American Educational Research Journal
Month XXXX, Vol. XX, No. X, pp. 1–35
DOI: 10.3102/0002831220945266
Article reuse guidelines: sagepub.com/journals-permissions
© 2020 AERA. http://aerj.aera.net

The Effects of Providing and Receiving Peer Feedback on Writing Performance and Learning of Secondary School Students

Yong Wu
Christian D. Schunn
University of Pittsburgh

Research has shown that engaging students in peer feedback can help students revise documents and improve their writing skills. But the mechanistic pathways by which skills develop have remained untested: Does receiving and providing feedback lead to learning because it produces more extensive revision behavior or is such immediate implementation of feedback unnecessary? These pathways were tested through analyses of the relationships between feedback provided and received, feedback implemented and overall revisions, and improved writing quality in a new article. Overall, the number of revisions predicted growth in writing ability, and both amount of received and provided feedback were associated with being more likely to make revisions. However, providing feedback was also directly related to growth in writing ability.

Keywords: feedback implementation, learning, peer feedback, peer review, revisions

Feedback generally plays an important role in helping students learn (Gielen & De Wever, 2015; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Topping, 2009). However, it is difficult for teachers to provide timely feedback to large numbers of students, especially in writing (Applebee &

Yong Wu is a researcher at the Learning and Research Development Center at the University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA 15260; *e-mail: yongwu@pitt.edu*. She specializes in EFL/ESL writing, responding to student writing, and technology and writing.

Christian D. Schunn is a professor of psychology and of learning sciences and policy as well as a senior scientist at the Learning Research & Development Center at the University of Pittsburgh. Originally trained as a cognitive psychologist, his current research is interdisciplinary in perspective and extends across four core areas: STEM reasoning and learning, web-based peer interaction and instruction, neuroscience of complex learning, and engagement and learning.

Langer, 2011), because providing feedback requires a considerable amount of time and effort. The workload problem is most strongly felt when teachers have many sections of a class or large classes. In order to reduce teacher workload and make it possible for students to receive detailed and immediate feedback, many teachers look to the use of peer review as an alternative or additional method for providing feedback. More interestingly, prior studies suggest that students actually benefit more from receiving multipeer feedback than feedback from a single teacher (K. Cho & Schunn, 2007). Further, when students actively engage in the reciprocal process of peer review, their concepts and knowledge about both writing and subject matter are further developed (Jonassen et al., 1995), and learning is enhanced (Falchikov, 2001). Peer review is also beneficial for developing students' audience awareness, fostering social skills such as learning how to provide and accept critical comments, justifying one's own position, and declining nonproductive suggestions (Topping, 2009). Because of these advantages, peer review is recommended as a high leverage practice for writing instruction (S. Graham & Perin, 2007a; Topping, 2009).

However, a number of concerns about peer review have been raised. First, some students worry about harming interpersonal relationships or the negative effects of power relations between students on feedback content (Topping, 2009). Such problems can be solved by anonymous peer review, which decreases bias and enables students to evaluate others' work in a nonthreatening environment (Lin et al., 2001; Patchan et al., 2018). Second, both students and teachers worry about peers having sufficient expertise to provide valid feedback (Lin et al., 2001). But again, well-structured rubrics and incentives embedded in online peer review systems, especially for honest and effortful participation in the review process, can also produce feedback with high reliability and validity (Patchan et al., 2018; Pearce et al., 2009; Sadler & Good, 2006; Schunn et al., 2016).

Although a number of challenges to and benefits of peer review have been uncovered over the past 20 years, relatively little is known about how peer review produces learning. Three challenges are responsible for this surprising lack of knowledge about a practice that is pervasive across grade levels, disciplines, and countries. First, most studies (e.g., Beason, 1993; K. Cho & MacArthur, 2010; Nelson & Schunn, 2009; Patchan et al., 2016; Tsui & Ng, 2000) have looked at *performance* (improvements in or final quality of a document after peer review) rather than at *learning* (whether writing in new tasks improves). It is likely assumed that whatever benefits a given document receiving feedback should also eventually benefit learning, but there are theoretical reasons to doubt this assumption (reviewed below) and therefore this assumption must be tested. In the present study, we look at both performance and learning and their interrelationship. Second, peer review centrally involves two different components that could be responsible for learning: receiving feedback from peers versus

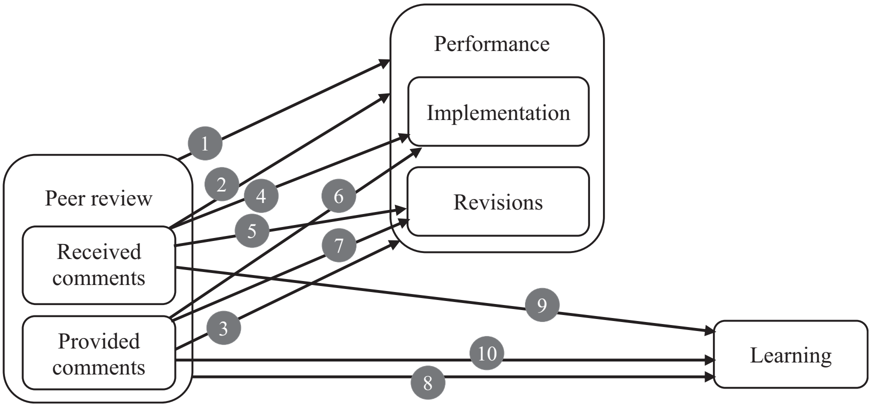

Figure 1. **A graphical summary of previously studied associations between the two main components of peer review (alone or together), the two ways of studying performance effects (alone or together), and learning effects.**

providing feedback to peers (Tsivitanidou et al., 2011). However, prior investigations of each component alone involved somewhat artificial experimental conditions (e.g., K. Cho & MacArthur, 2011; Lundstrom & Baker, 2009). In the present study, we test their separate effects on performance and learning using a statistical regression approach. Third, it is difficult to robustly and systematically measure the details of the peer review process in the context of a data set that is large enough to robustly show evidence of learning since learning is typically a slow process with small effects from a single learning experience (e.g., one round of peer review). In the present study, we examine student reviewing (receiving and providing peer feedback), their writing performance (peer feedback implementation and general revisions), and writing in a new task across multiple schools and multiple classes per school.

In the next section, we first review prior studies examining effects of peer review on performance (including implementing received feedback and making revisions more generally) and then review the literature examining learning effects. We focus the review on studies of the specific effects of receiving or providing given the focus of the current study. This review of past studies is shown graphically in Figure 1. To address the gaps in the literature, we then present a new study that simultaneously examines performance and learning in a large data set to robustly uncover the relationship of receiving and providing feedback on writing performance and learning, where writing performance includes revisions in response to specific feedback and general revisions, and learning refers to the general improvement from one writing task to a subsequent one (e.g., K. Cho & MacArthur, 2011; Van Beuningen et al., 2012).

Studies of Peer Review and Writing Performance

At the most general level, several studies have documented overall improvements in document quality as a result of overall peer review activities (Path 1 in Figure 1; Y. H. Cho & Cho, 2011; Gielen & DeWever, 2015; Zhang et al., 2017). Several studies have then also shown that these general document improvements appear to be the result of both receiving comments (Path 2; Huisman et al., 2018; Lu & Law, 2012; Paulus, 1999; M. Yang et al., 2006) and providing comments (Path 3; Huisman et al., 2018; Lu & Law, 2012; Philippakos & MacArthur, 2016). However, these studies leave open whether the document improvements hide the nature of the behaviors of the students: Do the changes in the document come from implementing received feedback or more generally engaging in revision work on the document?

Looking more specifically at effects on implementing received feedback or making revisions, many studies have not-surprisingly found that students who receive more feedback will then implement more changes based on this feedback (Path 4; Beason, 1993; Berggren, 2015; Nelson & Schunn, 2009; Patchan et al., 2016; Tsui & Ng, 2000; Wichmann et al., 2018). Peer feedback recipients have also been found to make more revisions not necessarily tied to particular feedback (Path 5; K. Cho & MacArthur, 2010). Turning to providing effects, providing feedback has also been linked to implementing more of the received feedback (Path 6; Berggren, 2015) and more revisions overall (Path 7; Y. H. Cho & Cho, 2011). These studies involved a mixture of correlational, survey, interviews, and experimental designs. Across these studies, there is pretty good evidence supporting the benefits of providing and receiving feedback on improving the documents receiving feedback. Further, the amount of benefit of providing versus receiving appears to be roughly similar when the effects are directly contrasted (Huisman et al., 2018).

Importantly, several of these studies found that the results varied by language level (higher level like argument, evidence, genre awareness vs. lower level like grammar and spelling). For example, Y. H. Cho and Cho (2011) and Berggren (2015) found the main benefit of providing feedback was on higher level aspects of writing, not on lower level aspects. Receiving feedback was sometimes found to produce improvements for higher levels (Lundstrom & Baker, 2009), sometimes for lower levels (Huisman et al., 2018; Wichmann et al., 2018), but sometimes showed no benefit at all on document quality (Y. H. Cho & Cho, 2011). It may be that students need help in making sense of all the feedback they receive from peers (Wichmann et al., 2018). Nicol et al. (2014) suggested that receiving feedback helps students focus more on areas that need improvement and develop a readers' perspective, whereas providing feedback enables students to think critically, apply criteria, and reflect on their own work. Further, students appear to benefit from both reviewing weak documents that have the same problems they made in their own document as well as from reviewing stronger documents that act as models for how to improve (Schunn et al., 2016). In other words, reviewing could help with both higher and lower levels of writing, but it depends on the focus of the criteria and the match to the areas needing improvement.

Studies of Peer Review and Learning to Write

While most studies discuss their findings of improved documents as being highly relevant to learning, only a few studies have looked directly at learning, defined here as writing better first drafts of later documents. Some work has examined the overall effect of peer review on student learning (Path 8; Nicol et al., 2014; Schunn et al., 2016). Two studies have observed benefits of receiving feedback on later writing ability (Path 9; Lundstrom & Baker, 2009; Wichmann et al., 2018) and three studies have observed benefits of providing feedback on later writing ability (Path 10; K. Cho & MacArthur, 2011; Lundstrom & Baker, 2009; Philippakos & MacArthur, 2016). Again, larger benefits were observed for higher level than lower level aspects of writing (Lundstrom & Baker, 2009). Interestingly, larger gains were observed in the students who were initially the weaker writers (Lundstrom & Baker, 2009).

Because there are relatively few studies of learning, it is important to note weaknesses in the evidence thus far. Lundstrom and Baker (2009) divided students into receivers (received feedback from peers, did not themselves do reviewing, and used received feedback to revise papers) and providers (reviewed, did not receive feedback, and did not revise papers). Here receivers revised others' writing rather than their own writing based on peer feedback which is somewhat artificial. In the other two studies of providing effects (K. Cho & MacArthur, 2011; Philippakos & MacArthur, 2016) students provided feedback without receiving any feedback, which is a useful experimental condition but "the ecological validity was necessarily limited" (K. Cho & MacArthur, 2011, p. 79). Similarly, Wichmann et al.'s (2018) experimental study involved strong control at the cost of ecological validity: Received feedback came from trained tutors rather than normal peers in the class, and learning was measured by comparing students' problem detection and correction skills from pre- to posttest. Another problem with all four of these studies is that separating providing and receiving processes deprives students from learning in that they do not just benefit from providing explanations and producing critical reviews but also from knowing deficiencies in their work and interpreting readers' needs (Nicol et al., 2014). Zhang et al. (2017) showed that revisions were especially likely when comments provided and comments received align, although they did not track these effects into learning outcomes.

As summarized in Figure 1, the effects of providing and receiving peer feedback have been studied on many outcomes and generally positive effects are found using various methods. However, the figure also makes salient the missing link between performance and learning: Does implementation or revision work lead to learning or are the effects of peer feedback independent of whether students act on the feedback? The present study focuses on the relationship between performance and learning. Because much of the prior research found the results varied by higher versus lower level aspects of writing (e.g., Berggren, 2015; Y. H. Cho & Cho, 2011; Liou & Peng, 2009), we examine these relationships separately within those two levels.

Unlike prior peer feedback studies that often focused on specific types of feedback and specific types of revisions, the present study analyzes feedback and revisions at a more aggregate level in order to be able to examine the larger system of actions and outcomes at once. The approach was to look at the number of received and provided feedback, and then correlate them with writing performance (amount of implemented feedback, amount of revisions) and improvements in writing quality on a new writing task. Different from previous experimental research designs, a more ecologically valid, correlational design was chosen in which both receiving and providing comments were involved in peer review. Due to the lack of prior research examining sources of learning, such a correlational design is useful for exploring the mediational relationship of various writing performance effects of peer review on learning outcomes; later experimental work can then focus on the specific component(s) receiving support from the correlational study. Three major research questions are addressed:

Research Question 1: Does both amount of provided and received peer reviews predict improved writing performance, particularly, both for implementation of received comments versus all revisions made?

Research Question 2: Does both amount of provided and received peer reviews predict student learning?

Research Question 3: Are the learning effects mediated by performance effects and, if so, via implementation or revision?

We examine the three research questions separately with high- and low-level aspects of writing (i.e., providing and receiving high- and low-level feedback, implementation of high- and low-level feedback, high- and low-level revisions, learning on high- and low-level dimensions) since the prior literature often finds different effects at each level.

The mediation (third) research question is both theoretically and pragmatically important. Theoretically, it speaks to different learning pathways (discussed in further detail below). Pragmatically, it speaks to the importance of requiring revisions at all or requiring students to specifically act on provided feedback. Teachers are often faced with the pragmatic dilemma of whether to encourage revision plans or even to require revisions at all because these come at the cost of having fewer writing tasks.

We specifically investigate these questions in the context of an Advanced Placement (AP) course (i.e., AP Language and Composition). This course is meant to be equivalent to a first-year college writing course, and therefore, sits on the boundary between secondary and tertiary education. This AP course has the highest annual enrollment among all AP courses, reflecting pressure to broaden access to experiences that improve college readiness (see College Board, 2018). At the same time, it has one of the lowest successful performance levels, with fewer than half of the students taking this course performing well enough to receive college credit, likely pointing to a general weakness in writing instruction at the high school level. These concerns call for more research on writing instruction with high school students.

By purposely sampling classes from high schools that vary in whether they predominantly serve high- or low-income families, the present study extends the generalizability of the research finding. First, students from lower income high schools commonly face challenges associated with the transition from high school to a more advanced curriculum. Helping such students write expands their access to universities, and also "gives them an edge for advancement in the workforce, and increases the likelihood they will actively participate as citizens of a literate society" (S. Graham & Perin, 2007b, p. 28). The skills they developed in peer review in the AP writing course can help them become an independent writer in the future. In addition, research on peer review in writing is scarce at the high school level (Schunn et al., 2016).

Schools serving many low-income families (typically measured by having Title I[1] status in the United States; U.S. Department of Education, 2018) are less likely to offer students AP courses. But there are now many programs in place to broaden participation in AP in such schools. For example, although there is a fee to take the exam, many schools serving low-income students and some states pay the fee for students. Nonetheless, because of the limited number of AP classes offered in Title I schools, these students may be less well prepared to engage with coursework at this level and have lower self-efficacy (Rymer, 2017). There are also higher rates of academic disengagement in Title I schools, which might influence norms for participation in peer review. Each of these aspects (less preparation, lower self-efficacy, lower participation norms) might influence the amount and type of feedback they give, as well as the ways in which they use the feedback they receive. At the same time, students from Title I schools might have a critical need to grow from the AP English course given how few other AP

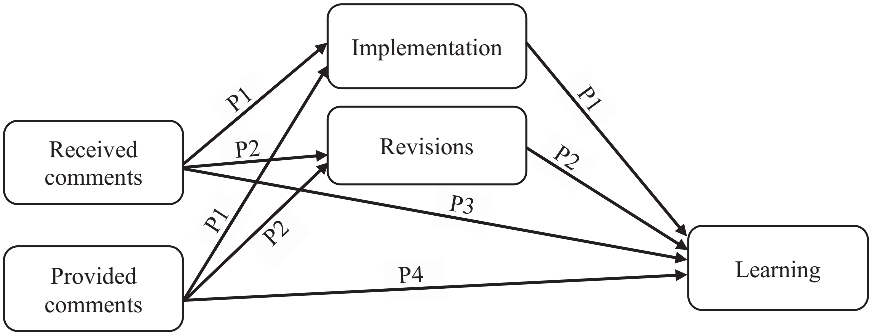

Figure 2. **Currently tested pathways for learning from peer review. Pathway 1 (P1): Students learn by implementing comments after peer review (including receiving and providing comments). Pathway 2 (P2): Students learn by making revisions after peer review (including receiving and providing comments). Pathway 3 (P3): Students learn by receiving comments without implementing them. Pathway 4 (P4): Students learn by providing comments.**

courses they will experience in high school in order to be ready to engage successfully in college.

Theoretical Framework

As noted above, peer review is a complex process involving both acts of providing and receiving feedback. Understanding the underlying learning pathways is vital to maximizing the benefits of peer review. In the current study, a conceptual model of peer review and learning is being proposed (see Figure 2) involving five major components—providing feedback, receiving feedback, implementation of changes based on peer comments, all revisions, and learning. Figure 2 highlights the four hypothesized pathways being tested as part of Research Question 3.

The first two hypotheses are concerned with how practice produces learning. The first hypothesis (Pathway 1 in Figure 2) comes from studies of expertise development that have found that feedback is important for developing expertise (Ericsson, 2006; Ericsson et al., 1993). Peer feedback draws students' attention to areas for improvement, and by implementing feedback in revisions, students practice problem-solving strategies and improve their writing ability. The second hypothesis (Pathway 2) is based on research of improvement via routine practice (Anderson, 1982). Students make revisions after peer review, perhaps triggered by peer feedback or other sources such as friends, parents, or automated computer feedback, and students simply learn to write by revising (regardless of source). If providing feedback and receiving feedback lead to learning through writing performance, the mediation could be through implementation or revisions or both.

On the other hand, providing and receiving feedback may lead more directly to learning. The third hypothesis (Pathway 3) is drawn from studies of feedback-based learning, which assumes that students learn by receiving feedback (Hattie & Timperley, 2007; Kluger & DeNisi, 1996), without necessarily implementing peer feedback or making any changes to the draft. For example, students may understand and learn from the praise and problems indicated by peers, but may not take actions in response to peer feedback due to time limitations or already being satisfied with the overall likely document grade.

The fourth hypothesis (Pathway 4) is based on studies of learning by observation (Schunk & Zimmerman, 1997, 2007). Learning takes place in the process of providing feedback because students observe advantages and weaknesses in peers' essays. Because they construct solutions for the problems, students transfer what they learn from evaluating their peers' work and providing feedback to new writing tasks. Students report learning from observing strengths and weaknesses in their peers' writing (Nicol et al., 2014; Schunn et al., 2016). However, these effects may be weak due to forgetting, lack of applicability to their own areas of weakness, or complexity of the problem.

It is worth noting that the most important difference between the prior research and the current study is that the former focused only on whether peer review led to learning, while the current study aims to answer both whether students learn from peer review (including receiving and providing) and also how they learn from peer review (directly or via revision/implementation work). The current study is not generally testing these broad learning pathways, but rather the open question is more specific to peer review: Which pathways are the primary sources of learning from peer review?

Method

Participants

The participants for this study consisted of 185 students from two U.S. high schools who were taking the same AP course, AP Language and Composition. Sixty participants came from a Title I school. The remaining 125 participants were from a non–Title I school serving many middle- and high-income families. All participants from a given school were taught by a single teacher across multiple sections; their teachers had agreed to implement shared writing tasks using a shared peer assessment tool at a similar time of year.

The participants' age ranged from 16 to 19 years, with the average being 17.1 years (*SD* = 0.5). Of the 185 participants, 57% were female (3% did not report gender). Among those students reporting their race/ethnicities, White students were the largest group (59%). Among the non-White participants, 31% were Asian, 5% African American, 4% Hispanic/Latinx. 19% of participants chose not to report their race/ethnicities. The composition of each school's participants significantly varied by race/ethnicity, $\chi^2(4) = 20.5$; $p < .001$: the non–Title I school had a higher percentage of Asian students, and the Title I school had a higher percentage of White students (see Supplemental Appendix B in the online version of the journal).

Materials

Peer Assessment Tool

Peerceptiv (K. Cho & Schunn, 2007; Schunn et al., 2016) is an online peer assessment program used by a large number of students in high schools and universities throughout the United States and around the world. As a classroom instructional tool, *Peerceptiv* can be used for formative assessment, allowing students to submit multiple drafts of one document and revise their document based on the peer comments given to the drafts. Within the system, teachers assign writing tasks and specify reviewing assignment details including submission deadlines, number of reviews, reviewing dimensions, and so on. Student writers submit first drafts online, and the program randomly and anonymously distributes each paper to a specified number of student reviewers. On a reviewing form, the reviewers provide diagnostic written comments along with analytic scores on different aspects of writing as specified by the teacher. The rating rubrics usually include details for the rating levels in student-friendly terms. Each reviewer is required to offer at least one written comment on each given dimension of writing, and there are usually suggestions for useful information to include in the comments. To further improve the comment quality, student writers are asked to rate the helpfulness of the comments they received based on a 5-point scale and explain their ratings to the reviewer. There is also grading accountability for accurate ratings and helpful comments (K. Cho & Schunn, 2007; Patchan et al., 2018).

Writing Tasks

Data were collected from two consecutive evidence-based, analytical-writing tasks that are a core part of the AP curriculum and more generally is a common area of struggle for secondary students (National Center for Education Statistics, 2012). Each of these two tasks asked students to read a one-page persuasive writing passage and then write a well-developed essay analyzing the rhetorical strategies used in the source passage. For example, they were required to describe what rhetorical strategies were used, support their descriptions and analysis of rhetorical strategies with evidence taken from the source text, and talk about how the rhetorical strategies connected to the overall thesis. The two tasks, drawn from prior end-of-course AP exams, were similar text-based argument writing tasks with the same requirements and identical rubrics. The only difference was different sources of passages being analyzed, which were selected to be of roughly similar length, difficulty, and reading levels. The first source passage was about the separation between people and nature because of technology, and the second passage was about the effect of migrations. As part of a high-stakes exam, the writing tasks are carefully designed by the College Board to be of equal difficulty. Further, expert scoring of these tasks in a larger study using these writing tasks across more schools found performance on these tasks to have nearly identical mean scores (Schunn et al., 2016).

Procedure

In general, AP courses involve some standardization of curriculum overall and assessments of writing in particular, which makes it easier to create a similarly structured experience across schools. To further reduce implementation variability which would add variance in this regression study, a carefully structured protocol was used in all classes. The protocol was based in best practice to increase the likelihood that learning could be observed. For the period of the study, the teachers participating in the project were provided with shared assignments, shared peer review rubrics, training on the use of the reviewing system, and protocols for training students in conducting peer review. Students wrote a first draft of the first writing task and turned in their first drafts to the online program by a specified deadline. The program distributed essays to peers across classrooms within a school randomly in a double-blind fashion; each student was required to review four peer essays for a given draft using a set of rubrics shared across all classes. Following best-practice, a detailed, rubrics-based comment form was provided to students (see Supplemental Appendix A in the online version of the journal); the rubrics were adapted from ones used by expert AP scorers to make them more student friendly (Schunn et al., 2016).

An in-class discussion at the beginning of the reviewing period was used to provide training for students on the peer review task and the peer review system. The teacher shared two sample essays with all the students. Students read the first sample essay, and then were shown example comments for it that were generally unhelpful versus generally helpful (e.g., specific and constructive) and discussed as a class what made reviews helpful. Then students read the second sample essay and completed a review with a partner in class using the assigned reviewing prompt. The class as a whole discussed the comments and ratings that were generated. At this point, the rating scales used in reviewing were discussed and students received calibration

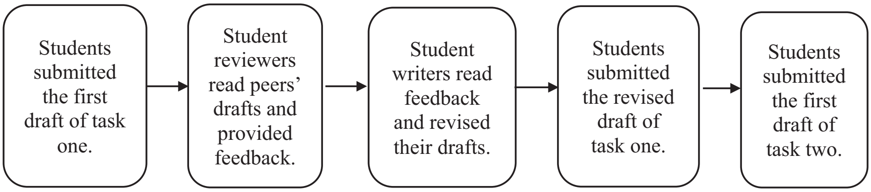
Students submitted the first draft of task one.
Student reviewers read peers' drafts and provided feedback.
Student writers read feedback and revised their drafts.
Students submitted the revised draft of task one.
Students submitted the first draft of task two.

Figure 3. **Peer review and writing tasks that students completed as part of the study.**

comments. Note that high-level comments were emphasized during the peer-review training because addressing high-level problems (e.g., explanation, argument) is more complex than addressing low-level problems (e.g., spelling, grammar).

Reviewers then had 1 week to provide comments and rated the essay based on given rubrics. The rubrics directed reviewers to assess eight dimensions of quality on a 7-point scale (see Supplemental Appendix A in the online version of the journal). For each dimension, they wrote comments and provided a quantitative rating. Students were required to provide at least one comment on each dimension. Then writers received peer comments, revised their draft, and submitted the revised draft to the program. To increase the quality of the reviews, the system provided a grade for the accuracy of peer ratings (based on being consistent with other reviewers) and for the helpfulness of peer comments (based on helpfulness ratings made by the authors). Finally, students completed a second writing task after a 1-week interval, using the same rubrics as with the first writing task. The peer review procedure is presented in Figure 3.

Measures

Essay quality, comments, and essay revisions in response to comments were systematically coded by multiple raters who were iteratively trained and continuously assessed for interrater reliability. To establish reliability levels and increase effective reliability of the resulting data, all data were exhaustively coded by at least two coders, and disagreements were resolved through discussion with a third coder present (Belur et al., 2018).

Writing Quality

The quality of the essays was determined by two trained writing experts who had years of teaching experience and used the same peer review rubrics (see Supplemental Appendix A in the online version of the journal). The expert rater training began with discussing the rubrics. A random subset of 50 essays were then rated, and differences between two experts' scores greater than 1.5 were discussed and resolved; a mean score was used for analysis in the low-conflict cases. The interrater reliability was substantial (Anthony & Joanne, 2005) for both high-level scores (k = .73) and low-level scores (k = .70). Then the raters independently rated the rest of the articles. The improvement in writing quality from the first draft of the first writing task to the first draft of the second writing task serves as the primary measure of learning in this study.

Total Number of Comments

In order to determine how many comments students provided and received, the first step of comment coding was to segment the peer comments into idea units. An idea unit refers to a comment made on one particular aspect of the student's writing. It might be a few words, a sentence, or multiple sentences having a unified intended aim. For example, there are two idea units in this peer comment: "This thesis is incomplete. It just copies the prompt and does not make any further analysis. It is not even paraphrased. // The word 'creat' is misspelled, which impedes understanding of the thesis." The sentences before the double slash mark are one idea unit, explaining a high-level problem. The second idea unit is the last sentence, which suggested a low-level revision.

Although students as reviewers were given multiple textboxes to provide separate comments within each of the eight fine-grained rubric categories (see Supplemental Appendix A in the online version of the journal), comments were further segmented by the researchers since reviewers often submitted comments with multiple ideas in one comment box. In particular, all comments were carefully read and were segmented depending on the problems identified. When later coding comments for type, scope, and implementation, the coders further segmented the comments including more than one problem. This resegmentation occurred rarely (2% of cases), suggesting that the prior segmentation step was a highly reliable process. The full segmentation process yielded a total of 6,507 idea units. Then a total number of comments per author and per reviewer were calculated from these data.

The scope of a peer comment was coded as high-level or low-level according to the type of the problem the comment intended to touch (see Table 1). High-level comments focused on thesis, argument, rhetorical strategies, evidence for claims, explaining evidence, and organization. Low-level comments considered smaller details including control of language and conventions. Then the total number of received and provided comments per individual were calculated separately for high- and low-level comments.

Table 1
Peer Feedback Coding Scheme

Category	Definition	Example
Type of feedback (k = .91)		
Praise	Purely evaluative remarks on good features of writing	You use quotes as evidence, which is good.
Summary	Statements of what the writer had done	The author started to analyze paradox first in his or her first paragraph, followed by rhetorical questions and imagery last.
Implementable comments	Revision-oriented comments that could trigger revisions	There are three rhetorical strategies highlighted, but they are a little simple, as they are just rhetorical devices.
Scope of feedback (k = .91)		
High-level	Comments with regards to thesis, arguments, rhetorical strategies, organization, evidence, and explanation.	The evidence you chose for the first body paragraph is very good; however, you only have one point for evidence. I would try to find more evidence.
Low-level	Comments with regards to control of language and conventions.	You misspelled the word “this” so just make sure you change that.
Implementation of feedback (k = .58)		
Implemented	Comments that were incorporated in the revision	**Peer feedback**: You did not analyze any rhetorical devices. You just stated them and provided no evidence. You should provide specific examples in your essay. *The writer added more specific examples in the revision.*
Not implemented	Comments that were not incorporated in the revision	**Peer feedback**: There is no evidence to be supported. Include evidence so that it may be supported.

(continued)

Table 1 **(continued)**

Category	Definition	Example
		The writer did not add any evidence in the later draft.
Vague implementation	Comments that were too vague to determine whether they were implemented or not	**Peer feedback**: The use of language was well but not as sophisticated as I would want.
Revision focus (k = .67)		
High-level	Revisions that altered the meaning of the original text	**Original:** The author supports his argument through anecdotes and hypothetical scenarios in order to demonstrate the growing barrier between humans and the natural world around them. **Revised:** The author supports his argument through anecdotes, hypothetical scenarios, and ironic tone in order to demonstrate the growing barrier between humans and the natural world around them.
Low-level	Revisions that did not alter the original meaning of the original text	**Original:** Humans are constantly striving to create and improve technology, trying to reach perfection. **Revised:** Humans are always striving to create and improve technology, trying to reach perfection.

Number of Comments Implemented in Revision

The segmented comments were further divided into praise, summary, or implementable comments that identified problems to improve (see Table 1). A total of 3,605 comments were found to be implementable. Two writing experts who taught undergraduate composition for multiple years coded the actual implementation of all implementable comments, that is, whether the author implemented a change based on the comment in their second draft. First, the revisions were highlighted by comparing the students' first and revised drafts using MS Word's Compare Document tool. Format changes were excluded. Then, evidence that the students had implemented the peer comments was identified by matching the revisions with the comments. A comment was labeled as Implemented if a revision was attributable to it. If no revision was found to be influenced by the comment, it was determined to be Not Implemented. Twelve percent of the implementable comments were coded as vague implementation because they were stated in such vague terms that they could not be coded for implementation—the vast majority of these vague comments were general statements about low-level problems (e.g., "There were many grammatical problems"). The comments coded as vague implementation were excluded from analysis. Coding of implementation at the comment level had moderate reliability (k = .58). To raise the effective reliability, all comments were double-coded and disagreements were resolved through discussion. Further, the data were analyzed at the level of number of implemented changes per author, further reducing the effect of coding noise. Based on comment focus, two measures were created: number of high-level implementations and number of low-level implementations.

Number of Revisions

Since received comments could produce one or many revisions in a document, and comments given to other students could also lead to revisions, the amount of first-to-second draft revising was separately coded, independent of comments received. In other words, both revisions triggered by peer comments and extra revisions were summed. Each separate revision was identified and coded for focus. A low-level revision was defined as one that did not alter the meaning of the original text, while a high-level revision involved changing the meaning of the original (see examples in Table 1). The MS Word Compare Documents function was used to facilitate revision coding. The first author and two trained research assistants independently labeled each revision as low-level or high-level. The reliability of coding each revision into high versus low was moderate (k = .67), and again was exhaustively double-coded to raise the effective reliability. Two measures were created: number of high-level revisions and number of low-level revisions.

Data Analysis

The purpose of the study was to test unique associations between the reviewing behaviors, the revision behaviors, and learning outcomes, separately for high-level and low-level aspects of writing. For this purpose, the use of multiple regression suffices. Structural equation modeling (SEM) was not used because the sample size was too small for SEM (Fritz & MacKinnon, 2007), and instead other techniques were used to address sample size in the mediation analyses. Also note that since there were only two schools, nesting students within schools in a hierarchical linear model also conveys no advantages for modeling the data (McNeish & Stapleton, 2016).

However, given the wide variety of variables included coming from different time points, listwise deletion for missing values could be problematic. Four variables had some missing values reflecting missing documents: Task 1 high-level and low-level scores (0.5%), and Task 2 high-level and low-level scores (4.9%). While Little's MCAR (missing completely at random) tests conducted in SPSS 25 suggested that the data using only significant predictors of learning could be treated as missing at random, a broader test of all variables found a significant violation of this assumption for high-level writing variables, $\chi^2(9) = 24.8$, $p < .005$ (Little, 1988). Missing data were initially addressed using EM (expectation maximization). EM sequentially uses the observed data to predict missing values (Expectation step), and then the full data set (observed and estimate) is used to build new prediction equations (Maximization step). Missing values are reestimated using the new prediction equations and the process repeats until the covariance matrix does not change (J. W. Graham, 2009). The analyses were also rerun using multiple imputation (10 imputations using SPSS 25). All the analyses produced the same results (same variables significant/nonsignificant; only minor variation in beta-weights). Since the two approaches are equivalent in this context and expectation maximization is already familiar in peer feedback research (e.g., Strijbos et al., 2010), we reported the findings using EM for missing values.

All of the main analyses involved multiple regression, but the particular form of the regression varied depending on outcomes because the distribution assumptions of linear regression were frequently violated. Negative binomial regression (NBR) and zero-inflated Poisson regression (ZIP) models were used when the outcome was implementation or revisions (including both high- and low-levels). Implementations and revisions are count variables, which are frequently right-skewed and cannot be normalized by transformation (e.g., log or square root). Such data are often modeled using Poisson regression (Coxe et al., 2009; Hilbe, 2007). However, Poisson regression requires the mean and variance to be roughly equal, and diagnostic tests of this assumption (i.e., the likelihood-ratio test) applied to implementations and revisions (for both high-level and low-levels) found that mean and variance to be unequal: for high-level implementation, $\chi^2(1) = 929.2$,

$p < .001$; for low-level implementation, $\chi^2(1) = 34.8, p < .001$; for high-level revisions, $\chi^2(1) = 484.3, p < .001$; and for low-level revisions, $\chi^2(1) = 495.8$, $p < .001$. When the assumption of means equal variance is violated, NBR can provide a better fit to the data (Coxe et al., 2009; Hilbe, 2007).

However, another problem in count data can be an unusually large number of zeros (e.g., having two modes in the data, with one being at zero), and the implementation and revision distributions showed some evidence of this pattern. If there are excessive zeros in the count distribution (sometimes operationally defined as at least 10% of the data; Blevins et al., 2015), a better fit to the data can be obtained using a ZIP or zero-inflated negative binomial (ZINB) models (Coxe et al., 2009; Hilbe, 2007).

It is important to note that using zero-inflated models should be decided based on theoretical grounds, because these models essentially assume two different kinds of processes are taking place: something that causes zero/nonzero counts and then something that influences the count when it is nonzero (e.g., something leading a student to revise at all, and then another thing leading a student to revise more once they have decided to revise at all). If no prior theory supports zero-inflated models, NBR models can be used (Allison, 2012; S. Yang et al., 2017). In the current case, zero-inflated models were considered plausible for both implementations and revisions. For implementation, some students may have received no implementable comments to act upon or they may have decided their peers were not useful sources of comments. For revisions, some students may have decided that doing revision was not worthwhile or necessary. In those cases, the models had two sets of regression results: one set predicting zero implementation/revision and a second set predicting amount of implementation/revision.

In sum, ZIP and ZINB models were tested when the outcome was implementations or revisions. The models were compared and a final model was selected based on best fit to the data defined as the smallest AIC (i.e., Akaike information criterion; Coxe et al., 2009; Patchan et al., 2016). The full set of models are presented in Supplemental Appendix E (in the online version of the journal) to show the consistency of findings across models based on different assumptions; which variables were important predictors of outcomes were largely consistent across models. The final, selected models predicting implementations and revisions based on model fit (i.e., AIC) are presented in Table 2.

Linear regression models were conducted when the outcome was high-level or low-level learning (first draft score of Task 2, controlling for first draft score of Task 1). According to the scatterplots of studentized residuals and unstandardized predicted value, the data satisfied the homoscedasticity assumption. The linear regression models were also screened for outliers according to leverage, studentized residuals, and Cook's *D* statistics (Aguinis et al., 2013; Fox, 1991). Examination of tolerance and variance inflation factor (VIF) for each of the independent variables showed that no

Table 2

Regression Results From Best-Fitting Models Predicting High-Level and Low-Level Implementations and Revisions

	Implementations				Revisions			
	High-Level (Model 1)		Low-Level (Model 2)		High-Level (Model 3)		Low-Level (Model 4)	
Predictors	B (SE)	e^b	B (SE)	e^b	B (SE)	e^b	B (SE)	e^b
Probability of not implementing or not revising (binary equation: odds of always 0)								
Title I school	0.09 (0.45)	1.10	−1.55 (1.40)	0.21	0.81 (0.96)	2.25	−4.30 (3.69)	0.01
Task1 score	−0.28 (0.24)	0.76	−0.34 (0.32)	0.71	−0.82 (0.58)	0.44	−0.97 (0.51)	0.38
Received comments	−0.10 (0.03)	0.91***	0.00 (0.09)	1.00	−0.09 (0.05)	0.91*	0.09 (0.13)	1.10
Provided comments	−0.05 (0.03)	0.95*	0.07 (0.07)	1.07	−0.16 (0.07)	0.85*	−0.30 (0.23)	0.74
Received comments × Title I school	—	—	−0.36 (0.32)	0.70	−0.10 (0.10)	0.90	−0.49 (0.45)	0.61
Provided comments × Title I school	—	—	−0.72 (0.33)	0.49*	0.47 (0.24)	1.59	−1.06 (0.82)	0.35
Quantity of implementation or revisions (count equation)								
Title I school	−0.48 (0.16)	0.62**	−0.52 (0.29)	0.59	−0.11 (0.35)	0.90	−0.56 (0.25)	0.57*
Task1 score	−0.37 (0.07)	0.69***	−0.08 (0.11)	0.92	−0.09 (0.10)	0.91	−0.08 (0.12)	0.92
Received comments	0.02 (0.01)	1.02*	0.13 (0.02)	1.14***	−0.01 (0.02)	0.99	0.10 (0.03)	1.10**
Provided comments	−0.01 (0.01)	0.99	0.01 (0.03)	1.02	0.00 (0.01)	1.00	0.02 (0.03)	1.02
Received comments × Title I school	—	—	0.07 (0.13)	1.07	0.06 (0.05)	1.06	0.20 (0.10)	1.22*
Provided comments × Title I school	—	—	−0.22 (0.08)	0.80**	0.06 (0.04)	1.06	−0.00 (0.08)	1.00

Note. Received comments and provided comments are centered.

*p < .05. **p < .01. ***p < .001.

predictor variable had a VIF greater than 2.4. Therefore, multicollinearity was not problematic.

Overall, three sets of statistical models were conducted: predicting revisions, predicting implementations, and predicting learning. In each case, the predictors included the number of comments students received and provided, with control variables of Task 1 score and school. The learning models also included number of revisions and number of implemented comments. Models also included interactions of the key predictors with school to formally test the consistency of effects across the two schools. All of these models were conducted separately at the high-level and low-level writing levels (e.g., high-level comments predicting high-level revisions, high-level implementations, and high-level learning). Unstandardized coefficients for linear regression models and e^b (interpretable as odds ratios) for zero-inflated models were reported as measures of effect sizes. All these zero-inflated regression models were conducted in R (see R syntax in Supplemental Appendix G in the online version of the journal). Linear regression models were conducted in SPSS 25.

As a final step, mediation of effects of comments on learning via revision or implementation was tested via a bootstrapping technique, which is recommended to test mediated effect with small to moderate samples (Shrout & Bolger, 2002). Mediation analyses included two control variables (Task 1 score and School), two predictors of interest (the number of received and provided comments), the mediator (the number of revisions/implementations) and learning. All of these models were conducted separately for high-level and low-level aspects of writing (e.g., models without and with interaction terms of the number of received/provided comments with school to test if indirect effects [if there were any] were consistent across schools). The PROCESS macro 3.4 for SPSS 25 was used to determine the significance of the indirect effects (Hayes, 2017), and, in particular, it employed 5,000 bootstrapped samples to estimate confidence intervals. If the 95% CI (confidence interval) does not include zero, the indirect effect is significant.

Results

Overall, students appeared to improve their writing over the studied window of time: mean first-draft ratings were higher for the second writing task than for the first writing task (see Supplemental Appendix C in the online version of the journal), for both high-level scores, paired *t*-test $t(184) = 4.32$, $p < .001$, Cohen's $d = 0.33$, and low-level scores, paired *t*-test $t(184) = 3.36$, $p < .001$, $d = 0.26$. The greater gains in high-level scores are consistent with the greater focus on high-level issues in comments and a higher rate of implementation for high-level comments than low-level comments in revisions (see Supplemental Appendix C in the online version of the journal). It is likely that high-level aspects of writing received more attention than did low-level issues because the criteria provided to students focused more on high-level aspects.

As an initial examination of which peer review factors might be associated with the outcome variables (implementation, revision, and second task writing scores), Pearson correlations were examined (see Supplemental Appendix D in the online version of the journal). Within a given level (high-level or low-level aspects of writing), none of the predictors were highly correlated with one another, reducing concerns about collinearity in the analyses. Implementations and revisions were highly correlated with one another as one would expect, but not so highly that they were redundant measures. High-level versus low-level aspects of writing within each measure were also generally correlated at moderate-to-high levels, but again not so highly that the two levels should be collapsed for analysis.

There were similar patterns in the correlations of predictors to outcome variables across levels (see Supplemental Appendix D in the online version of the journal). For both high and low levels, implementation was significantly correlated with receiving comments, and revisions were significantly correlated with providing comments and with receiving comments. For both high-level and low-level aspects of writing, the score of the second writing task was significantly correlated with providing comments and revisions. Thus, revision performance was generally correlated with providing and receiving comments, while learning appeared to be correlated more specifically with providing comments and revisions. Next, we report regression analyses that adjust for the moderate levels of covariance among the predictors.

Relationship of Providing and Receiving Comments to Writing Performance

High-Level Aspects of Writing

Both receiving and providing more comments decreased the odds of being among those who never implemented high-level comments (see binary equation of Model 1 in Table 2). Receiving more high-level comments also resulted in a higher number of implementations. For high-level revisions (see Model 3 in Table 2), receiving and providing more comments also decreased the odds of being among those who made no high-level revisions (see binary equation of Model 3 in Table 2). No variables predicted the quantity of high-level revisions (see count equation of Model 3 in Table 2). Thus, the effects were not always identical across the count and logistic equations (comparing Models 1 and 3), suggesting that their separate treatment is important. For example, both received and provided comments were significantly related to whether or not students implement high-level comments or make general revisions, but only received comments explained the variance in the quantity of high-level implementations (see Table 2).

Low-Level Aspects of Writing

Similar to high-level aspects of writing, the number of low-level comments received predicted the number of level implementations made, but to an even larger extent (see Model 2 in Table 2). Everything else about low-level writing was different from what was found to significantly predict high-level implementation and revision. For example, in contrast to having multiple main effect predictors at the level of making any high-level implementation or revision, there were none for low-level implementation or revision. Instead, especially for low-level revisions, there were more predictors of the amount of work. For example, the number of received comments predicted the number of revisions (see Model 4 in Table 2).

Another salient difference for low-level writing was the significant interactions of number of comments with school. Not only did students from the Title I school make fewer low-level revisions after controlling for numbers of comments they received and provided, there were also interactions of school with numbers of provided and received comments. For example, Title I students were more positively influenced by the number of comments they received in terms of making more revisions. Interestingly, the interaction of school with number of provided comments was mixed: For Title I school students, providing more low-level comments was positively correlated with making any implementations (as a dichotomous outcome), but negatively correlated with number of implementations (as a count outcome). This effect will be discussed in more detail in the "General Discussion" section.

Relationship of Providing and Receiving Comments to Learning

For both high-level and low-level learning, amount of revisions and amount of provided comments were significant predictors (see Table 3 and Figure 4). Note that providing comments significantly predicted low-level learning ($p < .01$) on its own, but this direct relationship became only marginally significant ($p = .06$) when the interaction term of provided comments and school was included. Since revisions were predicted by provided and received comments, there was the possibility of indirect effects via revisions. The mediation analyses revealed that the indirect effects were not statistically significant in the case of received and provided high-level comments (see Models 1 and 2 in Supplemental Appendix F in the online version of the journal). For low-level learning, by contrast, received comments had a significant indirect effect (see Models 3 and 4 in Supplemental Appendix F in the online version of the journal). In no case did the interactions with school produce significant direct, indirect, or total effects. Further, the total effect of providing comments on low-level learning was statistically significant across models including or excluding interactions. Note that high- and low-level implementations were not included in mediation analyses because they did not predict learning significantly.

Table 3

Multiple Regression Analysis Predicting Learning Outcomes

	High-Level Learning			Low-Level Learning		
Predictor Variables	B	SE	*B*	B	SE	*B*
1 (Constant)	3.19	0.26		3.45	0.29	
Task 1 score	0.38	0.06	0.45***	0.37	0.06	0.44***
Title I school	−0.14	0.10	−0.09	−0.28	0.10	−0.19**
	$R^2 = .22$, $F(2, 182) = 25.9$, $p < .001$			$R^2 = .23$, $F(2, 182) = 27.6$, $p < .001$		
2 (Constant)	2.69	0.29		3.41	0.30	
Task 1 score	0.41	0.06	0.50***	0.36	0.06	0.42***
Title I school	0.26	0.14	0.18	0.05	0.13	0.03
Received comment	0.01	0.01	0.09	−0.01	0.02	−0.03
Provided comments	0.01	0.01	0.16*	0.03	0.02	0.14
Received comments × Title I school	0.01	0.01	0.08	0.04	0.04	0.09
Provided comments × Title I school	0.00	0.01	0.03	0.06	0.03	0.15
Implementation	0.01	0.01	0.15	−0.01	0.03	−0.03
Revisions	0.02	0.01	0.18*	0.02	0.01	0.19*
	$R^2 = .36$, $\Delta R^2 = .14$, $\Delta F = 6.3$			$R^2 = .32$, $\Delta R^2 = .09$, $\Delta F = 3.71$		
	$F(8, 176) = 12.3$, $p < .001$			$F(8, 176) = 10.3$, $p < .001$		

Note. Received comments and provided comments are centered.
*$p < .05$. **$p < .01$. ***$p < .001$.

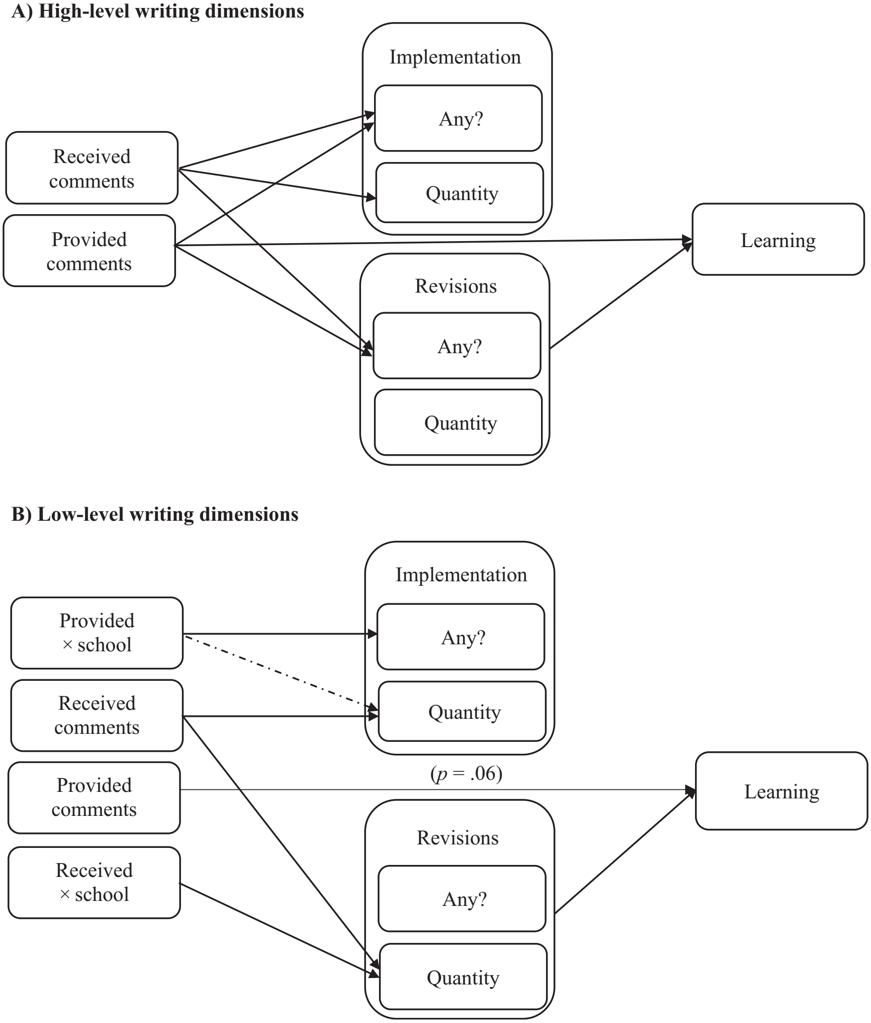

Figure 4. **Revised peer review and learning model for (A) high-level writing dimensions and (B) low-level writing dimensions. Dotted line indicates a negative effect.**

These reported analyses treated the high-level and low-level aspects of commenting, performance, and learning as completely separate from one another. While commenting to each level happened in different prompts, revising may have been connected. Therefore, follow-up data analyses were also conducted to investigate the influence of receiving and providing both high- and low-level comments (independent variables) on high-level scores or low-level scores of the second task (dependent variables). The results supported the initial approach of treating the two levels separately: (a) receiving high-level comments alone predicted high-level implementation and (b) receiving low-level comments alone predicted low-level implementation and low-level revisions.

General Discussion

The writing challenges of the 21st century, coupled with consistently poor results from national assessments of writing performance in high school students, have led researchers and educators to call for an increased focus on improving students' writing abilities, especially for argument-based writing (National Center for Education Statistics, 2012). Peer review has been identified as an efficient means for helping students write, but prior research has predominantly focused on the effects of peer review on students' writing performance of the same writing task (i.e., students revise their drafts on receiving peer feedback). Whether peer review can improve learning (i.e., writing in a new writing task) and the pathways through which peer reviewing helps students learn to write were therefore unclear. Understanding the effects of peer review on learning and identification of factors that mediate peer review and learning is important in guiding teachers' effective use of peer review.

A Revised Model of Peer Review and Learning

Figure 4 summarizes the findings regarding significant predictors of writing performance and learning. Note that the connections for the binary outcomes have been sign-reversed since double-negatives are confusing (e.g., the regression result of *providing more comments was negatively related to not making any revision* was sign-reversed to be *providing more comments was positively related to making any revision).* Overall, the results indicated that both providing and receiving feedback predicted performance and learning, and in slightly different ways for high- versus low-level aspects of writing. Further, the patterns of results of high versus low were consistent with different learning pathways (learning by observation and learning by routine practice).

The Relationships of Providing and Receiving Comments to Document Improvements

High-Level Aspects of Writing

The findings regarding performance on higher level aspects of writing were partially consistent with the theoretical framework (see Figure 4A). Received and provided feedback increased the probability of implementing feedback and making general revisions. This finding is different from Y. H.

Cho and Cho (2011), who found that the effects of receiving feedback on revised draft quality were limited but those of providing feedback were significantly positive. Higher level writing issues (as studied here) are more likely to be shared across authors, and therefore the provided high-level comments should be more likely to overlap with received comments for high-level issues. By receiving and providing high-level feedback targeting similar problems, students develop a better understanding of their weaknesses and are motivated to revise to narrow the gap between their current and desired performance (Nicol & Macfarlane-Dick, 2006).

It is possible that receiving feedback was the only significant predictor of the quantity of implementations perhaps because implementation is operationally so closely tied to receiving feedback. When students receive feedback, they focus on the areas for improvement. However, neither receiving nor providing high-level feedback seemed to increase the quantity of revisions.

Low-Level Aspects of Writing

For lower level aspects of writing, the findings were slightly different (see Figure 4B). The interaction term of provided feedback and school positively predicted the probability of implementing any feedback, but negatively predicted the quantity of implementations. In other words, compared with non–Title I school students, Title I school students were more likely to decide to implement feedback but less likely to implement more comments as they provided more feedback. As one possible explanation, providing more feedback could encourage students to decide to implement any feedback, perhaps because providing feedback helped them develop their ability to identify and solve problems (Nicol et al., 2014). However, when it came to the phase of actual implementation, the utility of providing feedback diminished as students provided more feedback. Another possible explanation relates to differences in context. For example, Title I school students might be less motivated to engage in peer review than their non–Title I counterparts. They might think they have benefited from providing feedback so that they did not need to do actual implementations. Students from Title I schools are rarely studied in peer review research, but the topic is important: They received/provided less feedback and responded less on receiving feedback than did students at the non–Title I school. Future research could investigate how student characteristics (e.g., motivation) influence peer review and writing performance. Received feedback predicted the quantity of low-level implementations significantly potentially because receiving more feedback helped students know their low-level problems.

Received feedback predicted the quantity of low-level revisions significantly for both schools, with a larger effect for Title I school students. Students from the Title I school might need more feedback to help them know their weaknesses. Provided feedback did not predict the probability of making any revisions and the quantity of revisions perhaps because students thought that they could learn from providing feedback, and it was not necessary to make low-level revisions.

Past research has focused exclusively on either implementation or revisions without treating the findings as conceptually different from one another, and the current research suggests that this produces an incomplete picture that potentially explains differences in findings across studies. Some of the differences may have been due to the difficulty in coding implementation for vague comments, for example, for low-level issues. However, since received comments predicted implementation and revisions, the vague-comments problem is unlikely the main driver of the different patterns of results for high-level issues.

The Relationships of Providing and Receiving Comments to Learning to Write

High-Level Aspects of Writing

For high-level aspects of writing, students appeared to learn to write by making revisions triggered or not triggered by received feedback. After accounting for the revisions, there was no additional predictiveness of implementations. Making general revisions enables students to use what they learn from peer review to identify and solve other problems that are not identified by reviewers, and thus, develop their detection and correction skills (M. Yang et al., 2006).

Further, there was a significant residual direct effect of providing comments as found by others in studies of providing feedback (K. Cho & MacArthur, 2011; Greenberg, 2015; Lundstrom & Baker, 2009). High-level writing consists of a broad range of component skills (e.g., knowing how to explain the provided evidence, knowledge of logical and clear organization of the essay). In terms of diversity of skills practiced, providing feedback, especially when it involves many different written objects, may provide more opportunities to practice a diverse range of skills than receiving feedback on a particular writing product. Similarly, providing feedback to others might present opportunities to practice detection of writing issues (Chen, 2010; Lundstrom & Baker, 2009), and improving detection skills might be especially important for writing skills. The effects of receiving and providing feedback on learning were not mediated via revisions, perhaps because the effect of number of comments was on making any revisions whereas it was amount of revisions that predicted learning.

Low-Level Aspects of Writing

Similar to high-level aspects of writing, revisions and providing comments were predictors of learning for low-level aspects of writing.

However, the indirect effect of receiving comments on learning via revisions was significant in the case of low-level writing. Different from high-level aspects of writing, the number of received comments predicted the quantity of revisions, and thus, there was aligned at the mediator level: quantity of revisions.

Interestingly, implementations did not significantly predict learning for both high-level and low-level aspects of writing. Why did implementations not predict learning? It is unlikely that no learning occurs via implementing comments after peer review, given the large amount of research supporting the pathway to learning across a wide range of domains (Astin, 1993; Ericsson et al., 1993; Kellogg & Whiteford, 2009). However, it may be that students benefit more from making general revisions than from implementations. While receiving feedback can provide students with performance goals that are just beyond their current performance but addressable (Ericsson, 2006), making general revisions may provide students with more opportunities to utilize what they learn in revisions beyond making only the specific changes suggested by reviewers.

These patterns of the learning effects on low- and high-level dimensions are most consistent with learning by routine practice and learning by observation (Couzijn, 1999; Schunk & Zimmerman, 1997, 2007). In other words, the pattern of results is most consistent with the studies of expertise development, in which students improve by practicing their revising skills after receiving and providing feedback. Students improve through practice that is influenced by immediate feedback from peer reviewers and other sources. Students may also learn to write by providing feedback because it helps them transfer what they learn from detecting problems in others' work/suggesting solutions for the problems, in addition to developing a better understanding of the readers' perspectives (K. Cho & MacArthur, 2011).

Implications for Practice

The findings of the current study provide useful lenses for examining and maximizing the benefits of peer review for secondary school students and beyond. First, they provide additional basic support to adopting peer review in writing instruction at secondary schools. Although peer review has already been identified as a high leverage practice for writing instruction, secondary school students are generally provided with few opportunities to write and receive feedback on their writing (Kiuhara et al., 2009). The current findings reveal that peer reviewing not only helps secondary school students improve a draft but also enhances student learning, particularly, from the activities of providing comments and making revisions after receiving comments.

Second, two learning paths have been identified as relevant to learning from peer feedback: learning by routine practice and learning by observation. Students do not narrow their focus on only the feedback they receive, but they appear to take a more active part in writing processes by making general revisions (in response to feedback received and provided) and directly from providing feedback, so that "observational and emulative learning leads to self-control by the learner, automaticity of the cognitive skill, and ultimately to self-regulation" (Van Steendam et al., 2010, p. 318). The current results suggest that particular instructional variations of revision tasks may help optimize student learning: (a) teachers should require students to submit revisions (Boud, 2000), rather than only assigning more writing tasks without revisions (Nicol & Macfarlane-Dick, 2006); (b) students should be encouraged to make general revisions beyond the received feedback rather than just acting on received feedback; and (c) students should be encouraged to provide concrete and specific peer feedback in practice rather than vague feedback. Clear feedback with more information can help authors understand and use peer feedback. Pragmatically, the teacher can emphasize that the general revisions, as well as the received and provided feedback, will help improve future writing performance.

Third, the current study was conducted in an online peer assessment system in the context of a specific kind of writing task, with a detailed, student-friendly rubric that was well aligned to the writing task, and incentives for students to provide accurate and constructive peer feedback. Well-structured peer review training was provided to help students develop a better understanding of the rubrics and how to provide adequate feedback. The amount and nature of learning might have been different if the writing task was less well-specified (making for less overlap in provided feedback with own writing), the evaluation criteria were overly general or poorly aligned (making the feedback less relevant or effective), the incentives for high-quality feedback were weak (making for less constructive reviewing), or the peer review was in a traditional face-to-face format (making anonymous and efficient multipeer feedback difficult). That is, the current findings are rooted in strong writing instruction practices, and there might be weaker performance or learning effects in less optimal writing instruction conditions.

Fourth, the current study has included a broader focus than was found in prior peer feedback research so that it plays an important foundational role within the larger research space. For example, it investigated the relationships between receiving, providing, feedback implementation, general revisions, and learning, with consideration of a contextual factor (i.e., school title). Based on this study, future research can investigate follow-up learning questions such as how implementations and revisions might influence students' later providing feedback behaviors.

Caveats and Conclusion

Several caveats should be considered. First, the current study examined students from two secondary schools in an AP writing course in the United

States. Future research should examine the generalizability of the results in other learning contexts with students at other levels of education and taking different peer assessment forms. Similarly, how receiving and providing peer feedback matter for students taking non-AP courses can also be examined. AP students may benefit more from peer review because AP courses generally involve a more homogeneous group of students who are especially motivated, hardworking, and advanced, while non-AP students are more likely to vary in academic performance, motivation, interests, and so on. In addition, feedback effects may be mediated by other variables than implementations or revisions, such as student characteristics (e.g., motivation, self-efficacy, thinking styles; Hattie & Timperley, 2007; Lin et al., 2001; Shute, 2008). Students with intrinsic versus extrinsic motivation, high versus low self-efficacy, or different thinking styles may engage differently in peer review. Further, many of the AP students from the Title I school might not come from low-income families. It would be helpful to directly examine various components of students' socioeconomic status (e.g., family income, family education levels) to develop a deeper understanding of the observed interactions of peer feedback with school context.

Second, more support should be provided for Title I school students, who were found to be less likely to respond to feedback and make fewer revisions than the non–Title I school students. The observed differences might be related to a variety of factors (e.g., motivation, engagement, prior experiences with peer feedback, level of teaching resources, school learning atmosphere; Hattie & Timperley, 2007; Irvin et al., 2011; Shute, 2008). Although it is more difficult to change the contextual factors, students' academic motivation and engagement in writing could be improved. For example, additional training can be provided to help Title I school students provide more clear and meaningful feedback (i.e., with supporting details; Van Steendam et al., 2010).

Third, the reciprocal nature of peer review should be taken into consideration when discussing its effects. Because students switch between the roles of assessors and assessees in peer review, what students provide as assessors and how they respond to peer feedback as assessees might change as peer review goes on (Tsivitanidou et al., 2011). Since this peer review process is dynamic, reflective, and interactive, future research can examine "reciprocal" effects between providing and receiving peer feedback acts (e.g., how providing feedback changes receiving feedback and vice versa).

Fourth, a number of low-level comments were too vague to be coded for implementation. The corresponding reduction in the amount of detected low-level implementation might have influenced the results, perhaps underestimating the benefits of implementation. On the other hand, the same challenge we faced in coding for implementation could have been a challenge for peers receiving vague comments: Exactly how to implement changes in the document (or learn from the comments) might have been unclear.

Finally, the current study is based on correlational analysis, which does not necessarily imply causality. However, correlational analyses are useful for ruling out causes through lack of correlations. Therefore, future experimental research, which is difficult to conduct on many variables at once, can focus on the significant variables highlighted in the current study. Further, future research can track whether the feedback-to-implementation process varies by more fine-grained distinctions. Initial analyses were conducted separately on this data set for each of the specific dimensions and found to produce roughly similar results. However, given multicollinearity issues among specific dimensions within high-level writing, a larger data set would be needed to strongly examine specificity or generality of learning within and between more specific aspects of writing.

Notes

The second author is a co-inventor of the peer review system used in the study. The current project was partially funded by grant R305A120370 from the Institute of Education Sciences.

[1] Schools with at least 40% of enrolled students coming from low-income families are eligible to receive Title I funds (U.S. Department of Education, 2018).

References

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*(2), 270–301. https://doi.org/10.1177/1094428112470848

Allison, P. D. (2012). *Logistic regression using the SAS system: Theory and application*. SAS Institute.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369–406. https://doi.org/10.1037/0033-295X.89.4.369

Anthony, J. V., & Joanne, M. G. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, *37*(5), 360–363.

Applebee, A. N., & Langer, J. A. (2011). A snapshot of writing instruction in middle schools and high schools. *English Journal*, *100*(6), 14–27.

Astin, A. W. (1993). *What matters in college: Four critical years revisited*. Jossey-Bass.

Beason, L. (1993). Feedback and revision in writing across the curriculum classes. *Research in Teaching of English*, *27*(4), 395–422. https://www.jstor.org/stable/40171241

Belur, J., Tompson, L., Thornton, A., & Simon, M. (2018). Interrater reliability in systematic review methodology: Exploring variation in coder decision-making. *Sociological Methods & Research*. Advance online publication. https://doi.org/10.1177/0049124118799372

Berggren, J. (2015). Learning from giving feedback: A study of secondary-level students. *ELT Journal*, *69*(1), 58–70. https://doi.org/10.1093/elt/ccu036

Blevins, D. P., Tsang, E. W. K., & Spain, S. M. (2015). Count-based research in management: Suggestions for improvement. *Organizational Research Methods*, *18*(1), 47–69. https://doi.org/10.1177/1094428114549601

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, *22*(2), 151–167. https://doi.org/10.1080/713695728

Chen, C. W. (2010). Graduate students' self-reported perspectives regarding peer feedback and feedback from writing consultants. *Asia Pacific Education Review*, *11*(2), 151–158. https://doi.org/10.1007/s12564-010-9081-5

Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, *20*(4), 328–338. https://doi.org/10.1016/j.learninstruc.2009.08.006

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. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education*, *48*(3), 409–426. https://doi.org/10.1016/j.compedu.2005.02.004

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

College Board. (2018). *Program summary report*. https://secure-media.collegeboard.org/digitalServices/pdf/research/2018/Program-Summary-Report-2018.pdf

Couzijn, M. (1999). Learning to write by observation of writing and reading processes: Effects on leaning and transfer. *Learning and Instruction*, *9*(2), 109–142. https://doi.org/10.1016/S0959-4752(98)00040-1

Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of Personality Assessment*, *91*(2), 121–136. https://doi.org/10.1080/00223890802634175

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683–703). Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.038

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406. https://doi.org/10.1037/0033-295X.100.3.363

Falchikov, N. (2001). *Learning together: Peer tutoring in higher education*. Routledge Falmer.

Fox, J. (1991). *Regression diagnostics: An introduction*. Sage. https://doi.org/10.4135/9781412985604

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *18*(3), 233–239. https://doi.org/10.1111/j.1467-9280.2007.01882.x

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

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Graham, S., & Perin, D. (2007a). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, *99*(3), 445–476. https://doi.org/10.1037/0022-0663.99.3.445

Graham, S., & Perin, D. (2007b). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools—A report to Carnegie Corporation of*

New York. Alliance for Excellent Education. https://media.carnegie.org/filer_public/3c/f5/3cf58727-34f4-4140-a014-723a00ac56f7/ccny_report_2007_writing.pdf

Greenberg, K. P. (2015). Rubric use in formative assessment: A detailed behavioral rubric helps students improve their scientific writing skills. *Teaching of Psychology*, *42*(3), 211–217. https://doi.org/10.1177/0098628315587618

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.

Hilbe, J. M. (2007). *Negative binomial regression*. Cambridge University Press. https://doi.org/10.1017/CBO9780511811852

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

Irvin, M. J., Meece, J. L., Byun, S., Farmer, T. W., & Hutchins, B. C. (2011). Relationship of school context to rural youth's educational achievement and aspirations. *Journal of Youth and Adolescence*, *40*(9), 1225–1242. https://doi.org/10.1007/s10964-011-9628-8

Jonassen, D., Davidson, M., Collins, M., Campbell, J., & Haag, B. B. (1995). Constructivism and computer-mediated communication in distance education. *American Journal of Distance Education*, *9*(2), 7–26. https://doi.org/10.1080/08923649509526885

Kellogg, R. T., & Whiteford, A. P. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist*, *44*(4), 250–266. https://doi.org/10.1080/00461520903213600

Kiuhara, S. A., Graham, S., & Hawken, L. S. (2009). Teaching writing to high school students: A national survey. *Journal of Educational Psychology*, *101*(1), 136–160. https://doi.org/10.1037/a0013097

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Lin, S. S. J., Liu, E. Z. F., & Yuan, S. M. (2001). Web-based peer assessment: Feedback for students with various thinking-styles. *Journal of Computer Assisted Learning*, *17*(1), 420–432. https://doi.org/10.1046/j.0266-4909.2001.00198.x

Liou, H. C., & Peng, Z. Y. (2009). Training effects on computer-mediated peer review. *System*, *37*(3), 514–525. https://doi.org/10.1016/j.system.2009.01.005

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202. https://doi.org/10.2307/2290157

Lu, J. Y., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science*, *40*(2), 257–275. https://doi.org/10.1007/s11251-011-9177-2

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, *18*(1), 30–43. https://doi.org/10.1016/j.jslw.2008.06.002

McNeish, D., & Stapleton, L. M. (2016). The effect of small sample size on two level model estimates: A review and illustration. *Educational Psychology Review*, *28*(2), 295–314. https://doi.org/10.1007/s10648-014-9287-x

National Center for Education Statistics. (2012). *The nation's report card: Writing 2011* (NCES 2012-470). Institute of Education Sciences, U.S. Department of Education. https://nces.ed.gov/nationsreportcard/pdf/main2011/2012470.pdf

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., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199–218. https://doi.org/10.1080/03075070600572090

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., & Clark, R. (2018). Accountability in peer assessment: Examining the effects of reviewing grades on peer ratings and peer feedback. *Studies in Higher Education*, *43*(12), 2263–2278. https://doi.org/10.1080/03075079.2017.1320374

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

Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*, *8*(3), 265–289. https://doi.org/10.1016/S1060-3743(99)80117-9

Pearce, J., Mulder, R., & Baik, C. (2009). *Peer review: Case studies and practical strategies for university teaching*. Centre for the Study of Higher Education, University of Melbourne. https://apo.org.au/sites/default/files/resource-files/2010-01/apo-nid20259.pdf

Philippakos, Z. A., & MacArthur, C. A. (2016). The effects of giving feedback on the persuasive writing of fourth- and fifth-grade students. *Reading Research Quarterly*, *51*(4), 419–433. https://doi.org/10.1002/rrq.149

Rymer, K. R. (2017). *Assessing self-efficacy to improve impoverished students' education* [Unpublished doctoral dissertation]. Carson-Newman University.

Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, *11*(1), 1–31. https://doi.org/10.1207/s15326977ea1101_1

Schunk, D. H., & Zimmerman, B. J. (1997). Social origins of self-regulatory competence. *Educational Psychologist*, *32*(4), 195–208. https://doi.org/10.1207/s15326985ep3204_1

Schunk, D. H., & Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Reading & Writing Quarterly*, *23*(1), 7–25. https://doi.org/10.1080/10573560600837578

Schunn, C. D., Godley, A. J., & DiMartino, S. (2016). The reliability and validity of peer review of writing in high school AP English classes. *Journal of Adolescent & Adult Literacy*, *60*(1), 13–23. https://doi.org/10.1002/jaal.525

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*(2), 422–445. https://doi.org/10.1037/1082-989X.7.4.422

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Strijbos, J. W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*, *20*(4), 291–303. https://doi.org/10.1016/j.learninstruc.2009.08.008

Topping, K. J. (2009). Peer assessment. *Theory Into Practice*, *48*(1), 20–27. https://doi.org/10.1080/00405840802577569

Tsivitanidou, O. E., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction*, *21*(4), 506–519. https://doi.org/10.1016/j.learninstruc.2010.08.002

Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, *9*(2), 147–170. https://doi.org/10.1016/S1060-3743(00)00022-9

U.S. Department of Education. (2018). *Improving basic programs operated by local educational agencies (Title I, Part A)*. https://www2.ed.gov/programs/titleiparta/index.html

Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, *62*(1), 1–41. https://doi.org/10.1111/j.1467-9922.2011.00674.x

Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction*, *20*(4), 316–327. https://doi.org/10.1016/j.learninstruc.2009.08.009

Wichmann, A., Funk, A., & Rummel, N. (2018). Leveraging the potential of peer feedback in an academic writing activity through sense-making support. *European Journal of Psychology of Education*, *33*(1), 165–184. https://doi.org/10.1007/s10212-017-0348-7

Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, *15*(3), 179–200. https://doi.org/10.1016/j.jslw.2006.09.004

Yang, S., Harlow, L. I., Puggioni, G., & Redding, C. A. (2017). A comparison of different methods of zero-inflated data analysis and an application in health surveys. *Journal of Modern Applied Statistical Methods*, *16*(1), 518–543. https://doi.org/10.22237/jmasm/1493598600

Zhang, F., Schunn, C. D., & Baikadi, A. (2017). Charting the routes to revision: An interplay of writing goals, peer comments, and self-reflections from peer reviews. *Instructional Science*, *45*(5), 679–707. https://doi.org/10.1007/s11251-017-9420-6

Manuscript received April 15, 2019
Final revision received May 29, 2020
Accepted June 30, 2020