Journal of Educational Psychology

The Nature of Feedback: How Peer Feedback Features Affect Students' Implementation Rate and Quality of Revisions

Melissa M. Patchan, Christian D. Schunn, and Richard J. Correnti

Online First Publication, April 7, 2016. http://dx.doi.org/10.1037/edu0000103

CITATION

Patchan, M. M., Schunn, C. D., & Correnti, R. J. (2016, April 7). The Nature of Feedback: How Peer Feedback Features Affect Students' Implementation Rate and Quality of Revisions. *Journal of Educational Psychology*. Advance online publication. http://dx.doi.org/10.1037/edu0000103

Journal of Educational Psychology

© 2016 American Psychological Association
0022-0663/16/$12.00 http://dx.doi.org/10.1037/edu0000103

The Nature of Feedback: How Peer Feedback Features Affect Students' Implementation Rate and Quality of Revisions

Melissa M. Patchan, Christian D. Schunn, and Richard J. Correnti
University of Pittsburgh

Although feedback is often seen as a critical component of the learning process, many open questions about how specific feedback features contribute to the effectiveness of feedback remain—especially in regards to peer feedback of writing. Nelson and Schunn (2009) identified several important features of peer feedback in their nature of feedback model. In the current study, we test an updated theoretical model that includes a broader set of features and considers not only students' likelihood of implementing a comment but also the quality of their revisions. To empirically test the updated theoretical model, we analyze over 7,500 comments from 351 reviewers to 189 authors. Each comment was coded for the presence of praise, a problem description, a suggested solution, localization, focus (i.e., low prose, high prose, substance), implementation, and revision quality. To account for the cross-classified nesting of data, we used a 2-level, cross-classified, hierarchical logistic regression model. Only 2 feedback features increased students' likelihood of implementation (i.e., overall praise and localization), while several feedback features reduced students' likelihood of implementation (i.e., mitigating praise, solutions, and high-prose comments). Overlapping feedback features affected students' ability to revise and in opposing directions from their effect on likelihood of implementation: Revisions were less likely to improve the quality of their paper when implementing comments that included a specific location in the text, but they were more likely to improve the quality of their paper when implementing comments that focused on high-prose and substance issues. Implications of these findings are discussed.

Keywords: peer feedback, writing, peer assessment, feedback implementation, revision quality

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Although feedback is often seen as a critical component of the learning process (Azevedo & Bernard, 1995; Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Chi & Wylie, 2014; Graham, Hebert, & Harris, 2015; Koedinger, Corbett, & Perfetti, 2012), how features (e.g., its timing, complexity, or content) contribute to the effectiveness of feedback is largely disputed (Molloy & Boud, 2014; Mory, 2004; Shute, 2008). Many open questions about effectiveness remain, especially in particular contexts. Furthermore, as the use of peer feedback during writing assignments increases, an understanding of how various feedback features affect students' use of peer feedback is critical to improving the advice given to students about how to construct useful feedback. Therefore, the goal of the current study was to further examine the effect of several important features of peer feedback, and in particular how peer feedback affects college students' likelihood to implement a comment and the quality of the revisions within Writing in the Discipline courses.

The research on feedback in general has a long history. At least two classic explanations for how feedback in general improves learning also apply to peer feedback: motivation that increases a general behavior (Brown, 1932; Symonds & Chase, 1929) and information that can be used by a learner to change performance in a particular direction (Pressey, 1926, 1927). For an example of the motivating function, praise is regularly included in models of good feedback (Grimm, 1986; Nilson, 2003; Saddler & Andrade, 2004), despite the lack of evidence supporting the effectiveness of praise (Ferris, 1997; Kluger & DeNisi, 1996). As an example of the additional information function, specific descriptions of the weaknesses or suggestions regarding how to improve performance can help a writer target problematic knowledge or skills (Gielen, Peeters, Dochy, Onghena, & Struyven, 2010).

Nature of Feedback Model: Revisited

In the context of peer feedback on college writing, Nelson and Schunn (2009) proposed a theoretical model of feedback that described the relationships among several feedback features, potential internal mediators, and the likelihood of implementing the comments (see Figure 1). In their nature of feedback model, they posited that these feedback features might be psychologically different—that is, some features (i.e., summarization, specificity, explanations, and scope) are more cognitive in nature, while other features (i.e., praise and mitigation) are more affective in nature. This distinction was especially important for explaining why these features might influence a writer's implementation of feedback. In particular, predictions were formulated from general psychological research for two potential mediators: understanding (e.g., Bransford & Johnson, 1972; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Novick & Holyoak, 1991) and agreement (e.g., Ilgen, Fisher, & Taylor, 1979). Understanding was expected to influence

Melissa M. Patchan, Christian D. Schunn, and Richard J. Correnti, Learning Research & Development Center, University of Pittsburgh.

Correspondence concerning this article should be addressed to Melissa M. Patchan, who is now at the Department of Learning Sciences and Human Services, West Virginia University, PO BOX 6122, Morgantown, WV 26506. E-mail: melissa.patchan@gmail.com

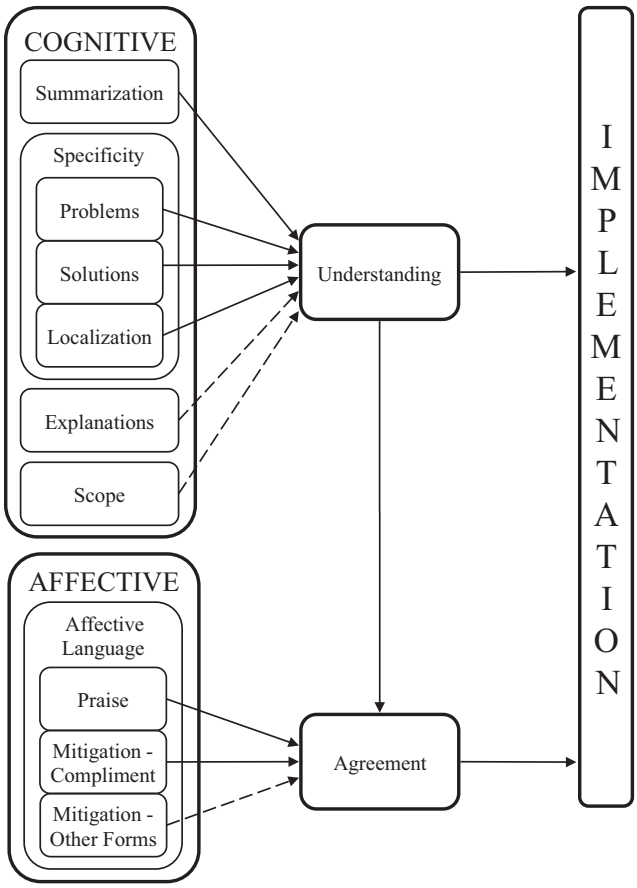

Figure 1. Proposed nature of feedback model (Nelson & Schunn, 2009).

how a writer responded to feedback in two ways: First, by understanding a comment, a writer might be more likely to implement a concrete suggestion, or second, by understanding a comment, a writer might be more likely to come up with a solution to a comment that lacked an explicit suggestion. In contrast to the especially cognitive nature of understanding, agreement was argued to have an affective component. Therefore, Nelson and Schunn (2009) predicted that the affective features would influence implementation by first influencing whether a writer agreed with the comment. To test these predictions, they randomly selected 24 students from a large introductory history course. For their in-depth analysis of 1,073 peer review comments, they conducted a series of chi-square and analysis of variance tests comparing the rate of implementation in feature-present and feature-absent comments. They concluded that several features influenced writers' implementation of feedback via understanding—that is, a comment was more likely to be implemented if the writer understood the problem, and certain features like summarization, solutions, and localization were more likely to improve a writer's understanding of the problem, while explanations of the problem were more likely to hurt the writer's understanding.

In the current study, we will also explore the effects of these feedback features on students' likelihood of implementation (i.e., whether the student implemented the comment). Although studies that have examined teacher provided feedback found that students implement most of the provided suggestions (Ene & Upton, 2014), studies that focused on student provided feedback found lower rates of implementation (Nelson & Schunn, 2009; Patchan, Hawk, Stevens, & Schunn, 2013). Note that teacher or peer feedback implementation has been previously examined in terms of *uptake* by applied linguists (Esmaeili & Behnam, 2014; Zoghi & Nikoopour, 2013), but we will use the term *implementation* for consistency. We will also further explore the effects of these feedback features on the quality of their revisions (i.e., whether the revision improved the quality of the draft). Below is a set of predictions that considers both likelihood of implementation and revision quality. These predictions include a broader set of features based on an expanded and updated literature review relative to the Nelson and Schunn (2009) study. This expanded literature review primarily focuses on studies examining the effects of feedback on adult learners, typically at the undergraduate level.

Praise

Praise is defined as a comment that describes a positive feature of the paper. Advice for many different ages of learners about how to provide helpful feedback often suggests including praise (Grimm, 1986; Nilson, 2003; Saddler & Andrade, 2004), and one common criterion of peer feedback quality is whether praise was offered (Gielen et al., 2010; Prins, Sluijsmans, & Kirschner, 2006). College students often report preferring to receive a balance between praise and criticism (Agius & Wilkinson, 2014; Parkes, Abercrombie, & McCarty, 2013), and teachers also assert that praise is important. However, these beliefs are not always reflected in teachers' commenting styles—that is, teachers' feedback to college students often lack any praise and usually includes fewer praise comments than students' feedback to peers (Cho, Schunn, & Charney, 2006; Orrell, 2006; Patchan, Charney, & Schunn, 2009).

Despite students' and teachers' reported beliefs about praise, there is some debate on its usefulness. Although there is some evidence in studies of college students that praise may help build trust (Dohrenwend, 2002; LeBaron & Jernick, 2000) and motivate the receiver to respond to criticism (Bienstock et al., 2007; Hesketh & Laidlaw, 2002), too much emphasis on the positive features may lead learners of all ages to overlook criticism and overestimate their performance (Richardson, 2004; Schlecht, 2008; Shute, 2008). Even though praise seems to evoke positive emotions from students that may be motivating, it lacks the information necessary to reinforce a particular behavior—that is, learners of all ages may not repeat what was done well in the future if they do not know what was specifically done well in the first place (Voerman, Korthagen, Meijer, & Simons, 2014). As a result, the motivating effects of praise must be examined on the likelihood of implementing other critical feedback. Therefore, when considering the impact of praise on students' likelihood of implementation, we predicted that the presence of praise would increase the likelihood of a student implementing other informational comments that are received.

Turning to the quality of revisions, studies conducted in second language writing (Ferris, 1997) as well as in clinical note-writing (Parkes et al., 2013) found that praise almost never impacted performance quality and in general, the effect size of praise is typically quite small (Cohen's $d = 0.09$; Kluger & DeNisi, 1996).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Therefore, when considering the impact of praise on students' ability to revise, we predicted that the presence of praise would have no effect on the quality of the revisions.

Solutions

A *solution* is defined as a comment that suggests how to fix a problem or improve the quality of the text. Students often reported a strong preference for receiving solutions (Agius & Wilkinson, 2014; Ferris, 1995). However, in a recent multiple-case study of 10 second language learners, the majority of the students reported a preference for the teacher to mark the errors and label them by type but not provide corrections (Ferris, Liu, Sinha, & Senna, 2013). Moreover, despite the teachers' indications that providing solutions was important for students' writing ability development (Orsmond & Merry, 2011), students received a surprisingly low amount of solutions from their teachers (Duncan, Prowse, Wakeman, & Harrison, 2007; Weaver, 2006). When comparing the commenting styles of teachers and peers, college students received almost three times more solutions from their teacher than from their peers (Cho et al., 2006). However, commenting styles also varied among teachers: A college writing instructor provided almost twice as many solutions as a content instructor with the amount of solutions provided by peers falling between the writing instructor and content instructor (Patchan et al., 2009).

The majority of prior research on the effectiveness of solutions has been focused on second-language learners. Typically, receiving solutions helped these learners improve their writing performance (AbuSeileek & Abualsha'r, 2014; Bitchener, 2012; Bitchener, Young, & Cameron, 2005; Nelson & Schunn, 2009; Shintani, Ellis, & Suzuki, 2014; Sugita, 2006; Tseng & Tsai, 2007; Van Beuningen, De Jong, & Kuiken, 2012). However, solutions were not equally helpful under all conditions. For tenth graders, receiving solutions that focused on incompleteness during the early part of the writing process was more helpful than receiving solutions that directly corrected errors, which had either no effect when provided earlier in the writing process or were harmful when provided later in the writing process (Tseng & Tsai, 2007). Receiving solutions was more helpful for lower proficient writers, while receiving descriptions of the problems only was sufficient for advanced learners (Bitchener, 2012). For secondary students, receiving solutions was more helpful for improving writers' grammatical errors, while receiving descriptions of the problems was more helpful for improving writers' nongrammatical errors (Van Beuningen et al., 2012). Based on this prior research that focused heavily on second language learners, we can make the tentative predictions: (a) The presence of a solution may increase the likelihood of a student implementing a particular comment, and (b) a revision made based on a comment with a solution may be more likely to improve the quality of the paper than a revision made based on a comment without a solution. However, it may be that solutions do not play such a critical role for more advanced writers, possibly because solutions provided by peers on more complex problems (rather than the more basic writing issues of relevance to second language learners) are either not well conceived or not well described.

Localization

When comments are made directly in the margins of the document, it is relatively clear whether the problem occurs. However, such marginalia often leads to a significant amount of low-level prose comments (Ene & Upton, 2014), and some comments refer to multiple locations, and thus peer commentaries often are asked to include or focus on 'end comments' (Ferris et al., 2013). However, end comments require extra work by the review or author to determine the relevant text location(s). A *localized* comment is defined as a comment that explicitly refers to the location of the issue. This feature is also commonly recommended in guides for college students about providing helpful feedback on writing (Nilson, 2003). In a study of end comments, college students and instructors did not differ in the amount of localized comments they provide—localization occurred in about half the comments (Patchan et al., 2009). Litman and colleagues (Nguyen, Xiong, & Litman, 2014; Xiong, Litman, & Schunn, 2012) have developed tools that automatically detect the absence of localization in peer feedback, pushing both students and instructors to provide additional localization.

Localization is likely to be particularly important when the problematic behavior could occur in multiple locations (e.g., lengthy writing assignments). Surprisingly little research has examined the impact of localized comments in writing. In one study, college writers were more likely to understand the problem after receiving a localized comment, which resulted in a greater likelihood of implementation (Nelson & Schunn, 2009).

Localization has been examined in other domains. Within the study of intelligent tutors, researchers have examined the impact of error-flagging on performance and learning. Error-flagging refers to the marking of errors without providing additional information about what the error is or how to fix it. Error-flagging consistently helps students complete problems more quickly (Corbett & Anderson, 1991, 2001). Students who received error-flagging feedback on tests scored higher than those who did not receive error-flagging feedback (Kumar, 2010, 2012). However, students did not perform better than those who received immediate feedback or demand feedback (Corbett & Anderson, 1991, 2001). In general, error-flagging did not improve learning—that is, performance on transfer tasks did not differ between conditions (Hausmann et al., 2013; Kumar, 2010). Students did not express a preference for error-flagging over other types of feedback (Corbett & Anderson, 1991, 2001).

In a series of studies, Goodman and Wood have explored the effect of feedback specificity on performance and learning. While training undergraduates how to manage employees, localized feedback positively influenced performance but had no direct effect on learning (Goodman & Wood, 2004; Goodman, Wood, & Chen, 2011; Goodman, Wood, & Hendrickx, 2004). Two possible reasons that may have contributed to the lack of an effect on learning were explored. First, learning outcomes were dependent upon the learning opportunities available to the participants, and while localized feedback improved participants' managerial decisions (i.e., participants made more correct decisions about how to manage simulated employees), it also limited their learning opportunities to how to deal with employees who demonstrated good performance (i.e., as participants made better managerial decisions, their simulated employees were more likely to performed

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

well; Goodman et al., 2004). Second, although processing explicit information was positively related to learning, localized feedback negatively affected how much students processed explicit information (Goodman et al., 2011).

On the basis of these diverse prior studies, when considering the impact of localization on students' likelihood of implementation, we predicted that students would be more likely to implement localized comments than comments that were not localized. The likely impact on quality of revision is less clear. On the one hand, localization may improve understanding and thus revision quality. On the other hand, localization may limit which possible revisions to complex problems are explored, thereby reducing revision quality.

Focus of Feedback

The focus of feedback is defined as the topic of the issue being described in the feedback (e.g., grammar, word choice, clarity, transitions, accuracy of content). The focus of feedback is especially relevant for Writing in the Discipline environments, where different kinds of knowledge, such as disciplinary knowledge (i.e., subject matter knowledge and genre knowledge), rhetorical knowledge, and writing process knowledge, are essential to effective writing (Bazerman, 2008; Beaufort, 2004; Prior, 2006). Writers, who vary in both writing ability and disciplinary knowledge, might respond differently to feedback about prose issues (i.e., a comment that addresses writing issues—such as clarity of main ideas and transitions between those ideas) versus substance issues (i.e., a comment that addresses issues with the content of the writing—such as missing or inaccurate content). Moreover, because novice writers are less likely to naturally attend to global issues (Hayes, Flower, Schriver, Stratman, & Carey, 1987), distinguishing between low-level prose issues (e.g., grammar, word choice, spelling) and high-level prose issues (e.g., transitions, appropriate evidence, inclusion of counterarguments) is also important. Research on writing pedagogy broadly also considers these distinctions using terms like *lower order concerns* versus *higher order concerns* and *form-focused feedback* versus *meaning-based feedback* or *content-based feedback*. Therefore, three foci of feedback will be considered in this study: *low prose*, *high prose*, and *substance*.

In a survey of 1,000 students conducted by Turnitin (2013), an online provider of services for plagiarism-prevention, online grading, and peer review, students reported that feedback focusing on particular high-prose issue (i.e., thesis/development) was the most valuable feedback from teachers, but they also preferred receiving feedback focusing on another high-prose issue (i.e., composition/structure) and a low-prose issue (i.e., grammar/mechanics). The students' perception of valuable feedback matches the types of feedback found to be provided by instructors—that is, instructors tend to provide feedback focusing more on high-prose issues than low-prose issues (Ene & Upton, 2014; Patchan, Schunn, & Clark, 2011), and tutors at writing centers are strongly encouraged to focus on higher order concerns rather than lower order concerns, suggesting that educators value feedback focused on high-prose and substance issues more so than feedback focused on low-prose issues (Nakamaru, 2010). However, the focus of feedback is likely to depend on prior knowledge of the feedback provider and the receiver. For example, when comparing a college content instructor's feedback to a writing instructor's feedback, the content instructor provided more solutions focused on substance than the writing instructor or peers, while the writing instructor provided more solutions focused on high prose than the peers, who in turn provided more than the content instructor (Patchan et al., 2009). In a study of college teacher perspectives, teachers of mostly mainstream first-year writing courses often reported providing more language-focused feedback for second language learners than for native English speakers (Ferris, Brown, Liu, & Stine, 2011). Among peers, high-ability reviewers and low-ability reviewers provided similar amounts of high-prose comments; however, high-ability reviewers provided more substance and low-prose comments than low-ability reviewers on papers of lower quality (Patchan et al., 2013).

The prior research on the effects of the different types of focus has supported the obvious expectation that the implementation rates of each focus depended on the amount received—that is, students who received more feedback that focused on high prose also implemented more feedback that focused on high prose (Ene & Upton, 2014), and similarly, students who received more feedback that focused on low-prose and substance issues also implemented more feedback that focused on low-prose and substance issues (Patchan et al., 2013). Similar effects were found on draft quality. For third graders, the amount of content feedback received predicted the quality of content in the final draft, and the amount of surface feedback received predicted the quality of writing mechanics in the final draft (Matsumura, Patthey-Chavez, Valdes, & Garnier, 2002). In contrast to these benefits of content-based feedback and despite teachers' and students' preference for high-prose and substance feedback, Zohrabi and Rezaie (2012) found added value to low-prose feedback—that is, English as a foreign language (EFL) students wrote higher quality posttest essays after receiving both form-focused and meaning-based feedback in comparison to EFL students who received just meaning-based feedback.

Because much of the prior research did not address whether focus was related to likelihood of implementation or whether these revisions improved quality, we can make only tentative, pragmatic predictions based on typical areas of strength and weakness in student writing and revision. Specifically, we predicted that the likelihood of implementing a low-prose comment would be higher than the likelihood of implementing high-prose or substance feedback because the low-prose issues are easier to address. Furthermore, when considering the impact of the focus of feedback on students' ability revise, we predicted that a revision made based on a comment that focused on high prose or substance would more likely improve the quality of the paper than a revision made based on a comment that focused on low prose.

Amount of Feedback

Assignments requiring peer feedback often involve receiving feedback from multiple peers, which can amount to a large number of comments for the writer to process (e.g., a mean of 48 comments per document in Nelson & Schunn, 2009). Given the sheer volume of feedback received, implementing multiple peer feedback may be especially difficult from a cognitive load perspective. Writing, in general, and revision, in particular, have previously been described as high-cognitive-load tasks (Hayes et al., 1987; Kellogg, 1994; Kirkland & Saunders, 1991). Cognition and learn-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

ing can often be negatively influenced when the student's cognitive capacity is exceeded by the learning task at hand (Paas, Renkl, & Sweller, 2004; van Merrienboer & Sweller, 2005). Therefore, when considering the impact of the amount of feedback on students' likelihood of implementation, we predicted that students who received a larger number of comments or longer comments might implement fewer comments than students who received fewer comments or shorter comments.

Similarly, when sifting through a vast number of comments, the writer may use less effective strategies when choosing which comments to implement (Ariely, 2000; Iselin, 1988). For example, a writer may revise using only the comments that are easier to implement (e.g., comments focused on low prose) or ignore long comments, which may have little to no impact on the draft quality. Therefore, when considering the impact of the amount of feedback on revision quality, we predicted that revisions made by students who received a larger number of comments or longer comments might be less likely to improve the quality of the draft in response to any given comment.

Summary of Hypotheses

Several hypotheses were derived from this literature review. First, praise was expected to increase a writer's likelihood of implementation but have no effect on one's ability to revise. Second, solutions were expected to either increase or have no effect on a writer's likelihood of implementation, and similarly, they were expected to improve or have no effect on one's ability to revise. Third, localization was expected to increase a writer's likelihood of implementation, but expectations in regards to one's ability to revise were less clear—that is, localized comments could either improve or reduce one's ability to revise. Fourth, low-prose comments were more likely to be implemented than high-prose or substance comments, but high-prose and substance comments were more likely to lead to greater improvements. Finally, the amount of feedback was expected to decrease both a writer's likelihood of implementation as well as one's ability to revise.

Current Study

The goal of the current study was to further examine the effect of several important features of peer feedback by testing a revised theoretical model that includes a broader set of features (i.e., focus of feedback and amount of feedback) and considers both students' likelihood of implementation and the quality of the revision. Moreover, the current study involves a more sophisticated analytic technique that better controls for covariation among features at the comment, author, and reviewer levels. Prior research might have misrepresented the contributions of particular features because it rarely considered co-occurrence among features. For example, Ene and Upton (2014) also examined the amount of different feedback features (i.e., topic/focus, corrective/negative vs. noncorrective, direct vs. indirect, and explicit vs. implicit) and the likelihood of implementing of those comments. However, they did not examine the relationships among the features (e.g., whether the effectiveness of explicit vs. implicit comments vary by focus). It may be that explicit comments are provided more often for low-prose problems, and these are easier to implement. Thus, apparent benefits of explicit comments may just be a matter of a confound with focus. Such confounds could occur at the comment level (i.e., a comment about low prose is likely to be explicit) or they could occur at the author level (i.e., documents receiving many explicit comments could have proportionally many low-prose issues being noted). By controlling for the covariation among the various possible predictions at both levels, we provide a better estimate of the true effects of each feedback feature. We will further highlight this confound by comparing a statistical model with the target feature alone to a statistical model with all the features of feedback.

Method

Overview

The current study was part of a larger study that examined multiple aspects of why students learn from peer assessment, including the benefits for the author of *receiving* feedback (Patchan & Schunn, in press) and the benefits for the reviewer of *providing* feedback (Patchan & Schunn, 2015), in contrast to the current focus on the relative effectiveness of different forms of peer feedback. To examine the effectiveness of the different features and focus of peer feedback, the effects of amount, features, and focus of comments received on the implementation rate and the revision quality were observed.

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 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.

Participants

Multiple research studies were implemented in this class. From the 838 students enrolled in the class, 432 were randomly selected to participate in a study of matching versus mismatching reviewer-author ability pairings. Within those students, coding was done exhaustively for all reviews received on a feasible but still large subset to optimize on analyses of the effect on authors (i.e., $N = 189$). Documents were selected on the basis of maximizing availability of supporting data (i.e., documents of authors who completed several surveys were given a higher priority in the selection process over documents of authors who did not complete any surveys), although this selection process had little effect on document quality ($M_{selected} = 3.8$, $SD = 0.5$; $M_{not\ selected} = 3.7$, $SD = 0.6$), $t(428) = 1.42$, $p = .16$, and, thus, the selected participants are broadly representative of the larger class. The current study reanalyzes that large coded dataset.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This sample represented students (77% female) at all levels with a predominance of less advanced students (i.e., 58% freshmen, 27% sophomores, 10% juniors, and 5% seniors) as well as a great variety of majors (i.e., of the declared majors: 30% social sciences, 30% natural sciences, 14% engineering, 13% education, 6% computer science, and 6% business). Although specific information about the participants' linguistic or cultural background was not collected, approximately 8% of the students at this particular university were international students.

Procedure

Participants completed three main tasks: (a) wrote a first draft, (b) reviewed peers' texts, and (c) revised own text based on peer feedback. At the end of the first month of the semester, participants had one week to write their first draft and submit it online using the Web-based peer review functions of turnitin.com. For this writing task, they were expected to produce a three-page paper in which they evaluated whether MSNBC.com, a U.S. digital news 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 that were either from peers with lower writing ability or peers with higher writing ability. 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 Scholastic Assessment Test (SAT) verbal,[1] SAT writing, and 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. The SAT verbal scores (i.e., a more general cognitive ability measure, Frey & Detterman, 2004) and the SAT writing scores (i.e., a more specific ability measure) were highly correlated ($r = .71$) such that it would not be possible to examine effects of each separately in the data analyses. To further validate this measure of writing ability, the first draft quality score (see the Coding Process: Quality of Writing section for more details; rubric details are in Appendix A) was compared between the high authors and low authors. An independent *t* test on the sample of 189 participants revealed a significant difference in author ability: The high authors ($N = 93$, $M = 17.0$, $SD = 3.8$) produced higher quality first drafts than the low authors ($N = 96$, $M = 14.8$, $SD = 3.2$), $t(187) = 4.33$, $p < .001$, $d = 0.63$.

Participants received no formal training on how to review their peers' papers. Rather, they were given a handout with the assignment details (see the Review Support Structures section). In addition to the assignment handout instructing students to generate end comments, the turnitin.com peer review functions also 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. Finally, the reviews were anonymous—that is, a pseudonym was used to identify both the writer and the reviewer.

Finally, participants were able to access the peer feedback online once the reviewing deadline had passed. The participants were given one week to revise their draft based on the peer feedback. The teaching assistants provided final grades for the paper.

Review Support Structures

Participants were provided with a detailed rubric to use for the reviewing task, which shaped what comments were available for analysis in this study. 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 argument was and did 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 counterarguments 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 counterarguments, 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 toward global writing issues (Wallace & 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

Several different methods of calculating interrater reliability were used because there were different types of data. For

[1] The SAT 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 United States typically require a first year composition course, and the university in the present study requires two semesters of composition.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

categorical data (e.g., coding of peer feedback, implementation, revision quality), we will primarily report Cohen's kappa. Because some of the kappa values were in the low range, we also included percent agreement. For continuous data (e.g., quality of writing), we will report consistency using the intraclass correlation coefficient (ICC), which has a similar interpretation to Cronbach's alpha.

Quality of writing. Two outside writing experts (i.e., rhetoric graduate students who taught at least eight semesters of undergraduate composition) rated the quality of the first drafts. The rubric used by the participants was elaborated for the expert coders to examine quality at a more fine-grained level (see Appendix A for rubric details). For example, the students' dimension of "how well this paper was organized" was further divided into two dimensions for the experts: "how well this paper was organized around a main idea" and "how well transitions connected paragraphs." A similar 5-point scale (1 = *very poor* to 5 = *very good*) was used. The final interrater reliability was high (ICC = 0.84; Shrout & Fleiss, 1979).

Coding peer feedback. The peers' written feedback was coded to determine how different features and focus of comments affected the implementation rate and the revision quality. Five undergraduate research assistants were trained to code each feature and focus. Because some of the reliabilities were moderate, comments were exhaustively double coded, and pairs of coders were required to come to consensus on all disagreement cases to improve effective reliability and reduce coding noise—kappa values are presented for each type.

First, the feedback was segmented into comments based on idea units because reviewers frequently commented about multiple issues within one dimension (e.g., transitions, use of examples, word choice). An idea unit was defined as a contiguous comment that refers to a single issue. To illustrate, the feedback in Figure 2 was segmented into six comments. Although the reviewer numbered the comments, several of these pieces of feedback focused on multiple issues. As one example, the second (as numbered by the peer) piece of feedback described two different flow problems:

> *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 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 sentences and it disrupts the flow of the paper.*

(a) One of the sentences seemed out of place, and (b) subject variables were only briefly mentioned and seemed like a randomly inserted topic. We segmented this piece of feedback into its two conceptual comments.

Second, each comment (N = 7,641) was coded for the presence/absence of four independent features (see Appendix B for definitions and examples of each code): summary (κ = .60; percent agreement was 95%), praise (κ = .90), problems (κ = .85), and solutions (κ = .90). All comments that were previously coded as either problem or solution (i.e., criticism comments) were coded for the presence/absence of localization (κ = .63; percent agreement was 92%) and the focus (i.e., low prose, high prose, or substance, κ = .54; percent agreement was 78%). Many issues can involve both high prose and substance; these comments were always coded as substance comments. Finally, comments with summary statements (i.e., paraphrase, rating, structure, κ = .95), praise (i.e., highly positive, moderately positive, no descriptive, rubric, no change), and localization (i.e., quote, text, topic, general, κ = .92) were further broken down into different types. Figure 3 illustrates the relationship between the feedback provided, seg-

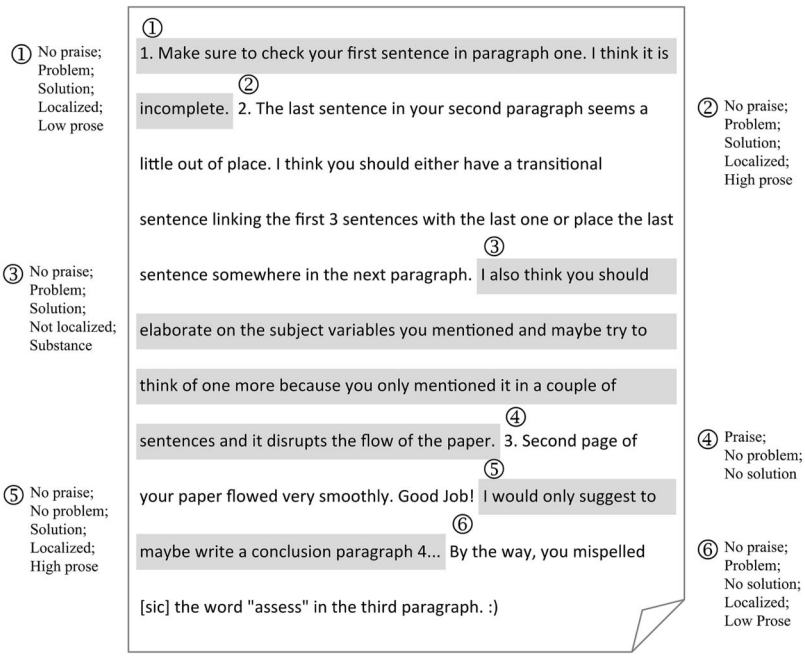

Figure 2. Example of how one piece of feedback was segmented and coded.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

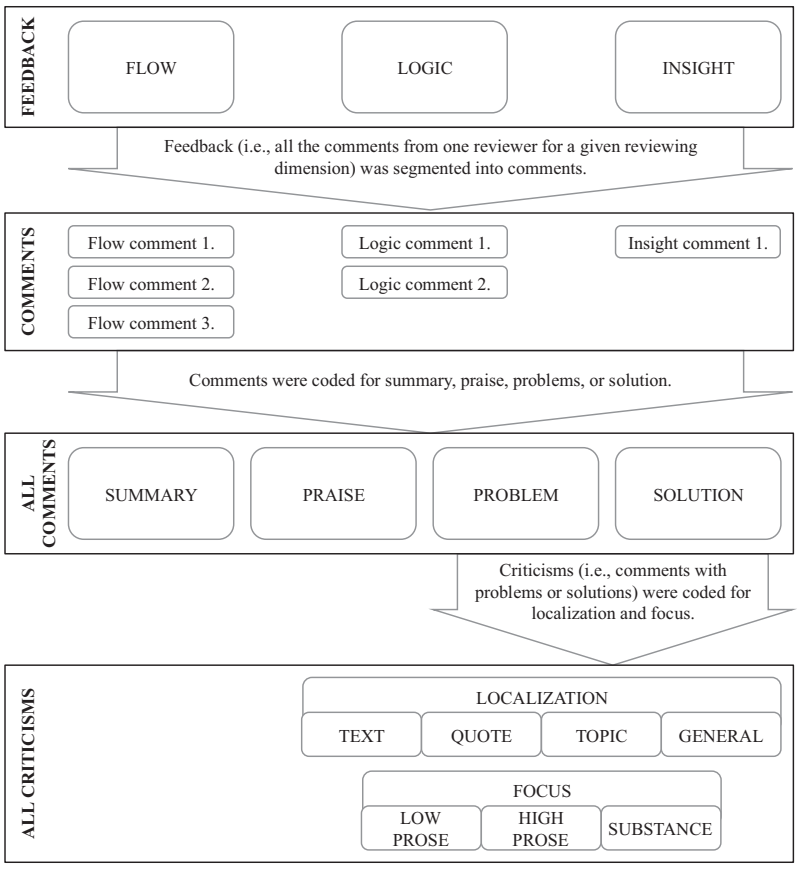

Figure 3. Comment segmenting and hierarchical coding process.

mented comments, and the types of feedback coded. Figure 2 demonstrates how the segmented comments in one piece of feedback were coded.

Implementation rate. To examine the implementation rate of the comments, the same two writing experts coded whether the writer implemented a revision that addressed the issue identified in each criticism comment ($\kappa = .74$; percent agreement was 89%). Microsoft Word's "compare documents" function applied to the second versus first drafts was used to facilitate this process. As long as the writer appeared to attempt a revision based on the comment, it was coded as implemented (see Appendix C for examples). A small percentage of comments (5%) were excluded from analysis for being too vague or unclear to determine whether they were implemented (e.g., "If anything, the paper should be a little spruced up").

Revision quality. For comments that were implemented, the same two writing experts rated the quality of the revision that was associated with the implementation, with moderate interrater reliability ($\kappa = .58$; percent agreement was 83%). This quality rating was on a binary scale: A rating of 0 indicated either no change in the quality of the paper or a decrease in the quality of the paper (a rare outcome, and hence collapsed with no change), and a rating of 1 indicated an increase in the quality of the paper (see Appendix C for examples).

Statistical Models

We conducted a series of two-level, cross-classified, hierarchical logistic regression models with two different outcomes. In the first set of statistical models we were interested in examining the effect of several important features of peer feedback on students' likelihood of implementation (i.e., *implement a comment*). In the second set of statistical models, we examined only implementable comments to determine which features of peer feedback enhanced the writers' ability to revise (i.e., *revision quality*). Some features of peer feedback are likely to affect implementation and revision quality at the comment level—for example, a writer may be more likely to implement a comment with mitigating praise. However, similar predictions cannot be made for a comment with only praise (i.e., one that lacks a description of the problem or potential solution) because there is no problem to be revised. Therefore, praise can only have an effect on implementation and revision quality at the writer level (i.e., influence revisions in response to the other comments received). The examined features were divided into comment-level variables (i.e., mitigating praise, solutions, localization, and focus) and writer-level variables (i.e., praise).

Adding further complexity to the structure of our data, the dependent variables were not independent observations. The 189 writers received 2,714 implementable comments (892 of which were implemented) from 351 reviewers. Further, the data were not

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

cleanly nested—that is, comments were nested within both writers and reviewers, but reviewers were not nested within writers or vice versa. Therefore, to account for the cross-classified nesting of data, we used a two-level, cross-classified, hierarchical logistic regression model (Raudenbush & Bryk, 2002), where comments were nested in reviewer–writer pairs. We used a full penalized quasi-likelihood method of estimation in HLM 7 (Raudenbush, Bryk, & Congdon, 2004). All models successfully converged, and the typical number of macroiterations was around 200 for implementation and under 50 for revision quality, with the exception of the partial model for focus, which required 8,938 macrointeractions before converging. Below we describe the first statistical model examining the effects of peer feedback on students' likelihood of implementation,[3] where we have comments at Level 1 for 1. . .*ijk* cases nested within cells cross-classified by 1. . .*j* writers designated as rows and 1. . .*k* reviewers designated as columns.

Level 1 model. We include 1. . .*p* comment-level predictors (i.e., word count and feedback features) at Level 1. The Level 1 model is

$$\eta_{ijk} = \pi_{0jk} + \sum_{p=1}^{P} \pi_{pjk} a_{pijk} \quad (1)$$

where η_{ijk} are the log-odds estimates of the likelihood of implementing a comment *i* from a given writer-reviewer pair, π_{0jk} is the average log-odds estimates of the likelihood of implementing a comment within the cell *jk*, π_{pjk} ($p = 1, 2 \ldots, P$) are comment-level coefficients for each predictor *p* from a given writer-reviewer pair *jk*, and α_{pijk} are comment-level peer feedback predictors for each comment *i* from a given writer-reviewer pair *jk*.

Level 2 model. We include 1. . .*r* writer-level predictors[4] (i.e., writer ability, the score for first draft quality, and total received comments variables such as total word count of received comments, number of comments, number of implementable comments, and amount of praise comments) and one reviewer-level variable (i.e., reviewer ability) for each of the *p* Level 1 predictors. Writer ability was measured using the composite measure described above in the procedure. Reviewer ability was operationalized as the writing ability of the reviewer (i.e., writing ability composite measure for each reviewer). The Level 2 model uses writer-level variables and reviewer-level variables, as well as the interaction between writer ability and reviewer ability as predictors for each of the (π_{pjk}) comment-level coefficients. The general form of the statistical model is

$$\pi_{pjk} = \theta_0 + \beta_p X_k + \sum_{r=1}^{R_p} \gamma_{pr} W_{rj} + \delta_{pjk} Z_{jk} + b_{00j} + c_{00k} \quad (2)$$

where θ_0 is the model intercept (i.e., the expected log-odds estimates of the likelihood of implementation π_{pjk} when all predictors are set to zero), β_p is the fixed effect of reviewer ability (X_k), γ_{pr} are the fixed effects of writer-level predictors (W_{rj}, $r = 1. . .R$), δ_{pjk} is the fixed effect of the interaction between writer ability and reviewer ability (Z_{jk}), and b_{00j} and c_{00k} are residual writer and reviewer random effects, respectively, on π_{pjk}, after taking into account X_k, W_{rj}, and Z_{jk}. We assume that $b_{00j} \sim N(0,\tau_{b00})$ and $c_{00k} \sim N(0,\tau_{c00})$ and that the effects are independent of each other.

Results and Discussion

Table 1 summarizes all the predictors that were included in the statistical models. Most comment-level variables (i.e., mitigating praise, problem/solution, localization, and focus) were dichotomous variables added as uncentered variables in the statistical models. One of the control variables (i.e., word count) was also at the comments level, but because it was continuous, it was grand-mean centered. Writer-level and reviewer-level variables (i.e., overall praise and additional control variables) were grand-mean centered. The statistical models also controlled for writer ability, reviewer ability, the interaction between writer and reviewer ability, first draft quality, and amount of feedback received (i.e., number of comments, number of implementable comments, word count).

Table 2 summarizes the relations among the predictors as well as between the predictors and the outcome variables (i.e., implementation and revision quality). Because the predictors are dichotomous, we reported tetrachoric correlations. Given the large samples, many of the relationships were statistically significant and thus collectively may shape the outcomes of the regression analyses, but the majority of these relations were relatively small correlations (i.e., $r < |.3|$). There were several very strong correlations, reflecting within-category alternatives: Comments with problems only, comments with solutions only, and comments with both problems and solutions were negatively correlated, and comments focused on low prose, comments focused on high prose, and comments focused on substance were negatively correlated. Importantly, the correlations across broad categories (i.e., praise, comment features, localization, focus) were not large in magnitude, which suggests that the inclusion of predictors across broad categories in the full model is possible. Also note that the correlations between localization and each type of focus (i.e., low prose, high prose, and substance) were negligible, which suggests that localization was not likely to be confounded with focus. Finally, the correlations between the predictors and revision quality foreshadow the findings from the partial models—that is, there was a strong negative association with localization and focus was a major predictor correlated with revision quality. Given the magnitude of the relationship between low-prose comments and revision quality ($r = -.57$), both high-prose and substance comments will likely be positive predictors of revision quality when low-prose comments is the reference.

In addition to the full statistical models (i.e., control variables plus all of the feedback features), we also examined several partial statistical models (i.e., control variables plus only a subgroup of the feedback features) to further explore how different features add explanatory power.

[3] The second statistical model (not described here) is exactly the same with two exceptions. First, the outcome changes because the focus is on the relationship between features of peer feedback and revision quality. Second, because the outcome changes, the only comments examined in this model were the ones that were actually implemented by writers. However, the nesting structure, variance components, predictors and coefficients remain the same in both models.

[4] In preliminary analyses, we also controlled for writer demographics, but the inclusion of these predictors did not change the coefficients in any appreciable way. Therefore, we removed them from the model, and we report the more parsimonious results in Table 4.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 1
Summary of Predictors in the Two-Level, Cross-Classified, Hierarchical Logistic Regression Models

Predictors	Level	Description
Praise		
Praise only	Writer	Number of praise only comments
Mitigated criticism	Comment	Whether a comment also included praise
Comment features		
Problem only	Comment	Whether an implementable comment only described the problem
Solution only	Comment	Whether an implementable comment only offered a solution
Problem + Solution	Comment	Whether an implementable comment described a problem and offers a solution
Localization		
Text	Comment	Whether an implementable comment mentioned a specific location (e.g., first paragraph on Page 2)
Quote	Comment	Whether an implementable comment included a quote from the text
General	Comment	Whether an implementable comment mentioned a general location (e.g., introduction)
Topic	Comment	Whether an implementable comment mentioned the topic whether the issue occurred
Focus		
Low prose	Comment	Whether an implementable comment focused on an issue with the literal text choice
High prose	Comment	Whether an implementable comment focused on high-level writing issues (e.g., clarity)
Substance	Comment	Whether an implementable comment focused on missing, incorrect, or contradictory content
Control variables		
Writer ability	Writer	Composite of the writer's SAT verbal, SAT writing, and final grades in the first and second semester composition courses
Reviewer ability	Reviewer	Composite of the reviewer's SAT verbal, SAT writing, and final grades in the first and second semester composition courses
First draft quality	Writer	Sum of the expert ratings of first draft quality across eight dimensions each with a 5-point scale
Number of comments	Writer	Total number of comments received
Number of implementable comments	Writer	Total number of implementable (i.e., criticism) comments received
Word count	Comment	Number of words in a given comment

Note. SAT = Scholastic Assessment Test.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Given that the dependent variables are binary, there is no estimate for Level 1 variance, which prevents us from reporting proportion of explained variance for each model. Instead, we compare the fit of each partial and full model to a baseline model that only includes the control variables using the Akaike information criterion (AIC), which is a conservative fit index that penalizes for additional parameters. Overall, the full statistical model for implementation was a good fit to the data and significantly reduced AIC-adjusted deviance from the baseline model: $\chi^2(10) = 51.11$, $p < .001$. For revision quality, the AIC-adjusted deviance increased between the baseline model and full model. For the partial models predicting revision quality, however, significant reductions in AIC-adjusted deviance were observed when particular features of peer feedback were added separately.

We will discuss the results of the statistical models grouped by each important feature of peer feedback (including praise, problem/solution, localization, and focus). The frequency of each feature varied, so we will first report the prevalence of each feature (see Table 3)—the descriptive information refers to the aggregate for the writer (i.e., number of times each dummy is "1" nested within writers). Next, we will describe the effects of that feature on implementation (see Table 4) and revision quality (see Table 5). Because the full and the partial statistical models generally found similar results, we will discuss the effects controlling for all other explanatory variables (i.e., results of the full statistical model). Finally, we will discuss results from the partial statistical models when there were discrepancies from the full statistical model.

On average, students received a total of 40 comments ($SD = 9$) from their peers. Only 41% of the comments were implementable (i.e., described a problem or offered a solution to a problem; $M = 16$ implementable comments; $SD = 7$). In their revisions, students typically tried to incorporate about 33% of these implementable comments ($M = 5$ comments were implemented; $SD = 5$). Thus, on average, only five of 40 comments (roughly 13%) led to writer revisions from their first to second drafts.

Probabilities for implementing a comment were calculated directly from the statistical model estimates.[5] For example, the base probability[6] of a comment being implemented was calculated using the following formula:

$$\frac{1}{1 + exp^{-intercept}}. \quad (3)$$

Substituting the predicted log-odds estimate for (θ_0) of -0.79 into Equation 3, we observe a 31% likelihood of implementation adjusting for all predictors in the statistical model. In other words, holding all other predictors constant, for every comment a writer received they were only likely to implement that comment in 31% of all opportunities. The change in the probability of implementing

[5] Probabilities for the likelihood of improving the quality of the text were calculated using the exact same formula for coefficients from the second model.

[6] Base probability is the probability when all of the dummy variables equal zero.

Table 2
Feedback Features: Descriptive Statistics and Tetrachoric Correlations

	Measure	*M*	*SD*	1	2	3	4	5	6	7	8
1	Mitigating praise	.19	.39								
2	Problem only	.45	.50	.15*							
3	Solution only	.37	.48	−.15*	−.98**						
4	Problem + Solution	.17	.38	−.01	−.92**	−.89**					
5	Localized	.25	.43	−.05	−.20**	−.04	.34**				
6	Low prose	.18	.38	−.19*	−.19**	.10	.14†	.20**			
7	High prose	.51	.50	.15*	.17**	−.11	−.10	.01	−.94**		
8	Substance	.32	.46	−.04	−.06	.05	.01	−.17**	−.87**	−.98**	
9	Implementation	.33	.47	−.13*	−.01	−.08	.13*	.25**	.14*	−.13*	.04
10	Revision quality	.21	.41	−.05	.08	.08	−.20	−.42**	−.57**	−.01	.35**

Note. $N = 2{,}714$ for correlations among predictors and between predictors and implementation (matrix above dotted line); $N = 892$ for correlations between predictors and revision quality (correlations below dotted line).
† $p < .05$. * $p < .01$. ** $p < .001$.

a comment given a particular feature of peer feedback was calculated using the following formula:

$$\frac{1}{1 + exp^{-intercept+predictor}}. \quad (4)$$

Therefore, to determine the change in likelihood for each additional comment containing praise for a writer, the predicted log-odds estimate for β_1 of 0.10 was added to the intercept and substituted into the equation. For one additional praise comment beyond the mean, the writers, on average, had a higher likelihood of implementing comments (33%). In other words, a writer was 2% more likely to implement a comment for each additional praise comment received.

With the exception of praise, all other predictor variables of interest in our statistical models are dichotomous (also, note that all control variables are continuous). Therefore, the relative magnitudes of their effects (i.e., the change in likelihood of implementation for our first model and the change in likelihood of improved quality after revision for our second model) are directly comparable. We summarize the magnitudes of the results for our significant covariates visually in Figure 4, with thicker arrows corresponding to greater change in probability of the outcome.

Praise

Given the motivational aspect of praise, comments including praise were expected to increase the likelihood of a writer implementing comments. However, the presence of praise was not expected to influence the quality of the revisions. Praise was commonly used in peer feedback—that is, 64% of all comments included some form of praise. Most of these comments (52% of all comments) included only praise (e.g., "I enjoyed the analogy of what the author stated at the end of how correlation is not causation. It really helped me see what the author was trying to get at when they stated the example of the ice cream and drowning being correlated, very smart!"). Praise was also used to mitigate a criticism comment in 12% of all comments (e.g., "The last paragraph does a great job explaining why causal conclusions cannot be made from the study, but the author doesn't really relate this to the article by Choney.").

As predicted, praise affected students' likelihood of implementation (see Column 6 in Table 4) but not the quality of their revision (see Column 6 in Table 5). Interestingly, mitigating praise had a different influence on implementation than praise only comments. As noted above, writers were 2% more likely to implement comments per additional praise comment received. By contrast, a comment with mitigating praise was 10% less likely to be implemented than a comment without mitigating praise. Meanwhile, receiving praise comments in either form had no effect on the likelihood of improving the quality of the text. The results from the partial statistical models were consistent with the full statistical models (see Column 2 in Tables 4 and 5), but the partial statistical model was not a significantly better fit than the baseline model, for implementation: $\chi^2(2) = 2.34$, $p = .31$, for revision quality: $\chi^2(2) < 1$, $p = .71$.

Unlike the finding in the Nelson and Schunn (2009) study, praise affected students' likelihood of implementation, but the relationship was complex. These results suggest that although praise overall could have a motivating effect on a writer when controlling for other features, praise being used to mitigate or soften the blow of a critique may have the unintended effect of allowing the writer to overlook the problem with his or her text.

Problem/Solution

Prior research, focusing heavily on second language learners, has demonstrated that the presence of a solution was expected to increase the likelihood of both implementing comments and improving the quality of the paper. However, if solutions do not play a critical role for more advanced writers, the presence of solutions might not influence either one. Of implementable comments, 45% described only a problem, 37% only offered a solution, and 17% both described a problem and offered a solution. Having only a solution embedded in a comment was marginally significantly related to the likelihood of implementing the comment. Students were 4% less likely to implement a comment that only offered a solution than a comment that only described a problem. Furthermore, the ignored comments tended to be at a higher level: 46% of the ignored comments suggested how to fix a high-prose issue and 36% of the ignored comments suggested how to fix a substance problem. Receiving a solution did not affect the likelihood of improving the quality of the text.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 3
Descriptive Statistics of the Amount, Features, and Focus of Comments Received From Peers

Measure[a]	*M*	*SD*	Minimum	Maximum
Praise				
Praise only	25.7	8.1	7	55
Mitigated criticism	3.1	1.9	0	8
Comment features				
Problem only	7.5	5.0	0	27
Solution only	6.1	4.1	0	21
Problem + Solution	2.9	2.4	0	14
Localization				
Text	2.5	2.8	0	17
Quote	.7	1.1	0	5
General	1.3	1.2	0	6
Topic	1.0	1.3	0	7
Focus				
Low prose	2.9	2.5	0	15
High prose	8.3	4.5	0	23
Substance	5.2	3.5	0	25
Control variables				
Writer ability	−.04	.82	−2.27	2.11
Reviewer ability	−.01	.80	−2.38	2.11
First draft quality	15.8	3.6	8	29
Number of comments	40.4	9.3	17	70
Number of implementable comments	16.5	7.4	2	42
Total word count	722	214	317	1,803

[a] Descriptive statistics are for implementable comments.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Unlike the finding in the Nelson and Schunn (2009) study, students in the current study were not more willing to revise after receiving a comment with a solution (see Column 3 in Table 4). Moreover, receiving a solution did not influence the likelihood of improving the quality of the draft (see Column 3 in Table 5). The results from the partial statistical models were more similar to the finding in the Nelson and Schunn (2009) study, suggesting the prior finding was an artifact of confounds in the comment features. A comment was 7% more likely to be implemented if it included both a description of the problem and a suggested solution. However, the partial statistical model was not a significantly better fit than the baseline model, for implementation: $\chi^2(2) < 1$, $p = .62$, for revision quality: $\chi^2(2) = 2.57$, $p = .28$.

Similar to the Nelson and Schunn (2009) study, the addition of solutions led to more revisions but only in the absence of other feedback features. When accounting for the other features, the effect disappears. These results suggest that when controlling for the other features, solutions that are provided by peers do not seem to play as critical a role.

Localization of Critiques

Prior research from a diverse range of contexts supported our prediction that a writer would be more likely to implement a localized comment. However, it was less clear how localization would affect a student's ability to revise. If localization improved understanding, it might also help a student improve the quality of the paper. However, if exclusive revision on localized comments crowded out a focus on more substantive issues, it would likely reduce revision quality. Only 25% of the implementable comments were localized. As expected, on average, students were 17% more likely to implement a comment that was localized than a comment that was not localized.

Interestingly, the effect varied by type of localization, with most of the localization techniques being much more effective. More specifically, a comment was approximately 29% more likely to be implemented if the localization mentioned a specific location in the text (11% of implementable comments), identified quotes within the text (4% of implementable comments), or mentioned a general location that could be found in any paper, such as the introduction (4% of implementable comments) than if no localization was present. However, if the localization described the topic being discussed in the paper (3% of implementable comments) there was no significant benefit of localization. Although comments were more likely to be implemented if localized, localization in general did not increase the likelihood of improved revision quality. The only type of localization that significantly predicted improvements in revision was localization that mentioned a specific location in the text, which actually decreased the likelihood of an improvement in quality by 2%. The results from the partial statistical models were generally consistent with the full statistical models (see Column 4 in Tables 5 and 6). In the partial statistical model, localization that identified quotes within the text also significantly decreased the likelihood of an improvement in quality by 1%. These partial statistical models were a significantly better fit than the baseline model, for implementation: $\chi^2(4) = 40.17$, $p < .001$, for revision quality: $\chi^2(4) = 9.42$, $p = .05$.

Overall, these results suggest that localization is an important feature for students' likelihood of implementation. Similar to Nelson and Schunn (2009), students were more likely to implement a comment if it was localized. However, the results leading to a decrease in revision quality adjusting for all other variables suggests that localized comments may limit the extent to which other more substantive revisions to complex problems were addressed.

Table 4
Effects of the Amount, Features, and Focus of Comments Received From Peers on Implementation

Implementation N = 2,714	Control variables			Praise			Problem/Solution		
Comment-level variables	Coef.	*p*	*SE*	Coef.	*p*	*SE*	Coef.	*p*	*SE*
Mean outcome in logits (γ00)	−.96	<.001	.11	−.86	<.001	.12	−.99	<.001	.13
Log-odds Δ in outcome due to:									
Control variables									
Writer ability	.02	.87	.14	.03	.85	.14	.02	.88	.14
Reviewer ability	.00	.98	.13	−.01	.93	.13	.01	.94	.13
Writer × Reviewer ability	.23	.15	.16	.22	.17	.16	.23	.15	.16
First draft quality	.01	.22	.01	.01	.24	.01	.01	.19	.01
Number of comments	.00	.97	.00	.00	.02	.00	.00	.96	.00
Number of implementable comments	.00	.30	.00	.00	.01	.00	.00	.20	.00
Word count (comment-level)	.00	.51	.00	.01	.18	.00	.00	.67	.00
Praise									
Praise only				.00	.01	.00			
Mitigated criticism				−.64	<.001	.14			
Comment features (reference: problem only)									
Solution only							−.08	.51	.11
Problem + Solution							.31	.03	.14
Localization (reference: not localized)									
Text									
Quote									
General									
Topic									
Focus (reference: low prose)									
High prose									
Substance									
Model fit statistics									
Deviance	7,476.01			7,470.67			7,471.04		
AIC	7,496.01			7,494.67			7,495.04		

Implementation n = 2,714	Localization			Focus			Full Model		
Comment-level variables	Coef.	*p*	*SE*	Coef.	*p*	*SE*	Coef.	*p*	*SE*
Mean outcome in logits (γ00)	−1.24	<.001	.12	−6.47	<.001	1.02	−.77	<.001	.18
Log-odds Δ in outcome due to:									
Control variables									
Writer ability	.04	.79	.14	−.16	.49	.23	.18	.17	.13
Reviewer ability	.00	.99	.13	−.07	.77	.25	−.11	.40	.13
Writer × Reviewer ability	.26	.11	.16	.28	.35	.30	.17	.28	.16
First draft quality	.01	.18	.01	.01	.44	.01	.02	.93	.22
Number of comments	.00	.94	.00	.00	.29	.00	−.09	.01	.03
Number of implementable comments	.00	.32	.00	.00	.67	.00	.09	.01	.03
Word count (comment-level)	.00	.59	.00	.05	<.001	.01	.00	.86	.00
Praise									
Praise only							.09	.01	.04
Mitigated criticism							−.50	<.001	.14
Comment features (reference: problem only)									
Solution only							−.22	.06	.12
Problem + Solution							.02	.88	.15
Localization (reference: not localized)									
Text	1.25	<.001	.15				1.20	<.001	.16
Quote	1.40	<.001	.24				1.24	<.001	.25
General	.99	<.001	.23				1.12	<.001	.24
Topic	−.05	.86	.31				−.07	.83	.31
Focus (reference: low prose)									
High prose				.66	.55	1.09	−.54	<.001	.15
Substance				2.48	.02	1.03	−.12	.43	.15
Model fit statistics									
Deviance	7,427.844			6,532.555			7,404.900		
AIC	7,455.84			6,556.55			7,444.90		

Note. Coef. = coefficient; AIC = Akaike information criterion.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 5
Effects of The Amount, Features, and Focus of Comments Received From Peers On Revision Quality

Revision Quality $N = 892$	Control variables			Praise			Problem/Solution		
Comment-level variables	Coef.	*p*	*SE*	Coef.	*p*	*SE*	Coef.	*p*	*SE*
Mean outcome in logits (γ00)	−2.06	<.001	.19	−2.01	<.001	.19	−2.01	<.001	.23
Log-odds Δ in outcome due to:									
Control variables									
Writer ability	.15	.55	.24	.11	.66	.25	.15	.55	.24
Reviewer ability	−.34	.14	.23	−.37	.11	.23	−.34	.14	.23
Writer × Reviewer ability	.38	.16	.27	.38	.17	.27	.38	.17	.27
First draft quality	−.11	.77	.37	−.07	.86	.38	−.10	.80	.37
Number of comments	−.01	.77	.02	.03	.57	.06	−.01	.73	.02
Number of implementable comments	.04	.17	.03	.00	.99	.06	.04	.17	.03
Word count (comment-level)	.00	.57	.01	.00	.78	.01	.00	.93	.01
Praise									
Praise only				−.05	.47	.07			
Mitigated criticism				−.31	.34	.32			
Comment features (reference: problem only)									
Solution only							.09	.73	.25
Problem + Solution							−.35	.27	.32
Localization (reference: not localized)									
Text									
Quote									
General									
Topic									
Focus (reference: low prose)									
High prose									
Substance									
Model fit statistics									
Deviance	2,275.500			2,272.188			2,274.06		
AIC	2,295.50			2,296.18			2,298.066		

Revision Quality $N = 892$	Localization			Focus			Full Model		
Comment-level variables	Coef.	*p*	*SE*	Coef.	*p*	*SE*	Coef.	*p*	*SE*
Mean outcome in logits (γ00)	−1.70	<.001	.19	−4.41	<.001	.53	−3.84	<.001	.56
Log-odds Δ in outcome due to:									
Control variables									
Writer ability	.13	.60	.25	.11	.65	.26	.08	.80	.30
Reviewer ability	−.30	.19	.23	−.33	.18	.24	−.04	.90	.30
Writer × Reviewer ability	.35	.21	.27	.39	.16	.28	.27	.41	.33
First draft quality	.01	.97	.38	−.05	.89	.39	−.42	.43	.53
Number of comments	−.01	.72	.02	−.01	.82	.02	.05	.48	.07
Number of implementable comments	.05	.09	.03	.04	.18	.03	−.08	.28	.07
Word count (comment-level)	.00	.81	.01	−.01	.20	.01	.00	.91	.01
Praise									
Praise only							−.05	.53	.08
Mitigated criticism							−.37	.28	.34
Comment features (reference: problem only)									
Solution only							.12	.66	.27
Problem + Solution							−.02	.96	.35
Localization (reference: not localized)									
Text	−2.00	<.001	.45				−1.53	.002	.48
Quote	−1.58	.01	.59				−1.25	.07	.67
General	−.53	.26	.47				−.52	.31	.51
Topic	−.74	.29	.70				−.51	.50	.76
Focus (reference: low prose)									
High prose				3.04	<.001	.53	2.24	<.001	.55
Substance				2.25	<.001	.52	2.69	<.001	.53
Model fit statistics									
Deviance	2,258.077			2,229.69			2,295.555		
AIC	2,286.07			2,253.699			2,325.55		

Note. Coef. = coefficient; AIC = Akaike information criterion.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

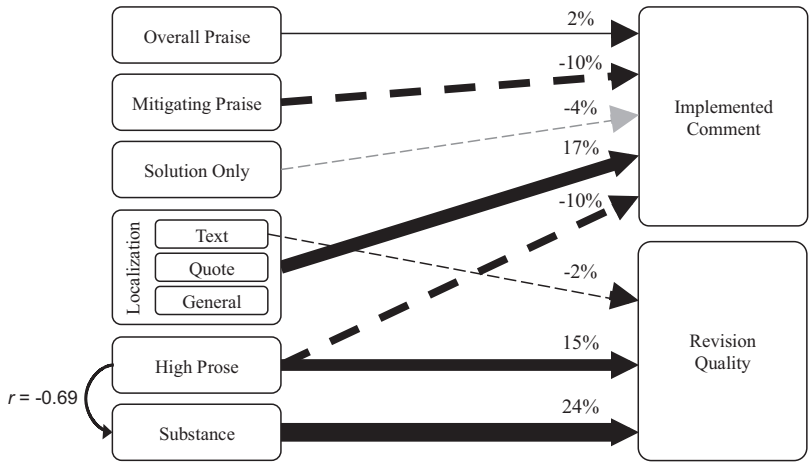

Figure 4. Updated nature of feedback model. The relative thickness of the lines indicates the probability of implementation/improved quality; gray line indicates a marginal effect ($p = .07$); dotted line indicates a negative relationship; curved line indicates a significant relation among predictors.

Focus of Critiques

Students received comments that focused on low prose relatively infrequently (18% of all comments)—perhaps due to the instructions provided to reviewers. More often, students received comments that focused on substance (32% of all comments) or high prose (51% of all comments).

Only high-prose comments had an influence on the likelihood of a comment being implemented. Students were 10% less likely to implement comments that focused on high prose than comments that focused on low prose. However, in contrast to low-prose comments, when high-prose comments were implemented, they led to a 15% higher likelihood of making a revision that improved the quality of the text. Furthermore, when substance comments were implemented, they led to a 22% higher likelihood of making a revision that improved the quality of the text. Importantly, the results from the partial statistical model for implementation were not consistent with the full statistical model (see Column 5 in Table 4). When not controlling for the other feedback features, high-prose comments appeared not to affect the implementation rate; rather students were 2% more likely to implement substance comments than comments that focused on low prose. By contrast, the results from the partial statistical model for revision quality were consistent with the full statistical model (see Column 5 in Table 5). These partial statistical models were a significantly better fit than the baseline model, for implementation: $\chi^2(2) = 939.46$, $p < .001$, for revision quality: $\chi^2(2) = 41.81$, $p < .001$.

In the absence of the other features, students appeared to place a higher value on the substance comments. However, when other features are accounted for, this affect disappears and students tended to avoid high-prose comments. Note that substance comments were less likely to be localized, and students were more likely to implement localized comments. Thus, when controlling for localization, the relationship between substance comments and localization likely cancelled out the positive effect of localized comments on implementation, and substance comments were no longer a significant predictor. By contrast, high-prose comments were more likely to be mitigated, and students were less likely to implement mitigated comments. Thus, when controlling for mitigation, the relationship between high-prose comments and mitigation likely amplified the negative effect of mitigated comments on implementation, and high-prose comments were more likely to be avoided. Moreover, that the revisions addressing high-prose and substance issues lead to greater improvements in the text quality (see Table 5, Focus and Full model columns) suggests that when writers attend to these substantive areas in their revision, they are more likely to improve their writing.

Additional Predictors: Influence of the Amount of Feedback

Several control variables were included in the statistical models (i.e., writer ability, reviewer ability, an interaction of writer ability by reviewer ability, the total number of comments received, the number of implementable comments received, the total word count of all the comments received, and the length of each comment). In the baseline model, none of these control variables significantly predicted whether a student implemented a comment (see Column 1 in Table 4) or whether a student improved the quality of their draft (see Column 1 in Table 5). Despite this lack of significance, we left these control variables in the models because they are theoretically important and the lack of simple effects appears to have been the result of lack of adjustment of correlations with other predictors.

In the full statistical model, the amount of feedback significantly predicted whether a student implemented the comment. Specifically, for each additional comment above the mean, the likelihood of implementing comments was on average 2% lower. By contrast, each additional *implementable* comment beyond the mean, on average, increased the likelihood of writers implementing comments by 2%. The amount of feedback did not influence revision quality. In the partial statistical models, the significance of the control variables varied. The only partial model in which the amount of feedback significantly predicted implementation was for praise. Additionally, the length of the comments became a significant predictor in the partial model for the focus of feedback.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

In both cases, the log-odds estimate was so low that the change in probability of implementing a comment is negligible.

Simply adding more comments did not promote a greater likelihood of implementation but having more implementable comments did. Perhaps implementable comments were perceived as more constructive and increased student motivation to implement, whereas additional comments without a clear prescription were seen as de-motivating. These results suggest that too much feedback could interfere with a student's likelihood of implementation. However, the more criticism a student receives, the more willing he or she is to revise. Perhaps, when the student receives too much praise and not enough criticism, the student may either over estimate his or her performance or experience an overload in working memory (Paas, Renkl, & Sweller, 2004; van Merrienboer & Sweller, 2005)—that is, while sifting through all the praise comments, important criticisms may be overlooked.

General Discussion

Updated Nature of Feedback Model

The current study provided new information about the effectiveness of feedback features on not only college students' likelihood of implementation of peer feedback but also their ability to revise effectively in response to the feedback (see Figure 4). Only two feedback features promoted students' likelihood of implementation: A writer was more likely to implement a comment if he or she received more than an average amount of praise, and, at the comment-level, he or she was more likely to implement a comment when the comment was localized. Several feedback features were associated with a lower likelihood of implementation: A writer was less likely to implement a comment that included mitigating praise, offered a solution, or focused on high-prose issues. Fewer feedback features affected students' ability to revise, and these effects contrasted the effects found for implementation. For example, although localized comments were more likely to be implemented, a writer was less likely to improve the quality of his or her paper by implementing those comments. Similarly, even though high-prose comments were less likely to be implemented, a writer was more likely to improve the quality of his or her paper by implementing comments that focused on high-prose issues. Similarly, the revision quality also improved when writers implemented comments focused on issues of substance as opposed to low prose.

We revisited Nelson and Schunn's (2009) nature of feedback model by examining the effectiveness of peer feedback in a new context using more sophisticated statistical analyses and increased statistical power due to a larger sample size. Moreover, we added to the theoretical model by examining a broader set of feedback features (i.e., focus of feedback and amount of feedback) and their effect on not only implementation but also revision quality. Only one of the effects found in the original study was replicated—that is, students were more likely to implement a comment that was localized. In writing, localization can be especially helpful when a problem could occur in multiple locations (e.g., the use of transitions may not be consistent throughout the document). Without localization, a writer might not know how to address a comment that only mentions that there is a problem with some of the transitions. Localization in this context acts similarly to error-flagging within the study of intelligent tutors (Corbett & Anderson, 1991, 2001; Kumar, 2010, 2012) and to the study of feedback specificity involved in training management students (Goodman & Wood, 2004; Goodman et al., 2011; Goodman et al., 2004). These results are also consistent with the work on feedback in the form of marginalia, where students implement most of the suggestions teachers make in the margins of their papers (Ene & Upton, 2014; Ferris, 1997).

There were several interesting dissimilarities between the original theoretical model and the updated theoretical model. First, with a much larger sample and the use of more sophisticated statistical analyses, we were able to detect several additional relationships between feedback features and implementation. The original theoretical model uncovered the importance of problem understanding—that is, the presence of solutions, summarization, and localization increased problem understanding and the presence of an explanation about the problem decreased problem understanding (Nelson & Schunn, 2009). Here, we did not have measures of underlying factors such as understanding or agreement. Although understanding the underlying mechanism is important, identifying significant relationships is just as important for improving instruction in writing, and for guiding future research on underlying mechanisms.

Another difference between the original theoretical model and the updated theoretical model is the effect of praise. In the original theoretical model, praise was not significantly related to implementation or any of the tested mediators (Nelson & Schunn, 2009). In the updated theoretical model, praise was significantly related to implementation. Consistent with prior literature, overall praise may have motivated students to address the problems detected by their peers, as indicated by the increased likelihood of implementation in the current study (Bienstock et al., 2007; Hesketh & Laidlaw, 2002). However, not all praise was useful. Also consistent with prior literature, when praise was used to mitigate a criticism comment, it undermined the severity of the problem, and students were more likely to ignore the problem (Richardson, 2004; Schlecht, 2008; Shute, 2008). This inconsistency with the original theoretical model may indicate that some construct other than problem understanding mediates the relationship between praise and implementation. Nelson and Schunn (2009) posited that agreement could possibly mediate this relationship; however, they did not find any evidence to support their hypothesis.

Finally, the effect of solution in the updated theoretical model conflicts with the findings from the original theoretical model. In the original theoretical model, students were more likely to implement a comment that included a solution (Nelson & Schunn, 2009). However, in the updated theoretical model, there was only a marginal effect of solutions, and moreover, it was in the opposite direction—that is, students were less likely to implement a comment if it only included a solution. Furthermore, the ignored comments tended to be at a higher level. Again, these results are consistent with prior research on more advanced writers (Bitchener, 2012) and when the focus of revisions is errors beyond the grammatical level (Van Beuningen et al., 2012). Ene and Upton (2014) also speculated that substance feedback at more advanced stages of writing would likely be more complex and, thus, more difficult for students to implement. Perhaps, when given only a solution for a higher-level issue, a student does not have enough information to recognize the problem that the reviewer was suggesting to fix. Without understanding what the problem was or

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

why the existing prose was problematic, the students in our sample chose to ignore the comment.

The current study also expanded on the original theoretical model in several ways. First, we more deeply coded several of the features examined in the original theoretical model. While coding for localization, we noticed that students used different techniques to localize a comment. These techniques can range from very specific to more general. Three of these techniques were important markers for the likelihood of implementing a comment: text (i.e., mentioning a specific location—usually with a page and paragraph number), quote (i.e., using a quote to identify where the issue occurred), and general (i.e., mentioning a general location such as the introduction or conclusion). Given the short length of the writing assignment, general localizations typically referred to the first paragraph (introduction) or the last paragraph (conclusion), making this type of localization relatively specific. These results suggest that the more specific the comment is when identifying the location of the issue, the more likely a student would be to include that comment in their revision. We also examined other ways of doing praise (e.g., highly positive, moderately positive), and patterns similar to the ones reported were found. For simplicity, we only included overall praise in the updated theoretical model.

An additional feedback feature (i.e., feedback focus) was included in the updated theoretical model. The focus of feedback is especially relevant for Writing in the Discipline environments. We expected that high-prose and substance issues would be more difficult for students to address, and thus, they would be more likely to ignore these comments. Indeed, students were less likely to implement comments that focused on high-prose issues. These findings confirm Ene and Upton's (2014) conclusion. We also included several control variables in the updated theoretical model, such as writer ability, reviewer ability, first draft quality, number of comments, number of implementable comments, and word count. More feedback in general was not as helpful to students as receiving more implementable comments. These results are consistent with the finding that writers may use less effective strategies when sifting through a substantial amount of comments to decide which comments to implement (Ariely, 2000; Iselin, 1988).

Perhaps more surprising was the lack of effect for writer ability or reviewer ability. Despite substantial evidence in the literature regarding developmental differences in revising for large differences in expertise (e.g., Hayes et al., 1987), neither writing ability nor reviewer ability predicted the likelihood of a student implementing a comment or the likelihood of a revision improving the quality of the document. Of course, the range of writing ability within a given class is likely to be smaller than ability spans studied in developmental or expertise research.

The most substantial extension of the updated theoretical model was the influence of feedback features on the quality of the revision. Only three feedback features affected students' ability to revise. First, although localized comments were more likely to be implemented, comments that were localized by including a specific location in the text were less likely to improve the quality of the paper. There are several possible interpretations: (a) It may be that localized comments led students to consider revision only in one instance when the problem actually occurred in multiple locations, (b) it may be that less important issues even within problem focus types are more localizable, or (c) it may be that students focusing on localized comments simply chose to implement some comments, and the ones they chose did not lead to higher revision quality. Future research should examine this finding in greater depth. Second, although high-prose comments were less likely to be implemented, revisions associated with high-prose comments were more likely to improve the quality of the paper. Similarly, revisions in response to substance comments were also more likely to improve the quality of the paper. It is not entirely surprising that high-prose and substance issues are more likely to affect the overall document quality than low-prose issues, but it was not necessarily the case that students would be able to repair high-prose or substance issues with sufficiently high success rates on the basis of peer feedback.

Practical Implications

The current study provided evidence to support common reviewing recommendations. Advice about how to provide helpful feedback often suggests including praise (Grimm, 1986; Nilson, 2003; Saddler & Andrade, 2004) and localization (Nilson, 2003). However, the use of praise and localization should be used with caution. Although praise overall seems to motivate students to revise, mitigating praise appears to interfere with the revisions students should be making. These findings suggest that students might benefit more from reviewing tasks that include a separate dimension where students can comment about what was done well in the paper. Praise may be more useful when it is more specific and not interspersed within the critical comments.

Localization also had contradicting effects on implementation and revision quality. Comments that included a specific location of the issue (via the location in the text, quote, or identifying a general location such as the introductory paragraph or concluding paragraph) were more likely to be implemented but were less likely to improve the quality of the paper.

Finally, given the impact high-prose and substance comments have on revision quality and the difficulty students seem to have dealing with these higher level issues, an increase in support may be needed to promote the use of these comments. Perhaps localization could be used to support students' implementation of high-prose and substance comments.

Caveats and Future Directions

There are several caveats that should be considered. The updated theoretical model of peer feedback was primarily based on a literature review of adult learners and the evidence provided by the current study also involved adult learners. Therefore, caution must be taken when considering younger learners, and future research should be conducted to determine how younger learners respond to feedback in general and peer feedback specifically. Additionally, we would like to note that the ways features influence feedback implementation could be influenced by the way comments are shown to authors (e.g., end comments vs. marginalia) and the ways in which students make changes to documents (e.g., revising handwritten documents vs. electronic documents). This updated model focuses on revising electronic documents using end comments. Future research should extend the model by examining other factors that may interact with how feedback features influence implementation and revision quality.

There are two cautions to the findings about focus that should be considered. First, the rubric used to assess writing quality primar-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

ily focused on high-prose and substance. Thus, the findings that only high-prose and substance comments were more likely to improve the quality of the paper may be a function of the rubric, although extensive low-prose issues can influence document understanding. Future research should include a broader writing quality rubric that incorporated a dimension for low prose. Second, focus had the lowest interrater reliability (i.e., $\kappa = .54$). One may wonder whether the focus category should have only two categories rather than three categories. Indeed, the interrater reliability was higher when considering just lower order comments versus higher order comments, which reflects the existence of a gray area between the categories. However, there are important conceptual reasons for using the three categories, especially in the context of Writing in the Discipline courses that strongly value issues of substance per se. Moreover, some of the results are different by the three categories, reinforcing the need to make the three-way distinction and providing predictive validity evidence that our coding was sufficiently robust. Future research should more closely examine these focus categories to determine how to more reliably categorize the focus of feedback.

Next, there are several limitations of the updated model that should be addressed in future research. One limitation is that we did not test for potential mediators. The original theoretical model found that relevant feedback features affected problem understanding, which affected implementation. Future research should incorporate this possible mediator along with other potentially relevant mediators (e.g., agreement with comment, motivation, memory load). Another limitation to the updated theoretical model is that it is also based on correlational analyses, although relatively sophisticated correlational analyses that attempt to adjust for a wide range of possible confounds. Future research should replicate the findings in experimental settings where the presence of each feedback feature is manipulated (e.g., some students receive feedback that is localized and some students receive feedback that is not localized).

Finally, we have identified several areas for future research. One feedback feature that has been overlooked is the distinction between implicit and explicit feedback. This distinction might be difficult to code reliably. Moreover, while some researchers strongly believe this distinction is important to consider (Li, 2010), others argue that all written feedback must be considered explicit (Ellis, 2009; Sheen, 2010). Thus, we have decided to exclude this feature from the current study and suggest that future research should further examine how to make this distinction reliably—perhaps with multiple categories rather than just as a dichotomous feature. Future research should also follow up on the negative relationship between feedback amount and implementation. There was not enough evidence from the current study to provide a suggestion regarding the optimal length of peer feedback. Therefore, future research should more closely examine the effects of the amount of feedback to support more a concrete recommendation. In addition, future research should examine how interventions can be utilized to improve the quality of students' comments and the consequential effects. Litman and colleagues (Nguyen & Litman, 2014; Nguyen et al., 2014; Xiong et al., 2012) have developed tools that automatically detect the absence of localization and solutions in peer feedback. These tools can be used to influence the comments that students produce.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

References

AbuSeileek, A., & Abualsha'r, A. (2014). Using peer computer-mediated corrective feedback to support EFL learner's writing. *Language Learning & Technology, 18,* 76–95.

Agius, N. M., & Wilkinson, A. (2014). Students' and teachers' views of written feedback at undergraduate level: A literature review. *Nurse Education Today, 34,* 552–559. http://dx.doi.org/10.1016/j.nedt.2013.07.005

Ariely, D. (2000). Controlling the information flow: Effects on consumers' decision making and preferences. *Journal of Consumer Research, 27,* 233–248.

Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research, 13,* 111–127. http://dx.doi.org/10.2190/9LMD-3U28-3A0G-FTQT

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61,* 218–238. http://dx.doi.org/10.3102/00346543061002213

Bazerman, C. (2008, November). U.S. writing across the curriculum and writing in the disciplines in context. In D. Alamargot, J. Bouchand, E. Lambert, V. Millogo & C. Beaudet (Eds.), *Proceedings of the International Conference From France to Quebec: Writing in All Its Forms*. Poitiers, France.

Beaufort, A. (2004). Developmental gains of a history major: A case for building a theory of disciplinary writing expertise. *Research in the Teaching of English, 39,* 136–185.

Bienstock, J. L., Katz, N. T., Cox, S. M., Hueppchen, N., Erickson, S., & Puscheck, E. E., & the Association of Professors of Gynecology and Obstetrics Undergraduate Medical Education Committee. (2007). To the point: Medical education reviews—Providing feedback. *American Journal of Obstetrics and Gynecology, 196,* 508–513. http://dx.doi.org/10.1016/j.ajog.2006.08.021

Bitchener, J. (2012). Written corrective feedback for L2 development: Current knowledge and future research. *TESOL Quarterly, 46,* 855–860. http://dx.doi.org/10.1002/tesq.62

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing, 14,* 191–205. http://dx.doi.org/10.1016/j.jslw.2005.08.001

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11,* 717–726. http://dx.doi.org/10.1016/S0022-5371(72)80006-9

Brown, F. J. (1932). Knowledge of results as an incentive in school room practice. *Journal of Educational Psychology, 23,* 532–552. http://dx.doi.org/10.1037/h0074392

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145–182. http://dx.doi.org/10.1207/s15516709cog1302_1

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49,* 219–243. http://dx.doi.org/10.1080/00461520.2014.965823

Cho, K., Schunn, C. D., & Charney, D. (2006). Commenting on writing: Typology and perceived helpfulness of comments from novice peer reviewers and subject matter experts. *Written Communication, 23,* 260–294. http://dx.doi.org/10.1177/0741088306289261

Corbett, A. T., & Anderson, J. R. (1991). *Feedback control and learning to program with the CMU Lisp Tutor*. Paper presented at the 1991 American Educational Research Association (AERA) Annual Meeting, Chicago, IL.

Corbett, A. T., & Anderson, J. R. (2001). *Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement, and attitudes*. Paper presented at the ACM CHI'2001 Conference on Human Factors in Computing Systems, Seattle, WA.

Dohrenwend, A. (2002). Serving up the feedback sandwich. *Family Practice Management, 9,* 43–46.

Duncan, N., Prowse, S., Wakeman, C., & Harrison, R. (2007). "Feedforward": Improving students' use of tutors' comments. *Assessment & Evaluation in Higher Education, 32,* 271–283. http://dx.doi.org/10.1080/02602930600896498

Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal, 63,* 97–107. http://dx.doi.org/10.1093/elt/ccn023

Ene, E., & Upton, T. A. (2014). Learner uptake of teacher electronic feedback in ESL composition. *System, 46,* 80–95. http://dx.doi.org/10.1016/j.system.2014.07.011

Esmaeili, F., & Behnam, B. (2014). A study of corrective feedback and learner's uptake in classroom interactions. *International Journal of Applied Linguistics & English Literature, 3,* 204–212. http://dx.doi.org/10.7575/aiac.ijalel.v.3n.4p.204

Ferris, D. R. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly, 29,* 33–53. http://dx.doi.org/10.2307/3587804

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

Ferris, D. R., Brown, J., Liu, H., & Stine, M. E. A. (2011). Responding to L2 students in college writing classes: Teacher perspectives. *TESOL Quarterly, 45,* 207–234. http://dx.doi.org/10.5054/tq.2011.247706

Ferris, D. R., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing, 22,* 307–329. http://dx.doi.org/10.1016/j.jslw.2012.09.009

Frey, M. C., & Detterman, D. K. (2004). Scholastic Assessment or *g*? The relationship between the Scholastic Assessment Test and general cognitive ability. *Psychological Science, 15,* 373–378. http://dx.doi.org/10.1111/j.0956-7976.2004.00687.x

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

Goodman, J. S., & Wood, R. E. (2004). Feedback specificity, learning opportunities, and learning. *Journal of Applied Psychology, 89,* 809–821. http://dx.doi.org/10.1037/0021-9010.89.5.809

Goodman, J. S., Wood, R. E., & Chen, Z. (2011). Feedback specificity, information processing, and transfer of training. *Organizational Behavior and Human Decision Processes, 115,* 253–267. http://dx.doi.org/10.1016/j.obhdp.2011.01.001

Goodman, J. S., Wood, R. E., & Hendrickx, M. (2004). Feedback specificity, exploration, and learning. *Journal of Applied Psychology, 89,* 248–262. http://dx.doi.org/10.1037/0021-9010.89.2.248

Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal, 115,* 523–547. http://dx.doi.org/10.1086/681947

Grimm, N. (1986). Improving students' responses to their peers' essays. *College Composition and Communication, 37,* 91–94. http://dx.doi.org/10.2307/357386

Hausmann, R. G. M., Vuong, A., Towle, B., Fraundorf, S. H., Murray, R. C., & Connelly, J. (2013). *An evaluation of the effectiveness of just-in-time hings*. Paper presented at the 16th International Conference on Artificial Intelligence in Education (AIED 2013), Memphis, TN.

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 Reading, writing, and language learning* (pp. 176–240). New York: Cambridge University Press.

Hesketh, E. A., & Laidlaw, J. M. (2002). Developing the teaching instinct, 1: Feedback. *Medical Teacher, 24,* 245–248. http://dx.doi.org/10.1080/01421590220140991l

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organization. *Journal of Applied Psychology, 64,* 349–371. http://dx.doi.org/10.1037/0021-9010.64.4.349

Iselin, E. R. (1988). The effects of information load and information diversity on decision quality in a structured decision task. *Accounting, Organizations and Society, 13,* 147–164. http://dx.doi.org/10.1016/0361-3682(88)90041-4

Kellogg, R. T. (1994). *The psychology of writing*. Oxford, UK: New York: Oxford University Press.

Kirkland, M. R., & Saunders, M. A. P. (1991). Maximizing student performance in summary writing: Managing cognitive load. *TESOL Quarterly, 25,* 105–121. http://dx.doi.org/10.2307/3587030

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,* 254–284. http://dx.doi.org/10.1037/0033-2909.119.2.254

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36,* 757–798. http://dx.doi.org/10.1111/j.1551-6709.2012.01245.x

Kumar, A. N. (2010, June). *Error-flagging support for testing and its effect on adaptation*. Paper presented at the 10th International Conference on the Intelligent Tutoring Systems (ITS), Pittsburgh, PA.

Kumar, A. N. (2012, June). *Limiting the number of revisions while providing error-flagging support during tests*. Paper presented at the 11th International Conference on the Intelligent Tutoring Systems (ITS), Chania, Crete, Greece.

LeBaron, S. W. M., & Jernick, J. (2000). Evaluation as a dynamic process. *Family Medicine, 32,* 13–14.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60,* 309–365. http://dx.doi.org/10.1111/j.1467-9922.2010.00561.x

Matsumura, L. C., Patthey-Chavez, G. G., Valdes, R., & Garnier, H. (2002). Teacher feedback, writing assignment quality, and third-grade students' revision in lower- and higher-achieving urban schools. *The Elementary School Journal, 103,* 3–25. http://dx.doi.org/10.1086/499713

Molloy, E. K., & Boud, D. (2014). Feedback models for learning, teaching and performance. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology,* (4th ed., pp. 413–424). New York, NY: Springer Science+Business Media http://dx.doi.org/10.1007/978-1-4614-3185-5_33

Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (2nd ed., pp. 745–783). Mahwah, NJ: Erlbaum.

Nakamaru, S. (2010). Lexical issues in writing center tutorials with international and U.S.-educated multilingual writers. *Journal of Second Language Writing, 19,* 95–113. http://dx.doi.org/10.1016/j.jslw.2010.01.001

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

Nguyen, H. V., & Litman, D. (2014, June). *Improving peer feedback predictions: The sentence level is right*. Paper presented at the 9th Workshop on Innovative Use of NLP for Building Educational Applications (BEA), ACL 2014 Workshop, Baltimore, MD.

Nguyen, H. V., Xiong, W., & Litman, D. (2014, June). *Classroom evaluation of a scaffolding intervention for improving peer review localiza-*

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

tion. Paper presented at the 12th International Conference on Intelligent Tutoring Systems (ITS), Honolulu, HI.

Nilson, L. B. (2003). Improving student peer feedback. *College Teaching, 51,* 34–38. http://dx.doi.org/10.1080/87567550309596408

Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 398–415. http://dx.doi.org/10.1037/0278-7393.17.3.398

Orrell, J. (2006). Feedback on learning achievement: Rhetoric and reality. *Teaching in Higher Education, 11,* 441–456. http://dx.doi.org/10.1080/13562510600874235

Orsmond, P., & Merry, S. (2011). Feedback alignment: Effective and ineffective links between tutors' and students' understanding of coursework feedback. *Assessment & Evaluation in Higher Education, 36,* 125–136. http://dx.doi.org/10.1080/02602930903201651

Paas, F. G. W. C., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32,* 1–8. http://dx.doi.org/10.1023/B:TRUC.0000021806.17516.d0

Parkes, J., Abercrombie, S., & McCarty, T. (2013). Feedback sandwiches affect perceptions but not performance. *Advances in Health Sciences Education: Theory and Practice, 18,* 397–407. http://dx.doi.org/10.1007/s10459-012-9377-9

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,* 124–152. http://dx.doi.org/10.17239/jowr-2009.01.02.2

Patchan, M. M., Hawk, B., Stevens, C. A., & Schunn, C. D. (2013). The effects of skill diversity on commenting and revisions. *Instructional Science, 41,* 381–405. http://dx.doi.org/10.1007/s11251-012-9236-3

Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: How students respond to peers' texts of varying quality. *Instructional Science, 43,* 591–614. http://dx.doi.org/10.1007/s11251-015-9353-x

Patchan, M. M., & Schunn, C. D. (in press).Understanding the benefits of receiving peer feedback for text revision: Relations between author and reviewer ability. *Journal of Writing Research*.

Patchan, M. M., Schunn, C. D., & Clark, R. J. (2011). Writing in natural sciences: Understanding the effects of different types of reviewers on the writing process. *Journal of Writing Research, 2,* 365–393. http://dx.doi.org/10.17239/jowr-2011.02.03.4

Pressey, S. L. (1926). A simple device which gives tests and scores: And teaches. *School and Society, 23,* 373–376.

Pressey, S. L. (1927). A machine for automatic teaching of drill material. *School and Society, 25,* 549–552.

Prins, F. J., Sluijsmans, D. M. A., & Kirschner, P. A. (2006). Feedback for general practitioners in training: Quality, styles, and preferences. *Advances in Health Sciences Education: Theory and Practice, 11,* 289–303. http://dx.doi.org/10.1007/s10459-005-3250-z

Prior, P. (2006). A sociocultural theory of writing. *Handbook of writing research, 54*–66.

Raudenbush, S. W., & Bryk, A. S. (2002). Models for cross-classified random effects. *Hierarchical linear models: Applications and data analysis methods* (2nd ed., pp. 373–398). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2004). *HLM 6 for Windows* [Computer software]. Skokie, IL: Scientific Software International, Inc.

Richardson, B. K. (2004). Feedback. *Academic Emergency Medicine, 11,* e1–e5. http://dx.doi.org/10.1111/j.1553-2712.2004.tb01913.x

Saddler, B., & Andrade, H. (2004). The writing rubric. *Educational Leadership, 62,* 48–52.

Schlecht, K. D. (2008). Feedback. *International Anesthesiology Clinics, 46,* 67–84. http://dx.doi.org/10.1097/AIA.0b013e31818623f3

Sheen, Y. (2010). Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition, 32,* 203–234. http://dx.doi.org/10.1017/S0272263109990507

Shintani, N., Ellis, R., & Suzuki, W. (2014). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning, 64,* 103–131. http://dx.doi.org/10.1111/lang.12029

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428. http://dx.doi.org/10.1037/0033-2909.86.2.420

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

Sugita, Y. (2006). The impact of teachers' comments types on students' revision. *ELT Journal, 60,* 34–41. http://dx.doi.org/10.1093/elt/cci079

Symonds, P. M., & Chase, D. H. (1929). Practice vs. motivation. *Journal of Educational Psychology, 20,* 19–35. http://dx.doi.org/10.1037/h0069958

Tseng, S.-C., & Tsai, C.-C. (2007). On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education, 49,* 1161–1174. http://dx.doi.org/10.1016/j.compedu.2006.01.007

Turnitin. (2013). Closing the gap: What students say about instructor feedback. Retrieved from http://go.turnitin.com/what-students-say-about-teacher-feedback

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,* 1–41. http://dx.doi.org/10.1111/j.1467-9922.2011.00674.x

van Merrienboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17,* 147–177. http://dx.doi.org/10.1007/s10648-005-3951-0

Voerman, L., Korthagen, F. A. J., Meijer, P. C., & Simons, R. J. (2014). Feedback revisited: Adding perspectives based on positive psychology. Implications for theory and classroom practice. *Teaching and Teacher Education, 43,* 91–98. http://dx.doi.org/10.1016/j.tate.2014.06.005

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

Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education, 31,* 379–394. http://dx.doi.org/10.1080/02602930500353061

Xiong, W., Litman, D., & Schunn, C. (2012). Natural language processing techniques for researching and improving peer feedback. *Journal of Writing Research, 4,* 155–176. http://dx.doi.org/10.17239/jowr-2012.04.02.3

Zoghi, A., & Nikoopour, J. (2013). The interface of error types, teacher's feedback, and students' uptake. *International Journal of Applied Linguistics & English Literature, 3,* 54–62. http://dx.doi.org/10.7575/aiac.ijalel.v.3n.1p.54

Zohrabi, M., & Rezaie, P. (2012). The role of form-focused feedback on developing students' writing skill. *Theory and Practice in Language Studies, 2,* 1514–1519. http://dx.doi.org/10.4304/tpls.2.7.1514-1519

(*Appendices follow*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix A

Experts' Writing Quality Rubric

Rating	Description
	Flow
How well the paragraphs were developed	
5–Very Good	All paragraphs stated a point and developed it
4–Good	Most paragraphs stated a point and developed it
3–Fair	Some paragraphs stated a point and developed it. All paragraphs introduced a topic, but may not state an explicit point
2–Poor	Some paragraphs stated a point *or* introduced a topic, but did not develop it
1–Unsatisfactory	No paragraphs stated a point and/or paragraphs shifted topics frequently.
How well transitions connected paragraphs	
5–Very Good	Strong transitions between all paragraphs
4–Good	Strong transitions between most paragraphs
3–Fair	Transitions between most paragraphs, but some were weak
2–Poor	Weak transitions between some of the paragraphs
1–Unsatisfactory	No transitions between paragraphs
How well this paper was organized around a main idea	
5–Very Good	All paragraphs were connected to the main point
4–Good	Most paragraphs were connected to the main point
3–Fair	Some paragraphs were connected to the main point
2–Poor	Most paragraphs were not connected to the main point
1–Unsatisfactory	No main point explicitly stated
	Argument logic
How well the author evaluated the MSNBC article	
5–Very Good	All points were supported by concrete evidence or examples
4–Good	Most points were supported by concrete evidence or examples
3–Fair	Some points were supported by concrete evidence or examples
2–Poor	Few points were supported by concrete evidence or examples
1–Unsatisfactory	No support was provided
How well the author explained causal conclusions	
5–Very Good	Provided a complete and clear explanation (i.e., A $\rightarrow$ B; B $\leftarrow$ A; C $\rightarrow$ [A $\rightarrow$ B])
4–Good	Provided a complete and somewhat clear explanation
3–Fair	Provided complete but unclear explanation
2–Poor	Provided an incomplete explanation
1–Unsatisfactory	No explanation was provided
How well the author explained an alternative possibility	
5–Very Good	Provided an appropriate and clear alternative
4–Good	Provided an appropriate and somewhat clear alternative
3–Fair	Provided an appropriate alternative, but did not explain it
2–Poor	Provided an inappropriate alternative
1–Unsatisfactory	No alternative possibility was provided
Whether all the required information from the research article was accurately provided	
5–Very Good	The summary accurately included all of the required information
4–Good	The summary accurately included most of the required information
3–Fair	The summary accurately included some required information
2–Poor	The summary included little required information *or* the information was inaccurate
1–Unsatisfactory	No summary of the article
	Insight
How well the main point was connected to a larger issue	
5–Very Good	Main point was fully connected to a relevant larger issue throughout the whole paper
4–Good	Main point was connected to a relevant larger issue
3–Fair	Some points demonstrated an innovative analysis, but these points were not connected to a relevant larger issue
2–Poor	One point demonstrated an innovative analysis, but this point was not connected to a relevant larger issue
1–Unsatisfactory	No points demonstrated an innovative analysis

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix B

Peer Feedback Coding Scheme

Category	Definition	Example
	All comments	
Praise	A comment that described a positive feature of the paper	This writer showed perfect knowledge of the article being discussed.
Problem	Description of something wrong with the paper	The writer did not offer insight into causal and correlational relationships.
Solution	Suggestion for 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 (i.e., problems and solutions)	
Localization: A comment that describes the location of the issue		
Text	Mentioned a specific location to identify where the issue occurred	Also, consider rewording the first sentence in the third paragraph.
Quote	Used a quote to identify where the issue occurred	There's a slight contradiction in the conclusion of your paper when it states that "Overall, the media article was consistent with the data from the study," right after you discuss the problems with the translation from the research to the media article.
General	Mentioned a general location that could be found in any paper	The problem again is the lack of a strong introduction. The introduction does not really tell the reader the direction of the paper.
Topic	Mentioned the topic to identify where the issue occurred	They go off topic a little bit when talking about aggression that is somewhat irrelevant to the topic.
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 counterarguments, 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 trhe [*sic*] readers attention."
Substance	An issue with missing, incorrect, or contradictory content	I don't see where you stated the independent and dependent variables.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix C

Implementation and Revision Quality Coding Scheme

Rating	Comment	Draft 1 text	Draft 2 text
0	I would try and revise the last sentence in the last paragraph so that it makes more sense and you can get your point across.	Playing video games causes bad health because there is not any physical activity evolved. Gaming only consists of sitting and pressing buttons. It can cause weight gain and it probably affects the eyes. From the lack of movement and the junk food that is most likely involved, adults will gain weight easier then juveniles due to the fact that their metabolism has slowed down since their teenage years. *The fact that video gaming involves continuously looking at a screen, it most likely causes strain to the eyes.*	Playing video games causes bad health because there is not any physical activity involved. Although there are a few types of interactive gaming devices, typical gaming only consists of sitting and pressing buttons. It can cause weight gain and it probably affects the eyes. From the lack of movement and the junk food that is most likely involved, adults will gain weight easier then juveniles due to the fact that their metabolism has slowed down since their teenage years. In turn, this study does not actually have any one correct answer to it because of the fact that casual conclusions cannot be formed from correlative studies. *The fact that video gaming involves continuously looking at a screen, it most likely causes strain to the eyes. All I all, the overall outcome of gaming has more on a negative effect on the human body than a positive effect.*
1	In the first sentence of the third to last paragraph: "correlational studies cannot end in causal conclusions because the relationship between the two variables is always presently unknown," what does "the relationship between the two variables is always presently unknown" mean?	Correlational studies cannot end in causal conclusions because *the relationship between the* two variables is always presently unknown. The variables being studied may have the group studied in common, but that cannot prove whether one causes the other or even whether they relate at all. Correlated variables usually occur alongside each other in time, not one after another, thus there is no scientifically valid cause and effect.	Correlational studies cannot end in causal conclusions because *the relationship between the two variables is always presently unknown. At any one period in time, variables with a correlational relationship can exist simultaneously. For instance, a gamer can be both depressed and fat before, during, or even after playing video games. In a correlational study, the variables are not dependent upon each other's existence in time, much unlike variables with casual natures.* The variables being studied may have the "group studied'" in common, but that cannot prove whether one causes the other or even whether they relate at all. Correlated variables *are capable of occurring* alongside each other in time, *and are* not *limited to* one after another, thus there is no scientifically valid cause and effect *on that basis alone.*

Received March 27, 2015
Revision received October 27, 2015
Accepted November 8, 2015 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.