Instructional Science (2021) 49:365–394
https://doi.org/10.1007/s11251-021-09546-5

ORIGINAL RESEARCH

# From plans to actions: A process model for why feedback features influence feedback implementation

**Yong Wu**[1,2] · **Christian D. Schunn**[2]

Received: 3 March 2019 / Accepted: 24 May 2021 / Published online: 25 June 2021


**Abstract**
Implementing peer feedback in revisions is a complex process involving first planning to fix problems and then actual implementing feedback through revisions. Both phases are influenced by features of the peer feedback itself, but potentially in different ways, and yet prior research has not examined their separate role in planning or the mediating role of planning in the relationship of feedback features and implementation. We build on a process model to investigate whether feedback features had differing relationships to plans to ignore or act on feedback versus actual implementation of feedback in the revision, and whether planning mediated the relationship of feedback features and actual implementation. Source data consisted of peer feedback comments received, revision plans made, and revisions implemented by 125 US high school students given a shared writing assignment. Comments were coded for feedback features and implementation in the revision. Multiple regression analyses revealed that having a comment containing a specific solution or a general suggestion predicted revision plans whereas having a comment containing an explanation predicted actual implementation. Planning mediated the relationship to actual implementation for the two feedback features predicting plans, suggestion and solution. Implications for practice are discussed.

**Keywords** Feedback features · Implementation · Peer review · Planning · Revision

## Introduction

Peer feedback involves students exchanging information about their performance with the aim of narrowing the gap of their current performance and the desired performance (Panadero et al., 2018; Shute, 2008). Peer feedback is increasingly included in a variety of educational settings for different purposes (e.g., summative and/or formative purposes, collaborative learning) (Kluger & DeNisi, 1996; Topping, 1998; van Gennip et al., 2010).

✉ Yong Wu
yongwu@pitt.edu

[1] Center for Research on Technology-Enhanced Language Education, School of Humanities, Beijing University of Posts and Telecommunications, Beijing 100876, China

[2] Learning Research and Development Center, University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA 15217, USA

When the student-to-teacher ratio is large, peer feedback is commonly used to make it possible for students to receive timely and extensive feedback (Cho & Schunn, 2007; Gielen & De Wever, 2015a; Huisman et al., 2018). Peer feedback also improves the learning process by having students take a more active role by also giving feedback (Li et al., 2020; Zheng et al., 2020). Therefore, teachers are generally encouraged to use peer feedback in their instruction to enhance student learning (Gielen & De Wever, 2015a; Huisman et al., 2018; Topping, 1998).

Peer review, the broader process, consists of numeric scores (often called peer assessment) and qualitative written or oral comments (often called peer feedback). In the present study, peer feedback is used to refer to qualitative written comments about strengths and weaknesses of student writing performance for the purpose of formative assessment. Existing research on peer feedback has tended to focus on the factors that influence the actual implementation of peer feedback or the quality of revised drafts (Leijen, 2017; Lu & Law, 2012; Nelson & Schunn, 2009; Patchan et al., 2016; Tseng & Tsai, 2007; Wu & Schunn, 2020a). There is much evidence to suggest that peer feedback, when appropriately structured, can have a favorable impact on student writing (Cho & Schunn, 2007; Gielen & De Wever, 2015a). By contrast, little research exists on factors that influence students at the planning stage, and the role of planning on the relationship of feedback features and implementation. Investigating factors that play roles at different stages of revising will build a more detailed picture of the thinking process that students undergo in the process of feedback implementation. This can advance understanding of the underlying mechanisms of learning through implementing feedback and help to find more optimal ways to support students' implementation and learning. Therefore, the present study seeks to identify feedback features associated with students' plans and with actual implementation, and test mediated effects of planning on the relationship of feedback features and implementation.

## Theoretical background

The present study aims to investigate the effects of peer feedback features on students' (i.e., authors/feedback receivers) plans to revise and on their actual implementation. We argue why feedback features might influence student's decisions to act upon or ignore the feedback they receive, and why these features might be differentially influential across planning and actual implementation phases. The present study also explores mediated effects of planning on feedback features and actual implementation. We argue that planning likely mediates the relationships of some feedback features to actual implementation. This process model, coupled with the specific research findings regarding key feedback features, can shed some light on the underlying mechanisms for why students implement or ignore peer feedback, which can be quite useful to improving writing instruction. In pursuit of this goal, we provide an overview of key prior research on feedback-based revision process and past work on critical feedback features.

### Planning vs. implementation in feedback-based revision processes

For feedback to strongly contribute to learning, a number of researchers have suggested that it must be translated into action (Hattie & Timperly 2007; Topping, 1998; Winstone et al., 2017). This translation into action is a complex, multi-dimensional, and challenging process (Gielen & De Wever, 2015a; Kluger & DeNisi, 1996; Narciss, 2008; Winstone

et al., 2017), with one core aspect having students try to implement the feedback they receive. Although making revisions in response to peer feedback is not necessarily equal to learning, the revision process through which students respond to peer feedback has been shown to contribute to learning to write (Wichmann et al., 2018; Wu & Schunn, 2020b).

Foundational studies of writing processes (Flower & Hayes, 1981; Flower et al., 1986; Hayes et al., 1987) provided an important insight into feedback-based revision process, although the studies were originally developed to explain authors' revisions based on their internal evaluation (i.e., self-revisions). Cognitively speaking, writing involves planning, translating, and reviewing (Flower & Hayes, 1981), with reviewing involving two sub-processes: evaluating and revising the text (Flower et al., 1986; Hayes et al., 1987). In revising, writers must evaluate what they write, diagnose potential problems, decide whether to change, and select a strategy to revise (Chenoweth & Hayes, 2001; Cotos, 2014).

These sub-processes involved in self-revisions are also likely observed when writers are making changes in response to peer feedback. In a different context, physicians responding to peer feedback, Sargeant et al. (2009) investigated the decision-making processes in responding to feedback using interviews. They found multiple stages involved in the evaluation of received feedback, a stage involving the decision whether to accept and act, and the actual action to improve and change. In the writing context, Wichmann et al. (2018) simplified this responding-to-peer-feedback process into two stages: *plans to fix* (i.e., a receiver reads a comment and then decides whether it will be responded) and *actual implementation* (i.e., a receiver works on the text and finally addresses/ignores the problems detected by reviewers). Wichmann et al. (2018) studied these two stages by asking receivers to answer questions about each received comment: were they planning to use the comment, and have they implemented the comment or not? They found that feedback planned to be used also tended to be actually implemented.

### Common peer feedback features related to planning vs. implementation

One factor that has been regularly examined as a predictor of whether students implement the peer feedback they receive involves the features embedded in how the feedback is given (Huisman et al., 2018; Strijbos et al., 2010), for example whether the comment includes constructive recommendations in addition to pointing out the problem. Feedback features have been examined using many different categorization systems, such as binary classification of simple and elaborated components (e.g., Narciss, 2008; Strijbos et al., 2010), cognitive and affective features (e.g., Lu & Law, 2012; Nelson & Schunn, 2009), informative and suggestive elaborations (e.g., Gielen & De Wever, 2015a), classifications of form vs. content (e.g., Elizondo-Garcia et al., 2019), function-focused classification (e.g., Huisman et al., 2018). These high-level binary classifications often name more specific feedback features in their lower-level details. For example, in Lu and Law (2012), cognitive features include identification, explanation, suggestions; affective features include criticism comments and positive comments. In Narciss (2008) and Strijbos et al. (2010), simple feedback components include outcome-related information about the correct answer, and performance, whereas elaboration feedback components provide extra information such as information about revisions, errors, and task requirements.

We begin with the general four-level framework (personal evaluation, product evaluation, learning process, and self-regulation) identified in Hattie and Timperley's (2007) meta-analysis of feedback effects. We then focus on specific features that tend to commonly occur in peer feedback that correspond with three of the levels (personal, product,

and learning process) that are most relevant to peer feedback: mitigating praise, identification, suggestion, solution, and explanation. We examine these features in terms of the information included in peer feedback to make predictions for when they will separately influence planning and actual implementation.

Within peer feedback, the most common kind of feedback at the personal level involves praise. Although commonly recommended as an important part of feedback given to students, feedback including only personal evaluation, such as praise, is less effective than feedback about the task and how to perform the task (Cheng et al., 2015; Nelson & Schunn, 2009), in part because it does not include information related to improvement. At more global levels, praise can promote students' work on other (critical) comments because the praise inspires them (Tseng & Tsai, 2007), highlighting the motivational aspect of feedback. At the level of feedback features within comments that do contain issues to address, there is mitigating praise (i.e., praise is included in a negative comment). This feedback feature could soften the criticism and improve agreement with the criticism and thereby improve implementation rates (Wu & Schunn, 2020a). However, Wu and Schunn (2020a) observed a negative effect of overall praise and Patchan et al. (2016) observed a negative influence of mitigating praise on peer feedback implementation, perhaps because praise weakened the perceived severity of the problems.

At the product level, peer feedback often explicitly *identifies* the problem and/or includes constructive advice (e.g., general suggestions, specific solutions). Some studies did not observe any effects of explicit identification of the problem on peer feedback implementation (e.g., Leijen, 2017; Nelson & Schunn, 2009). However, other studies found that identification is important for peer reviewers (e.g., Cho & MacArthur, 2011; Lu & Law, 2012), perhaps because it enhances reviewers' cognitive processing. Peer feedback often contains constructive advice on how to address problems. Correspondingly, receiving comments including constructive advice has been found to positively predict comment understanding (Wu & Schunn, 2020a) and implementation (Leijen, 2017; Nelson & Schunn, 2009). However, Patchan et al. (2016), using more sophisticated statistical techniques, did not find effects of including constructive advice on implementation. This surprise lack of an effect may be related to the specificity of the constructive advice. However, most prior research on peer feedback did not differentiate general suggestions from concrete solutions (e.g., Cho & MacArthur, 2011; Leijen, 2017; Lu & Law, 2012; Nelson & Schunn, 2009; Patchan et al., 2016).

At the learning level, a common and potentially important peer feedback feature involves reviewers providing detailed or brief explanations for identified problems. Explanatory feedback has been found to be predictive of students' willingness to improve based upon it (Gielen & De Wever, 2015b; Gielen et al., 2010; Huisman et al., 2018) because it improves the clarity of feedback (Gielen et al., 2010) and thus ambiguity and obscurity are avoided. Gielen et al. (2010) found that explanations are even more important than feedback accuracy for learning. However, negative influences of explanatory feedback have also been observed in some studies (e.g., Nelson & Schunn, 2009; Tseng & Tsai 2006), perhaps because novice reviewers could not provide clear explanations.

Little research has investigated what feedback features predicted planning. Wichmann et al. (2018) did an experimental study to investigate whether planning to fix facilitated feedback implementation, but they did not examine what particular comments influenced planning to fix or actual implementation of feedback. Perceived ease of implementation might matter for both planning and implementation, but detailed support for revision (e.g., with an explanation) might matter more for implementation. Note that having similar factors predict planning and implementation or seeing correlations between planning and implementation does not

establish a causal role of earlier planning on implementation. It could be that at implementation students recapitulate a decision-making process that ignores earlier planning because it was forgotten or inconvenient to use. A mediational analysis of feedback features to implementation via earlier planning would provide stronger evidence of the causal role of initial planning activities.

Based upon the details found within prior taxonomies and Hattie and Timperley's (2007) framework, we analyze five feedback features: (1) mitigating praise, (2) identification, (3) explanation, (4) suggestions, and (5) solutions. The first three features analyzed in the present study are similar to the existing categories. In prior studies, researchers have investigated the effects of mitigating praise (called "mitigation" in Elizondo-Garcia et al., 2019), identification (called "evaluation" in Huisman et al., 2018), explanation (called "didactic feedback" in Tseng & Tsai, 2007). The last two features analyzed in the present study go beyond existing categories. Prior research generally coded them as one type called "revisions" in Huisman et al. (2018), or "solutions" in Elizondo-Garcia et al. (2019). However, suggestion and solution were coded and analyzed separately in the present study because specificity of (general) suggestions and (specific) solutions may have different effects on implementation.

In sum, a wide range of feedback features have been examined in terms of their effects on peer feedback implementation, and five of them regularly occur and have sometimes been found to predict implementation: (1) mitigating praise, (2) identification, (3) explanation, (4) suggestions, and (5) solutions. No prior studies have examined how these feedback features play out separately at the planning and feedback implementation stages, or tested the mediating role of an earlier planning stage on implementation processes. Such an analysis could help build a better understanding of why these features matter, as well as help predict when they should matter and thus when to intervene.

## The present study

The central focus of the current study was to investigate the relative strengths of different peer feedback features in predicting planning and actual implementation of peer feedback, and the mediated effect of planning on the relationships of feedback features and actual implementation. The conceptual model of the relationships of feedback features, planning, and actual implementation is presented in Fig. 1.

The exploratory research questions, along with specific hypotheses for each question, were:

(1) Which feedback features predict planning vs. actual implementation?

**Hypothesis 1 (H1)** Feedback features differentially predict planning vs. actual implementation.

(2) Does planning mediate the relationships of some feedback features to actual implementation?

**Hypothesis 2 (H2)** Planning mediates the influence of multiple feedback features on actual implementation.

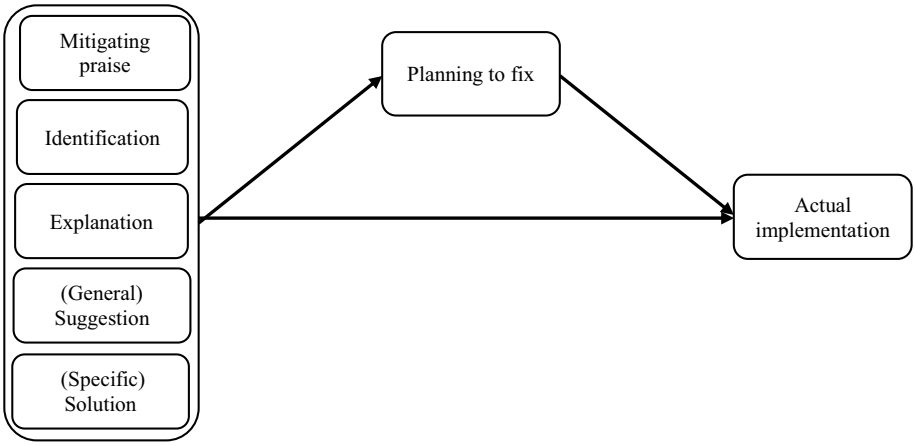


**Fig. 1** Feedback model depicting the potential relationships between feedback features, plans to revise and actual implementation

## Methods

### Overview

To test the hypotheses, the approach was to (1) systematically code features of received peer comments; (2) harvest plans to ignore vs. implement each comment in a revision; (3) code actual implementation of each comment within revisions; and (4) conduct multiple regression analyses of relationships of the features with the plans to fix and actual implementation. Different from prior research on feedback that analyzed factors influencing students' implementation by means of survey or interviews, the current study focused on comment-level features, testing statistical relationships between specific features of each comment, students' planning and implementation, while controlling for possible confounding features such as the grade given to the first draft, and whether the feedback was about higher vs. lower-level aspects of writing.

### Participants

Data were selected and systematically coded for 125 secondary school students (62% female; 4% did not report gender) drawn randomly from Advanced Placement (AP) Language and Composition classrooms within nine different schools distributed across the United States. Their age ranged from 16 to 19 years ($SD = 0.6$), with a mean of 17.3. They came from a variety of ethnicities, with the most frequent being Caucasian (65%), Asian (22%), Hispanic/Latino (9%), and African American (3%). Fifteen percent of the participants did not report their ethnicities.

AP (Advanced Placement) courses allow students to take university-equivalent courses in high school. AP Language and Composition, which is meant to be equivalent to a first-year writing seminar, is among the most highly-enrolled AP courses, with well over a half million students taking the end-of-course exam every year (College Board, 2018).

Teachers were recruited through a variety of mailing lists, seeking teachers who had previously taught this AP course and were willing to implement a common writing assignment with peer review. Schools serving low income students (i.e., Title I schools) were heavily recruited to improve generalizability of the findings: four of the nine teachers were from schools serving predominantly low-income families.

## Materials

### Peer reviews

Students conducted peer reviews using a widely-used online peer assessment tool, *Peerceptiv* (Cho & Schunn, 2007; Schunn et al., 2016). The program includes tools to improve feedback quality and peer review process. For example, students could also evaluate the quality of their peers' feedback. In the currently studied version, they could also plan their revisions in response to comments.

Students provided and received feedback using the program. Each student's first draft submission within a teacher's classroom was randomly assigned to four peer reviewers within the same teacher's writing classes. Upon receiving the drafts, the reviewers gave positive and negative comments to peers' work, and provided numeric feedback using provided rubrics. They were asked to give at least one comment on each writing dimension within the given rubrics. These rubrics were adapted to be more student friendly versions of the rubrics used by AP (Advanced Placement) expert raters in judging the essay component of the end-of-course exam (see Appendix A). The adapted rubrics include eight dimensions on a 7-point scale (1 = poor, 7 = excellent): thesis, argument, rhetorical strategies, evidence for claims, explaining evidence, organization, control of language, and conventions. Each rubric contains a specific commenting prompt. Teachers also provided a brief in-class training session for students on how to make effective comments according to the rubrics and how to use the program.

As a scaffold for planning, students were asked to create a revision plan using a tool embedded into the peer review system (see Fig. 2). In the current study, we use this process as the data source for student plans; pilot analyses indicated the tool had relatively little effect on student's actual implementation (66% vs. 62%, based on coding implementation for over 3800 comments across conditions). We do not take this as evidence that planning is irrelevant to implementation but rather that students did not find this tool useful to planning beyond other methods they would have used if they did not have the tool. The tool was pre-populated with all the comments an author had received. The author was invited to explicitly decide for each comment whether or not they would fix a problem or ignore the comment. If they chose to fix the problem, they could enter as free text what they were going to do to improve the draft in the plan column. If they chose to ignore a comment, they could select a reason from a drop-down menu or add a comment which further explained why it was ignored.

As a final step, students revised the paper and uploaded the revision to the system. The full peer review process is presented in Fig. 3 (see more information about the system in Appendix B). Teachers read and graded only the revised drafts. Students were given assignment credit for having completed the peer reviews and completing the revision plan. The peer review was conducted anonymously because anonymity enables assessors to provide honest and critical feedback with freedom (Panadero, 2016; Panadero & Alqassab, 2019). In addition, measures were taken to give assessors

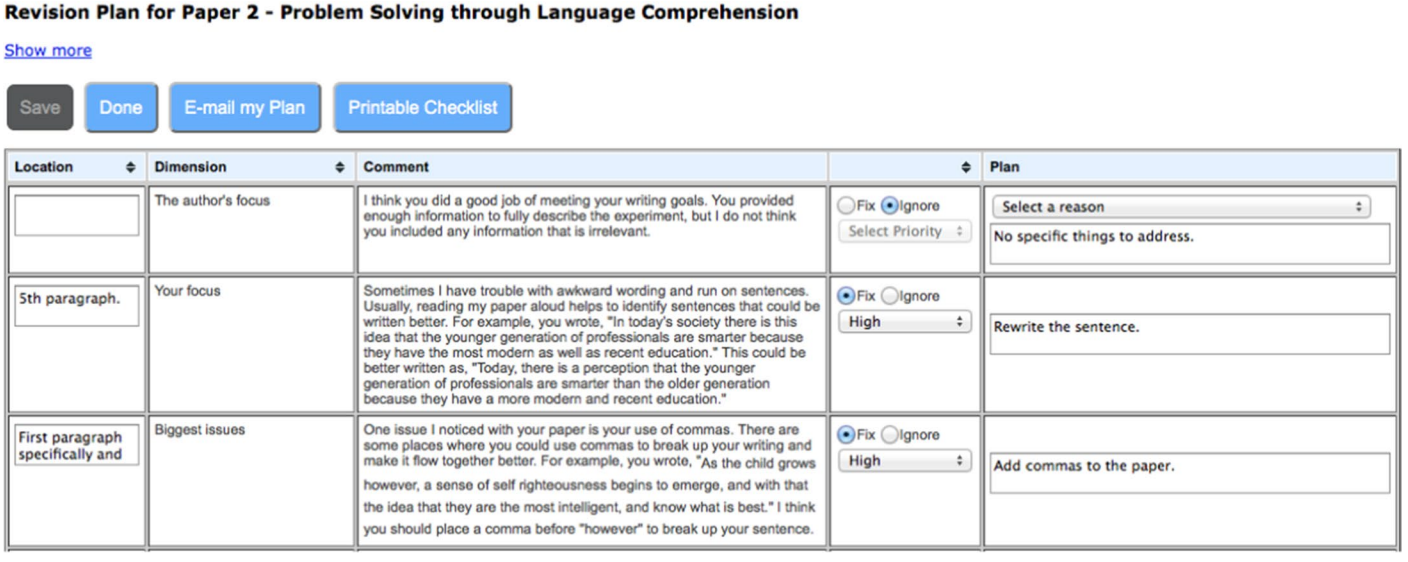


**Fig. 2** Revision planning interface

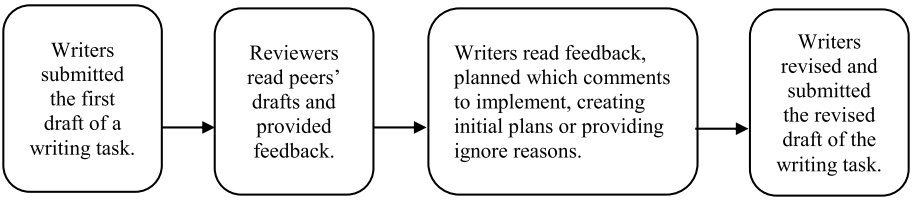


**Fig. 3** Peer review process

accountability pressures to provide quality feedback (e.g., students were invited to rate the helpfulness of the feedback they received) (Panadero & Alqassab, 2019; Patchan et al., 2017).

### Writing task

Data were collected from a typical writing task in this class, one that is closely aligned to the end-of-year standardized assessment for this AP course: read a one-page persuasive essay and then write a rhetorical analysis of the strategies used by the essay author. In particular, students were required to indicate which specific rhetorical strategies were used by the essay author to support their primary claims and to provide text evidence of those strategies from the source passage.

### Measures

Feedback features within each comment were coded by four research assistants; each comment was randomly distributed to two coders. Plans to fix were coded by two research assistants. Implementation of peer comments was coded independently by two writing experts who taught undergraduate writing for years. Disagreements between coders were resolved through discussion (Huisman et al., 2018). The whole coding process is presented in Fig. 4. The measures are summarized in Table 1.

### Fix and ignore plans

The dataset was taken from a larger study which divided students into two groups; only half of the students were given the Revision Planning tool, and only those students were included in the current study. These students received a total of 13,021 comments. However, not all students provided the tool actually completed the fix/ignore task. Some ignored the task entirely and some only partially completed the tasks. In the end, about 61% of the comments ($N = 7988$) had plans (i.e., explicitly selected fix or ignore decisions). However, some comments only contained praise, summaries, or a very vague comment, so that the authors had nothing to implement in response. The goal of the study was to examine implementation processes for those comments that did have something implementable, and therefore these unimplementable comments were removed, leaving 5753 implementable comments with fix/ignore plan information.

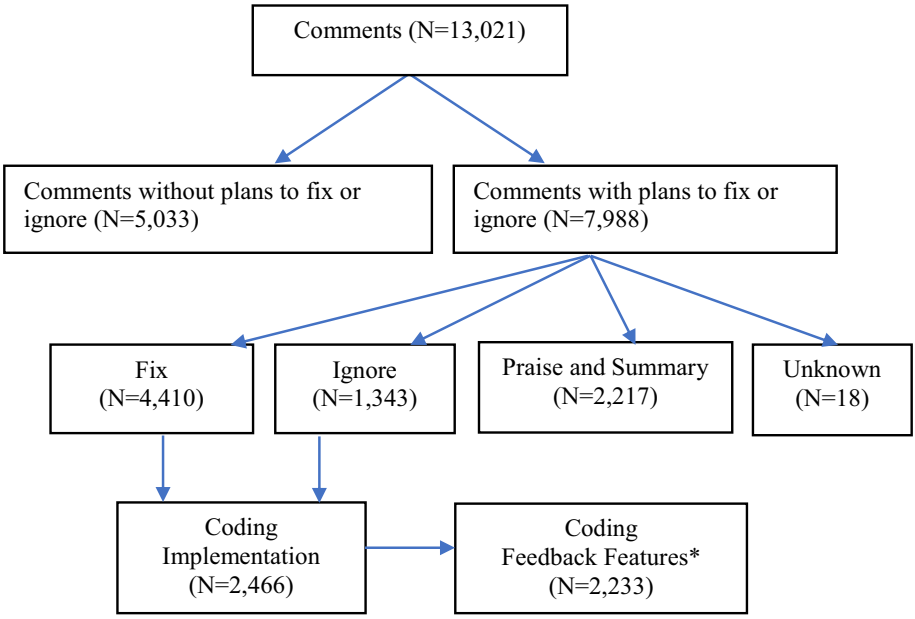


**Fig. 4** Peer feedback coding process. *Note. 2233 out of 2466 implementable comments were finally analyzed in regression analysis because vague comments could not be coded for implementation

Because of the unnecessarily large size of the dataset and the labor-intensive nature of coding comment features and comment implementation, the remaining coding and analysis was conducted on data from a random selection of 125 participants given the planning tool (2466 comments, 43% of the available comments) distributed across the classes. Note that the selected 125 participants versus the excluded 155 participants did not differ significantly in terms of rate of feedback focus (88% versus 84% of the comments were high-level respectively), planned fix rates (76% of the high-level comments would be fixed for both groups; 82% versus 84% of the low-level comments would be fixed), or first draft quality (5.36 versus 5.30); additional details on each of those measures are presented below.

Rather than relying purely on the radio button selection in the interface (fix vs. ignore), the contents of the reason/plan were further coded to insure of the radio button matched the student's intent. In particular, the plan variable was coded as Fix when a student said they were going to fix something in the paper or Ignore when they had no plan to make a change (*Kappa* = 0.92).

## Feedback features

Features of peer feedback were analyzed to identify the information involved in peer feedback. Each implementable comment was coded for five feedback features: mitigating praise (Kappa = 0.85), identification (*Kappa* = 0.81), explanation (*Kappa* = 0.80), solution (*Kappa* = 0.76), and suggestion (*Kappa* = 0.79) (see Table 2 for definitions and examples).

**Table 1** The measures examined in the study, including whether they were binary or continuous, whether they were features at the comment or writer level, and a description of how they were defined

| Variables | Type | Level | Description |
|---|---|---|---|
| Outcome variables | | | |
| Plans to fix | Binary | Comment | Whether a student planned to fix (1) or ignore (0) |
| Actual implementation | Binary | Comment | Whether a comment was implemented (1) or not (0) |
| Feedback features | | | |
| Mitigating praise | Binary | Comment | Whether a negative comment included mitigating praise |
| Identification | Binary | Comment | Whether a problem was identified explicitly |
| Explanation | Binary | Comment | Whether an implementable comment included an explanation |
| Suggestion | Binary | Comment | Whether an implementable comment included a general suggestion for revision |
| Solution | Binary | Comment | Whether an implementable comment included a specific solution to implement |
| Control variables | | | |
| Title I School | Binary | Writer | Whether a student came from Title I school (1) or not (0) |
| Age | Continuous | Writer | The age of a student |
| Gender | Binary | Writer | Whether a student is female |
| Feedback focus | Binary | Comment | Whether a comment is high-level or low-level |
| Comment length | Continuous | Comment | Number of words in a comment |
| Number of comments | Continuous | Comment | Total Number of comments |
| Number of implementable comments | Continuous | Writer | Total number of implementable comments |
| Total of praise comments | Continuous | Writer | Total number of positive comments |
| First draft quality | Continuous | Writer | Mean peer ratings across four peers and all dimensions |

### Additional factors

Several factors that might influence planning and actual implementaion were included in the models as control variables (see Table 1): school title, comment focus, comment lentgh, authors' and reviewers' 1st draft quality, number of comments, number of praise comments, age, and gender. Students from Title I schools and students with better 1st draft quality might be less likely to plan to fix or implement feedback in revisions because of motivational factors (Hughes, 2012; Lu & Law, 2012; Wu & Schunn, 2020a). In terms of reviewer effects, students might be more likely to respond to feedback from peer reviewers of high writing ability because the reviewers could provide better quality feedback. Two additional demographic variables, age and gender, were included because students of different ages and genders might engage in peer review differently.

Comments made about higher-level aspects of writing were expected to involve more effort to implement than comments about lower-level aspects of writing. Therefore, the feedback focus of every comment was coded (*Kappa* = 0.91): High-level if the feedback comment was concerned with thesis, argument, rhetorical strategies, evidence for claims, explaining evidence, and organization; and Low-level if the feedback focused on control of language and conventions.

Based on the prior research (e.g., Patchan et al., 2016), the total number of comments received (reflecting possible information overload effects), the number of praise comments (reflecting overall tone or need to revise), and comment length (reflecting possible depth of elaboration effects) are expected to be associated with planning and actual implementation.

### Feedback implementation in revisions

The 2466 implementable comments were further coded for whether they were implemented in the revised drafts. First, differences between first and revised drafts were identified using the Compare Documents function within Microsoft Word. Format changes were ignored since formatting was not formally part of this writing task. Second, the observed differences were inspected and compared to each peer comment. If a difference could be related to a feedback comment, that comment was coded as "Implemented"; a comment was coded as "Not Implemented" if the student did not appear to make any revision in response to it (*Kappa* = 0.76). Nine percent of the comments were coded as vague (and removed from further analysis) because the comment was so general/vaguely stated that it was not possible to identify whether the comment was implemented or not.

### Quality of writing

Revision plans are often influenced by a perceived need to improve, especially in terms of the score received on the first draft; students might be unlikely to make changes if they are already receiving high ratings. Since their first draft scores were based on the peer ratings, we used the mean peer scores received by each writer as the indicator of the perceived need to revise. However, it should be noted that the mean rating and expert ratings are correlated at levels similar to expert-expert reliability levels, based on expert scoring of a randomly selected subset of more than 300 essays from the larger dataset.

**Table 2** Peer feedback coding scheme, including definitions, examples, and inter-rater reliability

| Category | Definition | Example |
|---|---|---|
| Feedback focus (*Kappa* = 0.91) | | |
| High-level | Comments with regards to thesis, arguments, rhetorical strategies, organization, evidence, and explanation | Analyzed very well, but should use more devices. There were plenty of devices embedded in this work. The devices were well explained and had examples from the work. Just need a little more. |
| Low-level | Comments with regards to language of control and style | Bill board is actually billboard. Make sure you check your spelling when writing. |
| Implementation of feedback (*Kappa* = 0.76) | | |
| Implemented | Comments that were incorporated in the revision | Add in the rhetorical devices that will be discussed in the essay.<br>The writer added the rhetorical devices. |
| Not implemented | Comments that were not incorporated in the revision | No thesis present.<br>The writer did not add thesis in the later draft. |
| Vague | Comments that were too vague to determine whether they were implemented or not | I would try to list other sophisticated vocabulary to improve the credibility.<br>It's hard to determine whether the writer implemented it. |
| Feedback features | | |
| Mitigating praise (*Kappa* = 0.85) | Positive comments | There isn't a huge amount of textual evidence in what you have written but what is there is accurate and appropriate. |
| Identification (*Kappa* = 0.81) | To announce what is problematic, or what needs to be further developed | The author does not explain Louv's argument about the separation between people and nature. |
| Explanation (*Kappa* = 0.80) | To explain why it is problematic | The author supported the analysis of Louv's rhetorical strategies inadequately. They didn't really explain how Louv used the strategies. The author merely stated which strategy was used and left it at that. |
| Suggestion (*Kappa* = 0.79) | To provide general advice by giving the directions for changes | You could use different wording. Maybe using different words to shorten your statements. |
| Solution (*Kappa* = 0.76) | To provide specific advice by outlining alterations and corrections | In the first sentence of paragraph 2, the sentence does not make any sense. Change it to: In the first two paragraphs, Louv uses examples from real life to illustrate his point. |

### Data analysis

To test the first hypothesis, two sets of logistic regression analyses were conducted, using feedback features as predictors, but with different outcomes: (1) planning to fix; and (2) actual implementation. In addition to the feedback features of interest, the additional control variables were also included in the models (first alone and then with feedback features) to rule out these possible confounds as alternative explanations to the observed correlational relationships. To test the second hypothesis, mediation analyses were conducted to investigate mediated effects of planning to fix on the relationships of feedback features and actual implementation. Mediation analyses were done using the Binary Mediation program in Stata 15 (e.g., Forrester et al., 2021). This method is the adaptation of the Baron and Kenny approach (1986) for binary mediators. Five thousand Bootstrap samples were used to obtain 95% confidence intervals. The indirect effect is significant if the 95% confidence interval does not contain zero.

Before running the regressions, correlations among predictors were examined (see Table 3). A number of feedback features were correlated with the outcome variables and each other, therefore motivating the use of multiple regression to tease apart their separate effects on the outcomes. The number of implementable comments was highly correlated with the number of all comments ($r = 0.73$) and (negatively) with the number of praise comments ($r = -0.71$). Correspondingly, the Variance Inflation Factors (VIF) in the regressions were too high when the number of implementable comments was included. After removing this variable, no predictor variable had a VIF greater than 2.6, indicating no multicollinearity problems among the remaining factors. The correlations between the control variables and the two outcome variables were generally small ($r < |0.2|$). Interestingly, the two outcome variables, planning to fix and actual implementation, were only moderately correlated ($r = 0.28$), supporting the need to analyze them as separate outcomes. At the same time, plans were relevant to implementation: when students planned to address a problem, they actually made a change for 85% of comments, in comparison to making a change for only 43% of comments they initially planned to ignore.

Although a significance threshold of $p < 0.05$ has long been used in the social sciences and education research, there has long been disagreement about what the most appropriate significance level is (Benjamin et al., 2017). In the present study, a significance threshold of $p < 0.01$ was used for the regressions to obtain a better balance of Type I and Type II errors given the number of predictor variables being tested, the moderately large sample size (Murphy et al., 2014), and the exploratory nature of the current study.

## Results

### Feedback features predicting planning to fix versus predicting actual implementation (RQ1)

In line with the first hypothesis (H1), multiple feedback features predicted planning to fix and actual implementation differentially: solutions and suggestions significantly predicted planning to fix, while explanations significantly predicted actual implementation. The left half of Table 4 presented the results of the regressions predicting planning to fix, along with means and standard deviations for each predictor. A number of control variables were

**Table 3** Pearson correlations among control variables, feedback features, plans to fix, and actual implementation

| | Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Title I school | 1 | | | | | | | | | | | | | | | |
| 2 | Age | − .08** | | | | | | | | | | | | | | | |
| 3 | Gender (F = 1) | .09** | − .08** | | | | | | | | | | | | | | |
| 4 | Comment focus | − .02 | − .04 | − .04 | | | | | | | | | | | | | |
| 5 | Comment length | − .31** | .01 | − .02 | .02 | | | | | | | | | | | | |
| 6 | Authors' 1st draft quality | − .01 | − .06** | .02 | − .01 | .11** | | | | | | | | | | | |
| 7 | Reviewers' 1st draft quality | .14** | − .04 | .09** | − .02 | .21** | .14** | | | | | | | | | | |
| 8 | Number of comments | − .32** | .24** | − .18** | − .01 | − .03 | − .08** | − .20** | | | | | | | | | |
| 9 | Number of implementable comments | − .52** | .24** | − .25** | − .00 | .06** | − .46** | − .24** | .73** | | | | | | | | |
| 10 | Number of praise comments | .43** | − .11** | .18** | − .01 | − .09** | .64** | .15** | − .06** | − .71** | | | | | | | |
| 11 | Mitigating praise | − .02 | − .02 | .02 | .01 | .19** | .23** | .15** | − .16** | − .18** | .12** | | | | | | |
| 12 | Identification | .05* | − .03 | − .00 | − .04 | .16** | − .15** | .07** | − .05* | .00 | − .07** | − .12** | | | | | |
| 13 | Explanation | − .06** | − .05* | .04 | − .01 | .34** | − .08** | .08** | − .04* | .02 | − .08** | − .10** | .48** | | | | |
| 14 | Solution | − .17** | .02 | − .04 | − .14** | .26** | .06** | .05* | .08** | .12** | − .08** | − .03 | − .04 | − .02 | | | |
| 15 | Suggestion | − .23** | .05* | .02 | .06** | .15** | .05* | .02 | .04 | .12** | − .11** | .07** | − .30** | − .18** | − .15** | | |

**Table 3** (continued)

| | Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 16 | Planning to fix | − .04* | .02 | .05* | − .05* | .12** | − .06** | .10** | .00 | .04 | − .05* | .02 | -.01 | .05* | .11** | .08** | |
| 17 | Actual implementation | − .19** | − .03 | .04* | .00 | .15** | − .08** | .05* | − .02 | .11** | − .18** | .02 | .08** | .17** | .08** | .02 | .28** |

*$p < .05$, **$p < .01$, ***$p < .001$

significant predictors of outcomes, and collectively accounted for a significant part of the variance in outcomes, arguing for their importance in the analyses. Solutions and suggestions were the two peer feedback features that were significant predictors. Students were approximately twice as likely to fix the problems when a solution was included, and 71% more likely to fix when the comment included a suggestion. Including mitigating praise, identifying a problem explicitly, and providing an explanation were not significantly associated with planning to fix problems, although explanations were a small, marginally significant predictor.

The right half of Table 4 presented the results of the regressions predicting actual implementation. Providing explanations was the only feedback feature significantly predicted implementation directly, with a much larger effect size than at the planning stage. Students were 81% more likely to implement a comment if the problem was explained. The relationships of specific solutions and especially general suggestions with plans disappeared at the implementation stage. Mitigating praise and explicitly identifying problems continued to not be significant predictors of implementation, as was the case for predicting plans. Note that identification in isolation was correlated with actual implementation, but it did not add significantly as a predictor when the other predictive variables were included in the regression.

To test if explanation was a robust predictor of actual implementation, a follow-up analysis was conducted, contrasting comments only having a suggestion or solution ($N = 1{,}119$) and comments with suggestion/solution combined with an explanation ($N = 589$). Compared with the former, comments with both explanation and suggestive solutions were more likely to be implemented ($B = 0.55$, $SE = 0.15$, $odds\ ratio = 1.73$, $p < 0.001$), but the two groups of comments were not different significantly when predicting plans to fix ($B = -0.04$, $SE = 0.17$, $odds\ ratio = 0.96$, $p = 0.81$). The results supported the importance of explanations in motivating peer feedback, even in the case where authors have been given constructive advice.

### Mediated effects of planning to fix to actual implementation (RQ2)

H2 was partially supported (see Tables 5 and 6). The largest predictor of implementation was for planning to fix: students were almost 4 times as likely to implement a comment when they indicated they would do so at the planning stage. By contrast, another feedback feature, explanation, predicted actual implementation directly. Students were approximately 1.8 times as likely to implement a feedback comment with explanatory information. Mediation analyses showed that suggestion and solution both had statistically significant mediated effects to implementation via planning. In sum, three features were associated with implementation, two only indirectly via planning and one only directly (see Table 6 and Fig. 5).

## Discussion

The observed results supported differentiating planning from actual implementation. As support for treating them as fundamentally different phases, there were quite different features that predicted planning versus actual implementation. Even at the level of control variables, there was no overlap in terms of which variables were significant predictors of

**Table 4** Hierarchical regression analyses predicting plan to fix and actual implementation using control variables alone and all variables together

| Response variables | *M* | *SD* | Planning to fix | | | | Actual implementation | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Control variables | | All predictors | | Control variables | | All predictors | |
| Predictors | | | *B (SE)* | *Odds ratio* | *B (SE)* | *Odds ratio* | *B (SE)* | *Odds ratio* | *B (SE)* | *Odds ratio* |
| Feedback features | | | | | | | | | | |
| Mitigating praise | 0.48 | 0.5 | – | – | 0.06 (0.12) | 1.06 | – | – | 0.16 (0.11) | 1.17 |
| Identification | 0.73 | 0.45 | – | – | − 0.12 (0.14) | 0.88 | – | – | 0.09 (0.13) | 1.09 |
| Explanation | 0.4 | 0.49 | – | – | 0.27 (0.14) | 1.3† | – | – | 0.59 (0.13) | 1.81*** |
| Solution | 0.2 | 0.40 | – | – | 0.72 (0.17) | 2.05*** | – | – | 0.27 (0.14) | 1.31† |
| Suggestion | 0.67 | 0.47 | – | – | 0.54 (0.13) | 1.71*** | – | – | 0.04 (0.12) | 1.04 |
| Control variables | | | | | | | | | | |
| Title I school | 0.3 | 0.46 | − 0.18 (0.16) | 0.84 | − 0.06 (0.15) | 0.94 | − 0.71 (0.14) | 0.49*** | − 0.73 (0.14) | 0.48*** |
| Age | 17.3 | 0.61 | 0.07 (0.09) | 1.08 | 0.06 (0.09) | 1.06 | − 0.16 (0.08) | 0.85* | − 0.14 (0.08) | 0.87 |
| Gender (F = 1) | 0.57 | 0.5 | 0.21 (0.11) | 1.23 | 0.18 (0.11) | 1.20 | 0.31(0.10) | 1.36** | 0.29 (0.10) | 1.34** |
| Comment focus | 0.88 | 0.32 | − 0.34 (0.18) | 0.71 | − 0.27 (0.18) | 0.76 | − 0.08 (0.15) | 0.93 | − 0.01 (0.15) | 0.99 |
| Comment length | 54.42 | 36.88 | 0.01 (0.00) | 1.01*** | 0.00 (0.00) | 1.00 | 0.01 (0.00) | 1.01*** | 0.00 (0.00) | 1.00 |
| Authors' 1st draft quality | 5.26 | 0.74 | − 0.35 (0.11) | 0.71** | − 0.41 (0.12) | 0.66*** | − 0.22 (0.10) | 0.8* | − 0.19 (0.11) | 0.83 |
| Reviewers' 1st draft quality | 5.36 | 0.81 | 0.22 (0.07) | 1.25** | 0.21 (0.07) | 1.23** | 0.13 (0.06) | 1.14* | 0.13 (0.06) | 1.13† |
| Number of comments | 31.23 | 5.66 | 0.00 (0.01) | 1.00 | 0.00 (0.01) | 1.00 | − 0.02 (0.01) | 0.98 | − 0.02 (0.01) | 0.98 |
| Number of praise comments | 6.98 | 5.13 | 0.01 (0.02) | 1.01 | 0.02 (0.02) | 1.02 | − 0.04 (0.02) | 0.96** | − 0.04 (0.02) | 0.96** |
| Pseudo $R^2$ (Nagelkerke) | – | – | 0.04 | | 0.07 | | 0.10 | | 0.12 | |

†$p = .05$, *$p < .05$, **$p < .01$, ***$p < .001$

**Table 5** Regression analysis predicting the mediator and actual implementation together with the mediator, including only control variables and including all predictors together

| Response variables | Actual implementation (control variables + mediator) | | Mediator: planning to fix (control variables + feedback features) | | Actual implementation (all predictors) | |
|---|---|---|---|---|---|---|
| Predictors | *B (SE)* | *Odds ratio* | *B (SE)* | *Odds ratio* | *B (SE)* | *Odds ratio* |
| Feedback features | | | | | | |
| Mitigating praise | – | – | 0.06 (0.12) | 1.06 | 0.16 (0.11) | 1.17 |
| Identification | – | – | − 0.12 (0.14) | 0.88 | 0.13 (0.13) | 1.14 |
| Explanation | – | – | 0.27 (0.14) | 1.3† | 0.56 (0.13) | 1.76*** |
| Solution | – | – | 0.72 (0.17) | 2.05*** | 0.1 (0.15) | 1.11 |
| Suggestion | – | – | 0.54 (0.13) | 1.71*** | − 0.11 (0.13) | 0.89 |
| Control variables | | | | | | |
| Title I School | − 0.72 (0.14) | 0.49*** | − 0.06 (0.16) | 0.94 | − 0.77 (0.15) | 0.46*** |
| Age | − 0.2 (0.08) | 0.82* | 0.06 (0.09) | 1.06 | − 0.17 (0.09) | 0.85 |
| Gender (F = 1) | 0.28 (0.11) | 1.33** | 0.18 (0.11) | 1.2 | 0.28 (0.11) | 1.32* |
| Comment focus | − 0.00 (0.16) | 1.00 | − 0.27 (0.18) | 0.76 | 0.06 (0.15) | 1.06 |
| Comment length | 0.01 (0.00) | 1.01** | 0.00 (0.00) | 1.00 | 0.00 (0.00) | 1.00 |
| Authors' 1st draft quality | − 0.15 (0.11) | 0.86 | − 0.41 (0.12) | 0.66*** | − 0.09 (0.11) | 0.91 |
| Reviewers' 1st draft quality | 0.08 (0.07) | 1.08 | 0.21 (0.07) | 1.23** | 0.08 (0.06) | 1.08 |
| Number of comments | − 0.02 (0.01) | 0.98* | 0.00 (0.01) | 1.00 | − 0.02 (0.01) | 0.98 |
| Number of praise comments | − 0.05 (0.02) | 0.95** | 0.02 (1.70) | 1.76 | − 0.05 (0.02) | 0.95** |
| Mediator | | | | | | |
| Planning to fix | 1.33 (0.11) | 3.77*** | – | – | 1.36 (0.12) | 3.9*** |
| Pseudo $R^2$ (Nagelkerke) | 0.1 | | 0.07 | | 0.21 | |

†$p = .05$, *$p < .05$, **$p < .01$, ***$p < .001$, – = not included in the model

**Table 6** Total, direct, and indirect effects of the relationship of feedback features to actual implementation via planning

| | Total effect (SE) | Direct effect (SE) | Indirect effect (SE) | Indirect effect (95% CI) |
|---|---|---|---|---|
| Exaplanation | **0.17 (0.03)** | **0.15 (0.03)** | 0.02 (0.01) | [− 0.003, 0.04] |
| Solution | 0.06 (0.03) | 0.01 (0.03) | **0.05 (0.01)** | [0.03, 0.07] |
| Suggestion | 0.003 (0.03) | − 0.04 (0.03) | **0.04 (0.01)** | [0.02, 0.06] |

(1) Control variables: Title I school, Age, Gender, Comment length, Authors' first draft quality, Reviewers' first draft qulaity, Number of praise comments.

(2) Mitigating praise, Identification, Comment focus, and Number of comments predict neither planning to fix nor actual implementation, so they are not included in the mediation analyses.

(3) Bold denotes statistically significant effects

planning versus which features were significant predictors of actual implementation. When planning to fix was included as a mediator, it was a strong predictor of actual implementation, supporting the framing of planning to fix being on the path to revisions, but not identical to revisions. The theoretical and practical implications of these findings are discussed below.

## Feedback features predicting planning to fix versus predicting actual implementation (RQ1)

Two feedback features were associated with the likelihood of planning to make a change but not directly with actual feedback implementation: suggestions and solutions (see Table 4). Suggestions and solutions predicted planning because the two features provided information of how to solve the identified problem, and therefore may have seemed to make revisions easier (Nelson & Schunn, 2009).

Why were explanations not predictive of planning to fix? One possible reason could be related with perceived difficulty of translating explanatory feedback in revisions at the planning stage. If making a change is difficult or time consuming, it may be particularly important that students perceive a revision to be important and persist with the revision. Explanations may be key to motivating the revision at this point. By contrast, cost may be considered important at the planning stage because students might prefer an immediate and easier revision to a more important and difficult revision of larger reward that happens later (Green & Myerson, 2004).

The findings regarding feedback features that predicted implementation were partially consistent with prior research (Lu & Law, 2012; Nelson & Schunn, 2009; Patchan et al., 2016; Tseng & Tsai, 2007): only explanations appeared to be helping in persuading assessees to repair the problems (see Table 4). Explanatory comments provide detailed information on why it is a problem and the consequence of being a problem from a readers' perspective, which others have also argued to promote assessees' understanding of feedback and increase their willingness to use it (Gielen et al., 2010; Gielen & De Weaver 2015a, b; Huisman et al., 2018). Thus, the assessees are more likely to

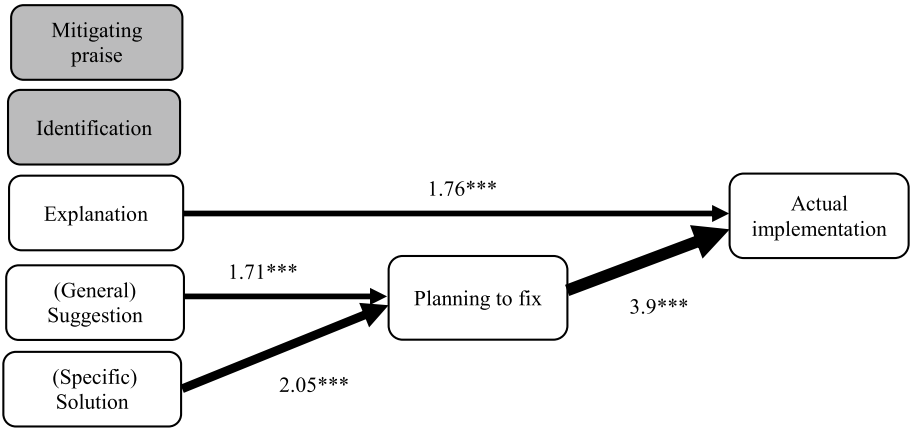


**Fig. 5** Revised feedback model for Planning to fix and Actual implementation. Line thickness corresponds with statistical strength of relationship in regression models (see Table 5)

be persuaded to make revisions when an explanation is included. The current findings extend past findings by showing that the positive role of explanations is specifically located at the implementation phase. Cheng and Tsai (2012) found that many students perceived high diversity in the interpretations of assessment criteria in peer feedback, which caused negative emotional responses. Reviewing essays involves subjective judgments even when detailed rubrics are included. Explanatory information can not only help clarify assessors' understanding of both the criteria and the identified problems, but also help assessees check shared knowledge of the assessment criteria and have a better understanding of the meaning of the comments.

When making a choice at the planning stage, students appear to implement comments with suggestions and solutions because the suggested solutions are expected to improve draft quality with minimal effort. However, at the implementation phase, the influence of feedback with explanations becomes more salient to implement revisions. Interestingly, there was an interaction: when suggestions or solutions are provided together with an explanation, students are more likely to use this more straightforward-to-implement kind of feedback. Although students appear to prefer seemingly easy-to-implement feedback at the planning phase, their preferences appear to alter with the passage of time so that students ultimately choose to address the feedback they both understand and are persuaded is important.

Interestingly, two feedback features predicted neither planning nor actual implementation: mitigating praise and identification. The finding about mitigating praise is contrary to common practical advice for giving feedback, but it does replicate findings supported by Nelson and Schunn (2009), who did not observe correlations between praise and implementation. On the one hand, it may be that this feature does not include information about the task and/or how to progress (Hattie & Timperley, 2007). On the other hand, it may be that mitigation sometimes reduces the perceived need to address the problem (Patchan et al., 2016). Indeed, Patchan et al. (2016) found a small negative effect of mitigating praise on implementation. At more global levels, the number of praise comments on the author level was negatively correlated with actual implementation, but had no effect on planning (see Table 4). This finding is consistent with Wu and Schunn (2020a), who found negative correlation between the number of praise comments and implementation. At the stage of planning, students may pay more attention to the comments that carry more information about their weaknesses. According to the meta-analysis on feedback (Hattie & Timperley, 2007), the effects of praise comments are limited because of little value to lead to learning goals. When it comes to actual implementation, the motivational effects begin to emerge. When students receive more praise comments, they are less likely to agree with the problems identified by reviewers, and less likely to implement received feedback (Wu & Schunn, 2020a).

Identification also did not predict planning or actual repairing of problems. No prior work had examined this variable at the planning stage. The lack of a relationship at the implementation stage replicated Nelson and Schunn (2009) but contradicted Lu and Law (2012). Although pointing out a problem explicitly allows students know what and where the problem is, it does not contain information about the nature of the problem (i.e., why it is problematic) and how to fix the problem, thus its persuasive effects may be limited.

### Mediated effects of planning to fix on feedback features and actual implementation (RQ2)

Suggestions and solutions indirectly predicted actual implementation through planning to fix. Why were suggestions and solutions not directly predictive of implementing feedback (similar to what Patchan et al., 2016 also found) but only mediated through planning? There are at least two major possible explanations. First, suggestions may have been too general such that students need to spend too much time figuring out how to actually implement the suggestion. Second, students might need more information to help them understand the problem underlying the suggestions and solutions. Feedback is much less likely to be implemented when students do not understand the comment (Lu & Law, 2012; Nelson & Schunn, 2009). In these two cases, the planning stage enables students to reflect on the received peer feedback and make a choice on whether to fix the problems. The finding points to an important function of planning: students are more likely to implement a feedback comment including suggestion or solution in revisions when they plan to do so at the planning stage. This finding is consistent with prior research by Wichmann et al. (2018), who found that students are more likely to make revisions when they are asked to use a sense-making table, which is another form of a planning tool.

The finding that explanation predicted implementation directly rather than through the mediator indicated the importance of including explanatory information in peer feedback. The finding is consistent with prior research. For example, Wu and Schunn (2020a) found that explanation directly predicts implementation rather than through the two perception mediators, understanding and agreeing with feedback. Gielen et al. (2010) found that explanations play a more significant role in helping students learn than feedback accuracy. Although explanations appear to be an important predictor in that they predict implementation directly, we also want to draw attention to an important motivational trade-off in the reviewing exchange: explanations may help assessees but they involve significant extra work for the assessors. It may also be the case that explanations help the assessors by deepening their learning from the reviewing process, but this benefit may not be salient to them. Additional research will be needed to see what supports or incentives need to be put into place to support assessors in more consistently including explanations.

### Implications for practice

The research findings are important for writing instruction practice and research. The study reveals that different feedback features are valued by students at different stages of revisions. Explanations should be a focal feature in peer feedback. Planning tools or other kinds of planning forms should be included in peer review. The current research makes a contribution to the existing literature of peer feedback in two major ways. First, few prior studies have investigated feedback features that influence students' plans to revise and their actual implementation. By comparing students' response to peer feedback at these two stages, the most important feedback features that might influence students to make revisions have surfaced. Second, different ways that feedback features influenced implementation have been discovered by including planning as a mediator. The findings suggest including planning stage in peer review would be worthwhile, which forms the basis for suggestions and solutions to improve implementation rates.

The findings of the present study provide helpful tips for encouraging students to implement received feedback. First, assessors should be more strongly encouraged to provide

feedback features that include explanatory information and constructive suggestions to solve the identified problems in order to positively influence feedback implementation. Explanations alone or together with other important features such as suggestions and solutions are strongly associated with implementation. In addition to reduce the difficulty of translating received feedback into revisions, these two features might also force the assessors to think more deeply about their comments, which could both improve the accuracy of the peer feedback (Wu & Schunn, 2020a) and improve the learning from providing feedback. Therefore, assessors should still be encouraged to provide clear and attainable suggestions/solutions with supporting information (i.e., explanations) and avoid overly-vague suggestions. Finally, future research on feedback should continue to separate (general) suggestions and (specific) solutions so that their differential effects can be identified.

Second, since explanations, suggestions, and solutions have been found to be important components of peer feedback, teachers' guidance should be directed toward emphasizing the significance of these features and instructing students to provide more explanatory and suggestive comments. For example, in peer review training, special attention should be given to explanatory feedback by including intervention questions such as what and why would you change (Gielen & De Weaver 2015a, b), or by prompting authentic feedback requests from writers (Voet et al., 2018). In online peer review system, Natural Language Processing and machine learning could be used to add automatic evaluations to peer comments as the reviewers do their work (e.g., detecting features of explanatory feedback), similar to prior work of this type focused on detecting solutions/suggestions in comments (Nguyen et al., 2016).

Third, planning stage could be included in peer review so that students have more opportunities to read and reflect on the received feedback. By including planning stage, assessees could evaluate feedback critically, rather than accepting it without thinking. Feedback, especially when it is inconsistent with expectations, can stimulate reflective thinking (Sargeant et al., 2009) and mindful reception (Bangert-Drowns et al., 1991).

Fourth, the relative need to critically evaluate the feedback depends somewhat upon its quality. For more expert feedback, students should critically evaluate the feedback to make sure they fully understand it and to deepen their learning from it. However, even teachers occasionally give bad advice, and the rate of errors will be higher in peer feedback since student reviewers are not experts in the field. Thus, critical evaluation becomes more important in the case of peer feedback. In the current context, where the students were in an advanced course, received carefully designed and detailed rubrics, and had some training on the rubrics, the rate of incorrect peer feedback was quite low: only 3% of the high-level comments and 1% of low-level were incorrect. In contexts where there are more errors, we suspect that feedback features will play an important role in helping authors distinguish correct from incorrect peer feedback. For example, assessees could better evaluate the quality of received feedback by receiving more explanatory information and constructive suggestions. Future research should be conducted to investigate the interaction of feedback features and feedback accuracy on implementation rates in contexts with higher rates of incorrect peer feedback.

## Limitations and considerations for future research

This study suffers from a few limitations. First, this study is inherently correlational, and it is possible that some of the feedback features were merely correlated with rather than influencing plans and revisions. Many control variables were included, but yet-unmeasured variables might serve as confounding variables. Correlational studies are useful first steps in identifying which variables are worth further testing in an experimental design. On a related point, the variables were limited to a range of feedback features that prior research had suggested as being critical to explain students' revisions. Other variables such as adequacy of feedback or complexity of suggestive revisions likely also account for implementation and should be included in future research.

Second, it is worth noting that many revisions were not conducted as planned in spite of the strong correlation between plans and actual implementation. The current study provides initial insights in understanding students' revising behavior at the planning and implementation phases, but it did not directly examine the slippage between the two phases. Future qualitative studies could be helpful to directly examine why students did not revise as planned, both for doing less than was planned and doing more than was planned.

Third, the present study focused on the planning stage and the implementing stage, so the findings can only be used to predict students' planning and actual implementation of peer feedback in revisions, but cannot be directly connected to learning outcomes. That is, the significant feedback features do not necessarily apply to student writing performance in a new writing task because making revisions does not equal to learning. Future research should examine the effects of feedback features on learning, in other words, whether making revisions on the basis of explanations, suggestions, and solutions also improves students' competence on specific dimensions of writing.

## Conclusions

The current study investigated the influence of feedback features on students' planning to fix problems and their actual implementation of feedback in revisions, as well as the mediated effects of planning on the relationships of feedback features and actual implementation. Very few studies have examined the effect of feedback features in different revising stages. The current study reveals that decisions at different stages appear to be made on substantially different bases. We found that suggestions and solutions mattered for actual implementation indirectly through planning and explanations mattered for actual implementation directly. In terms of the significant mediated effects of planning to fix, additional supports for the planning stage could be included in peer review classroom practices. Since actual implementation is likely to be especially important for learning, peer review training provided by instructors should emphasize the importance of explanations, suggestions, and solutions.

## Appendix A

Peer review rubrics

*Thesis* Did the author include a clear, specific thesis in his or her introduction?

7—The author's introduction includes a clear, specific thesis statement that connects Louv's rhetorical strategies with the argument he is making about the separation between people and nature.

6—6

5—The author's introduction includes a thesis, but the thesis does not make a specific or clear connection between Louv's rhetorical strategies and his argument about the separation between people and nature.

4—4

3—The author's introduction includes a thesis, but the thesis is overly general or simply a restatement of the essay prompt.

2—2

1—The author did not include a thesis in his or her introduction.

*Argument* Did the author accurately describe Louv's argument about the separation between people and nature?

7—The author accurately describes all of Louv's argument.

6—6

5—The author accurately describes most of Louv's argument.

4—4

3—In the majority of the essay, the author misunderstands Louv's argument.

2—2

1—The author does not address Louv's argument and instead writes about his or her own argument about the separation between people and nature.

*Rhetorical strategies* What rhetorical strategies did the author analyze in his or her essay?

7—The author analyses multiple, subtle rhetorical strategies that Louv uses accurately (such as appeal to a common cause, evoking nostalgia, or other sophisticated strategies).

6—6

5—The author analyses three or more obvious rhetorical strategies that Louv uses (such as using rhetorical questions, anecdotes, or other obvious strategies).

4—4

3—The author analyses only 1-2 obvious rhetorical strategies that Louv uses (such as rhetorical questions) or misunderstands Louv's strategies.

2—2

1—The author didn't write about Louv's rhetorical strategies (instead discussed a different topic, connected to personal experience, or just summarized Louv's piece).

*Evidence for claims* How strong is the textual evidence for each claim about Louv's rhetorical strategies?

7—Every claim has accurate evidence for all important aspects of the claim. Most evidence is conveyed through direct quotes.

6—6

5—5-Every claim has evidence, but some of the evidence is not accurate or not complete. Some evidence is conveyed through direct quotes.

4—4

3—Several claims are missing evidence, or most of the evidence is not accurate. Little or no evidence is conveyed through direct quotes.

2—2

1—No evidence is provided for any of the claims.

*Explaining evidence* Are the explanations of the textual evidence logical and thorough?

7—Explanations of all the evidence provided are thorough, logical and connected to the essay's thesis.

6—6

5—Explanations are sufficient, but not always thorough, logical, and clearly connected to the essay's thesis.

4—4

3—Explanations are simplistic, sometimes absent, or not clearly connected to the essay's thesis.

2—2

1—Explanations are missing or unrelated to the prompt (such as based in personal experience).

*Organization* Did the author organize his or her essay logically and clearly?

7—The essay has a clear organization with a logical progression of ideas and body paragraphs that are each focused on a single argument that connects back to the thesis.

6—6

5—The essay has a clear organization and progression of ideas, but the body paragraphs may sometimes be unfocused or not clearly connected to the thesis. The organization may be simplistic with formulaic transitions and a list-like progression of ideas.

4—4

3—The organization of the essay is difficult to follow in many places due to jumps in logic, lack of transitions, repetition, and lack of focused body paragraphs that connect to the thesis.

2—2

1—The essay is very disorganized with most ideas presented in random, repetitive, or illogical ways that make the author's argument and its connection to a thesis very difficult to understand.

*Control of language* How appropriate are the writing style and vocabulary for an academic essay?

7—Mature, sophisticated prose style, using specific academic terminology (such as pathos and ethos) and control of language.

6—6

5—Clear prose style with few lapses in academic word choice.

4—4

3—The prose generally conveys the writer's ideas but is inconsistent in controlling the elements of effective writing, such as academic word choice.

2—2

1—Simplistic style and vocabulary.

*Conventions* How well does the paper follow the conventions (grammar, punctuation, and spelling) of Standard Written English?

7—The paper follows the conventions of Standard Written English very well with very few or no errors.

6—6

5—The paper mostly follows the conventions of Standard Written English, but has about 1-2 error per paragraph. The errors don't interfere with your understanding the writer's ideas.

4—4

3—The paper does not consistently follow the conventions of Standard Written English and may include up to 3-5 errors per paragraph. In places, the errors make it hard to understand the writer's ideas.

2—2

1—In many sentences, the paper does not follow the conventions of Standard Written English. The errors make it very difficult to understand the write's ideas in many places.

## Appendix B

A typology of peer assessment in the present study

| | Variable | Range of variation |
|---|---|---|
| 1 | Curriculum area/subject | Advanced Placement Language and Composition |
| 2 | Objectives | Of staff and students<br>Time saving and cognitive/affective gains |
| 3 | Focus | Quantitative/ qualitative/formative |
| 4 | Product/output | Argumentative writing |
| 5 | Relation to staff assessment | Supplementary |
| 6 | Official weight | Contribute to assessee final grade |
| 7 | Directionality | Reciprocal |
| 8 | Privacy | Anonymous |
| 9 | Contact | Distance |
| 10 | Year | Same year |
| 11 | Ability | Similar ability |
| 12 | Constellation assessors | Groups |
| 13 | Constellation assessed | Groups |
| 14 | Place | Out of class |
| 15 | Time | Free time |
| 16 | Requirement | Compulsory for assessors/assessees |
| 17 | Reward | Course credit |

#### Declarations

**Conflict of interest** The second author is a co-inventor of the peer review system used in the study.

## References

Bangert-Drowns, R., Kulik, C. L., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213–238. https://doi.org/10.3102/00346543061002213

Benjamin, D. J., Berger, J., Johannesson, M., Nosek, B. A., Wagenmakers, E., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., & Johnson, V. (2017). Redefine statistical significance. *Nature human behaviour, 2*(1), 6–10. https://doi.org/10.31234/osf.io/mky9j

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182.

Cheng, K. H., Liang, J. C., & Tsai, C. C. (2015). Examining the role of feedback messages in undergraduate students' writing performance during an online peer assessment activity. *The Internet and Higher Education, 25*, 78–84. https://doi.org/10.1016/j.iheduc.2015.02.001

Cheng, K. H., & Tsai, C. C. (2012). Students' interpersonal perspectives on, conceptions of and approaches to learning in online peer assessment. *Australasian Journal of Educational Technology, 28*(4), 599–618.

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: generating text in L1 and L2. *Written Communication, 18*(1), 80–98. https://doi.org/10.1177/0741088301018001004

Cho, K., & MacArthur, C. (2011). Learning by reviewing. *Journal of Educational Psychology, 103*(1), 73–84. https://doi.org/10.1037/a0021950

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

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

Cotos, E. (2014). *Genre-based automated writing evaluation for L2 research writing: From design to evaluation and enhancement*. Palgrave Macmillan.

Elizondo-Garcia, J., Schunn, C., & Gallardo, K. (2019). Quality of peer feedback in relation to instructional design: A comparative study in energy and sustainability MOOCs. *International Journal of Instruction, 12*(1), 1025–1040.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387. https://doi.org/10.2307/356600

Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, diagnosis, and the strategies of revision. *College Composition and Communication, 37*(1), 16–55. https://doi.org/10.2307/357381

Forrester, S. N., Zmora, R., Schreiner, P. J., et al. (2021). Accelerated aging: A marker for social factors resulting in cardiovascular events? *SSM-Population Health*. https://doi.org/10.1016/j.ssmph.2021.100733

Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior, 52*, 315–325. https://doi.org/10.1016/j.chb.2015.06.019

Gielen, M., & De Wever, B. (2015). Structuring the peer assessment process: A multilevel approach for the impact on product improvement and peer feedback quality. *Journal of Computer Assisted Learning, 31*(5), 435–449. https://doi.org/10.1111/jcal.12096

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

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin, 130*(5), 769–792. https://doi.org/10.1037/0033-2909.130.5.769

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

Hayes, J. R., Flower, L. S., Shriver, K., Stratman, J., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Advances in psycholinguistics: Reading, writing and language* (pp. 176–240). Cambridge University Press.

Hughes, G. D. (2012). Teacher retention: Teacher characteristics, school characteristics, organizational characteristics, and teacher efficacy. *The Journal of Educational Research, 105*(4), 245–255. https://doi.org/10.1080/00220671.2011.584922

Huisman, B., Saab, N., Van Driel, J., & Van den Broek, P. (2018). Peer feedback on academic writing: Undergraduate students' peer feedback role, peer feedback perceptions and essay performance. *Assessment & Evaluation in Higher Education, 43*(6), 955–968. https://doi.org/10.1080/02602938.2018.1424318

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

Leijen, D. A. J. (2017). A novel approach to examine the impact of web-based peer review on the revisions of L2 writers. *Computers and Composition, 43*, 35–54. https://doi.org/10.1016/j.compcom.2016.11.005

Li, H., Xiong, Y., Zang, X., et al. (2020). Peer assessment in the digital age: A meta-analysis comparing peer and teacher ratings. *Assessment & Evaluation in Higher Education, 41*(2), 245–264. https://doi.org/10.1080/02602938.2014.999746

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

Murphy, K. R., Myors, B., & Wolach, A. (2014). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests* (4th ed.). . Routledge.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125–143). Erlbaum.

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

Nguyen, H., Xiong, W., & Litman, D. (2016). Instant feedback for increasing the presence of solutions in peer reviews. *Proceedings Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations (NAACL-HLT)*. San Diego, CA.

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of social and human conditions in assessment* (pp. 247–266). Routledge.

Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment & Evaluation in Higher Education, 44*(8), 1253–1278. https://doi.org/10.1080/02602938.2019.1600186

Panadero, E., Jonsson, A., & Alqassab, M. (2018). Peer feedback used for formative purposes: Review of findings. In A. Lipnevich & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback*.UK: Cambridge University Press.

Patchan, M. M., Schunn, C. D., & Clark, R. J. (2017). Accountability in peer assessment: examining the effects of reviewing grades on peer ratings and peer feedback. *Studies in Higher Education*. https://doi.org/10.1080/03075079.2017.1320374

Patchan, M. M., Schunn, C. D., & Correnti, R. J. (2016). The nature of feedback: How peer feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology, 108*(8), 1098–1120. https://doi.org/10.1037/edu0000103

Sargeant, J. M., Mann, K. V., Van der Vleuten, C. P., & Metsemakers, J. F. (2009). Reflection: A link between receiving and using assessment feedback. *Advances in Health Sciences Education, 14*, 399–410. https://doi.org/10.1007/s10459-008-9124-4

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

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

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

Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276. https://doi.org/10.3102/00346543068003249

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*(4), 1161–1174. https://doi.org/10.1016/j.compedu.2006.01.007

van Gennip, N., Segers, M., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction, 20*, 280–290. https://doi.org/10.1016/j.learninstruc.2009.08.010

Voet, M., Gielen, M., Boelens, R., & De Wever, B. (2018). Using feedback requests to actively involve assesses in peer assessment: Effects on the assessor's feedback content and assessee's agreement with feedback. *European Journal of Psychology of Education, 33*(1), 145–164. https://doi.org/10.1007/s10212-017-0345-x

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

Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2017). "It'd be useful, but I wouldn't use it": Barriers to university students' feedback seeking and recipience. *Studies in Higher Education, 42*(11), 2026–2041. https://doi.org/10.1080/03075079.2015.1130032

Wu, Y., & Schunn, C. D. (2020a). From feedback to revisions: Effects of feedback features and perceptions. *Contemporary Educational Psychology*, 60. https://doi.org/10.1016/j.cedpsych.2019.101826.

Wu, Y., & Schunn, C. D. (2020b). The effects of providing and receiving peer feedback on writing performance and learning of secondary school students. *American Educational Research Journal*, *58*, 492–526. https://doi.org/10.3102/0002831220945266.

Zheng, L., Zhang, X., & Cui, P. (2020). The role of technology facilitated peer assessment and supporting strategies: A meta-analysis. *Assessment & Evaluation in Higher Education, 45*(3), 372–386. https://doi.org/10.1080/02602938.2019.1644603