

Practicing Peer Feedback: How Task Repetition and Modeling Affect Amount and Types of Feedback over a Series of Peer Reviews

LUCY BRYAN, DAYNA S. HENRY, SARAH R. BLACKSTONE,
ANNA MARIA JOHNSON, AND LACIE KNIGHT

Providing feedback on peers' writing is a complex endeavor that engages several higher-order cognitive processes. While some evidence suggests that practice improves peer-review skills, more research is needed to understand how peer feedback changes with practice. The present study aims to (1) explore the impact of practice on the amount and types of feedback that students give in peer review and (2) investigate whether providing model feedback in addition to practice enhances students' development as peer reviewers. The researchers analyzed 3,761 comments provided by eighty students over the course of four peer-review sessions. Quantitative analysis of feedback quantity and qualitative analysis of feedback content revealed changes over time, including differences in the feedback of students who did and did not have access to model feedback, and differences in feedback from minimal, moderate, and heavy commenters. Practicing providing feedback throughout several rounds of peer review may help students generate more and higher-quality feedback, especially when paired with training in the form of reviewing model feedback.

Across many disciplines in higher education, instructors require students to provide feedback on their peers' coursework. In contrast with peer assessment, which asks students to rate or grade the work of their peers, peer feedback is generally understood to be process-oriented and formative (Elizondo-Garcia et al., 2019; Kasch et al., 2022). As an instructional method, peer feedback can facilitate learning for both the giver and the receiver, improving critical-thinking and problem-solving skills, enhancing knowledge of the subject matter, and deepening understanding of a task or creation process (Baker, 2016; Cho and Cho, 2011; Cho and MacArthur, 2011; Nicol et al., 2014; Patchan and Schunn, 2015; Vickerman,

2009). Peer feedback also offers a practical way to ensure that students receive personalized responses to their works-in-progress, a task that may not always be feasible for instructors, particularly those with heavy teaching loads and large class sizes (Elizondo-Garcia et al., 2019; Zong et al., 2021).

Much of the literature on peer feedback in higher education focuses on its use with writing assignments, including literature reviews, concept-application papers, term papers, research reports, and evaluation essays in a range of disciplines (Baker, 2016; Gao et al., 2019; Kelly, 2015; Huisman et al., 2018; Simpson and Clifton, 2015; Zong et al., 2021). In the context of undergraduate- and graduate-level writing assignments, peer feedback is often called peer review (Baker, 2016; Min, 2016; Reddy et al., 2021; Simpson and Clifton, 2015). The present study adds to this body of literature by investigating the impact of practice on the amount and types of feedback that undergraduate students give over the course of four writing assignments. Furthermore, this study investigates whether, in addition to practice, providing peer reviewers with model feedback from a teaching assistant enhances their development as peer reviewers.

Previous studies of peer review have produced typologies for classifying peer-review comments and investigating their effects and efficacy (Cho and Cho, 2011; Cho and MacArthur, 2011; Cho et al., 2006; Kelly, 2015; Nelson and Schunn, 2009; Patchan et al., 2016). Scholars have analyzed both quantitative and qualitative features of peer feedback, including number, length, focus, scope, and function of comments (Elizondo-Garcia et al., 2019; Huisman et al., 2018; Patchan and Schunn, 2015; Patchan et al., 2016; Zong et al., 2021). The findings of these investigations have important implications for how peer review is taught and delivered. For example, comment length appears to be positively associated with helpfulness (Zong et al., 2021). Additionally, feedback that identifies the location of a problem seems to improve the writer's understanding of the comment (Nelson and Schunn, 2009). Moreover, the presence of a solution in a review comment, particularly when paired with a description of the problem, appears to increase the likelihood that the writer will implement that feedback—though some research suggests that explanations of problems can interfere with understanding (Elizondo-Garcia et al., 2019; Nelson and Schunn, 2009; Patchan et al., 2016).

The literature of peer review has also offered insights into how instructors should implement peer-feedback processes in their classrooms (Min, 2016; Reddy et al., 2021; Topping, 2009; van den Berg et al., 2006). For example, research has shown that instructors can help their students develop peer-review skills by modeling how to give feedback on sample papers (Min, 2016; Topping, 2009). Research also supports the use of feedback groups in which writers receive commentary from multiple

peers, a practice that gives writers “an opportunity to compare their fellow students’ remarks, and to determine their relevance” (van den Berg et al., 2006, p. 34–35).

A number of researchers have acknowledged that providing peer feedback is a cognitively demanding task (Carless and Boud, 2018; Deiglmayr, 2018; Gielen and De Wever, 2015; Min, 2016; Reddy et al., 2021). Particularly when student peer reviewers are unfamiliar with the conventions of academic or disciplinary writing, they struggle to identify the issues that are most worthy of a writer’s attention (Kelly, 2015). Student peer reviewers may have difficulty detecting higher-order concerns, such as problems with organization, counter-arguments, audience awareness, and evidence (Baker, 2016; Crossman and Kite, 2012; Gao et al., 2019; Kelly, 2015). Instead of devoting their attention to those important issues, they may focus on “polishing” or “fixing” surface-level problems, such as typos or errors in grammar, spelling, and punctuation (Baker, 2016; Crossman and Kite, 2012; Gao et al., 2019; Kelly, 2015). As Krishneel Reddy et al. (2021) have pointed out, student peer reviewers are likely to require practice in order to realize “the full benefits of peer review” (p. 826). However, the vast majority of studies of peer feedback analyze a single instance of peer review. That said, the few studies that do explore the effects of practicing peer review indicate that task repetition improves students’ ability to provide helpful feedback, particularly when paired with training or guidance from an instructor (Gielen and De Wever, 2015; Reddy et al., 2021; Zong et al., 2021). Research from Zheng Zong et al. (2021) reveals some of the mechanisms through which students improve as reviewers over the course of multiple rounds of peer review. Their study found that over the course of six rounds of peer review, students “were more likely to provide helpful feedback after they received helpful feedback” themselves (p. 981). However, the strongest predictor of feedback helpfulness was, in fact, the total length of commentary that the peer reviewer provided in the previous round. In other words, the more feedback a student provided in one round of peer review, the more likely they were to provide helpful feedback in the next round.

While these studies show the promise of practice in developing students’ peer-review skills, more research is needed to understand how peer feedback changes over the course of multiple rounds of practice, as well as how additional variables, such as feedback models, influence students’ evolution as peer reviewers. The present study investigated how peer-review comments changed over the course of four peer-feedback assignments that took place in the span of five weeks. We also imposed an experimental condition that offered insights into the effects of modeling feedback: students in the experimental condition were able to view feedback offered by a teaching assistant (TA) before providing their own commentary, while students in the control condition were not. We advanced four research questions:

1. Is there a relationship between practice and the amount of feedback peer reviewers provide?
2. Is there a relationship between practice and the types of feedback peer reviewers provide?
3. Are minimal, moderate, and heavy commenters more or less likely to offer certain types of feedback?
4. Does the availability of model feedback from a TA influence the amounts and types of feedback peer reviewers provide?

Method

Course Setting and Participants

Participants were undergraduate health sciences majors enrolled in an upper-division course on health-behavior change at a regional university in the southern United States. We collected data from three sections of this course in the same academic year, one in the fall semester and two in the spring semester. All sections were taught by the same instructor, covered the same content, and included the same assignments. IRB approval (no. 18-0254) was obtained prior to the course ending, and consent procedures were handled by a member of the research team who was not the instructor of the course. Of the 115 students enrolled in the three sections of the course, 70% ($n = 80$) participated in this study. Students who did not provide informed consent and students whose group members did not provide informed consent were excluded from the study.

Procedures

Group Selection. In this course, students worked in groups of five throughout the semester on a series of assignments that concluded with a group paper and presentation on a health intervention. Prior to assigning groups, the instructor administered a survey and gathered information about students' academic performance and habits. The instructor then matched students with similar GPAs and work styles. This approach was intended to limit conflict and social loafing. A total of sixteen groups (eight from each semester) were included in this study.

Article-Summary Assignment. To prepare for the high-stakes group assignments at the end of the semester, each student individually completed a two-to-three-page summary of a scholarly article relevant to their group's topic and to a particular theory of health-behavior change. For this assignment, students had to cite and summarize the content of the article, identify how it employed the theory, and reflect on their

own learning. Along with the assignment instructions, students were provided with a copy of the instructor's grading rubric. Students submitted drafts of their article summaries according to staggered deadlines—one student per group per week. Each week, the writers submitting the assignment were instructed to post their summaries and a copy of the original article in a group discussion board, where they received feedback from their fellow group members. After receiving feedback, the writers had four days to revise and resubmit their article summaries, at which time the instructor graded them. This cycle was repeated five times over five weeks.

Peer-Review Instructions. Students were required to provide feedback on each of their group members' article summaries, so they provided four peer reviews over the course of five weeks. Peer reviews had to be submitted within three days of the submission of the article summary. For each peer review, students had to read both the article summary and the scholarly article it summarized. They were then instructed to provide feedback, questions, and comments on their peer's writing using the comment function in Microsoft Word. Peer reviewers were explicitly told to focus on the effectiveness of the writer's paraphrasing, their accuracy in interpreting the article, and the correctness of their citation. After reviewing the article summary, students were required to upload the document with their commentary to the discussion board, along with a reflection on what they had learned in the peer-review process. Their feedback and reflections were visible to the writer and to the other group members.

Evaluation and Grading of Peer Reviews. As a way to hold reviewers accountable for providing high-quality feedback and writers accountable for implementing it, writers were asked to evaluate the depth and utility of the feedback they received. Along with their revised article summaries, writers had to submit a list ranking their peer reviewers from most helpful to least helpful, supported by descriptions of why each peer's feedback was or was not useful. In addition to the peer reviewers' comments themselves, the course instructor used these evaluations to inform the grades for the peer reviews. The instructor assigned grades according to the number of comments, the quality of comments, and evidence that the peer reviewer had thoroughly read the original article. If the reviewer lost any points, the instructor provided summative feedback.

Experimental Manipulation

In the two sections of the course that ran in the spring, a key change was made to the procedures outlined above. Unlike the fall section of the course, each spring section had an undergraduate teaching assistant (TA) who had previously taken the course with the instructor. These TAs were required to provide comments on each article summary within twenty-four hours of its submission and to post their review to the

relevant discussion board. The goal was to provide reviewers in the group with model feedback before they had to submit their own feedback.

Materials

After grades were submitted at the end of the semester, the instructor examined the signed consent forms, removed identifications from the peer-reviewed article summaries of those who had consented, and supplied them as PDF files to the research team. The vast majority of the eighty participants submitted all four of the required peer reviews. However, six of them submitted only three, and one submitted only two. Hence, a total of 312 article summaries with peer-review comments were provided for analysis.

Coding Process

All documents were imported into NVivo12 for qualitative coding. A total of four coders performed the coding process. Participants were randomly assigned to coders so that all peer-review comments provided by a given participant were coded by two different coders. Although interrater reliability (calculated using Cohen's kappa) was very high in all categories, a third "master coder" reviewed the codes assigned by the two initial coders and resolved any discrepancies in order to generate a final dataset to be used for analysis.

Because peer reviewers provided the vast majority of comments via the comment function in Microsoft Word, comments were already broken into discrete units. Coders were instructed to code each comment as a single unit and to select all codes applicable to the material in that comment. Thus, it was possible for a single comment to be coded according to multiple classifications, or "nodes," within a single category. In the rare case that a peer reviewer provided a comment using the track-changes function, the coder was instructed to highlight the entire sentence and code it as a single unit. Similarly, in the rare case that a peer reviewer provided end comments within the document, the coder was instructed to code the entire block of commentary as a single unit. This approach resulted in a total of 3,761 discrete comment units.

Coding Categories

The coding approach involved classifying the mode, scope, and topic of each peer-review comment. This coding scheme drew upon the work of previous scholars, in particular Cho and Cho (2011), Cho and MacArthur (2011), Cho et al. (2006),

Kelly (2015), Nelson and Schunn (2009), and Patchan et al. (2016). The coding categories are briefly described below and elaborated in Appendix A.

Feedback Mode. Feedback mode describes the function of the peer feedback. Every comment unit coded met at least one of these five mode classifications: problem-detecting, advising, editing, justifying, and praising. The feedback mode category had near perfect interrater reliability, with percentage agreement between pairs ranging from 98.12% ($\kappa = .94$) to 99.67% ($\kappa = .99$).

Feedback Scope. Feedback scope describes a comment's degree of focus, indicating whether it addresses a specific instance of a problem or achievement, a holistic trend, or something in between. Every comment unit coded met at least one of these three scope classifications: local, mid-range, and global. The feedback scope category had near perfect interrater reliability, with percentage agreement between pairs ranging from 97.75% ($\kappa = .93$) to 99.56% ($\kappa = .98$).

Feedback Topic. Feedback topic describes the subject matter of a comment. Every comment unit coded met at least one of these nine topic classifications: accuracy; citations; clarity, precision, and wording; grammar, mechanics, formatting, spelling, and typos; idea development; paraphrasing; purpose; structure, organization, and flow; and wordiness and concision. The feedback topic category had near perfect interrater reliability, with percentage agreement between pairs ranging from 97.85% ($\kappa = .93$) to 99.46% ($\kappa = .97$).

Descriptive and Quantitative Measures

The following descriptive and quantitative measures were also included in the dataset.

Round of Feedback. Each document in the dataset was labeled with a review number designating whether it was from the first, second, third, or fourth round of feedback provided by the peer reviewer. This made it possible to view the data as a time series.

Paper Number. Each document in the dataset was labeled with a number designating when it had been submitted and received comments. Any given document, for example, could have been the first, second, third, fourth, or fifth article summary within the group to receive a peer review.

Group Number. Each document in the dataset was labeled with a number designating the group to which its writer and reviewer belonged. This made it possible to explore group effects.

Semester. Each document in the dataset was labeled with an *F* or *S* designating whether it came from the fall or spring semester. This made it possible to compare the feedback from participants in the experimental condition (spring) with the feedback from those in the control group (fall).

Comment Count. Each document in the dataset was assigned a number designating how many discrete comment units were present in the document.

Word Count. Each document in the dataset was assigned a number designating the aggregate word count of all of the comments in the document. Because there was more variability and a greater range in word count than in comment count, we referenced this measure when we wanted to analyze the amount of feedback provided.

Commenter Designation. Each participant in the study was labeled as a minimal, moderate, or heavy commenter based on the average number of words they provided per review. Their designation was determined using percentiles: participants in the 1st–32nd percentile were labeled as minimal commenters, those in the 33rd–65th percentile as moderate commenters, and those in the 66th percentile and above as heavy commenters.

Data Analyses

Research Question 1: The Relationship between Practice and Amount of Feedback Provided. In order to determine if there was a relationship between practice and the amount of feedback peer reviewers provided, we calculated average word counts for each round of feedback and ran a linear regression, controlling for round of feedback and semester, with word count as the dependent variable.

Research Question 2: The Relationship between Practice and Types of Feedback Provided. To explore any changes in feedback mode, scope, and topic over the course of the peer-reviews, we calculated the total number of comments coded at each node for each of the four rounds of feedback. We then ran an analysis of variance (ANOVA) to determine whether the number of comments coded at any of the feedback mode, scope, and topic nodes changed according to round of feedback.

Research Question 3: The Relationship between Commenter Designation and Types of Feedback Provided. Because the total number of comments varied by round, comment counts were not used as the basis for comparing the types of comments made by minimal, moderate, and heavy commenters. Instead, each reviewer received a *Y* (yes) or *N* (no) for each feedback type, indicating whether they had given any commentary that fell into each of the seventeen available feedback classifications under

mode, scope, and topic. Then, we used a chi-squared test to determine whether minimal, moderate, or heavy commenters were more or less likely than the others to provide feedback of each type.

Research Question 4: The Influence of Model Feedback on Amount and Types of Feedback Provided. We investigated whether the experimental condition—presenting students with model feedback, provided by a TA during every round of peer review—influenced the amount of feedback peer reviewers provided. We used a T-test to compare the average word counts given by reviewers during each semester. We also explored the influence of the experimental condition on feedback mode, scope, and topic. Looking at the proportion of reviewers who had or had not given any commentary within each of the seventeen available feedback classifications, we used a chi-squared test to determine whether participants in the spring semester were more or less likely than those in the fall semester to provide particular types of feedback.

Results

Research Question 1: The Relationship between Practice and Amount of Feedback Provided

Students saw a statistically significant increase ($p = 0.005$) in average word count per comment set between their first ($n = 171$) and second ($n = 237$) rounds of peer review (see Table 1).

Table 1.
Changes in average word count over four rounds of feedback.

Round of peer review	Average word count	Degree of change (β)	95% confidence interval	p -value
Round 1	171			
Round 2	237	66	17, 95	0.005
Round 3	203	26	-13, 66	0.2
Round 4	186	11	-28, 51	0.6

Research Question 2: The Relationship between Practice and Types of Feedback Provided

Peer reviewers also shifted their commenting strategies and the foci of their feedback over the four rounds of peer review (see Table 2). Changes occurred in each of the major coding categories, though not at every node. In terms of feedback mode, peer reviewers were more likely to offer praise in later rounds of feedback ($F = 7.646$,

$p < 0.001$), particularly in the last two rounds. While a greater proportion of comments contained editing during the first round and justifying during the second round, these differences were not statistically significant. Additionally, peer reviewers increased the scope of their comments over time. Students were more likely to offer mid-range comments after the first round of feedback ($F = 9.33, p < 0.001$) and global comments in the latter two rounds of feedback ($F = 3.17, p = 0.025$). Finally, peer reviewers were more likely to focus on idea development after the first round of feedback ($F = 5.09, p = 0.002$).

Table 2.
Number of comments, per person mean, coded at feedback mode, scope, and topic over four rounds of feedback.

	Round 1		Round 2		Round 3		Round 4	
	Number	Mean	Number	Mean	Number	Mean	Number	Mean
Feedback mode								
Advising	310	4.01	422	5.29	388	4.85	354	4.75
Editing	425	5.54	431	5.32	403	5.13	388	5.2
Justifying	118	1.56	175	2.15	131	1.66	125	1.69
Praising	88	1.14	149	1.87	180	2.23	167	2.25
Problem-detecting	188	2.48	256	3.19	188	2.39	189	2.55
Feedback scope								
Local	681	8.87	742	9.2	682	8.66	658	8.83
Mid-range	143	1.87	261	3.29	272	3.46	248	3.32
Global	32	0.41	33	0.43	53	0.61	49	0.67
Feedback topic								
Accuracy	39	0.51	50	0.62	35	0.44	38	0.51
Citations	63	0.84	73	0.9	71	0.9	61	0.81
Clarity, precision, wording	256	3.34	287	3.57	258	3.22	229	3.08
Grammar, mechanics . . .	215	2.81	192	2.35	205	2.53	235	3.15
Idea development	187	2.43	293	3.67	287	3.57	266	3.56
Paraphrasing	61	0.77	93	1.18	76	0.96	66	0.89
Purpose	79	1.01	107	1.35	125	1.59	102	1.39
Structure, organization, flow	52	0.7	86	1.08	82	1.04	72	0.97
Wordiness, concision	72	0.92	113	1.42	87	1.11	80	1.08

Research Question 3: The Relationship between Commenter Designation and Types of Feedback Provided

The percentages of comment units coded at each node were remarkably similar among minimal, moderate, and heavy commenters, but there were a few notable trends within these designations (see Table 3). The more comments peer reviewers provided, the more likely they were to advise ($\chi^2 = 10.58, p = 0.005$), to

justify ($\chi^2 = 25.78, p < 0.001$), and to offer comments that were mid-range in scope ($\chi^2 = 8.08, p < 0.018$). Additionally, reviewers who wrote more commentary were more likely to comment on wordiness and concision ($\chi^2 = 23.3, p < 0.001$). Finally, heavy commenters were more likely than minimal or moderate commenters to comment on structure, organization, and flow ($\chi^2 = 20.8, p < 0.001$).

Table 3.
Number and proportion of comments, by commenter designation, coded at feedback mode, scope, and topic.

	Minimal		Moderate		Heavy	
	Number	Proportion	Number	Proportion	Number	Proportion
Feedback mode						
Advising	280	32.4%	478	39.6%	716	42.3%
Editing	365	42.3%	525	43.5%	763	45.1%
Justifying	77	8.9%	161	13.3%	313	18.5%
Praising	170	19.7%	182	15.1%	231	13.6%
Problem-detecting	176	20.4%	267	22.1%	385	22.8%
Feedback scope						
Local	648	75.1%	891	73.8%	1235	73.0%
Mid-range	197	22.8%	291	24.1%	442	26.1%
Global	44	5.1%	59	4.9%	61	3.6%
Feedback topic						
Accuracy	28	3.2%	49	4.1%	85	5.0%
Citations	82	9.5%	94	7.8%	93	5.5%
Clarity, precision, wording	220	25.5%	323	26.8%	488	28.9%
Grammar, mechanics . . .	224	26.0%	250	20.7%	370	21.9%
Idea development	230	26.7%	356	29.5%	445	26.3%
Paraphrasing	60	7.0%	107	8.9%	130	7.7%
Purpose	94	10.9%	123	10.2%	200	11.8%
Structure, organization, flow	59	6.8%	66	5.5%	170	10.1%
Wordiness, concision	51	5.9%	108	8.9%	195	11.5%

Research Question 4: The Influence of Model Feedback on Amount and Types of Feedback Provided

Amount of Feedback. This study imposed an experimental condition in which half of the peer reviewers in the study (those who took the course during the spring semester) had the opportunity to view model feedback from a TA before submitting their own peer-review comments. On average, peer reviewers in the experimental condition generated about fifty more words per comment set than peer reviewers in the control condition. Reviewers in the fall cohort wrote an average of 175.4 words per round of feedback, while reviewers in the spring cohort wrote an average of 225.3 words per round of feedback ($t = -3.492, df = 309.42, p = 0.0005$). In fact, in every

round of feedback, reviewers in the spring semester produced average word counts that were higher than those produced by reviewers in the fall semester (see Figure 1). That said, after the boost in average word count that occurred in the second round of feedback, word counts in the experimental condition dropped in the subsequent two rounds, even dipping beneath the word count of round one in the final round. On the other hand, students in the control group only saw a dip in word count after the second round of feedback.

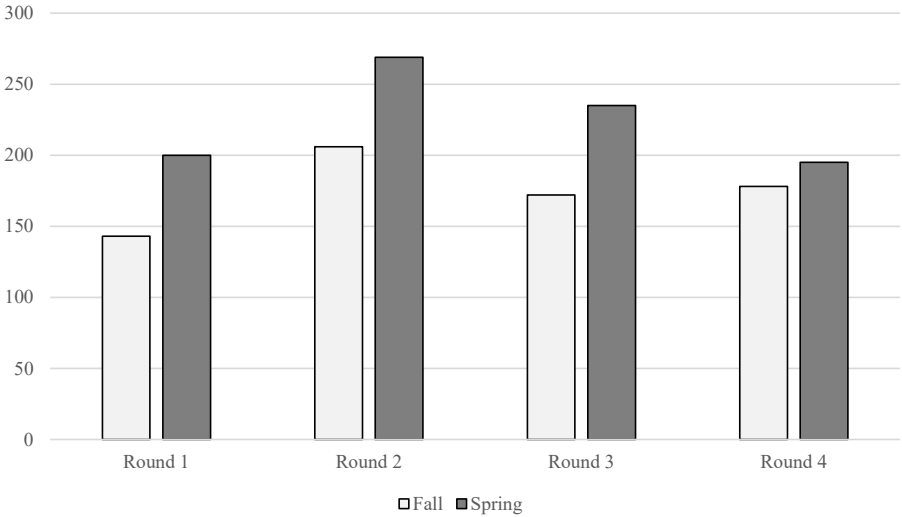


Figure 1. Average word count per participant, per round of feedback, and by semester.

Types of Feedback. Additionally, peer reviewers in the control and experimental conditions appeared to favor different types of feedback (see Table 4). Comments from reviewers in the spring semester were more likely to include problem-detecting ($\chi^2 = 6.74, p < 0.001$) and to focus on mid-range concerns ($\chi^2 = 12.92, p < 0.001$). Additionally, idea development ($\chi^2 = 7.03, p = 0.008$), paraphrasing ($\chi^2 = 19.49, p < 0.001$), purpose ($\chi^2 = 9.38, p = 0.002$), and wordiness and concision ($\chi^2 = 14.83, p < 0.001$) were more likely to be topics of commentary for reviewers in the experimental condition. There were also some trends in the control condition that were marginally significant. For example, peer reviewers who took the course in the fall semester appeared more likely to address citations ($\chi^2 = 2.73, p = 0.063$) and surface-level concerns, such as grammar and spelling ($\chi^2 = 2.21, p = 0.087$), in their comments.

Table 4.
 Number and proportion of comments, by semester, coded at feedback mode, scope, and topic.

	Fall semester (control group)		Spring semester (experimental group)	
	Number	Proportion	Number	Proportion
Feedback mode				
Advising	686	36.5%	788	41.8%
Editing	878	46.7%	775	41.1%
Justifying	271	14.4%	280	14.9%
Praising	285	15.2%	298	15.8%
Problem-detecting	373	19.9%	455	24.2%
Feedback scope				
Local	1456	77.6%	1318	70.0%
Mid-range	408	21.7%	522	27.7%
Global	76	4.0%	88	4.7%
Feedback topic				
Accuracy	71	37.8%	91	48.3%
Citations	140	7.5%	129	6.8%
Clarity, precision, wording	549	29.2%	482	25.6%
Grammar, mechanics . . .	508	27.1%	336	17.8%
Idea development	473	25.2%	558	29.6%
Paraphrasing	109	5.8%	188	10.0%
Purpose	183	9.7%	234	12.4%
Structure, organization, flow	128	6.8%	167	8.9%
Wordiness, concision	139	7.4%	215	11.4%

Discussion and Conclusion

The present study had two aims: (1) to explore the impact of practice on the amount and types of feedback that peer reviewers give and (2) to investigate whether providing peer reviewers with model feedback in addition to practice promotes their development as peer reviewers. Here, we discuss how the findings of our study advanced those aims.

Amount of Feedback

Our data revealed two important trends regarding the amount of feedback (measured in word count) that peer reviewers provided over the sequence of four peer-review sessions. First, the peer reviewers in our study wrote significantly more commentary in their second round of feedback than in their first. Next, peer reviewers in the experimental condition—that is, those who had the opportunity to view a model set of comments written by a TA before submitting their own feedback—wrote an average of 22% more words per peer review than those in the control group.

While it is impossible to separate task repetition from other influences, it seems likely that the learning gains enabled by practice played some role in the increase in word count that occurred during the second round of feedback. A single round of peer review may have been sufficient to allow reviewers to “automate” familiar aspects of the task—for example, using the comment function in Microsoft Word, toggling between a peer’s article summary and the scholarly article, or revisiting the assignment description to review criteria. As a result, peer reviewers may have been able to dedicate more cognitive resources to generating feedback in the second round. This explanation aligns with educational psychologists’ current understanding of the relationship between practice and learning as informed by cognitive load theory (Sweller et al., 2019).

It is worth considering these findings in conjunction with studies that have explored the effects of practice on feedback quality. Zong et al. (2021) have found that the amount of commentary a peer reviewer provides is a powerful predictor of feedback helpfulness in the subsequent round of peer review. This finding suggests that many reviewers in our study provided their most helpful commentary in their third round of feedback, as reviewers generally wrote the most feedback in round two. Zong et al.’s findings also lead us to believe that, at least in the latter three feedback rounds, peer reviewers in the experimental condition provided more helpful commentary than those in the control group, who by comparison wrote fewer words in every round of peer review. This inference is bolstered by the findings of Gielen and De Wever’s (2015) study, which indicates that the more guidance students receive about how to provide feedback, the higher quality their reviews become. We think it likely that having access to model feedback gave peer reviewers an enriched task representation; that is, reviewing model feedback improved their understanding of the expectations regarding peer-review comments and therefore their ability to fulfill those expectations.

Types of Feedback

As peer reviewers in the present study gained more practice giving peer feedback, numerous shifts occurred in the types of feedback they provided. We think it is helpful to view these findings through the lens of previous research, which indicates that student reviewers often favor “fixing” superficial problems that are local in scope at the expense of addressing recurring problems and higher-order concerns that more broadly and profoundly affect the paper (Crossman and Kite, 2012; Gao et al., 2019; Kelly, 2015). This tendency, which may be a result of the cognitive demands of peer review, demonstrates students’ inclination to focus on simpler or more familiar problems with clear solutions as opposed to issues of greater complexity.

Our results did document a strong focus on local concerns. In all of the feedback rounds, the majority of comments (78.3%) given by peer reviewers were local in scope. However, as peer reviewers got more practice, they tended to expand the scope of their comments to include problems (or successes) that affected multiple sentences, whole sections, or even the entire paper. Additionally, the proportion of comments focused on the clarity and precision of wording dropped in each successive round, while the proportion of comments focused on idea development increased. This trend suggests that practice facilitated movement from lower-order concerns toward higher-order concerns, presumably reflecting increased motivation or ability to engage in the cognitively demanding aspects of peer review. A possible explanation for this shift is that the more participants practiced providing feedback, the better they understood the aims and expectations of the assignments, enabling them to identify larger-scale problems in their peers' writing. It is also possible that the drafts submitted in later rounds were more clearly and precisely written than those in earlier rounds—potentially because the writers had benefited from the process of reviewing peers' drafts in previous rounds (Cho and MacArthur, 2011). That said, a significant portion of comments in all rounds, including the last one, devoted attention to grammar, mechanics, formatting, spelling, and typos. Additionally, some higher-order concerns, such as purpose and structure, only received slightly more attention in later rounds of feedback.

In our study, the example set by model feedback from a TA appeared to increase the likelihood that peer reviewers would comment on higher-order concerns. Peer reviewers in the control group were more likely than those in the experimental condition to comment on grammatical concerns and citations. Meanwhile, those in the experimental condition were more likely to focus on mid-range concerns and to comment on accuracy, idea development, purpose, and structure, organization, and flow. Similarly, peer reviewers who were designated heavy commenters (those in the 66th percentile and above for aggregate word count) were more likely to offer mid-range commentary and feedback on structure, organization, and flow than peers who were minimal or moderate commenters. Furthermore, peer reviewers of higher commenter designations were more likely to advise and to justify in their feedback—suggesting that they moved beyond “fixing” to offering more complex forms of commentary, such as instructions and explanations. These findings align with our understanding that higher word counts correspond with higher-quality commentary (Zong et al., 2021).

We wish to highlight a final finding regarding types of commentary: peer reviewers were most likely to offer praise in the third and fourth rounds of feedback. There are a number of possible explanations for this trend. First, it seems likely that the article summaries reviewed in these rounds were objectively better because their writers

had applied insights gleaned from earlier rounds of peer review. Second, peer reviewers in the later rounds may have been more sensitive to writers' desire for praise, having recently received peer feedback on their own article summaries. A third possibility is that practice improved peer reviewers' ability to perceive and articulate what was working well in their peers' papers. Whether or not this finding is the direct result of practice, it has interesting implications. Patchan and Schunn (2015) have proposed that the process of *giving* praise may help peer reviewers better understand or discover successful writing strategies. Thus, the presence of praise is a desirable trait in peer review, but our results indicate that it is less likely to happen without practice.

Practical Implications, Limitations, and Directions for Future Research

Our study's findings suggest that giving students the opportunity to practice providing feedback throughout several rounds of peer review may help them generate more and higher-quality feedback, especially when paired with training in the form of reviewing model feedback. While one round of practice produced some significant changes in the amount and types of commentary given, three or more rounds may further increase the likelihood that peer reviewers offer feedback related to higher-order and global concerns.

As is often the case with studies conducted in real-world educational settings, it is impossible to isolate the impact of practice and of feedback modeling from other potential influences. For example, peer reviewers in this study may have been motivated to change their feedback practices after receiving their professor's summative feedback and grade or after seeing commentary written by other group members. Additionally, this study did not account for students' dispositions and academic behaviors, which could have influenced their feedback practices. Despite these limitations, as Huisman et al. (2018) have pointed out, "the authenticity of the learning context" is vital "in determining the practical value of the research findings" (p. 964).

It is worth noting that our study did not assess feedback helpfulness; instead, we used word count as a corollary for helpfulness in subsequent feedback rounds, relying on previous research from Zong et al. (2021). Similarly, we did not explore the uptake of feedback or how the experience of peer review affected students' writing, as this would have involved qualitative analysis of student's writing, which we chose not to pursue.

One potential downside of the peer-review structure employed in this study is that students whose work was reviewed in earlier rounds had more opportunity to integrate what they had learned into their own commentary, while students whose work was reviewed in the final round were deprived of this opportunity. This inequitable distribution of learning opportunities could be mitigated by having students assume the role of writer multiple times (for example, by also submitting a revision

for peer review) in addition to acting as reviewer multiple times over the course of the peer-review sequence.

Because our study took place in a specific context—namely, an upper-division health sciences course—its findings may not be generalizable to other contexts. More research is needed to determine whether similar patterns emerge when students engage in a series of peer reviews in other disciplines and for other assignments. Future research might explore whether there are differences in feedback uptake following successive rounds of peer review. Additionally, researchers might explore the effects of peer-review series with structures different from the one we studied—for example, a series of three peer reviews in which students serve as both writers and reviewers in every round. Because research has underscored the important role that providing feedback can play in developing students as writers, it would be worthwhile to study whether students' growth as peer reviewers over a sequence of peer reviews correlates with their growth as writers over a sequence of writing assignments.

References

- Baker, K. M. (2016). Peer review as a strategy for improving students' writing process. *Active Learning in Higher Education*, 17(3), 179–192. <https://doi.org/10.1177/1469787416654794>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science*, 39(5), 629–643. <https://doi.org/10.1007/s11251-010-9146-1>
- Cho, K., & MacArthur, C. (2011). Learning by reviewing. *Journal of Educational Psychology*, 103(1), 73–84. <https://doi.org/10.1037/a0021950>
- Cho, K., Schunn, C. D., & Charney, D. (2006). Commenting on writing: Typology and Perceived helpfulness of comments from novice peer reviewers and subject matter experts. *Written Communication*, 23(3), 260–294. <https://doi.org/10.1177/0741088306289261>
- 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. <https://doi.org/10.1177/1469787412452980>
- Deiglmayr, A. (2018). Instructional scaffolds for learning from formative peer assessment: Effects of core task, peer feedback, and dialogue. *European Journal of Psychology of Education*, 33(1), 185–198. <https://doi.org/10.1007/s10212-017-0355-8>
- Elizondo-Garcia, J., Schunn, C., & Gallardo, K. (2019). Quality of peer feedback in relation to instructional design: A comparative study in energy and sustainability

- MOOCs. *International Journal of Instruction*, 12(1), 1025–1140. <https://doi.org/10.29333/iji.2019.12166a>
- Gao, Y., Schunn, C. D. D., & Yu, Q. (2019). The alignment of written peer feedback with draft problems and its impact on revision in peer assessment. *Assessment & Evaluation in Higher Education*, 44(2), 294–308. <https://doi.org/10.1080/02602938.2018.1499075>
- Gielen, M., & De Wever, B. (2015). Structuring the peer assessment process: A multilevel approach for the impact on product improvement and peer feedback quality. *Journal of Computer Assisted Learning*, 31(5), 435–449. <https://doi.org/10.1111/jcal.12096>
- Huisman, B., Saab, N., van Driel, J., & van den Broek, P. (2018). Peer feedback on academic writing: Undergraduate students' peer feedback role, peer feedback perceptions and essay performance. *Assessment & Evaluation in Higher Education*, 43(6), 955–968. <https://doi.org/10.1080/02602938.2018.1424318>
- Kasch, J., Van Rosmalen, P., Henderikx, M., & Kalz, M. (2022). The factor structure of the peer-feedback orientation scale (PFOS): Toward a measure for assessing students' peer-feedback dispositions. *Assessment & Evaluation in Higher Education*, 47(1), 15–28. <https://doi.org/10.1080/02602938.2021.1893650>
- Kelly, L. (2015). Effectiveness of guided peer review of student essays in a large undergraduate biology course. *International Journal of Teaching and Learning in Higher Education*, 27(1), 56–68.
- Min, H. (2016). Effect of teacher modeling and feedback on EFL students' peer review skills in peer review training. *Journal of Second Language Writing*, 31, 43–57. <https://doi.org/10.1016/j.jslw.2016.01.004>
- 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. <https://doi.org/10.1007/s11251-008-9053-x>
- Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education*, 39(1), 102–122. <https://doi.org/10.1080/02602938.2013.795518>
- Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: How students respond to peers' texts of varying quality. *Instructional Science*, 43(5), 591–614. <https://doi.org/10.1007/s11251-015-9353-x>
- Patchan, M. M., Schunn, C. D., & Correnti, R. J. (2016). The nature of feedback: How peer feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology*, 108(8), 1098–1120. <https://doi.org/10.1037/edu0000103>
- Reddy, K., Harland, T., Wass, R., & Wald, N. (2021). Student peer review as a process of knowledge creation through dialogue. *Higher Education Research & Development*, 40(4), 825–837. <https://doi.org/10.1080/07294360.2020.1781797>

- Simpson, G., & Clifton, J. (2015). Assessing postgraduate student perceptions and measures of learning in a peer review feedback process. *Assessment & Evaluation in Higher Education*, 41(4), 501–514. <https://doi.org/10.1080/02602938.2015.1026874>
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31, 261–292. <https://doi.org/10.1007/s10648-019-09465-5>
- Topping, K. J. (2009). Peer assessment. *Theory Into Practice*, 48(1), 20–27. <https://doi.org/10.1080/00405840802577569>
- van den Berg, I., Admiraal, W., & Pilot, A. (2006). Peer assessment in university teaching: Evaluating seven course designs. *Assessment & Evaluation in Higher Education*, 31(1), 19–36. <https://doi.org/10.1080/02602930500262346>
- Vickerman, P. (2009). Student perspectives on formative peer assessment: An attempt to deepen learning? *Assessment & Evaluation in Higher Education*, 34(2), 221–230. <https://doi.org/10.1080/02602930801955986>
- Zong, Z., Schunn, C. D., & Wang, Y. (2021). Learning to improve the quality peer feedback through experience with peer feedback. *Assessment & Evaluation in Higher Education*, 46(6), 973–992. <https://doi.org/10.1080/02602938.2020.1833179>

Appendix A

Table 5.
Coding categories with definitions and examples.

Classification	Definition	Example
Feedback mode		
Advising	Gives general direction or options for revision in response to a problem, concern, or error	"Add in some of the findings of the experiment here."
Editing	Supplies the actual deletion, punctuation, or language needed to resolve a problem	"Comma instead of semicolon"
Justifying	Justifies advice or an edit by describing the reviewer's reasoning or the intended outcome	"Consider including a brief definition of meaning. . . . What is a longitudinal cohort study? It would help in readers' comprehension."
Praising	Highlights something the author is doing well	"I agree with this statement, and I think it is an important take away from this journal article."
Problem-detecting	Indicates a problem, concern, or error	"The sentence is a little wordy."
Feedback scope		
Local	Narrowly focuses on a specific problem or achievement, typically a word or something sentence-level	"Use a different word here."
Mid-range	Discusses a problem or achievement that occurs multiple times or encompasses multiple variables, sentences, or paragraphs but does not apply to the entire composition	"I couldn't really find the main results and findings of this study within your paper. Maybe reference the Results section in the article and expand more on what was concluded."
Global	Holistically describes the product, highlighting problems or achievements that affect most or all of the paper	"Overall, I think your draft is really well put organized and has very few grammatical errors. I think you did a great job at paraphrasing and not taking too much detail from the article. It was easy for me to read the paper (meaning it had good flow)."
Feedback topic		
Accuracy	Makes comments and/or suggestions regarding the accuracy of the writer's interpretation of the article's purpose, methods, results, theories, etc.	"Although this might be true, I did not read any supporting information from the study for this claim."
Citations	Makes comments or suggestions regarding whether or not the citation follows APA guidelines	"Other than missing the page numbers, the citation looks good."
Clarity, precision, wording	Identifies content that generates confusion and/or suggests additions, substitutions, or changes that will increase clarity, precision, or effectiveness of language; alternatively, praises writer for clear, precise, or direct writing	"This part of the sentence sounds a bit off. I would maybe change it to 'the individuals were chosen purely based on geographical location.'"
Grammar, mechanics, formatting, spelling, typos	Makes comments or suggestions related to the rules of written language and the presentation of textual and visual elements	"Good sentence but watch your tense. You start in past tense, switch to present, and then switch back to past tense."

Idea development	Makes comments or suggestions regarding the ideas in the article summary; typically, identifies a need to add information, explanation, detail, or examples, but may offer praise for insights	“Great point! Maybe here you can elaborate on how the TPB was used in this study and why it was the most appropriate and accurate theory to use!”
Paraphrasing	Comments on the writer’s success or failure in paraphrasing content from the original article by employing different wording and sentence structures	“This sentence appears to be quite similar to the sentence on page 879. Maybe try and paraphrase it using a few more of your own words!”
Purpose	Gives feedback regarding the purpose of the assignment, paragraph, or type of writing (in this case, an article summary); may comment on how well or poorly the writer conformed to conventions, expectations, or instructions; may also comment on the general efficacy of the prose	“She was very critical on my conclusion when it was written like this. Specify exactly what you learned about this theory.”
Structure, organization, flow	Makes comments or suggestions about the order of ideas, flow of information, or paragraphing	“I would try to find a way to combine these paragraphs.”
Wordiness, concision	Identifies wordiness and/or suggests deletions or changes that will result in more concise phrases or sentences; alternatively, praises writer for concision	“To remain concise, you can eliminate these sentences.”