered solutions significantly improved their own writing quality post-review; their students often commented that peer review helped them consider audience and what they should and should not do in their own work. Along with course context, Cho and MacArthur's (2011) participants and methods aligned with this study in several other respects. Their 61 participants, enrolled in an entry-level physics course, were also predominately 1st or 2nd year students at a Research 1 university, and they were evenly divided between males and females. Their evaluative rubric consisted of instructional guidelines which also contained four main questions as well as several supplemental tasks and examples. This comparison preliminarily suggests that WAC/WI courses with comparable goals, tools and student demographics may procure similar categories of peer response across assignments, and that strategies for improvement may be effective across such classrooms. However, other research indicates that further experimentation is necessary to better understand what components are most important for generalizability. For example, some results of this study were consistent with Dominguez et al. (2012), who examined peer reviewer commentary from 39 participants in a mid-level engineering course, while others were markedly different (Table 4).

Additional research can define what factors have the greatest influence on differences between categories of commentary and if some responses remain consistent across classrooms outside the humanities. In order to do so, clarifying the peer review process and the supporting materials used is critical. For example, few results are consistent between this study and the writing-enriched courses analyzed by Beason (1993); however, no information about the type of peer review or rubrics given to students is provided, making it difficult to fully assess cause and effect. Topping (2010) offers an extensive list of procedural questions to address including, "Does the interaction involve guiding prompts, sentence openers, cue cards or other scaffolding devices? What extrinsic or intrinsic rewards are made available for participants?" (p. 343). These questions are especially important in order to realistically compare the few studies examining peer review in the context of WAC/WI courses.

Rubrics and Commentary

This study hypothesized that a rubric with guided questions would influence the categories of student commentary received, and changing the rubric's form did significantly affect the amount of comments in 4 of the 8 categories. Overall, the guided rubric had more questions and guidance and less problem detection and summation than the control rubric (Table 3). In addition, comments on the guided rubric were more equally spread across categories. Though guidance, summation and problem detection were the most common, praise and questions also had approximately one comment per rubric on average. Explanation, reader response and doubt were infrequent. On the control rubric, summation, problem

Table 2
The Percent of Student Responses in each of the Response Categories

Category	Examples	Student responses control rubric (%)	Student responses guided rubric (%)
Useful (reviewer)	Circled 'yes' (response sheet)	65.9	60.1
Useful (author)	"The peer review I received gave me insight as to how others perceived my paper"	48.2	46.7
Lack of time/effort	"My peer reviewer (I felt) did not give me a very detailed review" "…People rushed through their peer reviews"	20.8	23.7
Peer Inadequate	"The person who peer reviewed my paper did not seem to understand the assignment"	15.7	19.1
Depends on peer	"If the reviewer is basing their reviews off of false knowledge, then the review hurts you rather than helps you"	9.1	13.2
Vague/confusing	"Didn't really give me specific things I could change" "the person that reviewed it was not clear or made no sense"	9.6	9.9
Instructor better	"I would much rather have a teacher review it"	3.6	3.9
Already knew	"I already knew what I needed to fix and add"	3.0	2.6
Bad rubric	"Too detailed questions" "Rubric inadequate"	2.0	2.6
Harsh Grader	"I feel like my peer reviewer was too brutal"	1.5	2.0
Personal Inadequacy	"I didn't have the [right] paper or topic, and it was too short so I didn't get very much feedback"	1.5	2.0

detection and praise were the three common categories of commentary, with all other categories remaining infrequent (less than one comment per rubric on average). The rubrics did not differ in the categories of explanation, praise, reader response, doubt or the total number of comments received (Table 3). Thus, the guided rubric did succeed in facilitating feedforward responses and, when compared to the control rubric, had fewer instances of problem detection, a less useful category due to its lack of specificity (Nelson & Shunn, 2009). These results are consistent with Artemeva and Logie (2003), who state that guidelines in the form of questions and checklists help students provide commentary that addresses a wider variety of issues and problematic sections of the text.

However, only limited data suggest that students found the guided rubric to be more effective. When students were asked to compare their experiences with peer review in general to peer review in the course, 37% of students who used the guided rubric rated peer review as more effective in the course compared to 25% of students with the control rubric. In contrast, students' perceived rating of peer review effectiveness both in general (6.2 out of 10) and in the course (5.9 out of 10) did not differ based on rubric. Overall, approximately half (48%) of students commented that they thought peer review was useful. Reported reasons why peer review was ineffective remained consistent between rubrics, with the most cited reasons being lack of time/effort from reviewer, inadequate peer reviewer and vague/confusing review (Table 2). All of the other reasons for peer review being ineffective were utilized by less than 5% of the students (Table 2). No significant correlations were found between the ratings of effectiveness of the peer review in the course and the number and type of responses made by the reviewer.

Several of the study's outcomes may explain why students did not consistently find the guided rubric to be more effective. One reason is that the control rubric had the highest average number of comments in the summation category, a category of non-evaluative feedback that can allow students to detect mistakes without a negative value judgment. Ferris (1997)

Table 3
Difference Between Guided and Control Rubrics using ANCOVA

Response Categories	Mean Squared (Control)	SE	Mean Squared (Guided)	SE	F ratio	P
Problem Detection	2.31	0.16	1.58	0.19	6.40	**0.012**
Explanation	0.71	0.08	0.81	0.10	0.53	0.467
Praise	1.10	0.10	1.03	0.12	0.16	0.685
Guidance	0.57	0.11	1.71	0.13	31.71	**<0.001**
Question	0.15	0.10	1.22	0.12	36.08	**<0.001**
Summation	3.40	0.15	1.61	0.17	46.93	**<0.001**
Reader Response	0.04	0.02	0.08	0.02	1.18	0.279
Doubt	0.06	0.02	0.02	0.02	1.53	0.216
Total	8.35	0.24	8.06	0.28	0.45	0.501

Note: Degrees of Freedom equal 1 for all response categories. Significant p-values are highlighted in bold at $\alpha = 0.05$.

Table 4
Comparison of Student Responses during Peer Review in Different Science Classrooms

Current study Category	Percent (%)	Dominguez et al. 2012 Category	Percent (%)	Cho & MacArthur 2011 Category	Percent (%)
Problem detection & explanation	55.2	Problems	31.1	Problem detection	48.8
Praise	21.5	Praise	21.7	Praise	22.4
Guidance	23.3	Solutions	22.7	Solution suggestion	19.2
Summation	30.4	Summarization	5.5	NA	NA

indicates that providing summary promoted more substantial student revision, and Nelson and Schunn (2009) demonstrate that summarization positively affected students' understanding of the problems in the text. Another potential reason is the low level of explanation present in both rubrics. Leijen and Leontjeva (2012) found that directive comments, or statements commenting on specific changes exclusive to the paper, were a better predictor of implementation than mentioning solutions. Thus, the lack of specificity resulting from the low level of explanation across rubrics may have been frustrating to all students. The fact that many students cited a lack of reviewer time/effort and vague/confusing commentary as reasons for ineffective peer review supports this explanation. This study also focused on student attitudes rather than performance or learning, and it is possible that the guided rubric did positively affect student revision regardless of perceived effectiveness. Further studies are necessary in order to relate feedfoward to performance and determine what role student attitude plays in the process.

Research that quantifies student response to peer review provides additional measures for making peer review more effective. Artemeva and Logie (2003) cited similar frustrations to students in this study during peer review (e.g., dismissive attitudes of peers, peer incompetence and confusion), and suggest two improvements: having papers reviewed by more than one student and providing time for face-to-face interactions as well as written response. Several students also recommended that post-review discussions would be useful. One student remarked, "I believe this peer review was somewhat effective. [It] would have been more beneficial personally if we could discuss our papers with the reviewer after the peer review took place," and another stated, "I didn't actually talk to the person who graded me. I didn't have a chance to hear exactly what they meant." Two additional students provided similar statements. In addition, one student commented on the benefits of more than one reviewer: "I think it would have been more effective if multiple people peer reviewed your paper. That way more opinions would have been stated." All of these comments were received even though the questionnaire did not specifically ask how peer review might be improved, a fact that highlights their perceived importance to students. These suggestions are beneficial because they could also be

implemented without a significant increase in planning time for the instructor, an ongoing pragmatic concern.

Anonymity and Harsh Commentary

Only 16 of the 366 rubrics examined contained unnecessarily harsh commentary (12 of the control rubrics and 4 of the guided rubrics) in comparison to the 39% coded by Rysdam and Johnson-Shull (2011) and the < 0.5% coded by Nelson and Schunn (2009). Nelson and Schunn defined unnecessarily harsh commentary as criticism that is insulting instead of constructive, and Rysdam and Johnson-Shull (2011) defined it as "any comment that identified incorrectness without correcting, announced what the writing was not doing, and/or emphasized the negative with exclamation or other dramatics" (p. 4). Though Rysdam and Johnson-Shull (2011) did not separately categorize problem detection, the overwhelming majority of comments announcing failure were also mean, and characteristic examples included: "Unbelievably boring," "Follow instructions!", and "Overall the quality is poor. I can't even tell where to start correcting" (p. 7). Examples from this study included, "Looks like it was written this morning," "Needs smoother sentences!," and "It was hard to read and stay interested with it." Far from being what the student needs to hear, harsh commentary is unconstructive and negatively influences the effectiveness of peer review. For example, the author of a paper subject to one of the harsh reviewers gave peer review in the course an effectiveness rating of 3 out of 10, lower than his effectiveness score of peer review in general. The author's response also indicates he was affected by the comment: "I feel like my peer reviewer was too brutal. They said it looked like it'd been written that morning, mostly because of a few typos and unfinished citations." Many researchers and instructors warn against harsh commentary during peer review, regardless of the age and position of the reviewer (Belcher, 2009; Cho & MacArthur, 2011 Rosenfield & Hoffman, 2009).

The lack of harsh commentary in this study may be due to the fact that both authors and reviewers were identified on the rubric. For example, research indicates that even anonymous professional peer review can lead to unnecessarily cruel or ignorant comments not useful for revision (Rosenfield & Hoffman, 2009) and others have considered a move to open professional peer review to solve this problem (Walsh, Rooney, Appleby, & Wilkinson, 2000). In the current study, some students were quick to criticize their peers cruelly on the post-peer review questionnaire, which they knew was not going to be viewed by anyone in the class. The following comments were given even though the authors had not received any unnecessarily harsh commentary:

> [Peer review was ineffective] because who reviewed my paper was rude and not constructive at all.
>
> She kept asking/saying pointless things.
>
> The peer review I received was sub-par.
>
> My reviewer gave nothing but bad feedback and judging by her comments, doesn't understand how to read a paper.
> The person who reviewed mine obviously failed English in high school and had no idea what they were doing.
>
> Reviewer didn't know what they were talking about.
>
> I thought the peer review process wasn't actually effective…my reviewer stunk.

Therefore, though previous studies indicate that students prefer providing feedback anonymously to allow for honest assessment (Bostock, 2009), instructors must carefully consider whether or not students should be identified. For example, few students in this study indicated that they felt peer reviewers were afraid to be honest, and, contrary to expectations, some students stated that anonymous commentary may not be desired. One student remarked, "I feel that sometimes a random peer review will not always have a good effect fixing your own paper. If someone you know looks at your paper, he/she will give you the best ways on how to improve your paper."

Supporting materials, such as rubrics or other tools, and grade incentives may also keep unnecessarily harsh commentary at a minimum. Students in Rysdam and Johnson-Shull's (2011) study were trained in a peer review technique called AFOSP (focusing on a hierarchy of values: assignment, focus, organization, support, and proofreading) but were asked to write directly on drafts of the author's paper and were not graded on their responses. Nelson and Schunn (2009), who also had very low number of inflammatory comments, used anonymous peer review; however, students used an online peer review system (SWoRD) that allowed authors to directly evaluate reviewer helpfulness. Thus, if anonymous peer review is used in the classroom, a technique should be implemented to motivate students to provide constructive categories of response. In this study, aspects of the essay to comment on were explicitly outlined in the rubrics, and 5% of the final grade was based on providing useful peer review commentary. Furthermore, results preliminarily indicate that providing guiding questions may also help students remain cordial, because only 4 of the 16 rubrics with

unnecessarily harsh commentary were guided rubrics. Additional research is necessary to gauge the degree to which anonymity, supporting materials and grade incentives contribute to a reduction in cruel commentary.

Conclusions and Future Directions

This study supplements the literature examining peer review in higher education by providing one of the first empirical studies specifically analyzing commentary and student instructional rubrics in the context of WAC/WI courses outside of the humanities. The results indicate the categories of responses provided by students in science courses with analogous goals and participant demographics may be strikingly similar, and cognitive gains by reviewers may be most apparent. A guided rubric did procure significantly more guidance and questions and significantly less summation and problem detection than a similar control rubric, increasing the amount of useful feedforward commentary provided by students. However, most measures of perceived peer review effectiveness suggest that participants in this study found both rubrics equally useful, perhaps due to an increased number of summary responses with the control rubric or the infrequent use of explanations across both rubrics. Unnecessarily harsh commentary was rarely noted, indicating that anonymous peer review, a lack of supporting materials, such as rubrics, and failing to provide grade incentives may contribute to harmful categories of response. For example, when students provided comments that their peers were not going read on the post-peer review questionnaires, they were more likely to be cruel, and a study that had similarly low levels of inflammatory language to this one also provided tools for assessing and evaluating peer responses.

Including multiple reviewers, offering face-to-face interaction along with written peer responses, and identifying reviewers may all contribute to more positive attitudes post-peer review, and additional studies are required to better examine these strategies as well as other important aspects of the process. For example, this study did not compare drafts and final essays to determine what peer review comments were actually used by students, nor did it examine differences in performance between students with the control and guided rubric. Recent work has assessed the relationship between peer review techniques and writing quality in different contexts, including courses focused on foreign language and grade-school learners (Rahimi, 2013; Yu & Wu, 2013), and other investigators have examined the relationship between understanding, agreement and implementation in the history classroom (Nelson & Schunn, 2009). Investigating these associations further in science courses (see, for example, work by Mulder, Baik, Naylor, & Pearce, 2014) will allow for a more comprehensive understanding of peer review under the framework of WAC and WI classrooms. Furthermore, researchers such as Gielen, Peeters, Dochy, Onghena, and Struyven (2010) suggest that the type of commentary that significantly improves performance may also be the most difficult to teach, while Rahimi (2013) found that training increased the number of peer review comments used by students and overall writing quality. Thus, providing additional TA training and tutorials or a calibration process for students may also assist in improving the peer review process.

References

Artemeva, N., & Logie, S. (2003). Introducing engineering students to intellectual teamwork: The teaching and practice of peer feedback in the professional communication classroom. *Language and Learning Across the Disciplines, 6*(1), 62-85.

Beason, L. (1993). Feedback and revision in writing across the curriculum classes. *Research in the Teaching of English, 27*(4), 395-422.

Belcher, W. L. (2009). *Writing your journal article in 12 weeks: A guide to academic publishing success*. Thousand Oaks, CA: SAGE Publications.

Bostock, S. J. (2009). Motivation and electronic assessment. In A. Irons & S. Alexander (Eds.), *Effective Learning and Teaching in Computing* (pp. 86-99). New York, NY: RoutledgeFalmer.

Cho, K., & MacArthur, C. (2011). Learning by reviewing. *Journal of Educational Psychology*, *103(*1), 73-84.

Covill, A. E. (2012). College students' use of a writing rubric: effect on quality of writing, self-efficacy, and writing practices. *Journal of Writing Assessment*, 5(1).

Crossman, J. M., & Kite. S. L. (2012). Facilitating improved writing among students through directed peer review. *Active Learning in Higher Education, 13*(3), 219-229. doi: 10.1177/1469787412452980

Dominguez, C., Cruz, G., Maia, A., Pedrosa, D., & Grams, G. (2012). Online peer assessment: An exploratory case study in a higher education civil engineering course. In *Interactive Collaborative Learning (ICL), 2012 15th International Conference on Interactive Collaborative Learning (ICL),* Villach, Austria: 1-8.

Ferris, D. R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly,* 31(2) 315-339. doi: 10.2307/3588049

Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection diagnosis and the strategies of revision. *College Composition and Communication*, *37*(1), 16–55.

Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction, 20*(4), 280-290. doi:10.1016/j.learninstruc.2009.08.010

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction, 20*(4), 304-315. doi:10.1016/j.learninstruc.2009.08.007

Goldin, I. M., & Ashley, K. D. (2012). Eliciting formative assessment in peer review. *Journal of Writing Research, 4*(2), 203-237.

Goldsmith, M. (2003). Try feedforward instead of feedback. *Journal for Quality & Participation, 26*(3), 38-40.

Harris, J. R. (2011). Peer assessment in large undergraduate classes: An evaluation of a procedure for marking laboratory reports and a review of related practices. *Advances in Physiology Education, 35*(2), 178-87. doi: 10.1152/advan.00115.2010

Kelly, L. P. (1995). Encouraging faculty to use writing as a tool to foster learning in the disciplines through writing across the curriculum. *American Annals of the Deaf, 140*(1) 16-22. doi: 10.1353/aad.2012.0272

Leijen, D. A. J., & Leontjeva, A. (2012). Linguistic and review features of peer feedback and their effect on the implementation of changes in academic writing: A corpus based investigation. *Journal of Writing Research, 4*(2), 177-202.

Levine, R. E., Kelly, P. A., Karakoc, T., & Haidet, P. (2007). Peer evaluation in a clinical clerkship: students' attitudes, experiences, and correlations with traditional assessments. *Academic Psychiatry,* 31, 19-24. doi: 10.1176/appi.ap.31.1.19

Murtagh, L. & Baker, N. (2009). Feedback to feed forward: student response to tutors' written comments on assignments. *Practitioner Research in Higher Education, 3*(1), 20-28.

Mulder, R., Baik, C., Naylor, R., & Pearce, J. (2014). How does student peer review influence perceptions, engagement, and academic outcomes? A case study. *Assessment & Evaluation in Higher Education*, 39(6), 657-677. doi: 10.1080/02602938.2013.860421

Nelson, G., & Carson, J. (1998). ESL students' perceptions of effectiveness in peer response groups. *Journal of Second Language Writing, 7*(2), 113-131. doi:10.1016/S1060-3743(98)90010-8

Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: how different types of peer feedback affect writing performance. *Instructional Science,* 37(4), 375-401. doi: 10.1007/s11251-008-9053-x

Odom, S., Glenn, B., Sanner, S., & Cannella, K. A. S. (2009). Group peer review as an active learning strategy in a research course. *International Journal of Teaching and Learning in Higher Education, 21*(1), 108-117.

Pain, R., & Mowl, G. (1996). Improving geography essay writing using innovative assessment. *Journal of Geography in Higher Education, 20*(1), 19-31. doi: 10.1080/03098269608709341

Petkov, D., & Petkova, O. (2006). Development of scoring rubrics for IS projects as an assessment tool. *Issues in Informing Science & Information Technology, 3,* 499-510.

Popham, W. J. (1997). What's wrong—and what's right—with rubrics. *Educational Leadership*, *55*, 72-75.

Rahimi, M. (2013). Is training student reviewers worth its while? A study of how training influences the quality of students' feedback and writing. *Language Teaching Research*, *17*(1), 67-89. doi: 10.1177/1362168812459151

Rosenfield, D., & Hoffman, S. J. (2009). Holiday review. Snappy answers to stupid questions: An evidence-based framework for responding to peer-review feedback. *Canadian Medical Association Journal, 181*(12), 301-305.

Rubin, R. (2006). The academic journal review process as a framework for student developmental peer feedback. *Journal of Management Education, 30*(2), 378-398.

Rysdam, S., & Johnson Shull, L. (2011). The ink we leave behind. In C. Lowe & T. Williams (Eds.), *2010 WPA Conference Proceedings*. 15-18 July 2010, Philadelphia, PA: Council of Writing Program Administrators.

Topping, K. J. (2009). Peer assessment. *Theory into Practice, 48*(1), 20-27. doi: 10.1080/00405840802577569

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction, 20*(4), 339-343. doi:10.1016/j.learninstruc.2009.08.003

Van Zundert, M., Sluijsmans, D., & Van, M. J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20(*4), 270-279. doi:10.1016/j.learninstruc.2009.08.004

Vickerman, P. (2009). Student perspectives on formative peer assessment: An attempt to deepen learning? *Assessment & Evaluation in Higher Education, 34*(2), 221-230. doi: 10.1080/02602930801955986

Walsh, E., Rooney, M., Appleby, L. & Wilkinson, G. (2000). Open peer review: A randomised controlled trial. *The British Journal of Psychiatry: the Journal of Mental Science* 176:47-51. doi: 10.1192/bjp.176.1.47

Walvoord, M. E., Hoefnagels, M. H., Gaffin, D. D., Chumchal, M. M., & Long, D. A. (2008). An analysis of calibrated peer review (CPR) in a science lecture classroom. *Journal of College Science Teaching, 37(*4), 66-73.

Yu, F. Y., & Wu, C. P. (2013). Predictive Effects of Online Peer Feedback Types on Performance Quality. *Educational Technology & Society*, *16*(1), 332-341.

Zhu, W. (2001). Interaction and feedback in mixed peer response groups. *Journal of Second Language Writing, 10*(4), 251-276. doi:10.1016/S1060-3743(01)00043-1

LAUREN KELLY is working toward her M.A. in Rhetoric and Composition at Washington State University. Her research interests include scientific communication, writing and assessment across the curriculum, the role of peer and instructor commentary and creative writing.

Acknowledgements

I would like to thank Andrew Storfer, Linda Cook, and the TAs and students in Biology 102 for allowing me to perform this study and participating in the research. This project also would not have been possible without the statistical support of Richard Lamb and the editing and assistance of Lisa Johnson-Shull, Justin Poinsatte, Todd Butler and the graduate students in English 595.

Appendix A
Guided rubric.

"What questions do you have for the author? What steps might the author take to improve…?" were replaced with "Explain" on the control rubric.

Your name: _____ Author's name: _____
Directions: Actively read through the paper you've been assigned to peer-review. Make comments on the paper (in the margins etc.) and then fill out this peer review form. Return this form + the peer-reviewed rough draft during next week's lab (the week of 2/25)

Part 1: Content
1. Write down the author's thesis statement.

2. Is it clear and easy to find? YES NO

3. Is it stated at the end of the introduction and again in the conclusion? YES NO

4. Does the paper summarize the articles well in 1-2 paragraphs (1=weak or missing, 2=good, 3=strong)?

_____

What questions do you have for the author? What steps might the author take to improve his/her summary?

5. What are the key assumptions in the articles? Does the author present both sides of the ethical issue(s) (1=weak or missing, 2=good, 3=strong)?

#
_____

What questions do you have for the author? What steps might the author take to improve his/her assessment of the assumptions and ethical issues provided in the articles?

6. Does the author assess the validity of the conclusions made in both articles based on supporting data/evidence (1=weak or missing, 2=good, 3=strong)?

Questions to consider from rubric: Is the evidence supported by scientific experimentation? Is it only a single experiment? Are there conflicting data? Does the article overstate the issue based on the evidence? Are the conclusions well supported? Is the sample size large enough? Are the graphs accurate? Are there potentially studies that yield conflicting results in the literature? Are there true causative links established or are there simply correlations?

#
_____

What questions do you have for the author? What steps might the author take to improve his/her assessment of the evidence's validity/supporting data?

7. After reading this section, can you tell if the author trusts the articles? YES NO

8. Does the author provide his/her own opinion on the issue (in up to one page)? YES NO

9. Does he/she provide enough evidence to back up her opinion (1=weak or missing, 2=good, 3=strong)?

Questions to consider from the rubric: Identifies, appropriately, one's own position on the issue, drawing support from experience and information not available from the chosen article. (What additional information is needed? Are you aware of any conflicting studies? If so, what are they and what are the conclusions?)

_____

What questions do you have for the author? What steps might the author take to improve his/her opinion on the issue?

Part 2: Citations

1. Is there a works cited (bibliography) page? YES NO

2. Are there in-text citations for quotes and paraphrasing (If missing, please mark on paper)?
YES NO SOME