CrossMark

Understanding the benefits of providing peer feedback: how students respond to peers' texts of varying quality

Melissa M. Patchan[1] · Christian D. Schunn[1]

Abstract Prior research on peer assessment often overlooks how much students learn from providing feedback to peers. By practicing revision skills, students might strengthen their ability to detect, diagnose, and solve writing problems. However, both reviewer ability and the quality of the peers' texts affect the amount of practice available to learners. Therefore, the goal of the current study is to provide a first step towards a theoretical understanding about why students learn from peer assessment, and more specifically from providing feedback to peers. Students from a large Introduction to Psychological Science course were assigned four peers' papers to review. The reviewing ability of each student was determined, and to whom the students provided feedback was manipulated. The features and focus of the comments from a sample of 186 participants were coded, and the amount of each type was analyzed. Overall, reviewer ability and text quality did not affect the amount of feedback provided. Instead, the content of the feedback was affected by reviewer ability. Low reviewers provided more praise than high reviewers, whereas high reviewers provided more criticism than low reviewers. This criticism from high reviewers described more problems and offered more solutions, and it focused more often on high prose and substance. In the only significant reviewer ability × text quality interaction, high reviewers described more problems in the low-quality texts than in the high-quality texts, whereas low reviewers did not make this distinction. These results suggest that high reviewers and low reviewers may utilize different commenting styles, which could significantly impact the benefits of peer assessment.

Keywords Peer assessment · Providing feedback · Individual differences · Revision skills

✉ Melissa M. Patchan
mpatchan@pitt.edu

[1] University of Pittsburgh, Pittsburgh, USA

🎉 Springer

Peer assessment is the quantitative or qualitative evaluation of a learner's performance by another learner of the same status. It is typically implemented in classrooms with the intention of developing the knowledge or skill of all learners involved. Although a substantial amount of research has been conducted on peer assessment, these studies have focused primarily on the reliability and validity of quantitative evaluations (i.e., peer ratings) and the effectiveness of peer feedback in comparison to instructor feedback. In general, this research has found that peers are capable of providing ratings comparable to instructors' ratings (Falchikov and Goldfinch 2000), and that peer feedback is usually just as effective as an instructor's feedback (Topping 2005), if not more effective (Cho and Schunn 2007). One major limitation of the research to date is that there are few empirical studies that have systematically examined the mechanisms of peer assessment—that is, what factors mediate learning (Strijbos and Sluijsmans 2010). Furthermore, when researchers more deeply examined the effects of peer assessment, they often focused only on the benefit of receiving feedback from peers. What is overlooked is how much students learn from the process of providing feedback to peers. Therefore, the goal of the current study is to provide a first step towards a theoretical understanding about why students learn from peer assessment, and more specifically from providing feedback to peers.

Lundstrom and Baker (2009) powerfully illustrated the contrast between the benefit of providing feedback to peers and the benefit of receiving peer feedback. They examined whether second language learners' writing improved more from only providing feedback or only receiving feedback. Students experienced four peer-review training sessions throughout the course of a semester. During the training sessions, students were given sample essays, and they either practiced how to provide effective feedback (i.e., they only provided feedback and did no revising) or practiced how to revise the essay based on feedback they received (i.e., they only revised essays using this feedback). They found that providing feedback led to greater improvements from pretest to posttest than receiving feedback. To better understand why students learn from the process of providing feedback, we first review which aspects of peer assessment appear to help students and then we offer an explanation for why this benefit may occur.

Which aspects of peer assessment help skill development?

It is important to first set the instructional context: many students beginning college do not possess a proficient level of writing, and thus peer assessment in introductory course contexts actually involves instruction on the fundamental aspects of writing, rather than just disciplinary conventions. According to a recent national assessment of writing in the United States (National Center for Education Statistics 2012), more than 50 % of high school seniors demonstrated only a limited understanding of the basic knowledge and skills that are fundamental for competent writing and an additional 21 % of high school seniors performed below this basic ability level. For example, when writing to explain, students often included details that did not enhance the clarity or progression of ideas or provided explanations that were inconsistent in clarity and quality. Furthermore, many students used simple sentence structures and poor organization. Many of these deficits might come from poor revision skills. Hayes et al. (1987) observed that, in comparison to higher-ability writers, lower-ability writers were not able to detect as many problems, were less likely to attend to global issues, had fewer strategies for dealing with global issues, and were less

likely to choose effective strategies for revisions. Peer assessment can help writers improve these revision skills, through a variety of methods.

First, peer assessment may be helpful from just mere exposure to models of writing—that is, by reading other peers' texts, students may create a mental representation of each text, which could serve as models of successful or unsuccessful writing strategies. Using models has been shown to affect the content students choose to include in their papers and how they decide to organize that content (Charney and Carlson 1995). However, reading alone does not appear to be sufficient. Students who only read peers' texts wrote poorer quality texts than those who reviewed (i.e., read but also rated the quality and provided feedback) peers' texts (Cho and MacArthur 2011). One possible explanation for the Cho and MacArthur results is that their 'read only' students were not using goals appropriate to learning. Although all students in that study were trained to evaluate papers using the evaluation rubric, students in the 'read only' condition were instructed to read the example papers repeatedly until time was up. For these students, their reading goal was likely to be to comprehend the papers rather than to evaluate the papers. Readers, whose intention is to comprehend the text, do not attend to problems within the text unless those problems disrupt the reader's ability to comprehend the text (Hayes et al. 1987). Then readers will attend to certain problems (e.g., spelling errors, grammatical errors, errors of fact) only long enough to decipher the meaning of the text before they move on and forget the problems. Thus, mere exposure is unlikely to be a strong source of learning.

Another possibility could be that students learn from reading for evaluation. When reading for evaluation, readers adopt additional goals of problem detection, problem diagnosis, and searching for strategies to fix the problems (Hayes et al. 1987). These goals draw the readers' attention to a wider variety of problems and possibly useful discoveries that could be generalized to future writing. The evaluation could be completed at two levels: (1) to determine whether the text was sufficiently well written, and (2) to detect and diagnose problems and determine how to best fix those problems. Several studies (Lu and Law 2012; Wooley et al. 2008) have examined the effectiveness of each level of evaluation. Consistently, students who provided feedback (i.e., detected and diagnosed problems or offered solutions) outperformed students who only rated the quality of the peers' work, and simply rating peers' work sometimes had no benefit over a control condition that wrote without doing any peer assessment. Therefore, although rating peers' texts may increase students' writing ability under some circumstances, constructing comments appears to be the most effective evaluation activity in which students practice diagnosis and solution skills rather than just detection skills. Yet a critical question remains—why does providing feedback in general, and constructive criticisms in particular, help student skill development?

An identical elements approach to understanding the benefits of providing feedback

We turn to the classic Identical Elements Theory to create a framework for understanding how students transfer knowledge about writing to the task of providing feedback and back again to improve writing knowledge. Thorndike's Identical Elements Theory (Thorndike and Woodworth 1901) posited that success in a new situation depends on the number of shared stimulus–response elements (i.e., identical elements). The greater number of these identical elements, the more likely one would be successful in the new situation. However,

Thorndike, reflecting the behaviorist era, focused on behaviors in defining what constitutes an element. In a more modern cognitive incarnation of the Identical Elements Theory of transfer, Singley and Anderson (1989) proposed specific cognitive elements (i.e., chunks representing declarative "knowing that" knowledge and productions representing procedural "knowing how" knowledge) associated with their influential adaptive control of thought (ACT) theory. This incarnation of the Identical Elements Theory successfully accounts for learning rates and transfer (or lack thereof) in a wide variety of tasks (Singley and Anderson 1989; Anderson and Lebiere 1998). Therefore, to understand the benefits of providing feedback, we must identify the conceptually identical elements in both writing and providing feedback tasks.

As Flower and Hayes (1981) noted, "Writing is best understood as a set of distinctive thinking processes which writers orchestrate or organize during the act of composing." In their influential cognitive process theory of writing, they identified three major elements by observing think-aloud protocols of higher-ability and lower-ability writers: the task environment, the writer's long-term memory, and the writing processes, which included planning, translating, and reviewing. Flower, Hayes, and colleagues (Flower et al. 1986; Hayes et al. 1987) later elaborated on the reviewing process, which included defining the task, detecting problems, diagnosing problems, and selecting a revision strategy. These reviewing processes likely form the basis for tasks that involve providing feedback. In the remainder of this section, we examine the relationship of each of these processes to peer assessment. To foreshadow our framework, the ways in which stronger and weaker writers engage with these processes during providing feedback will shape what learning opportunities they have from providing feedback. Thus, we consider ability effects in our discussion of these processes.

Defining a task involves developing a deeper understanding of the task. This understanding may include writing and revising goals, which features in the paper need to be attended, and how the revision process should be approached (Flower et al. 1986; Hayes et al. 1987). Lower-ability writers begin the writing task, and more specifically the revision process, at a disadvantage because they do not develop appropriate task definitions. They approach revision thinking they merely need to correct errors. This perspective limits them to a very narrow view of the paper, and they often overlook the global goal of communicating their point to a larger audience. Therefore, lower-ability writers must learn what needs to be attended to in their papers. A well-developed peer assessment task could offer guidance. By providing well-defined criteria for providing feedback, students could gain an understanding of what features are important. Very little instruction is necessary to change what a writer focuses on in a reviewing task. For example, lower-ability writers often fail to make global revisions, but after eight minutes of instruction on making global revisions, students were able to make significantly more global revisions than those who did not receive this instruction (Wallace and Hayes 1991).

Detecting a problem involves perceiving differences between the text produced so far and the intended text (Flower et al. 1986; Hayes et al. 1987). Two explanations for possible difficulties with problem detection have been offered. In addition to inadequate task definitions, writers may be working with an inaccurate representation of their text. Writers in general have difficulty perceiving errors in their own writing compared to others' texts because while reading one's own text, errors are often automatically mentally corrected (Flower et al. 1986). Peer assessment reduces the problem of inaccurate representations of the text by having students evaluate peers' texts, and thus students have better opportunities to practice detecting problems.

Diagnosing a problem involves creating a representation of the problems detected (Flower et al. 1986; Hayes et al. 1987). Although diagnosis is not essential to revision, it is a preferred step for higher-ability writers. It is especially helpful when presented with an ill-defined problem or a problem in which the appropriate revision strategy is not obvious. Lower-ability writers tend to avoid diagnosing the problem, and instead choose to just delete or rewrite the problematic text. By not exploring the nature of these problems, lower-ability writers are limited in their knowledge about the kinds of writing problems that occur, making it more difficult to detect and solve these problems. By providing feedback to peers, students are forced to diagnose problems, which could not only increase their knowledge about a particular problem but also increase their awareness of the kinds of writing problems.

Strategy selection involves reacting to a detected problem (Flower et al. 1986; Hayes et al. 1987). In order to revise text, writers need to first decide which problems to solve and then choose which strategy to apply. Already limited in the number of problems detected, especially global problems, lower-ability writers utilize fewer revision strategies. Moreover, they often choose less effective strategies (e.g., delete problematic text). By providing feedback to peers, students are forced to come up with solutions to problems that if found in their own paper, they may normally ignore or address by just deleting the text. Through this practice of solution generation, they may discover new strategies for revision.

In sum, there are many overlapping cognitive processes between the task of revising and the task of providing feedback to peers. By taking an identical elements approach to explain why and under which circumstances providing feedback promotes writing ability, we frame the benefits of providing feedback in terms of what is being practiced (i.e., what feedback was produced). Nelson and Schunn (2009) identified several features of feedback that frequently vary across peer reviews, and these features frequently are influenced by writer ability (Patchan et al. 2009) and document quality (Patchan et al. 2013). Here we discuss these elements and their relationship to the component processes of reviewing and revising, and predicted effects on learning.

At the most basic level, feedback can be categorized as praise or criticism. Praise comments identify what a learner does well. Although they are often recommended as part of the best practices for feedback provision (Roediger 2007), the effect of receiving praise comments on performance tends to be quite small (Cohen's $d = .09$, Kluger and DeNisi 1996). Despite these weak learning effects of receiving praise, providing praise may be effective for assessors. By identifying what a peer does well, students may reinforce known successful strategies or discover new successful strategies. Lu and Zhang (2012) examined the kinds of comments that students provided to their peers and how they related to the performance on their own projects. Although a significant relationship between praise and performance was not found, the comments categorized as praise were very broad and vague (e.g., "Good job."). When examining more specific types of praise, Cho and Cho (2011) found that providing global praise about high-level writing (i.e., across multiple paragraphs) was positively related to the quality of texts. These results suggest that praise should be specific, global, and at a high-level to be effective.

Criticism comments identify where a learner needs improvement. By providing criticism comments, students have an opportunity to practice specific revision skills, such as detecting problems, diagnosing the problem, and selecting appropriate strategies to solve the problem. Consistently, the construction of criticism comments was positively related to student performance across several studies (Cho and Cho 2011; Inuzuka 2005; Li et al. 2010, 2012; Lu and Zhang 2012; Topping et al. 2013). However, similar to receiving feedback, not all types of criticism are equally effective.

Feedback specificity is often considered when evaluating the effectiveness of criticism. Feedback specificity varies along a continuum with outcome feedback only (i.e., whether an action was correct or incorrect) at one end of the continuum and highly specific feedback (i.e., describing problems and suggesting solutions) at the other end of the continuum. Receiving specific comments has been found to be more effective than receiving less specific comments (Ferris 1997). Because the construction of more specific comments involves practicing skills that are associated with revision, providing more specific comments is also expected to be more effective for the assessor. Several studies supported this hypothesis—students learned more after providing elaborated feedback that included descriptions of problems and scaffolded solutions (Li et al. 2010, 2012; Topping et al. 2013). However, these studies did not address whether providing criticism comments that describe problems is equally effective as providing criticism comments that offer solutions.

Not only are specific revision skills being practiced (i.e., identifying problems, diagnosing problems, suggesting solutions) when constructing peer feedback, the focus of the feedback (i.e., prose vs. domain content) could also affect whether gains will be seen in writing ability or content knowledge. Inuzuka (2005) identified several different categories in which students' tend to focus their comments. These categories varied from low-level prose issues (i.e., language usage) to high-level prose issues (i.e., coherence) and substance issues (i.e., factual errors). In the Inuzuka study, students whose comments focused on a variety of issues improved their writing more than those who only focused on a few different categories. The focus of the feedback provided is likely to affect what knowledge is reinforced or gained—that is, providing feedback about low-level prose will likely increase knowledge about low-level writing issues and providing feedback about high-level prose will likely increase knowledge about high-level writing issues. Therefore, it is important to examine which types of focus are the most beneficial to the provider. Cho and Cho (2011) examined the effectiveness of providing feedback about low-level writing issues versus high-level writing issues. The amount of high-level feedback provided to peers positively influenced the quality of the provider's own text, but only when the feedback was about issues contained within a single paragraph rather than across multiple paragraphs. Research examining the effect of providing feedback about substance issues is still needed, especially since this type of focus is likely to be related to writing-to-learn (i.e., comments about substance provide opportunities to increase domain knowledge). In particular, identifying and diagnosing problems in claims about disciplinary substance can involve unpacking reasons behind content knowledge, perhaps serving as a form of self-explanation, which has generally been found to be an effective method of learning disciplinary content knowledge (Chi 1996; Chi et al. 1989).

To understand why students learn from peer assessment, and more specifically from providing feedback to peers, we first identified conceptually identical elements in both writing and providing feedback tasks: (1) defining the task, (2) detecting problems, (3) diagnosing problems, and (4) selecting a revision strategy. Then we identified several types of peer feedback associated with positive effects on learning that could be used to examine the ways in which stronger and weaker writers might engage with these processes while providing peer feedback: (1) specific, global, and high-level praise, (2) elaborated feedback that describes the problem and scaffolds solutions, and (3) focusing on high-level writing issues. Several gaps in the literature were also found: (1) whether describing a problem is equally effective as offering a solution, and (2) whether focusing on substantive issues improves content knowledge.

Possible moderators of the effectiveness of providing feedback

Because peer assessment activities take place in different contexts and can be differentially structured, it is important for improving instruction through peer assessment to understand how students' practice opportunities can differ. As foreshadowed in much of the prior section, two salient factors will likely influence what feedback is produced: the ability of the reviewer and the quality of the texts to be reviewed (see Fig. 1).

As previously mentioned, Hayes et al. (1987) have identified many differences between higher-ability and lower-ability writers. Higher-ability writers (1) are able to detect more problems, (2) are more likely to attend to global issues, (3) are more likely to choose effective strategies for revisions, and (4) have more strategies for dealing with global issues. These skill differences are likely to affect what a student is able to detect, diagnose, and propose as solutions, and thus the features of the feedback provided are expected to vary by writer ability. Because higher-ability writers are expected to be able to detect and diagnose more problems and offer more solutions than lower- ability writers, they are likely to provide more criticism comments, and these comments should include more descriptions of the problems as well as offer more solutions. Therefore, higher-ability writers who provide feedback to peers will be referred to as *high reviewers*, and lower-ability writers who provide feedback to peers will be referred to as *low reviewers*. Not surprisingly, students' initial writing ability positively influenced the amount of comments they provided on local and global, high-level writing issues (Cho and Cho 2011).

Furthermore, these writer ability differences are likely to affect the quality of the texts to be reviewed—that is, papers written by higher-ability writers (i.e., *high-quality texts*) are likely to have more positive qualities and fewer problems than papers written by lower-ability writers (i.e., *low-quality texts*). In turn, a learner who is providing feedback to peers is expected to construct more praise comments and fewer criticism comments for
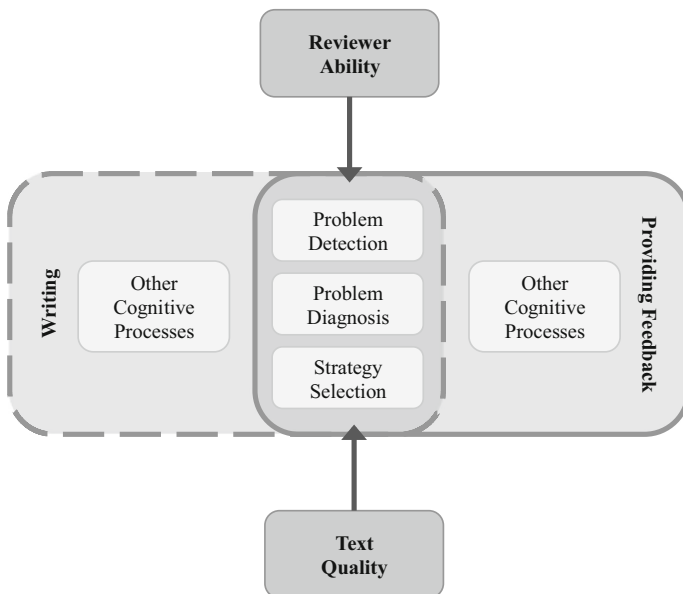

Fig. 1 Identical Elements Theory of providing feedback

high-quality texts than low-quality texts. Indeed, the quality of the peers' texts positively influenced the amount of local and global, high-level praise as well as negatively influenced the amount of global, high-level criticisms (Cho and Cho 2011).

What remains unknown is whether both high reviewers and low reviewers are equally able to distinguish the high-quality texts from low-quality texts. Therefore, it may be important to consider the interaction between reviewer ability and text quality (Patchan et al. 2013). Interviews and surveys with students and instructors revealed a common belief that the benefits of peer review are inherently asymmetrical across skill: stronger students must provide feedback to weaker students (Kaufman and Schunn 2011). In support of this belief, Patchan et al. (2013) found that high reviewers provided more feedback (i.e., provided more criticism in general, detected more problems, offered more solutions, commented more often on low prose issues and content issues) to low-quality texts than high-quality texts. Furthermore, the feedback provided by high reviewers was more effective for the low-quality texts than the high-quality texts. However, low reviewers did not differ in the amount or effectiveness of the feedback provided to low-quality texts and high-quality texts. These results might suggest that high reviewers were better able to distinguish between the high-quality and low-quality texts than were low reviewers. However, another interpretation seems more likely given the pattern of comments produced by instructors, who most certainly can distinguish between different quality texts. In the study by Patchan et al. (2009), only a writing instructor, and not a content instructor, detected and diagnosed more explicit problems in the low-quality texts than the high-quality texts. It is doubtful that the content expert was not able to distinguish the high-quality texts from the low-quality texts, so this pattern of results for low versus high ability reviewers is more likely to reflect the commenting style associated with specific levels of expertise. In general, this study makes salient the possibility that the content of reviews can be driven by beliefs about review content. If high reviewers and low reviewers have different beliefs about what constitutes a review, this difference would influence what they produce and hence learn from peer review.

It is also important to note that, in the Patchan et al. (2013) study, the multi-peer nature of peer feedback was not taken into account. Students reviewed multiple papers of varying quality at once, but the analyses treated the reviews of each text as independent. This assumption is likely flawed, as both the content of the reviews and the overall learning will likely be influenced by the contrast across texts. Without shared quality anchors provided by having to review both high-quality texts and low-quality texts, beliefs about the content of reviews may play a larger role in shaping the content of reviews. What is needed is a study in which students review only papers of a given quality to precisely estimate the effects of reviewer ability and text quality on the process of providing feedback.

The current study

The goal of the current study is to provide a first step towards a theoretical understanding about why students learn from peer assessment, and more specifically from providing feedback to peers. The Identical Elements Theory will be used as a framework to motivate why the nature of what is being produced during reviews is important for learning. Previous research has demonstrated that although reading and rating the quality of peers' texts are important activities in peer assessment, providing feedback appears to be the most effective activity. Moreover, the various features of feedback provided (i.e., type of

feedback: praise vs. criticism; features of criticism: problems vs. solutions; focus of feedback: low prose vs. high prose vs. substance) influence the effectiveness of providing feedback. Finally, reviewer ability and text quality moderate this effect. To extend these findings, the current study will systematically examine how reviewer ability and text quality jointly affect the kinds of comments produced using data from a new context in which students were specifically assigned to review papers of similar quality.

Method

Overview

The current study was part of a larger study that examined multiple aspects of why students learn from peer assessment, including the relative effectiveness of different forms of peer feedback (Patchan and Schunn, under review) and the benefits of *receiving* feedback for the author (Patchan et al. under review), in contrast to the current focus on the benefits of *providing* feedback for the reviewer. In order to describe the extent to which reviewer ability and text quality affect peer assessment, we determined the writing ability of each participant and then manipulated which participants were assigned to each document according to groups of reviewer ability and text quality. In other words, in a 2 × 2 between-subjects design, groups of participants of higher-writing ability (i.e., high re-viewers) or lower-writing ability (i.e., low reviewers) each provided feedback to either only groups of peers with higher-writing ability (i.e., high-quality texts) or only groups of peers with lower-writing ability (i.e., low-quality texts). To examine how reviewer ability and text quality affected peer assessment, the amount, features, and focus of comments provided were compared across the conditions.

Course context

This study was conducted in an Introduction to Psychological Science course at a large, public research university in the southeast United States. The specific class and assignment context was selected to represent an authentic writing assignment that occurred in a large, content course as part of the writing in the discipline (WID) program. This course was a popular general education course that students commonly took to meet one of their social science requirements. In addition, it was compulsory for not only all psychology majors, but also for a number of other majors as well, including education and nursing. Because this course was very large (i.e., 838 students), three sections were offered, each taught by a different lecturer. Students were also required to attend one of 24 different lab sections taught by 12 graduate student teaching assistants (TAs).

Participants

From the 838 students enrolled in the class, 432 were selected to participate in a study of matching versus mismatching reviewer-author ability pairings, and others participated in a different experimental manipulation. Coding was done exhaustively for all reviews re-ceived on a feasible subset of the 432 documents to optimize on analyses of the effects on authors. Documents were selected on the basis of maximizing availability of supporting data (e.g., completion of surveys by the author). The current manuscript reanalyzes that

large coded dataset from the perspective of reviewers, rather than from the perspective of authors. Included in the current analyses are 186 reviewers who completed four reviews and all four reviews were in the coded dataset. As author-reviewer mappings were random within the ability grouping, included reviewers are essentially a random subset of the original 288 participating in the experimental manipulation. Indeed, the selected 186 versus the excluded 244 participants were not significantly different (overall or within each of the four conditions) in mean paper quality, number of documents submitted, or number of surveys completed. Note that the Ns are slightly different across conditions due to by chance variation in how many reviewers met the all-four-reviews-coded criterion.

This sample represented students (77 % female) at all undergraduate years with a predominance of less advanced students (i.e., 57 % freshmen, 26 % sophomores, 10 % juniors, 5 % seniors, and 2 % other) as well as a great variety of majors (i.e., of the declared majors: 31 % social sciences, 28 % natural sciences, 18 % engineering, 12 % education, 7 % computer science, and 4 % business).

Design

A 2 × 2 between-subjects design was used. Reviewer ability and text quality were based on the participants' writing ability. First, the participants' writing ability was determined by a composite of four self-reported ability measures—that is, the average z-scores (i.e., student's score minus group mean divided by group standard deviation) of SAT verbal,[1] SAT writing, the final grades in the first and second semester composition courses.[2] This combination of measures provided a more generalizable ability measure that one can also obtain easily for future research or practical applications in the classroom.

Next, a median split was used to determine which students had higher writing ability and which students had lower writing ability. Indeed, relative to the U.S. ability standards, the two groups were above and below median performance levels (The College Board 2012). Further, there were grouping differences of 2.8 standard deviations (i.e., a very large effect size) on the composite measure, and there were also large group differences on each of the components of this composite measure (see Table 1). To further validate the composite measure, two writing experts (i.e., rhetoric graduate students with extensive writing teaching experience) rated the quality of the students' first drafts using a 5-point scale on eight dimensions focused on the flow, logic, and insight of the papers. The average score across all dimensions (i.e., min = 1; max = 5) was compared between the high writers and low writers. An independent t test revealed a significant difference in writer ability: the high writers ($M = 2.4$, $SD = .5$) produced higher quality first drafts than the low writers ($M = 2.0$, $SD = .4$), $t(187) = 4.74$, $p < .0001$, $d = .7$.

Finally, students with higher writing ability were considered high reviewers and high-quality texts, and students with lower writing ability were considered low reviewers and producing low-quality texts. These classifications were used to create four conditions: a high reviewer who reviewed a high-quality text ($n = 44$), a high reviewer who reviewed a low-quality text ($n = 46$), a low reviewer who reviewed a high-quality text ($n = 48$), and a

---

[1] The SAT (Scholastic Assessment Test) is a standardized test used for college admissions in the United States. It consists of three sections: the verbal section tests critical reading skills, the writing section tests problem detection skills and grammar and usage knowledge, and the mathematics section tests arithmetic operation, algebra, geometry, statistics, and probability knowledge.

[2] Universities in the U.S. typically require a first year composition course, and the university in the present study requires two semesters of composition.

Table 1 Summary of demographic and ability data by writer ability

	High-ability writer			Low-ability writer			t test	
	n	*M*	*SD*	*n*	*M*	*SD*	*p*	*d*
Demographics								
Gender[a]	93	75 %		95	78 %		.67	
Year in school[b]	93	84 %		96	82 %		.77	
Age	93	18.8	1.5	96	19.0	1.8	.45	
Ability measures								
Writing ability z-score	93	.64	.49	96	−.70	.47	<.0001	2.8
SAT verbal	79	599	55	70	489	55	<.0001	2.0
SAT writing	76	592	63	69	494	58	<.0001	1.6
1st semester grade[c]	70	4.4	.7	65	3.3	.6	<.0001	1.7
2nd semester grade[c]	49	4.2	.7	51	3.1	.6	<.0001	1.7

[a] % female

[b] % freshman + sophomore

[c] Composition grades were coded on a 5-point scale: 5—placed out; 4—A, 3—B, 2—C, 1—D or below. Missing data points included participants who did not take the composition course because it was not a required course ($n_{2nd\ semester} = 1$) and participants who were currently taking the course or will take it in the future ($n_{1st\ semester} = 54$; $n_{2nd\ semester} = 88$)

low reviewer who reviewed a low-quality text ($n = 48$). Although this method was not the most precise way to define reviewer ability and text quality, it was pragmatically required for creating the reviewing groups for this study and in future instructional applications. This decision decreases the power of this study, which could result in missing some relevant data patterns. However, there is little chance of making false claims, and the overall large number of participants means that the instructionally important patterns will generally be detectable. We believe that a lower powered study was a reasonable tradeoff for higher external validity (i.e., how reviewer ability would typically be determined).

The dependent variables included the draft quality improvement, number of comments received for each feature and focus, the number of implemented comments, and the quality of the revisions based on a peer's comment as described in the "Coding Process" section.

Procedure

Participants completed three main tasks: (1) wrote a first draft, (2) reviewed peers' texts, and (3) revised own text based on peer feedback. At the end of the first month of the semester, participants had 1 week to write their first draft and submit it online using the web-based peer review functions of turnitin.com.[3] For this task, they were expected to write a three-page paper in which they evaluated whether MSNBC.com, a US digital news

---

[3] The turnitin.com peer review functions primarily focused on generating end comments rather than marginalia. Reviewers were able to tag specific locations in the text that could be used in the end comment to indicate where a particular problem existed; however, this function was not obvious and most students did not use it. In addition, the specific commenting prompts were separate from the ratings prompts, which could allow one to create a reviewing assignment that utilized more fine-grained evaluation dimensions and broader commenting dimensions. Finally, the reviews were anonymous—that is, a pseudonym was used to identify both the writer and the reviewer.

provider, accurately reported a psychological study—applying concepts from the Research Methods chapter covered in lecture and lab in the prior week. After the first draft deadline passed, participants were assigned four papers to review based on the text quality condition they were assigned. Participants were able to access the peer feedback online once the reviewing deadline had passed. The participants were given 1 week to revise their draft based on the peer feedback. After each of the writing and reviewing tasks, participants completed a short survey about their experience.

The TAs and lecturers were available to answer questions and offer feedback to students if more help was requested. However, most students did not take advantage of this opportunity. The TAs also provided final grades for the paper.

Review support structures

Participants were provided with a detailed rubric to use for the reviewing task. The rubric included commonly-used general reviewing suggestions (e.g., be nice, be constructive, be specific) and specific guidelines, which described the three reviewing dimensions that have been applied in many disciplinary writing settings: flow, argument logic, and insight. For each commenting dimension, a number of questions were provided to prompt the reviewer to consider the paper using several particular lenses. The flow dimension focused on whether the main ideas and the transitions between the ideas were clear (e.g., Did the writing flow smoothly so you could follow the main argument? Did you understand what each of the arguments was and the ordering of the points made sense to you?). The argument logic dimension focused on whether the main ideas were appropriately supported and whether obvious counter-arguments were considered (e.g., Did the author just make some claims or did the author provide some supporting arguments or evidence for those claims? Did the author consider obvious counter-arguments, or were they just ignored?). The insight dimension focused on whether a perspective beyond the assigned texts and other course materials was provided (e.g., Did the author just summarize what everybody in the class would already know from coming to class and doing the assigned readings, or did the author tell you something new? Did the author provide an original and interesting alternative explanation?). The purpose of these specific guidelines was to direct the participants' attention primarily towards global writing issues (Wallace and Hayes 1991).

Finally, participants rated the quality of the papers using a 5-point scale (1–'Very Poor' to 5–'Very Good'). They rated six aspects of the paper within the three commenting dimensions of flow (i.e., how well the paper stayed on topic and how well the paper was organized), argument logic (i.e., how persuasively the paper made its case, how well the author explained why causal conclusions cannot be made from correlational studies, and whether all the relevant information from the research article was provided), and insight (i.e., how interesting and original the paper's conclusion was to the reviewer). For each rating, participants were given descriptive anchors to help with determining which rating was most appropriate.

Coding process

The feedback was coded to determine how the amount and type of comments varied as a function of reviewer ability and text quality. The coding scheme originally established by Nelson and Schunn (2009) was used to categorize the types of comments, with minor revisions about how the *type of feedback* was coded (i.e., praise, problem, and solution were considered independent features rather than mutually exclusive). Pairs of

undergraduate research assistants (RAs) coded all of the comments—Kappa values for exhaustive coding are presented for each dimension.

First, the feedback was segmented by idea unit into comments because reviewers frequently commented about multiple issues within one dimension (e.g., transitions, use of examples, word choice). A total of 8288 provided comments were coded and analyzed (see Appendix 1 for definitions and examples of each code). Second, each comment was coded for the presence/absence of three independent features: praise, problems, and solutions (Kappa = .92, .88, .92, respectively). Finally, all comments that were previously coded as either problem or solution (i.e., criticism comments) were coded for the presence/absence of localization (Kappa = .63; percent agreement was 92 %) and the focus (i.e., low prose, high prose, or substance—Kappa = .54; percent agreement was 78 %). Many issues can involve both high prose and substance; these comments were always coded as substance. Figure 2 illustrates the relationship between the feedback provided, segmented comments, and the types of feedback coded. An example of how one piece of feedback was segmented and coded can be found in Appendix 2.

Results AND discussion

Overview

The goal of the current study is to provide a first step towards a theoretical understanding about why students learn from peer assessment, and more specifically from providing feedback to peers. We systematically examined how reviewer ability and text quality jointly affect the kinds of comments produced. Each dependent variable (i.e., number of comments for each type, feature, and focus) was analyzed using a 2×2 between-subjects ANOVA with reviewer ability (i.e., high reviewers vs. low reviewers) and text quality (i.e., high-quality texts vs. low-quality texts) as between-subjects independent variables. In order to interpret how the learning opportunities may differ by reviewer ability and text quality, the unit of analysis was at the participant level—that is, the number of comments provided by each participant was summed. To tease apart the simple effects from significant interactions, independent t tests were performed comparing high-quality texts to low-quality texts for high reviewers and low reviewers separately.

Only results that were significant at $p < .05$ will be discussed in detail in the text. All descriptive and inferential statistics are reported in Appendix 3. As an indicator of effect size, eta squared (i.e., η^2—proportion of variance in the dependent variable accounted for by the independent variable(s) while controlling for other possible variables) was included for all ANOVAs—an η^2 of .01 is considered small, .06 is medium, and .14 is large (Cohen, 1988), and Cohen's d (i.e., mean difference divided by average standard deviation) was included for all t tests—typically, a Cohen's d of .3 is considered small, .5 is medium, and .8 is large (Cohen, 1977).

As an advance summary, there were several main effects of writer ability. High reviewers were more likely to construct comments that led to learning how to write better (i.e., practiced describing problems and offering solutions about high prose and substance issues). In general, neither the high reviewers nor the low reviewers produced different amounts of various comments between the high-quality texts and low-quality texts. There was one significant interaction between reviewer ability and text quality: although low reviewers did not differentiate in the amount of problems described in high-quality texts

and low-quality texts, high reviewers described more problems in low-quality texts than high-quality texts.

Amount of feedback

Overall, reviewer ability and text quality did not affect the amount of feedback provided by the students. The number of comments high reviewers ($M = 43.6$, $SD = 12.6$) provided was similar to the number of comments low reviewers ($M = 45.5$, $SD = 14.6$) provided, and these amounts did not differ by text quality. Similarly, the length of high reviewers' comments ($M = 829$, $SD = 362$) and the length of low reviewers' comments ($M = 778$, $SD = 291$) were not significantly different, and these amounts did not differ by text quality. The lack of an effect on the number of comments and the length of comments is convenient for in-depth analyses of these comments because a correction for amount or length is not needed. However, there were interesting differences in the content of these comments.

Type of feedback

First, we observed differences in the frequency of comments about things done well in the paper (i.e., praise) and comments about things that were wrong with the paper (i.e., criticism). Only reviewer ability affected the type of feedback provided (see Fig. 3). Low reviewers ($M = 30.8$, $SD = 12.8$) provided more praise than high reviewers ($M = 26.0$, $SD = 9.8$), $F(1, 182) = 8.17$, $p = .01$, $\eta = .04$. By contrast, high reviewers ($M = 20.0$, $SD = 12.1$) provided more criticism than low reviewers ($M = 16.2$, $SD = 8.1$), $F(1, 182) = 6.65$, $p = .01$, $\eta = .04$.

Surprisingly, these amounts did not differ by text quality. High-quality texts would likely have more things to praise, and low-quality texts would likely have more things to criticize. However, neither the high reviewers nor the low reviewers distinguished the quality of the texts in this way.

Together these results suggest that the amounts of praise and criticism are not influenced by an ability to detect problems in a text or differences between expected and perceived text quality because each of those factors would have predicted either main effects of text quality or interactions between reviewer ability and text quality. Rather, these results suggest that the amounts of praise and criticism may reflect general beliefs towards feedback content associated with reviewer ability (i.e., how praise-oriented or criticism-oriented feedback should generally be).

Features of Criticism

Next, we observed differences in the frequency of the criticism features—that is, comments that describe the problem or offer a solution. Although reviewer ability did not affect the presence of problems and solutions in a single comment, reviewer ability did affect how often students described a problem only or offered a solution only (see Fig. 4a). High reviewers ($M = 7.8$, $SD = 7.3$) offered more solutions than low reviewers ($M = 5.9$, $SD = 5.0$), $F(1, 182) = 4.38$, $p = .04$, $\eta = .02$, and high reviewers ($M = 8.8$, $SD = 7.0$) described more problems than low reviewers ($M = 7.1$, $SD = 4.8$), $F(1, 182) = 3.86$, $p = .05$, $\eta = .02$.

However, the effect of reviewer ability on the frequency of the problems described was driven by a significant interaction with text quality (see Fig. 4b). Specifically, low
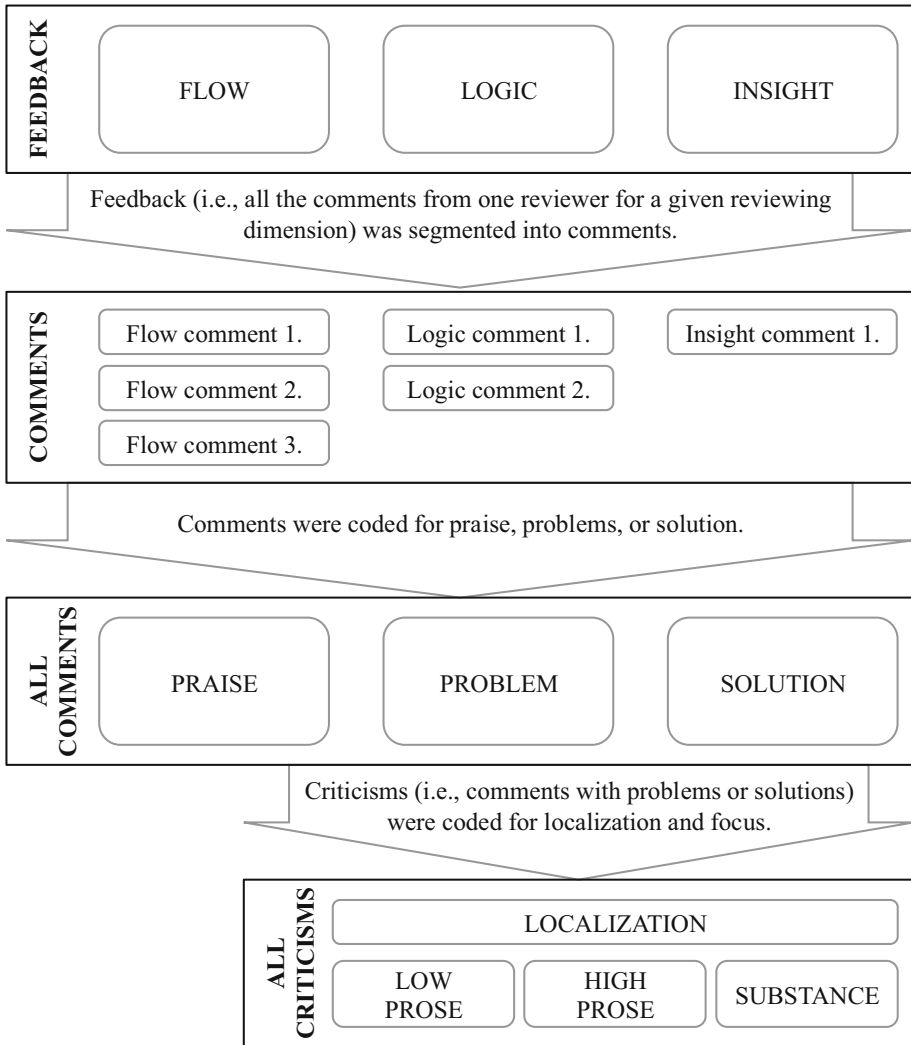
Fig. 2 Coding process

reviewers did not differ in the amount of problems described in the low-quality texts ($M = 6.9$, $SD = 4.7$) and high-quality texts ($M = 7.3$, $SD = 5.0$), and high reviewers described more problems in the low-quality texts ($M = 10.2$, $SD = 7.5$) than the high-quality texts ($M = 7.3$, $SD = 6.2$), $F(1, 182) = 3.75$, $p = .05$, $\eta = .02$. These results indicate that there may be a difference in the focus of problems that low-quality texts tended to have (i.e., problems with obvious solutions, so they only needed the problem described). By contrast, the simple main effect of reviewer ability on number of solutions likely reflects an expectation that solutions should be offered, rather than the ability to offer solutions, or else there would have been an interaction of reviewer ability and text quality. However, the nature of the problems being addressed may differ by reviewer or text quality, complicating this interpretation and is therefore considered next.

Focus of criticism

Finally, we observed differences in the frequency of comments focused on high prose issues and comments focused on substance issues. Again, only reviewer ability affected the focus of criticism (see Fig. 5). High reviewers ($M = 9.8$, $SD = 6.1$) provided more high prose comments than low reviewers ($M = 8.2$, $SD = 4.2$), $F(1, 182) = 4.88$, $p = .03$, $\eta = .03$. High reviewers ($M = 7.7$, $SD = 6.7$) also provided more substance comments than low reviewers ($M = 5.6$, $SD = 4.7$), $F(1, 182) = 6.31$, $p = .01$, $\eta = .03$. Similar to the type of feedback, neither the high reviewers nor the low reviewers distinguished the text quality by identifying more high prose or substance issues in the low-quality texts than the high-quality texts. This continued pattern of main effects of reviewer ability without effects or interactions with text quality again suggest the effects are based on personal beliefs of what feedback should include that was associated with reviewer ability rather than objective frequency of problems or ease at which problems can be detected.

Interestingly, both the low reviewers and the high reviewers provided the same number of low prose comments for low-quality texts and high-quality texts. Again, this lack of a difference by text quality is likely to result from a general commenting style associated with reviewer ability. The general focus on high prose was likely influenced by the re-viewing assignment; these students were instructed to only comment on low prose issues if they disrupted understanding of the paper. Therefore, students rarely commented on low prose issues ($M = 2.5$, $SD = 3.4$).

General discussion

Summary of results

The goal of the current study is to provide a first step towards a theoretical understanding about why students learn from peer assessment, and more specifically from providing feedback to peers. By systematically examining how reviewer ability and text quality jointly affect the kinds of comments produced, we were able to provide a more detailed look at the ways in which the peer review task will influence what students learn from providing feed-back to peers by. Although reviewer ability and text quality did not affect the amount of feedback provided (i.e., number of comments and length of comments), there were interesting effects on the content of the feedback. In general, there were several significant main effects of reviewer ability. Low reviewers provided more praise than high reviewers. By contrast, high reviewers provided more criticism than low reviewers. This criticism described more problems and offered more solutions. Furthermore, this criticism also focused more often on high prose and substance. There was one interesting interaction between reviewer ability and text quality—that is, high reviewers described more problems in the low-quality texts than in the high-quality texts, whereas low reviewers did not make this distinction.

Possible moderators of the effectiveness of providing feedback

Variations in commenting styles were observed with different levels of expertise (Patchan et al. 2009). Accordingly, the use of different commenting styles may result in different amounts of practice. Therefore, one possible moderator of the effectiveness of providing feedback examined in the current study was reviewer ability. High reviewers were

expected to be able to detect more problems, focus more often on high-level issues, possess more solutions to these problems, and better select the most effective solutions than low reviewers. Indeed, the results of the current study supported these expectations. However, these findings differed from the Patchan et al. (2009) study, which found that high reviewers only provided more feedback to low-quality texts. This study differed from the current study in one important way: the papers to be reviewed were randomly assigned to each writer, which resulted in reviewing both high-quality texts and low-quality texts. The different levels of quality was likely to be more apparent when so closely contrasted in time, and therefore the features of the comments were affected by this distinction. On the other hand, participants in the current study only reviewed high-quality texts or low-quality texts, so the contrast between the different levels of quality was not as evident. Taking the two studies together, it appears that relative quality more than absolute quality seems to drive comment content.

Another expected moderator of this learning effect examined in the current study was text quality. The quality of the paper being reviewed was expected to affect how much practice is available to a reviewer—that is, low-quality texts presumably have more problems than high-quality texts and thus provide more opportunities for problem detection, diagnosis, and selection of appropriate solutions. Surprisingly, no significant effects of text quality were found. Do these results indicate that the students were not able to distinguish between the low-quality texts and high-quality texts? Not necessarily. Even expert writers do not always describe more problems in low-quality texts than high-quality texts (Patchan et al. 2009). These results more likely reflect the writer's style of commenting. More specifically, certain features of feedback (e.g., describing problems) are considered important regardless of the quality of the paper, and consequently those features will likely occur equally often in feedback for low-quality texts and high-quality texts. The question about whether low-quality texts can offer more opportunities to practice revision skills than high-quality texts is still unanswered. Future research can address this question by focusing the students' task definition on identifying, describing, or solving as many problems as they can find throughout the papers. In doing so, one can then observe whether text quality affects the features of the feedback produced.

Theoretical contributions

Students consistently benefit more from providing feedback than any of the other reviewing activities during peer-review (Lu and Law 2012; Wooley et al. 2008). To frame why providing feedback in general, and constructive criticism in particular, is likely to help students develop their writing ability, we developed a framework using the Identical Elements Theory (Thorndike and Woodworth 1901; Singley and Anderson 1989). More specifically, we identified several elements that overlap across writing and providing feedback tasks—that is, in both writing tasks and while constructing feedback, students must detect problems and diagnose those problems or select appropriate solutions. This practice of revision skills while constructing feedback may be an important contributor to why students learn from the process of providing feedback to peers. Several theories of cognition recognize that skills can be acquired and refined by simply practicing the skill (Anderson et al. 2004; Logan 1988; Newell 1994; Newell and Rosenbloom 1981). Through practicing revision skills, students could strengthen their ability to detect, diagnose, and solve these problems, resulting in faster and more efficient retrieval of information about these problems while writing in the future. In other words, a theoretical contribution of the

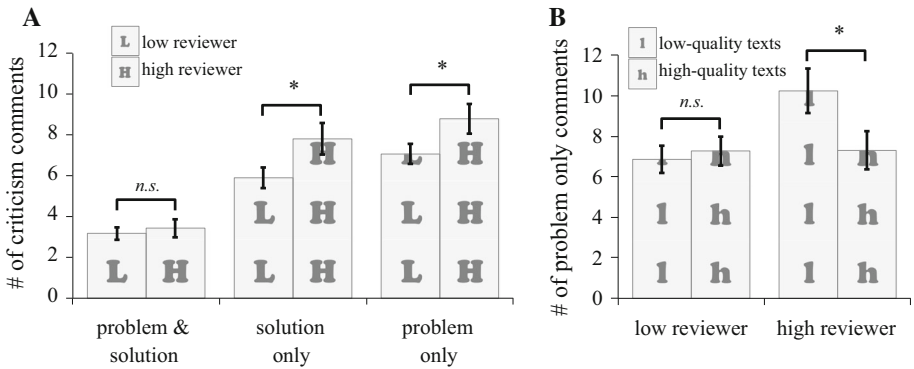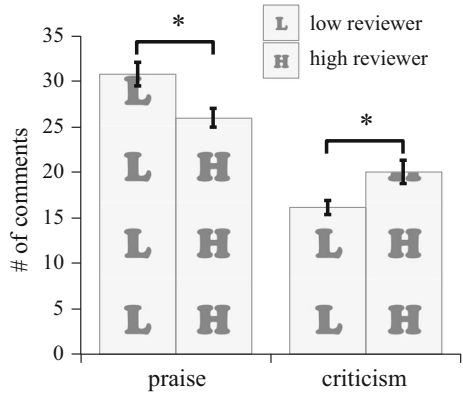**Fig. 3** Amount of each type of feedback as a function of reviewer ability

Fig. 4 **a** Amount of criticism features as a function of reviewer ability. **b** Amount of problem only comments as a function of reviewer ability and text quality

current work is to frame reviewing-to-learn as practice opportunities under an Identical Elements framework.

The purpose of examining the effects of reviewer ability and text quality was to describe how the practice opportunities might differ for individual students. Thus, we suggest that theories of reviewing-to-learn must consider the significant variation that occurs as a

Fig. 5 Amount of criticism focus as a function of reviewer ability
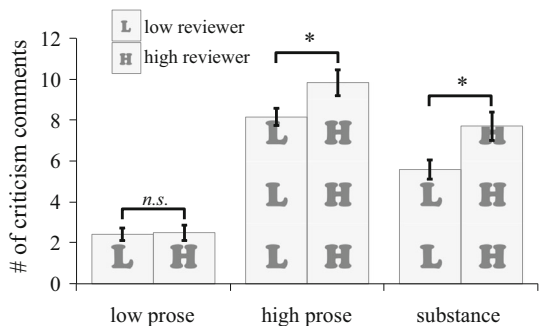
function of the relative (not absolute) quality of the texts being reviewed. More specifically, high reviewers provided more criticism that described problems and offered solutions about high prose and substance issues, and as a result, these students likely strengthen their revision skills more than the low reviewers. By systematically assigning only papers of a particular quality, the current study did a more thorough job of examining the effects of reviewer ability and text quality than the Patchan et al. (2013) study.

Caveats and future directions

There are a few caveats to these findings that must be considered. First, several method-ological decisions could have affected the power of this study. Given the instructional context of the current study, all students' texts needed to be reviewed regardless of their quality. Furthermore, students needed to be assigned peers' papers to review shortly after the deadline for the writing assignment. In order to accommodate these pragmatic issues, as well as for future instructional applications, we utilized an indirect measure of writing ability as a proxy for reviewer ability and text quality. In addition, we categorized students as high reviewers and low reviewers and texts as high-quality texts and low-quality texts by using a median split of the writing ability measure. Therefore, we may have missed some relevant data patterns because these decisions lowered the power of the study. Although we believe that a lower powered study was a reasonable tradeoff for higher external validity, future research should examine these measures more closely. For research purposes, direct measures of reviewer ability and text quality should be chosen, and for pragmatic purposes, the indirect measures should be validated.

Another caveat relates to the generalizability of these findings. One of the goals of the current study was to extend the results of the Patchan et al. (2013) study by systematically assigning only papers of a given quality to precisely estimate the effects of reviewer ability and text quality on the process of providing feedback. Given that the results of the current study differed from the Patchan et al. study, high reviewers may only provide more feedback overall if they are assigned papers of similar quality. Future research should more closely examine how a mix of quality changes the feedback provided by peers. Additionally, the peer review process was anonymous—that is, students did not know whether the texts they were reviewing came from high-ability writers or low-ability writers. The feedback provided by peers may differ if students know whose paper they are reviewing.

Finally, future research should consider the impact of these feedback features on learning—that is, do certain features promote learning more than others? Nelson and Schunn (2009) found that feedback with certain features (i.e., summary, solutions, localization) was more likely to be implemented. Future research should further examine whether the focus of feedback (i.e., low prose, high prose, substance) affects the implementation rate, and more importantly whether implementing specific types of feedback increases one's ability to write in the future. Furthermore, future research should determine whether increasing practice opportunities (i.e., the amount of problems described or solutions offered) is sufficient for learning or whether the specific problems being described or solved (i.e., describing or solving a problem that one also struggles with) has an impact on learning.

Practical implications

Based on the findings from the current study, students are likely to benefit equally from providing feedback to high-quality texts and low-quality texts as long as all the papers they

review are of the same quality. However, the level of student (i.e., high reviewer vs. low reviewer) could affect how much students benefit from providing feedback. Because high reviewers are likely to describe more problems and offer more solutions of both high prose issues and substantive issues than low reviewers, instruction with extra scaffolding may be necessary to increase the output of the low reviewers. For example, students may be instructed to mark all of the problems they detect in the text, but to only describe and offer solutions to seven of the problems for each reviewing dimension that affect the quality of the text the most. This instruction will help the low reviewers produce as much criticism as the high reviewers. Moreover, having students prioritize certain errors will not only help them understand what problems need attention but also provide them practice diagnosing and solving problems these problems.

Given the reciprocal nature of peer-review, all students are expected to receive more feedback from high reviewers. One way to balance the amount of feedback students receive would be to assign both high reviewers and low reviewers to review each paper. However, caution must be taken when assigning papers to be reviewed because the nature of the feedback is likely to change as a result of reviewing a mix of high-quality texts and low-quality texts.

Appendix 1

See Table 2.

Table 2 Peer feedback coding scheme

Category	Definition	Example
All comments		
Praise	A positive feature of the paper	It was a good job explaining the differences between the MSNBC article and the article from the scientific journal
Problem	Something wrong with the paper	The writer did not offer insight into causal and correlational relationships
Solution	How to fix a problem or improve the quality of the paper	Also, I would suggest writing a stronger conclusion to the end of the paper
Criticism comments only		
Localization	Where the issue occurred	
Low prose	An issue dealing with the literal text choice—usually at a word level	Where you say 'the hypotheses and whether those hypotheses were proven', I think you would say 'that hypothesis' or 'the hypothesis' because it's just one hypothesis
High prose	High-level writing issues (e.g., clarity, use of transitions, strength of arguments, provision of support and counter-arguments, insight)	I do not understand what the argument is as it isn't very clear." Another peer suggested, "use your own voice in order to capture the [sic.] readers attention
Substance	An issue with missing, incorrect, or contradictory content	I don't see where you stated the independent and dependent variables

Appendix 2

See Table 3.

Table 3 Example of segmentation and coding of one piece of feedback

① No praise; Problem; Solution; Localized; Low prose

② No praise; Problem; Solution; Localized; High prose

③ No praise; Problem; Solution; Not localized; Substance

④ Praise; No problem; No solution

⑤ No praise; No problem; Solution; Localized; High prose

⑥ No praise; Problem; No solution; Localized; Low Prose

①
1. Make sure to check your first sentence in paragraph one. I think it is
②
incomplete. 2. The last sentence in your second paragraph seems a

little out of place. I think you should either have a transitional

sentence linking the first 3 sentences with the last one or place the last
③
sentence somewhere in the next paragraph. I also think you should

elaborate on the subject variables you mentioned and maybe try to

think of one more because you only mentioned it in a couple of
④
sentences and it disrupts the flow of the paper. 3. Second page of
⑤
your paper flowed very smoothly. Good Job! I would only suggest to
⑥
maybe write a conclusion paragraph 4… By the way, you mispelled

[sic] the word "assess" in the third paragraph. :)

Appendix 3

See Table 4.

Table 4 Descriptive & Inferential Statistics: Amount, Features, and Focus of Comments

Text quality		Reviewer ability				ANOVA						t test			
		High reviewer[a]		Low reviewer[b]		Reviewer ability		Draft quality		Interaction		High reviewer		Low reviewer	
		M	SD	M	SD	p	η²	p	η²	p	η²	p	d	p	d
Word count	High-quality	779	291	778	291	.30	.01	.32	.01	.31	.01	.21	−.27	.99	.00
	Low-quality	876	417	777	293										
Comments	High-quality	42.2	11.7	45.2	16.0	.35	.00	.42	.00	.60	.00	.31	−.21	.85	−.04
	Low-quality	44.9	13.4	45.8	13.2										
Praise	High-quality	26.2	9.3	30.8	13.0	.005*	.04	.89	.00	.91	.00	.83	.05	.99	.00
	Low-quality	25.7	10.3	30.8	12.8										
Criticism	High-quality	18.0	11.5	15.6	7.7	.01*	.04	.09	.02	.36	.00	.13	−.33	.49	−.14
	Low-quality	22.0	12.4	16.7	8.6										
Problem and solution	High-quality	2.6	3.0	3.1	3.2	.66	.00	.10	.02	.14	.01	.06	−.40	.87	−.03
	Low-quality	4.3	5.1	3.2	2.9										
Solution only	High-quality	8.2	8.7	5.2	4.6	.04*	.02	.68	.00	.23	.01	.64	.10	.15	−.30
	Low-quality	7.5	5.7	6.6	5.3										
problem only	High-quality	7.3	6.2	7.3	5.0	.05	.02	.15	.01	.05	.02	.04*	−.43	.68	.09
	Low-quality	10.2	7.5	6.9	4.7										
Low prose	High-quality	2.4	2.7	2.2	2.6	.87	.00	.44	.00	.81	.00	.73	−.07	.43	−.16
	Low-quality	2.6	4.4	2.7	3.6										
High prose	High-quality	8.8	5.9	7.9	4.2	.03*	.03	.09	.02	.26	.01	.10	−.35	.63	−.10
	Low-quality	10.9	6.1	8.4	4.1										
Substance	High-quality	6.9	6.5	5.5	4.4	.01*	.03	.29	.01	.45	.00	.28	−.23	.79	−.05
	Low-quality	8.5	6.8	5.7	5.0										

[a] High reviewer: high-quality ($n = 44$); low-quality ($n = 46$)

[b] Low reviewer: high-quality ($n = 48$); low-quality ($n = 48$); * $p < .05$

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036–1060.

Anderson, J. R., & Lebiere, C. (Eds.). (1998). *The atomic components of thought.* Psychology Press.

Charney, D. H., & Carlson, R. A. (1995). Learning to write in a genre—what student writers take from model texts. *Research in the Teaching of English, 29*(1), 88–125.

Chi, M. T. H. (1996). Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology, 10*(7), 33–49.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explantions: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145–182.

Cho, Y., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science, 39*(5), 629–643.

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

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.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Earlbaum Associates.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287–322.

Ferris, D. R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly, 31*(2), 315–339.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387.

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

Hayes, J. R., Flower, L., Schriver, K. A., Stratman, J. F., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics* (Vol. 2, pp. 176–240)., Reading, writing, and language learning New York: Cambridge University Press.

Inuzuka, M. (2005, July). *Learning how to write through encouraging metacognitive monitoring: The effect of evaluating essays written by others.* Paper presented at the Annual Conference of the Cognitive Science Society, Stresa, Italy.

Kaufman, J., & Schunn, C. (2011). Students' perceptions about peer assessment for writing: their origin and impact on revision work. *Instructional Science, 39*(3), 387–406.

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.

Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology, 41*(3), 525–536.

Li, L., Liu, X., & Zhou, Y. (2012). Give and take: A re-analysis of assessor and assessee's roles in technology-facilitated peer assessment. *British Journal of Educational Technology, 43*(3), 376–384.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*(4), 492–527.

Lu, J., & Law, N. (2012). Online peer assessment: effects of cognitive and affective feedback. *Instructional Science, 40*(2), 257–275.

Lu, J., & Zhang, Z. (2012). Understanding the effectiveness of online peer assessment: A path model. *Journal of Educational Computing Research, 46*(3), 313–333.

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.

National Center for Education Statistics. (2012). The nation's report card: Writing 2011. Retrieved from http://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.

Newell, A. (1994). *Unified theories of cognition.* Cambridge: Harvard University Press.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale: Lawrence Earlbaum Associates.

Patchan, M. M., Charney, D., & Schunn, C. D. (2009). A validation study of students' end comments: Comparing comments by students, a writing instructor, and a content instructor. *Journal of Writing Research, 1*(2), 124–152.

Patchan, M. M., Hawk, B., Stevens, C. A., & Schunn, C. D. (2013). The effects of skill diversity on commenting and revisions. *Instructional Science, 41*(2), 381–405.

Patchan, M. M., & Schunn, C. D. (under review). Understanding the benefits of receiving peer feedback: A case of matching ability in peer-review.

Patchan, M. M., Schunn, C. D., & Correnti, R. J. (under review). The nature of feedback—revisited: How feedback features affect students' willingness and ability to revise.

Roediger, H. L. (2007). Twelve tips for reviewers. *APS Observer, 20*(4), 41–43.

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge: Harvard University Press.

Strijbos, J.-W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*(4), 265–269.

The College Board. (2012). SAT Percentile Ranks. Retrieved from: http://media.collegeboard.com/digitalServices/pdf/research/SAT-Percentile-Ranks-2012.pdf.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 8*(3), 247–261.

Topping, K. J. (2005). Trends in PEER LEARNING. *Educational Psychology, 25*(6), 631–645.

Topping, K. J., Dehkinet, R., Blanch, S., Corcelles, M., & Duran, D. (2013). Paradoxical effects of feedback in international online reciprocal peer tutoring. *Computers & Education, 61*, 225–231.

Wallace, D. L., & Hayes, J. R. (1991). Redefining Revision for Freshmen. *Research in the Teaching of English, 25*(1), 54–66.

Wooley, R. S., Was, C., Schunn, C. D., & Dalton, D. (2008). The effects of feedback elaboration on the giver of feedback. Paper presented at the Cognitive Science, Washington DC