Contemporary Educational Psychology 60 (2020) 101826

Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

From feedback to revisions: Effects of feedback features and perceptions

Yong Wu, Christian D. Schunn*

Learning Research and Development Center, University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA 15260, USA

ARTICLE INFO

Keywords:
Understanding
Peer feedback
Feedback features
Agreement
Mediators
Implementation

ABSTRACT

Although the effects of peer feedback have been studied from a number of perspectives, much remains to be learned about what leads students to act (or not) on their peers' feedback in revisions. The present study examined the relationship between peer feedback features, student perceptions as potential mediators (understanding versus agreement with the feedback), and the likelihood of students' implementation of the feedback. Peer feedback, back-evaluation comments, and revisions from 185 US high school students were analyzed. Investigated feedback features included four cognitive features (identification, explanation, solution, suggestion), and two affective features (mitigating praise, hedges). Logistic regression analyses revealed that: (1) both understanding and agreement with feedback predicted implementation; (2) presence of solutions predicted understanding of feedback; (3) mitigating praise predicted agreement of the problem; and (4) explanation and hedges predicted implementation separately from perception effects. Theoretical and practical implications of the findings are discussed.

1. Introduction

In writing instruction, peer review is a common pedagogical activity for evaluating students' writings (Yu & Hu, 2017). When a teacher has a large number of students, peer review is often adopted to help students improve their performance (Cho & Schunn, 2007; Graham & Perin, 2007) because it is difficult for the teacher to respond to every student as immediately and frequently as expected. In addition, a number of studies have found that students learn both from receiving peer feedback and from providing feedback (Huisman, Saab, Van Driel, & Van den Broek, 2018). Peer review, especially online peer review, has a number of advantages over teacher feedback in terms of timeliness, convenience, volume, and learner autonomy (Patchan, Schunn, & Clark, 2018; Topping, 1998).

However, for feedback to positively influence learning, it must be put into action (Topping, 1998; Winstone, Nash, Rowntree, & Parker, 2017). Students often find translating feedback into action to be complex, multi-dimensional, and challenging (Price, Handley, & Millar, 2011). Students' implementation of peer feedback in revisions can be influenced by various factors such as feedback content/features (Lu & Law, 2012; Nelson & Schunn, 2009; Tseng & Tsai, 2007), individual differences (Liu & Carless, 2006; Weaver, 2006; Winstone et al., 2017), as well as students' attitudes and perceptions towards peer feedback (Strijbos, Narciss, & Dünnebier, 2010; Van der Pol, Van den Berg, Admiraal, & Simons, 2008). However, open questions remain regarding critical feedback features and the mechanisms by which peer feedback does or does not result in students implementing peer comments in revisions. As a consequence, teachers and peer reviewers are left with relatively narrow guidance as to which features should be prioritized when conducting peer review training and peer review. The present study seeks to address this gap by identifying features and mediators associated with feedback implementation across a large sample of peer feedback, providing new insights into the mechanism that underlies the observed relationship between them.

2. Theoretical background

2.1. Prior research on feedback features

Because peer feedback is studied by many different educational research communities, a wide variety of terms have been used. We define peer feedback as comments (usually in written form) provided by peers in a class to the author regarding strengths and weaknesses of a document along with constructive recommendations for how to improve. A feedback feature refers to the structural components of feedback comments, such as whether they explicitly describe a problem or give praise, sometimes called feedback content (e.g., Strijbos et al., 2010). Implementation refers to students' incorporation of peer comments in revisions to the document.

Different feedback features are thought to influence students'

* Corresponding author.
E-mail address: schunn@pitt.edu (C.D. Schunn).

https://doi.org/10.1016/j.cedpsych.2019.101826

Available online 26 November 2019

feedback implementation in different ways (Nelson & Schunn, 2009; Strijbos et al., 2010). Hattie and Timperley (2007) meta-analysis investigated whether feedback impact varied as function of different information content in the feedback. The effect sizes of feedback were the highest when students received feedback about the task and how to perform the task effectively; the effect sizes were lower when feedback focused on goals; and the effect sizes were the lowest when students just received praise, reward, or punishment. In addition, Hattie and Timperley described multiple levels at which feedback can be performed, going beyond binary classifications like evaluative vs. informational feedback components (Narciss, 2008), or simple vs. elaborated feedback components (Narciss, 2008; Strijbos et al., 2010). At their first level, feedback can be performed at the task or product level (e.g., identifying whether work is correct or not, or giving corrective feedback). The second is at the process level with an emphasis on information processing or the learning process (e.g., giving explanations, suggestions, or solutions). The third level is concerned with personal evaluations such as praise. They argued that personal evaluations are the least effective and that providing too much feedback at one level may not be effective.

Building from these levels at which feedback can be directed, common peer feedback features can be classified as follows: identification of problems (level 1); explanation, suggestion, and solutions (level 2), and praise (level 3). Effect sizes of praise were found to be low in meta-analysis (Hattie & Timperley, 2007; Kluger & DeNisi, 1996), but these studies did not differentiate between overall praise (i.e., pure praise comments) and mitigating praise (i.e., praise attached to a criticism), in addition to not being focused on peer feedback, which may have different necessary features in contrast to expert feedback (e.g., needing to convey a sense of expertise). Both pure praise and mitigating praise occur commonly in peer feedback (Patchan, Schunn, & Correnti, 2016) and mitigating praise, by definition, is provided in the context of information that can shape improvement. Another related feedback feature not previously considered is the use of hedges (e.g., maybe or possibly) which peers might also use to soften the blow of negative comments. Whether identifying a problem, giving a suggestion, or giving praise, a peer comment can be made with confidence or with various forms of hedges that indicate reviewer uncertainty. However, students generally worry about the competence of their peers (Kaufman & Schunn, 2010), and comments suggestive of reviewer uncertainty may be ignored in implementation since they may give an impression of low competence (Hosman, 1989; Pennebaker, Mehl, & Niederhoffer, 2003).

To better understand the mechanism by which features influence implementation, feedback features can also be grouped into two types according to the kinds of information transmitted: information about the affective relationship (e.g., *I like it.*) and cognitive information such as information about facts, suppositions, beliefs (e.g., *I do not quite understand it.*) (Lee, 2013; Lu & Law, 2012; Nelson & Schunn, 2009). In particular, we propose a model in which the affective vs. cognitive nature of feedback features will determine the feedback influences on implementation by changing author perception of the feedback (understanding or agreement).

2.2. Prior research on feedback perceptions

Building on general research on feedback processing, in our proposed model, cognitive and affective features influence implementation via their effects on the students' perception of the feedback, particularly in terms of understanding of feedback and agreement with feedback (Hattie & Timperley, 2007; Ilgen, Fisher, & Taylor, 1979). Some of the prior research on students' perceptions of feedback relied on information collected by questionnaires (e.g., Huisman et al., 2018; Kaufman & Schunn, 2010; Lizzio & Wilson, 2008; Strijbos et al., 2010) and interviews (Tsui & Ng, 2000; Yang, Badger, & Yu, 2006) and therefore tends to emphasize student preferences rather than reveal effects of features on perceptions. For example, Kaufman and Schunn (2010), using interviews and surveys, found that the majority of students were very concerned whether peers have sufficient expertise to provide quality feedback. In addition, when students receive both teacher and peer feedback, they consider teacher feedback as more helpful and of better quality (Tsui & Ng, 2000), and are more likely to incorporate teacher feedback in revisions than peer feedback (Yang et al., 2006). Although trustworthiness may influence students' perceptions and use of peer feedback, concerns about trustworthiness and uncertainty of peer feedback stimulate "mindful reception" (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991) and are associated with a higher degree of learner autonomy and independence (Yang et al., 2006). Further, several studies have found that such high-level perceptions of the quality of peer feedback in general are not correlated with students' revision work (Kaufman & Schunn, 2010; Strijbos et al., 2010). Instead, the information that is included in feedback influences students' actions (Gielen, Peeters, Dochy, Onghena, & Struyven, 2010).

In general, using questionnaires and interviews may miss information regarding internal feedback processing. Mory (2004) encouraged researchers to "identify measurable variables that can reflect internal cognitive and affective processing of learners that might potentially affect how feedback is perceived and utilized" (p. 777). For example, Nelson and Schunn (2009) measured students' specific perceptions of peer comments in terms of agreement and understanding on the comment level by analyzing their back-evaluations of each feedback comment they received. This method provides more direct information regarding cognitive and affective feedback processing to develop a better understanding of the underlying mechanisms of feedback implementation.

2.3. Prior research on feedback features and perceptions to revision behaviors

Research linking feedback features to feedback perceptions and revision behaviors is scarce and with mixed results. Van der Pol et al. (2008) investigated the relationship between peer feedback features, students' perceptions of the received feedback (e.g., perceived importance of the feedback, usefulness of the feedback, and agreement with the feedback) and their implementation of the feedback. The study did not find significant correlations between feedback features and the receivers' perceptions of the feedback, but they did find significant correlations between feedback implementation and both perceived importance of the feedback and agreement with the feedback, but not with perceived usefulness. Van de Pol et al. argued that the lack of an effect of feedback usefulness on implementation may have resulted from inclusion of praise comments in the analysis because students perceived these compliments as useful but they require not revision.

Nelson and Schunn (2009) also investigated the relationship between feedback features, perceptions of feedback, and students' implementation of peer comments in revisions. Unlike Van de Pol et al., they included understanding of the problem as a feedback perception dimension, in addition to the similar dimension of agreement with the feedback. They found that understanding of the problem was the primary predictor of implementation and that comments including summarization and localization influence receivers' understanding of the problem. Presence of solutions also predicted implementation but not through either of the feedback perception variables they investigated, although this may have been a problem of statistical power given the small size of the dataset they analyzed. Both Nelson and Schunn (2009), Van der Pol et al. (2008) measured receivers' feedback perceptions on the comment level, but they had small sample sizes so that the conclusions drawn from those studies may have failed to detect separable effects of correlated feedback perception dimensions.

In a more controlled experiment investigating the effects of feedback features on students' perceptions of feedback and revisions made to a text based on the feedback, Strijbos et al. (2010) used

questionnaires to measure students' perceptions of peer feedback adequacy and willingness to use the feedback. The study participants were also required to revise a provided text that was created to include a range of text errors. Feedback features were found to influence receivers' perceptions of the feedback, but the feedback perceptions were uncorrelated to their revision performance, in direct opposition to the findings of Nelson and Schunn (2009), and of Van der Pol et al. (2008). However, the study by Strijbos et al. (2010) focused on surface features of writing, which were easier to improve. In addition, the surface errors that students needed to correct were artificially included into the original text by one of the researchers. That the study was not carried out in the context of authentic writing could also limit the effects of students' perceptions of feedback on their revision work.

Building upon the Strijbos et al. (2010) questionnaire approach, Huisman et al. (2018) investigated the relationship between feedback features, perceptions of the received feedback (perceived adequacy and willingness to improve), and writing performance (performance increase from the first draft to the final revised draft). Explanatory comments were correlated with both feedback perceptions dimensions, but once again students' perceptions of the feedback and their writing performance was not significantly correlated. Because the outcome was overall document improvement rather than specific revisions in response to specific comments, it could be the analysis was statistically under-powered due to the indirect measurement of change. In general, additional research is required to empirically investigate these questions in the context of authentic writing tasks but with sufficiently large sample size to test with sufficient power whether and which peer feedback features influence students' implementation of peer feedback.

2.4. Proposed feedback features-to-implementation model

The present study aims to examine the relationship between feedback features, students' (authors') perceptions of the feedback in terms of understanding and agreement with the received feedback, and implementation of the feedback in revisions across a large sample of peer feedback comments in an authentic writing task and learning. The model of feedback features, feedback perception, and feedback implementation that was tested is presented in Fig. 1. First, two key perception mediators, understanding of and agreement with the received feedback, are expected to predict students' likelihood of feedback implementation. Second, feedback features are conceptually divided into cognitive and affective features, with cognitive features expected to predict feedback understanding and affective features expected to predict feedback agreement. Third, there should be no direct relationships between feedback features and implementation beyond their mediation through understanding and agreement (i.e., the effects are predicted to be fully mediated). The next section presents the specific feedback features and feedback perception dimensions to be tested, as well as a description of important control variables to consider that may be confounding factors (i.e., correlated with feedback features, feedback perception, and implementation).

2.4.1. Affective and cognitive features

Identification of the problem. Critical feedback can address a problem explicitly (e.g., *The organization did not follow any train of thought.*) or implicitly (e.g., *Add in what rhetorical devices Louv used that will be discussed in your essay in the thesis.*). Explicit identification of a problem in feedback can help students understand what and where the problem is. However, a number of researchers have found no effects on feedback implementation of problem identification alone (Nelson & Schunn, 2009) or in direct contrast against providing suggestions or solutions (Bangert-Drowns et al., 1991). It may be that identification alone does not provide sufficient details and guidance (Weaver, 2006), whereas identification coupled with additional feedback elements is more helpful. Identification has been found to be more important for reviewer learning (Cho & MacArthur, 2011; Lu & Law, 2012), perhaps because it involves reviewers in more highly cognitive demanding performance (Lu & Law, 2012). Identification is expected to positively predict understanding of the feedback.

Explanations. Among the feedback features that are likely to influence understanding, perhaps the most significant one is whether an explanation is included in the feedback, whether the explanation is about problems or solutions. Explanations can provide more detailed information to improve understanding. Explanatory comments have been found to be positively related to students' understanding of peer feedback and students' willingness to respond to it (Gielen et al., 2010; Huisman et al., 2018). Explanations are argued to be even more important than accuracy of peer feedback for improving performance (Gielen et al., 2010), perhaps because explanation enhances mindful cognitive processing (Bolzer, Strijbos, & Fischer, 2014). However, some studies (Nelson & Schunn, 2009; Tseng & Tsai, 2007) found that explanations can have a negative impact on writing performance because novice writers are unable to offer clear explanations. But guidance with peer review rubrics may help improve the interpretability of the explanations. Therefore, explanations in peer feedback, when supported by a good rubric, are expected to positively influence students' understanding of the feedback and their likelihood of implementation.

Suggestions and Solutions. Even when students have understood the problem identified in a feedback comment, they may be unable to respond to the comment without additional advice on how to improve (Price et al., 2011). Indeed, peer feedback is often constructive: providing guidance on how to address identified weaknesses. The specificity of the constructive advice often varies. Sometimes reviewers provide a suggestion that is indirect or more general (e.g., add more evidence), while sometimes they provide a solution that is direct and very specific (e.g., add evidence X). Cho, Schunn, and Charney (2006) found that instructors are more likely to provide specific solutions than general suggestions whereas peers tend to offer the two constructive comment types at roughly equal rates. Importantly, most prior studies of peer feedback have not distinguished between general suggestions and specific solutions (e.g., Cho & MacArthur, 2011; Lu & Law, 2012), so that it is difficult to apply those research findings to the current model. It is important to distinguish the two types because they involve different engagement in cognitive activities and therefore are likely to have different effects on students. Students may be less likely to implement a more general suggestion because it is more difficult to understand and requires higher cognitive demand to translate the suggestion into a specific solution. Nelson and Schunn (2009) found that offering solutions was related to implementation of feedback, although not necessarily because it influenced understanding or agreement. However, that study did not separate solutions from suggestions, and further may have suffered from lack of statistical power. Providing concrete suggestions is expected to improve understanding of the constructive aspects of the comment.

Hedges. Another salient element within feedback is the inclusion of hedges (e.g., maybe, perhaps), and can be applied to both the problem identified in the comment and solutions or suggestions given in the comment. Peers are inherently non-experts in the domain, and sometimes confidence in their recommendations (by both provider and receiver) is low. This issue may be particularly frequent when a peer-review rubric forces students to comment on many aspects of a document, including less well understood aspects. Hedges in feedback, whether about the problem or the constructive elements, are likely to make the feedback less persuasive (i.e., reduce agreement of the problem and the solution) and therefore less likely to be implemented. Explicit hedging expressions may also make the feedback more complex and thus harder to understand (Baker & Bricker, 2010).

Mitigating Praise. Mitigating praise refers to positive feedback that is added directly to negative feedback to soften the criticism. Critical comments without praise are more likely to be associated with negative emotions that can lead students to reject the feedback and decrease the effectiveness of feedback interventions (Kluger & DeNisi, 1996). By

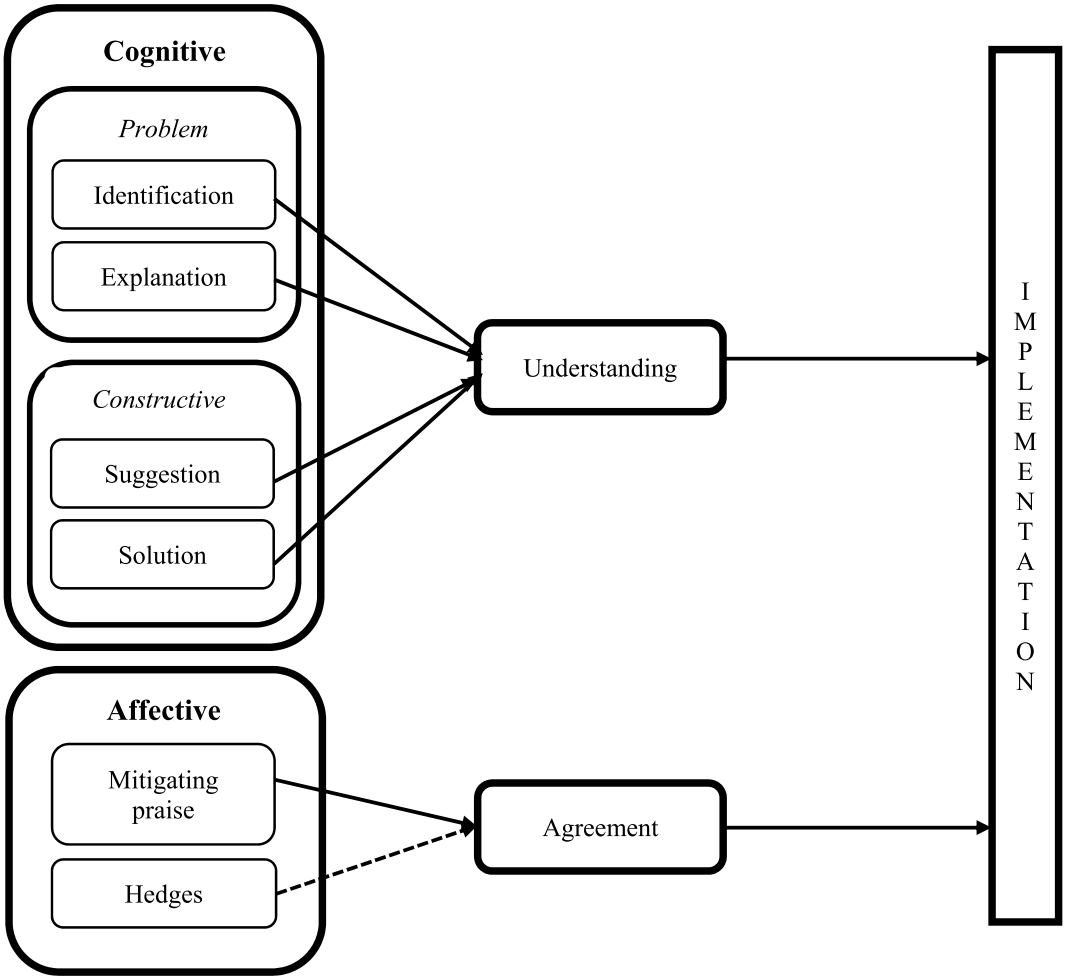

Fig. 1. Proposed feedback-to-implementation model.

contrast, mitigating praise is thought to decrease the potential negative effects of constructive criticism on students' self-esteem and motivation, and increase the likelihood of students' implementing negative feedback (Cho, Schunn, & Charney, 2006; Hyland & Hyland, 2001). Despite the prior claims about the benefits of mitigating praise, students may not use feedback that included mitigating praise to improve their documents because they do not take the problem seriously when the criticism is mitigated (Patchan et al., 2016). Nonetheless, mitigating praise is expected to influence agreement with the critical comment being mitigated.

2.4.2. Hypothesized mediators for feedback implementation

Two hypothesized feedback perception mediators are included in the present study: understanding feedback and agreement with feedback. Cognitive features (i.e., identification, explanation, solution, and suggestion) are expected to predict feedback understanding and the affective features (i.e., mitigating praise and hedges) predict feedback agreement.

Understanding feedback. Understanding feedback is important for students to take actions on the feedback (Gibbs & Simpson, 2004; Price et al., 2011). However, feedback is often not understood by students (Jönsson & Panadero, 2017). As noted by Higgins (2000), "many students are simply unable to understand feedback comments and interpret them correctly" (p. 1). Others have also found that some students do not have a clear understanding of what the comment means and what is needed to improve (Weaver, 2006), usually because of the contents of the feedback: it is vague, unclear, brief, or lacks useful information (Lizzio & Wilson, 2008). Further intervention to interpret feedback is often needed for receivers (Carless, 2006). Overall, lack of understanding will make the implementation of peer feedback less likely to happen (Price et al., 2011; Winstone et al., 2017). Therefore, because of its necessity for implementation and relatively high frequency of being problematic, understanding of feedback is expected to be a strong predictor of implementation in the present study.

Agreement with feedback. In spite of the frequently large influence of feedback on learning, a number of students do not use the feedback that they receive (Jönsson, 2013). One of the major reasons for not using feedback is that students do not agree with the feedback. Agreement with feedback is related with whether the content of the feedback message matches the learners' perceptions of the quality of the document being evaluated (Nelson & Schunn, 2009) or the learners' expectations of feedback (Hattie & Timperley, 2007; Ilgen et al., 1979). When students believe that the negative feedback or the suggested recommendations in the feedback will improve performance, they will be more likely to agree with and implement the feedback (Van der Pol et al., 2008). However, when students attribute the criticism to faulty or inappropriate identification and disagree with the feedback, they will ignore the feedback (Leung, Su, & Morris, 2001). Prior research has suggested that task performance is likely influenced by perceived feedback fairness (Chory & Westerman, 2009; Locke, 1968). Most directly, Van der Pol et al. (2008) found that students' agreement with peer feedback was significantly correlated with its use for revision. However, Nelson and Schunn (2009) found that understanding alone, and not agreement, predicted implementation; although they may not have had sufficient statistical power to see effects of both agreement and understanding. Strijbos et al. (2010) also did not find direct

relations with students' revision work and feedback perceptions. But one possible explanation could be that surface errors were artificially included so that effects of students' feedback perceptions on their revision work were limited. In the present study, agreement with feedback is expected to increase students' likelihood of implementation.

2.4.3. Contextual variables also predicting implementation

Other contextual factors above the level of the comment also influence feedback implementation. Although not the central focus of the current study, they must be measured and statistically controlled in analyses to reveal the unique relationships between features, perceptions, and implementation.

Students can receive a large number of comments in multi-peer review and each comment can vary in length. This variation in number of comments can have many effects on whether students make changes or not. If there are only a few critical comments, students may feel the document is basically fine and no changes are required. Alternatively, students can be overwhelmed by receiving too much critical feedback (Sweller, 1988). Patchan et al. (2016) found that the number of implementable comments, rather than the total amount of feedback of all types (e.g., criticism, praise), increased the number of revisions. In terms of feedback length, longer implementable comments generally contain more detailed information that can help students to understand and has been positively associated with revisions (Ferris, 1997). Based on the past research, both number of implementable comments and comment length are expected to be associated with both agreement and implementation.

Praise comments, different from mitigating praise included in an implementable comment, solely consist of positive comments about what the author has done. Praise alone does not likely trigger revision, but it is often argued to be helpful in promoting students' revision work because positive comments can create a more supportive learning environment in which students are inspired to learn (Lizzio & Wilson, 2008; Weaver, 2006). However, some researchers have argued that praise is ineffective because of its lack of information relevant to improvement and learning (Hattie & Timperley, 2007), especially for complex performance tasks (like writing) that require information on how to improve. Therefore, having a higher number of praise comments is expected to predict a lower rate of feedback implementation. Similarly, students are less likely to make any revisions if they received a high score on their first drafts or generally perceived their first draft as being of sufficiently high quality.

Students receive peer feedback within a context with established norms and expectations (Bangert-Drowns et al., 1991; Hattie & Timperley, 2007), which could include what kinds of feedback should be produced, how feedback should be processed, how much revision is expected, etc. These norms and expectations can come from the current teacher of the class in which the peer feedback is received, or it could come from prior educational experiences at that school. Schools with higher rates of poverty will often have more students that are less academically prepared, more teachers who are less experienced, and various broader environmental factors that undermine students' general academic motivation (Hughes, 2012; Irvin, Meece, Byun, Farmer, & Hutchins, 2011). Therefore, when data are collected across schools of varying demographics (as in the current study), it is important to control for whether the school enrolls relatively high numbers of low-income students or not.

3. Method

3.1. Participants

There were 185 participants in the study (102 Female), with a mean age of 17.1, ranging from 16 to 19 years. They were students enrolled in an Advanced Placement course in writing (AP Language and Composition) at two secondary schools in the United States. The schools were strategically selected to have varying writing abilities within and between schools (i.e., not have only high performing students). Sixty participants were from a lower-performing school serving many low-income families (i.e., were at a Title I school). The remaining 125 participants were from a higher-performing school serving middle- and high-income families. All sections within a school were taught by the same teacher. Both schools included participants performing at high and low levels on the writing task used in the study. According to student self-reports, a near majority of participants were White students (48%). Non-white participants included students who listed Asian (25%), African American (4%), and Hispanic/Latinx (3%) race/ethnicities; 19% of the participants chose not to report their race/ethnicity.

AP (Advanced Placement) courses are introductory college courses in a wide variety of subjects that are provided to high-school students within their school and culminate in nationally-administered exams. By taking AP courses, students can challenge themselves in subjects of strong interest to them, strengthen college applications, and reduce the number of courses they need to take in college. This AP course (AP Language and Composition) has the highest annual enrollment among all AP courses, reflecting a broad push in the US to include a broader cross-section of schools and students who have access to more rigorous coursework and become college ready (see College Board, 2018). The objective of this course was to provide students with more opportunities to practice critical writing skills and develop their writing proficiency. All participants from the same school were taught by a single teacher across multiple sections, and reviewing was implemented as students interacting across sections within a school. They make the writing and reviewing contexts closely comparable for the current study, and participants across schools were given the same writing assignment and conducted peer reviews using shared rubrics and processes at a similar time during the school year.

3.2. Materials

3.2.1. Writing assignment

The writing assignment asked students to read a one-page passage and write a rhetorical analysis essay, one of the three core genres taught in the AP course. In particular, students analyzed the rhetorical strategies used in the source passage and were required to support their analysis with specific examples from the source passage.

3.2.2. Peer reviews

Students used *Peerceptiv*, a widely-used online peer assessment program (Cho & Schunn, 2007). After students composed their first drafts, they submitted the essays to the online system. The system randomly distributed each essay to four peers within the school anonymously. Each student was required to review four peer essays for a given draft. They commented on each paper online and graded it based on eight seven-point rating rubrics (see Appendix A). The rubrics are an adaptation of the holistic rubric used in the AP exam, adapted to be more student friendly (Schunn, Godley, & DiMartino, 2016). Reviewers were required to provide at least one comment for each evaluation dimension. Teachers did not provide comments on these first drafts and students completed the reviews on their own. However, teachers did provide a short in-class training session to students on how to use the system, interpret the rating dimensions, and generally provide comments.

In addition, the student as authors back-evaluated each of the reviews they received, rated the helpfulness of the peer feedback using a five-point scale and providing a comment that explained their helpfulness rating. Finally, students revised their essays based on peer feedback and submitted the revised drafts to the system. Accuracy of ratings and helpfulness of reviews contributed to student grades, providing incentives for each reviewer to take their reviewing task seriously, and these incentives have been shown to produce longer reviews and more accurate ratings (Patchan et al., 2018).

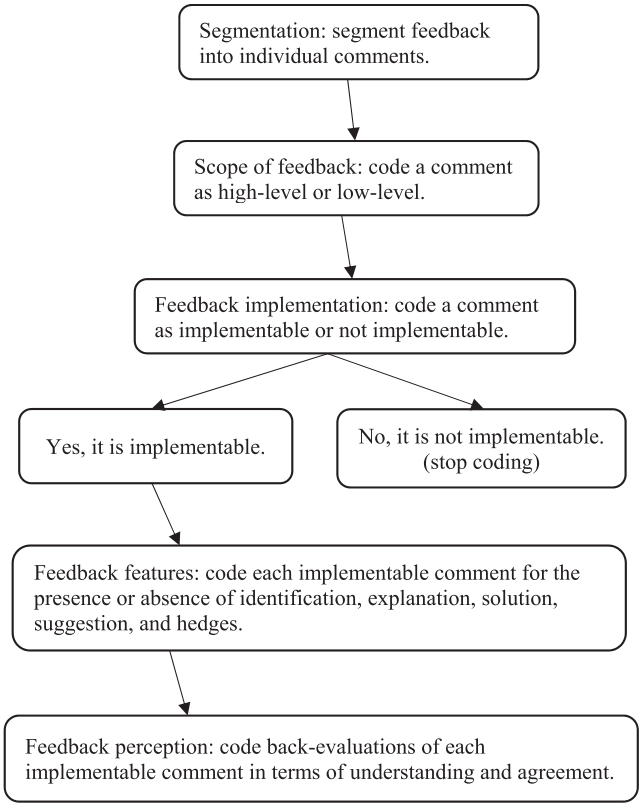

Fig. 2. Coding process of feedback features and mediators.

3.3. Measures

Measures used to examine the effects of peer feedback include measures applied to the first draft essays, essay revisions, peer comments, and back-reviews of peer feedback. The full feedback coding process is presented in Fig. 2. Multiple trained research assistants were involved in each coding dimension. Most Kappa coefficients were high (see Table 1), but a few were around 0.6, which is only considered moderate reliability (Anthony & Joanne, 2005). To raise the effective reliability of the analyzed data, all aspects were exhaustively double-coded and all disagreements were resolved through discussion with a third coder present.

3.3.1. Feedback segmentation, scope, and amount of feedback measures

First of all, peer comments were segmented into separate idea units because some comments focused on more than one problem. An independent idea unit was defined as raising and/or solving one problem on one dimension. For example, the following comment includes two idea units: *"I think that you misinterpreted the purpose of Louv's anecdote in the first body paragraph. I don't think that the anecdote provided a contrast of present and earlier time periods. The wording in the same paragraph could be changed to be a bit more sophisticated."* The first two sentences constitute an idea unit focused on a problem regarding the author's analysis (a high-level problem). The last sentence is a separate idea unit with a focus on control of language (a low-level problem). A total of 6507 idea units were produced as a result of the segmentation process.

The scope of each segment was then coded as High-level or Low-level (see Table 1). High-level comments were related to thesis, argument, rhetorical strategies, evidence for claims, explaining evidence, and organization, while low-level comments involved control of language and conventions. Two measures were created that summed across the reviews they received: number of high-level comments and number of low-level comments.

3.3.2. Feedback implementation in revision

To code implementation of the feedback, the segmented feedback comments were first categorized into non-implementable (summary or praise only comments) or implementable comments (see Table 1 for definitions and examples; Kappa = 0.91). A total of 3605 comments were found to be implementable. In terms of non-implementable comments, summary was excluded from further analysis, and praise was summed at the author level to be a control variable, i.e., the number of praise comments an author received. Another control variable, the number of implementable comments per author were also summed at the author level, in contrast to comment length of implementable comments, which was analyzed at the comment level.

To code whether the implementable feedback was actually implemented, students' first and second drafts were compared using Microsoft Word's Compare Document tool. Then, the implementable peer feedback was compared with these changes (Kappa = 0.58). If any change could be attributed to a feedback comment, the comment was coded as Implemented; if no changes could be matched with a peer comment, this comment was coded as Not Implemented. 12% of the implementable comments that were found to be too vague to identify for implementation were excluded from further analysis. The majority of these vague comments were low-level (see Table 1).

3.3.3. Feedback features

To examine the effect of peer feedback features on implementation, implementable comments were exhaustively double-coded by four research assistants for the presence or absence of the following features: identification (Kappa = 0.84), explanation (Kappa = 0.71), suggestion (Kappa = 0.76), solution (Kappa = 0.67), hedges for problems (Kappa = 0.72) and hedges for constructive comments (Kappa = 0.75), and mitigating praise (Kappa = 0.86). The definitions and examples of the feedback features are presented in Table 1 and Appendix B.

3.3.4. Feedback perception

Back-review comments were exhaustively double-coded for whether students agreed with or understood the problem/solution in each feedback comment. Since some students did not provide a back-evaluation, the effective number of back-reviews for analysis was 1951 for high-level comments and 242 for low-level comments. Because many low-level comments were too vague to identify for implementation and the number of the low-level back-reviews was small, back-reviews of low-level comments were not included in this analysis.

If the writer explicitly disagreed or partially agreed with the problem or solution, the back-review comment was coded as Not Agreed or Partially Agreed; otherwise, it was coded as Agreed (for problems Kappa = 0.79; for solutions Kappa = 0.66). If the writer explicitly stated that she/he did not understand or partially understood the problem or solution in the back-review comment, it was coded as Not Understood or Partially Understood; otherwise, it was coded as Understood (for problems Kappa = 0.71; for solutions Kappa = 0.64). Back-review coding created two measures: 1550 back-review comments for problems and 1229 back-review comments for constructive comments (i.e., containing a suggestion or solution). The definitions and examples of the feedback perception are presented in Table 1 and Appendix C.

3.3.5. Essay quality

Essay quality score was measured using the mean peer ratings across four peers for each dimension based on the rubrics. Students received this score as the only indicator of draft quality, and thus it serves as a reasonable proxy of the perceived need to revise. However, there is also reason to believe this mean rating was actually an adequate measure of essay quality. First, the overall setup of the peer assessment assignment made it likely that there would be high reliability and validity to these scores. Including a written description of the characteristics of responses within each rating reduces variation across raters

Table 1
Peer feedback coding scheme.

Category	Definition	Example
Type of feedback (Kappa = 0.91)		
praise	Purely evaluative remarks on good features of writing	On the other hand, in the second paragraph, you've done it perfectly! You've given about 2–3 quotes and a lot more analysis into the quotes which conveys your point more.
summary	Statements of what the writer has done	The author uses quotes to support what rhetorical devices the author used in his writing and then they also describe why they chose that quote and how that relates to the rhetorical device they saw.
implementable comments	Revision-oriented comments that could trigger revisions	In the third body paragraph, they begin to explain but never reference their third rhetorical strategy.
Feedback features		
identification (Kappa = 0.84)	To announce what is problematic, or what needs to be further developed	I did not notice a thesis in this essay.
explanation (Kappa = 0.71)	To explain why it is problematic	You talk of various rhetorical devices, and explain all of them, but I must object to your choice of hyperbole, which was frankly self-explanatory and is not worth a reader's time.
suggestion (Kappa = 0.76)	To provide general advice by giving the directions for changes	The thesis seems undeveloped. Express a little background information of the prompt that the author uses in the text.
solution (Kappa = 0.67)	To provide specific advice by outlining alterations and corrections.	The thesis is a little weak. Strengthen it by combining the last two sentences together.
hedges (problem) (Kappa = 0.72)	Hedges used in the identification or explanation	The quote in the second paragraph seemed awkward and a little forced.
hedges (solution) (Kappa = 0.75)	Hedges used in the suggestion or solution	I think maybe you could add a little more evidence for rhetorical strategy.
mitigating praise (Kappa = 0.86)	Being embedded in a negative comment to soften criticism	Your thesis is good, however I think it would be a good idea to introduce what rhetorical devices you will be analyzing in your body paragraphs.
Scope of implementable feedback (Kappa = 0.91)		
high-level	Comments with regards to thesis, arguments, rhetorical strategies, organization, evidence, and explanation	The author misinterprets the message in the first body paragraph which causes him/her to analyze the rhetorical strategies inaccurately.
low-level	Comments with regards to language of control and style	There should be a comma after "piece".
Implementation of feedback (Kappa = 0.58)		
implemented	Comments that were incorporated in the revision	Feedback: Also while you address rhetorical devices, you don't connect them to what they create or to a purpose. It's clear you understood his point based on your stance and tone, however you should try to connect your thesis to the purpose. *The author added two sentences at the end of Para. 4*: This quote proves how people and nature have separated over time with a personal experience. The detail of the memory illustrates the child-like fascination of nature and its beauty that still needs and deserves to be appreciated.
not implemented	Comments that were not incorporated in the revision	Feedback: There is no evidence to be supported. Include evidence so that it may be supported. *The author did not add any evidence in the later draft.*
vague	Comments that were too vague to determine whether they were implemented or not	Throughout the essay you were able to effectively control your language. More specific academic vocabulary would have been nice to see though.
Mediators		
agree with problem (Kappa = 0.79)	To agree with the reviewer about the problem mentioned in the comment	This is true. I need to work on explaining to how the rhetorical devices enhance the argument.
agree with solution (Kappa = 0.66)	To agree with the reviewer about how to solve the problem	I will elaborate more when it comes to explanations.
understand problem (Kappa = 0.71)	To express understanding of the problem mentioned in the comment	Thank you for the review, it helps. And yes, I know I didn't analyze rhetorical devices. I will fix it.
understand solution (Kappa = 0.64)	To express understanding how to solve the problem	Tell me what I need to do to improve. I understand. Thank you so much!

(Brown, Glasswell, & Harland, 2004; Moskal, 2000; Russell, Horne, Ward, Bettis, & Gikonyo, 2017). Using rubrics with descriptive scales that specify important strengths and typical weaknesses insures that students know what is expected from them (Andrade, 2000; Sadler, 1989). Using an average across four reviewers further increases reliability. Second, peer scores have been found to be reliable and valid in prior research (e.g., Cho, Schunn, & Wilson, 2006; Russell et al., 2017), especially when multiple peer reviewers are involved (Cho, Schunn, & Charney, 2006; Tseng & Tsai, 2007) and accountability features are used (Paré & Joordens, 2008). Further, as part of a larger study, randomly selected essays from these two schools and eight other schools were rated by trained expert raters based on the same rubrics. A correlation coefficient of 0.65 was found between the mean peer ratings and those of expert raters, which is similar to correlations previously observed between teachers using carefully designed rubrics (Brown et al., 2004; Cho, Schunn, & Wilson, 2006). In other words, the accurate and consistent scoring of essays is difficult (Brown et al., 2004), and the mean peer score has similar reliability and validity to single expert ratings. Students may be systematically more generous in their ratings than are teachers or experts, but such mean differences effects are not relevant to the current use of the ratings as covariates.

4. Results

Three sets of logistic regression analyses were conducted to test associations between feedback features, perceptions, and implementation: (1) from feedback perception (predictors) to implementation (response variable) to test which mediators were important; (2) from feedback features (predictors) to feedback perception (response variables); and (3) from feedback features and perceptions combined (predictors) to implementation (response variable) to test whether there were unexplained remaining direct effects from features to

Table 2
The variables examined in the regression models, 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			
Implementation	Binary	Comment	Whether an implementable comment was implemented
Control and contextual Variables			
Comment length	Continuous	Comment	Number of words in an implementable comment
Number of implementable comments	Continuous	Author	Total number of implementable comments an author received
Number of praise comments	Continuous	Author	Total number of positive comments an author received
First draft quality School	Continuous Binary	Author Author	Mean peer ratings across four peers and all dimensions Whether students came from Title I school
Comment Features			
Mitigating praise	Binary	Comment	Whether an implementable comment included mitigating praise
Identification	Binary	Comment	Whether a problem was identified explicitly in an implementable comment
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
Hedges	Binary	Comment	Whether a reviewer used hedges when offering an identification/explanation/suggestion/ solution
Mediators			
Understanding of the problem	Binary	Comment	Whether a student understood the problem
Understanding of the solution	Binary	Comment	Whether a student understood the solution
Agreement with the problem	Binary	Comment	Whether a student agreed with the problem
Agreement of the solution	Binary	Comment	Whether a student agreed with the solution

implementation. The analyses were done separately for the subset of comments describing problem (N = 1550) and subset of comments containing constructive advice (N = 1229). Theoretically, problems versus constructive advice comments could depend upon different features for implementation (based on Nelson & Schunn, 2009), and the coded perceptions were specific to the problem or constructive advice aspects of the comments. Further, empirically the two datasets showed differences in statistically significant relations.

Descriptive statistics for each variable are presented in Table 2. For problems, the feedback feature identification was present in almost all cases, so it was removed from data analyses of problem comments. In terms of feedback perception, partially understand/agree and completely understand/agree were collapsed because of the small number of partial understanding/agreement cases and because they showed similar patterns in terms of their effects on implementation. Since there were only two schools, it was not possible to implement a Hierarchical Linear Model with students nested within schools.

Tables 3 and 4 present the simple Pearson correlations between all the variables. For both problems and constructive comments, the correlation between understanding and agreement was small (r < 0.2), revealing that concerns about each aspect of a comment were quite independent. Within the feedback features (upper right), most correlations were small (i.e., the features were generally independent). For problems, the correlations among the features were all small (r < |0.2|). For constructive comments, there were two strong correlations: identification and explanation were positively correlated; and suggestion and solution were negatively correlated. Within the control variables, as to be expected, the number of praise comments and first draft quality were positively correlated, and number of implementable comments was negatively correlated with number of praise and first draft quality. For problems, school was negatively correlated with three control variables. In others words, students from Title I school received fewer and shorter implementable comments, more praise comments, and their first draft quality was lower than students from non-Title I school.

Finally, in terms of the correlations between the features and the mediators, both explanation and solution had significant correlations with understanding of the problem and the solution. Suggestion and mitigating praise had small but significant correlations with agreement of the problem, while none of the features were significantly correlated with agreement of the solution. However, because of the existence of some significant correlations with and between feedback features and control variables, a multiple regression was needed to parse out the separate contributions from each predictor. In addition, the existence of mixtures of negative and positive correlations suggests there may be suppression effects by which direct relations between variables was in fact larger than the simple correlations would suggest.

For each regression analysis, relevant assumptions were tested.

Table 3
Control variables, feedback features, mediators, and implementation: correlations for problems.

Variable		1	2	3	4	5	6	7	8	9	10	11	12
1	Comment length	1											
2	Number of implementable comments	0.12**	1										
3	Number of praise comments	−0.01	−0.62**	1									
4	First draft quality	0.08**	−0.33**	0.53**	1								
5	School	−0.38**	−0.45**	0.10**	−0.13**	1							
6	Explanation	0.42**	0.00	−0.05*	−0.08**	−0.08**	1						
7	Suggestion	0.28**	0.05	−0.04	0.03	−0.13**	0.01	1					
8	Solution	0.21**	−0.02	−0.00	0.07**	−0.05	0.03	−0.06*	1				
9	Mitigating praise	0.14**	−0.06*	0.14**	0.28**	−0.14**	−0.14**	0.03	−0.03	1			
10	Hedges	0.04	−0.07*	0.03	0.02	0.09**	0.09**	0.01	−0.02	0.05	1		
11	Understanding of the problem	0.11**	−0.01	−0.03	−0.03	0.01	0.06*	0.02	0.07**	0.03	0.00	1	
12	Agreement with the problem	0.07**	0.07**	−0.09**	−0.06*	−0.07**	−0.00	0.07**	0.00	0.06*	0.03	0.13**	1
13	Implementation	0.05	0.13**	−0.22**	−0.25**	−0.05	0.13**	0.03	−0.04	−0.08**	−0.03	0.18**	0.21**

Note. N = 1550 for correlations among feedback features, mediators and implementation. *p < .05, **p < .01.

Table 4
Control variables, feedback features, mediators, and implementation: correlations for constructive comments.

Variable		1	2	3	4	5	6	7	8	9	10	11	12	13
1	Comment length	1												
2	Number of implementable comments	0.14**	1											
3	Number of praise comments	−0.08**	−0.63**	1										
4	First draft quality	0.00	−0.38**	0.58**	1									
5	School	−0.31**	−0.41**	0.08**	−0.09**	1								
6	Identification	0.30**	0.12**	−0.16**	−0.21**	−0.02	1							
7	Explanation	0.47**	0.06*	−0.15**	−0.17**	−0.05	0.61**	1						
8	Suggestion	−0.05	0.03	0.03	−0.04	0.00	−0.05	−0.04	1					
9	Solution	0.18**	−0.02	−0.02	0.07*	−0.04	−0.05	−0.01	−0.69**	1				
10	Mitigating praise	0.08**	−0.12**	0.16**	0.27**	−0.06*	−0.23**	−0.20**	0.06*	−0.01	1			
11	Hedges	−0.04	−0.04	0.06*	0.15**	0.00	−0.09**	−0.09**	−0.12**	0.06*	0.17**	1		
12	Understanding of the solution	0.15**	−0.01	−0.03	−0.04	0.10**	0.05	0.09**	−0.09**	0.13**	−0.02	0.00	1	
13	Agreement with the solution	0.05	0.08**	−0.07**	−0.11**	−0.00	0.02	0.01	0.05	−0.00	0.01	−0.05	0.08**	1
14	Implementation	0.04	0.10**	−0.22**	−0.23**	0.04	0.15**	0.18**	0.01	−0.02	−0.05	−0.12**	0.16**	0.20**

Note. $N = 1229$ for correlations among feedback features, mediators and implementation. *$p < .05$, **$p < .01$.

Suggestion and solution were strongly (negatively) correlated (see Table 4) in the constructive comments dataset, so the predictor variable suggestion was removed from the covariates to overcome this multicollinearity problem. After removing this variable, no predictor variable had a Variance Inflation Factor greater than 2.2, suggesting there were no remaining multicollinearity problems. The results of logistic regression analyses are presented in terms of odds ratios in Tables 5 and 6. An odds ratio greater than 1 means a positive relationship between the predictor and the outcome (when holding all other variables constant), while an odds ratio below 1 implies a negative relationship between the predictor and the outcome (when holding other variables constant). An odds ratio of 1 means that there is no difference between the groups when other variables are included.

4.1. What feedback perceptions predicted implementation?

Both feedback perceptions (whether students understood or agreed with feedback) were significant predictors of implementation by the mediators for both problems and constructive comments. The effects were quite large (see the leftmost regression results in Tables 5 and 6). Students were 3.8 times as likely to implement a comment when they understood the problem than when they did not, with a similar effect of agreeing with the stated problem. The effect of understanding the solution was slightly smaller, but still large: 2.7 times as likely to implement when they understood the solution than when they did not. The largest effect was for agreeing with the solution: students were 4.4 times as likely to implement the solution when they agreed with it than when they did not.

4.2. What feedback features predicted feedback perceptions?

For both problems and constructive comments, the prediction model for understanding was consistent: concrete solutions were more likely to have the problems and solutions be understood, with a larger effect for understanding of the solution (see middle columns of Tables 5 and 6). In contrast, the problem and solution models for agreement were different: mitigating praise predicted agreement with the problem but not with the solution Students were 1.64 times as likely to agree with the problem when it included mitigating praise. Note that explanation in isolation was correlated with understanding of the problem and understanding of the solution, but this relationship disappeared when other variables were included. Identification, suggestion, and hedges were not significant predictors of feedback perception.

4.3. What feedback features directly predicted implementation?

To test whether any of the feedback features were related to implementation in ways not explained by the proposed perception mediators, models with features predicting implementation were run including the mediators (see rightmost columns of Tables 5 and 6). For problems, explanation directly predicted implementation of feedback in revisions (i.e., not through mediators). For constructive comments, two features—explanation and hedges—predicted implementation directly. In particular, students were approximately twice as likely to implement a change when the comment included an explanation, and less likely to implement a change when the suggestion/solution was made with hedges. Importantly, since all of these relationships were found with the inclusion of the mediating perception variables, these relationships with implementation are not because they influenced whether a comment was understood or influenced the level of explicit agreement the author had with the comment.

Mitigating praise in isolation was associated with implementation for problems, but not when other variables were included. Similarly, identification in isolation was associated with implementation for constructive comments, but not when other variables were included. These changes in patterns of association from simple correlations to multiple regression highlight the importance of controlling for shared variance with other features of a comment and other contextual variables.

4.4. Additional predictors

Five contextual and control variables (i.e., comment length, number of implementable comments, number of praise comments, authors' first draft quality, and school) were included in the statistical models. Four were significantly related to either feedback perception or implementation, arguing in favor of their inclusion in the models. Comment length predicted understanding of both the problem and the solution and also implementation for constructive comments. The number of praise comments was negatively correlated with implementation. Authors' first draft quality was also negatively correlated with implementation but with a larger relationship for implementation of problem comments. Although first draft quality and number of praise comments were correlated with one another, note that these effects held in a model that included both at once. That is, there appears to be separate effects of receiving high scores and large amounts of praise in demotivating changes in response to feedback. In addition, first draft quality was negatively correlated with agreement of the solution: authors with higher scores on the first drafts were much less likely to agree with the solutions/suggestions they received. Finally, students from the Title I school were less likely to implement feedback. We also tested whether the control variables had quadratic relationships to feedback implementation, but failed to find such relationships.

Table 5
Hierarchical regression analysis predicting mediators and implementation for problems.

Predictors	Response variables									
					Mediators					
			Implementation (control + mediators)		Understanding of the problem (control + cognitive features)		Agreement with the problem (control + affective features)		Implementation (all predictors)	
	M	*SD*	*B (SE)*	*Odds ratio*	*B (SE)*	*Odds ratio*	*B (SE)*	*Odds ratio*	*B (SE)*	*Odds ratio*
Control and contextual variables										
Comment length	56.48	27.38	0.00 (0.00)	1.00	0.02 (0.01)	1.02***	0.01 (0.00)	1.01	−0.00 (0.00)	1.00
Number of implementable comments	25.21	8.07	−0.02 (0.01)	0.98	−0.02 (0.02)	0.98	−0.01 (0.02)	0.99	−0.02 (0.01)	0.98
Number of praise comments	9.12	5.75	−0.05 (0.02)	0.95**	−0.03 (0.03)	0.98	−0.05 (0.02)	0.95*	−0.05 (0.02)	0.95**
First draft quality	4.96	0.99	−0.60 (0.09)	0.55***	−0.14 (0.13)	0.87	−0.22 (0.12)	0.80	−0.54 (0.09)	0.58***
School	0.21	0.40	−0.50 (0.19)	0.61*	0.37 (0.32)	1.45	−0.47 (0.28)	0.63	−0.50 (0.20)	0.60*
Feedback features										
Mitigating praise	0.42	0.49	–	–	–	–	0.49 (0.21)	1.64*	−0.10 (0.13)	0.91
Explanation	0.67	0.47	–	–	0.04 (0.23)	1.04	–	–	0.58 (0.15)	1.78***
Suggestion	0.48	0.50	–	–	−0.04 (0.22)	0.96	–	–	0.05 (0.13)	1.05
Solution	0.09	0.28	–	–	2.03 (1.02)	7.60*	–	–	−0.35 (0.22)	0.71
Hedges	0.24	0.43	–	–	–	–	0.28 (0.23)	1.32	−0.16 (0.14)	0.85
Mediators										
Understanding of the problem	0.93	0.25	1.34 (0.23)	3.81***	–	–	–	–	1.37 (0.23)	3.93***
Agreement with the problem	0.92	0.28	1.28 (0.20)	3.59***	–	–	–	–	1.34 (0.20)	3.83***
Pseudo R^2 (Nagelkerke)			0.19		0.06		0.05		0.20	

Note. "–" means that the variables were not included in the model. *$p < .05$, **$p < .01$, ***$p < .001$.

5. General discussion

Fig. 3 summarizes the findings regarding peer feedback features, feedback perception, and feedback implementation. Now the theoretical and practical implications of these results are discussed.

5.1. When do students implement peer feedback? (from perception to implementation)

The findings with regard to the perceptions-to-implementation relationship found here were consistent with the hypothetical model and prior research (Gibbs & Simpson, 2004; Lu & Law, 2012; Nelson & Schunn, 2009; Price et al., 2011): increased understanding and agreement of feedback (for both problems and solutions) predicted higher implementation rates (see Fig. 3). The cognitive process an author undergoes to implement feedback in revisions is initiated with the author's reading the feedback for understanding. Implementing feedback is difficult without first understanding what the feedback means (Price et al., 2011). When authors fail to understand a comment, they generally choose to reject it (Nelson & Schunn, 2009; Patchan et al., 2016).

In addition, when a student agrees with a comment (especially with the solution), he/she is more likely to include it in revisions. Students' perceptions of the feedback usefulness are generally thought to mediate the influence of feedback on performance (Harks, Rakoczy, Hattie, Besser, & Klieme, 2014). However, contrary to the present study, Strijbos et al. (2010) did not find correlations between students' perception of feedback usefulness and their revision behaviors. The major reason for the different findings may be that the task in their study focused on surface features that were easier to identify and correct, while the present study focused on high-level problems that are more challenging for students, especially secondary school students. Students' perceptions of feedback might be more important for them to address high-level problems than low-level problems.

5.2. When do students understand or agree with feedback?

As predicted, solutions (i.e., very specific recommendations for change) were understood more often than general suggestions. This finding is similar to that of Nelson and Schunn (2009) who also found a connection between including a solution and understanding of the problem. With concrete solutions, students likely knew how to solve problems in a specific way, and thus had a better understanding of the problem. Further, longer comments were also predictive of understanding of the problem and especially the solution. The one surprise is that including an explanation was not associated with understanding either problem or solution (which is discussed in further detail below). In any case, this pattern of results suggests more information is generally helpful for understanding and concrete exemplification is also helpful.

The finding that mitigating praise was significantly correlated with problem agreement provided evidence for the hypothesis. Providing mitigating praise is commonly practiced in peer review—in the present study, 47% of the implementable feedback included mitigating praise. By ameliorating the potential adverse attitude towards negative feedback (Cho, Schunn, & Charney, 2006), mitigating praise increased students' agreement with the problem. Note, however, that the positive relationship to praise was at the comment level. The number of praise comments was negatively associated with agreement (and also had a negative relationship with implementation). Thus, the general advice about a feedback sandwich in which the overall review begins and ends with positive comments may be misdirected; instead, it appears that balancing positive and negative within a comment may be more useful.

5.3. Unmediated predictive feedback features

Two feedback features were found to predict implementation

Table 6
Hierarchical regression analysis predicting mediators and implementation for constructive comments.

Predictors	Response variables									
	M	*SD*	Implementation (control + mediators)		Mediators				Implementation (all predictors)	
					Understanding of the solution (control + cognitive features)		Agreement with the solution (control + affective features)			
Control & contextual variables			*B (SE)*	*Odds ratio*	*B (SE)*	*Odds ratio*	*B (SE)*	*Odds ratio*	*B (SE)*	*Odds ratio*
Comment length	59.52	26.72	0.00 (0.00)	1.00	0.02 (0.01)	1.02***	0.01 (0.01)	1.01	−0.01 (0.00)	0.99*
Number of implementable comments	24.88	8.02	−0.02 (0.01)	0.98	0.02 (0.02)	1.02	0.04 (0.02)	1.04	−0.02 (0.01)	0.98
Number of praise comments	9.60	5.91	−0.06 (0.02)	0.94***	0.01 (0.03)	1.01	0.02 (0.03)	1.02	−0.06 (0.02)	0.94***
First draft quality	5.14	0.94	−0.39 (0.09)	0.68***	−0.11 (0.14)	0.89	−0.48 (0.19)	0.62**	−0.31 (0.10)	0.73**
School	0.16	0.36	−0.01 (0.22)	0.99	2.19 (0.50)	8.96***	0.36 (0.41)	1.43	−0.05 (0.23)	0.95
Feedback features										
Mitigating praise	0.50	0.50	–	–	–	–	0.36 (0.25)	1.44	0.25 (0.14)	1.28
Identification	0.66	0.47	–	–	−0.06 (0.25)	0.94	–	–	0.14 (0.17)	1.15
Explanation	0.46	0.50	–	–	0.16 (0.28)	1.18	–	–	0.72 (0.19)	2.06***
Solution	0.17	0.38	–	–	2.45 (0.72)	11.58**	–	–	−0.05 (0.18)	0.95
Hedges	0.40	0.49	–	–	–	–	−0.36 (0.25)	0.70	−0.40 (0.14)	0.67**
Mediators										
Understanding of the solution	0.91	0.29	1.01 (0.22)	2.74***	–	–	–	–	1.07 (0.22)	2.92***
Agreement with the solution	0.94	0.24	1.47 (0.27)	4.35***	–	–	–	–	1.53 (0.28)	4.63***
Pseudo R^2(Nagelkerke)			0.15		0.15		0.05		0.19	

Note. "-" means that the variables were not included in the model. *$p < .05$, **$p < .01$, ***$p < .001$.

directly rather than through one of the perception mediators: explanation and hedges. Similar to prior research (Gielen et al., 2010), students were more likely to implement feedback if explanations of problems were provided. However, contrary to expectation, there was no relationship with problem understanding (or solution understanding). This raises two mysteries: (1) Why were explanations not correlated with understanding of the feedback and (2) why did they predict implementation directly? For the first mystery, it may simply be that explanations, produced by peers, were often overly terse or themselves confusing. Further research should investigate the quality of the explanations found in the peer comments. Regarding the second mystery, it may be that the feedback-to-implementation process is not necessarily achieved through explicit understanding. In contrast to the commonly shared assumption that one cannot implement a feedback comment without explicitly understanding it, it may be that this process can also be automatic and implicit. Decision-making process is often influenced by past experience by implicitly changing choices and the influence is often neglected because decision makers are unaware of the process (Reder & Schunn, 1996). For example, the explanation might produce a deeper encoding of the feedback which is more likely to be remembered at the time of revision, without necessarily having a separate effect on understanding.

A similar story may occur for the effects of hedges. The finding that inclusion of hedges was negatively associated with implementation was in line with the more general hypothesis. But this relationship was expected to hold as a result of the author believing the solution to be clear (understanding) and correct (agreement). Instead, it appears the effect is more consistent with an effect on perceived expertise (Ilgen et al., 1979) or an implicit persuasion process (Horcajo, Briñol, & Petty, 2010): a peer comment without hedges is more likely to persuade the recipient without changing explicit understanding or agreement.

Hedges were not correlated with problem agreement or implementation. The absence of these relationships may result from the presence of suggestions or solutions. As an exploration of this hypothesis, two additional sets of logistic regression analysis were carried out for comments without suggestions or solutions (N = 720). The first analysis included control variables, affective features (i.e., mitigating praise and hedges) as predictors and agreement as the response variable. The second one included control variables, affective features (i.e., mitigating praise and hedges), a cognitive feature (i.e., explanation), and two mediators as predictors and implementation as the response variable. The additional analysis found that hedges did not predict agreement with the problem, but did predict implementation directly (B = -0.49, SE = 0.21, *odds ratio* = 0.62, p = 0.02). Students were less likely to implement a comment with hedges for problems when they did not receive any suggestions or solutions. In addition, understanding (B = 1.55, SE = 0.32, *odds ratio* = 4.72, $p < 0.001$) and agreement with the problem (B = 1.39, SE = 0.28, *odds ratio* = 3.99, $p < 0.001$) were found to be significantly associated with implementation. When a solution or suggestion is added, the perceived expertise of the reviewer might be re-established. Further, in the present study, hedges were framed in terms of the function of uncertainty. Because hedges have various meanings and perform multiple functions (Hyland, 1996; Lee, 2013), future research could determine whether specific functions of hedges (e.g., uncertainty, politeness) result in lower implementation.

5.4. Additional predictors of perception and implementation

Four control and contextual variables were found to be significantly predictive of feedback perception. Comment length positively predicted understanding of the feedback. Longer comments including more information appeared to help students understand the feedback. In addition, first draft quality was negatively associated with agreement with the provided solutions. It is unclear whether higher ability writers are better able to detect poor advice (and hence disagree with it) or whether higher ability writers are just more resistant to their peers' advice.

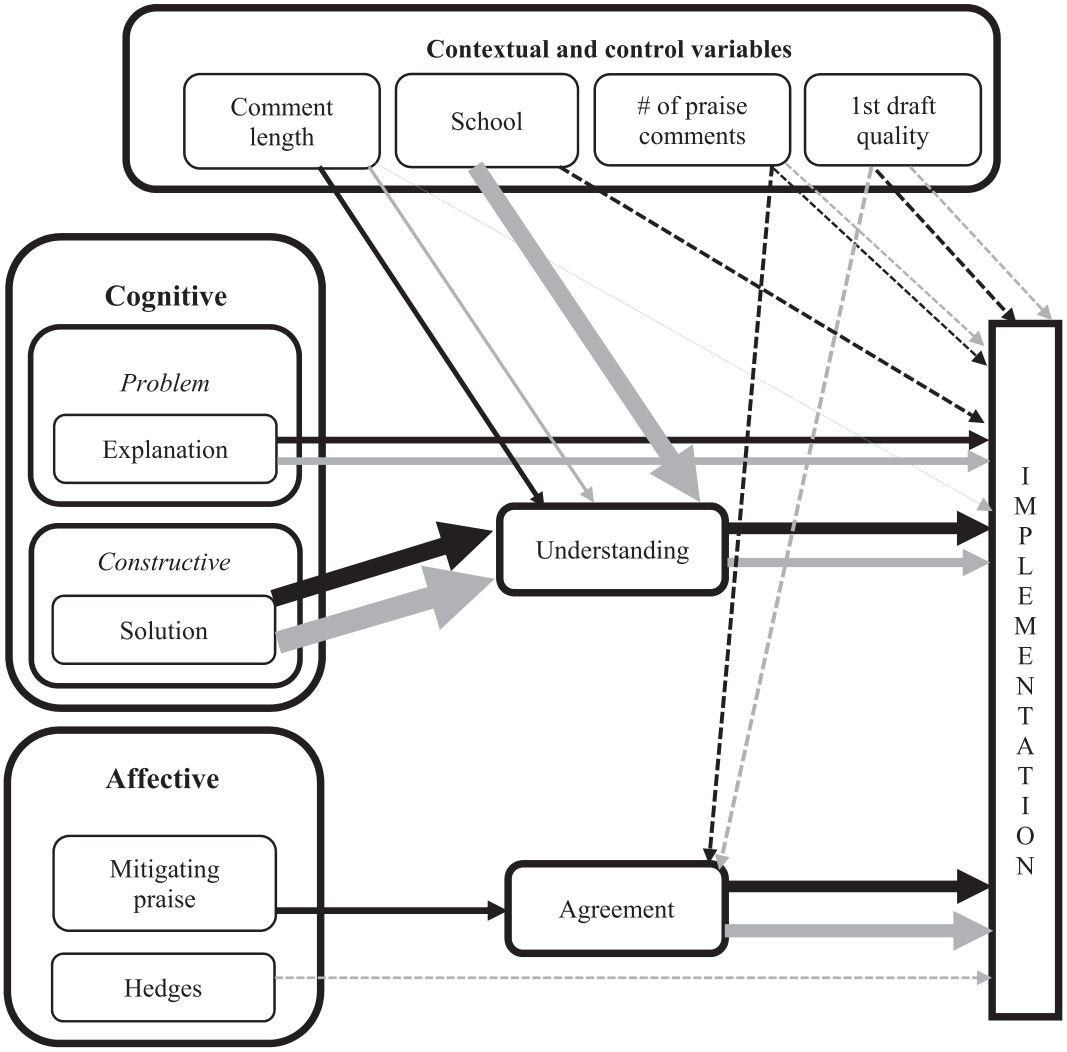

Fig. 3. Revised Feedback Model for Problem comments (black lines) and Constructive comments (grey lines). Line thickness corresponds with statistical strength of relationship in regression models; dotted lines indicate negative relationships.

Similarly, students were less likely to agree with the problem if they received more positive comments. Interestingly, students from Title I school were more likely to understand the solutions than were the non-Title I school students, which may reflect the level of complexity of the solutions being offered.

Comment length negatively predicted implementation for constructive comments. Title I school students were less likely to implement feedback. Note that comment length and school in isolation were not correlated with implementation, but it added as a predictor significantly when feedback features were included in the regression. In addition, the total number of praise comments and first draft quality negatively predicted students' implementation of feedback. Each of these factors could be indicative of contextual motivational effects: when the first draft is perceived to be strong by the writer, then they will be less motivated to make changes. This kind of effect could also explain why the students from the Title I school were less likely to implement feedback.

6. Implications and conclusions

6.1. Implications for practice

Feedback has long been understood to play an important role in task performance (Hattie & Timperley, 2007; Ilgen et al., 1979; Kluger & DeNisi, 1996), but the mediating influence of students' perception of feedback has often been ignored in empirical research (Harks et al., 2014). The present study contributed to a deeper understanding of how feedback influences implementation through students' perceptions. The research findings provide some implications for teachers in designing and using peer assessment. First, strategies should be developed for modifying students' perceptions because perceptions are associated with students' implementation of feedback in revisions. For example, teachers can lead classroom discussions to surface concerns about peer feedback and perhaps both address those concerns and push students to provide feedback that will be better received. In addition, teachers can provide guidance on peer review rubrics that remind students of the importance of providing feedback with detailed explanation of the problem and provide alternative solutions on how to improve writing (Patchan et al., 2018). Furthermore, affective factors should be considered. Constructive criticisms (from teachers or peers) do not consistently reach students because negative affective reactions form a kind of filter. Students are more likely to ignore comments if they trigger negative emotions (Ryan & Henderson, 2017). In peer assessment, students should be reminded to provide positive feedback to soften negative comments via mitigation to establish a friendly and non-threatening environment in which students are more likely to accept feedback (Young, 2000). However, hedges appear not to be a good way to soften negative comments. Hedges included in implementable

feedback may make it challenging for authors to interpret reviewers' intentions due to lack of certainty and clarity (Baker & Bricker, 2010). But teachers could help students interpret hedges, for example, to consider hedges more as politeness strategy than uncertainty. Students could also be encouraged to provide more information (e.g., suggestions, solutions) to remove the adverse effects of hedges.

More support should be targeted for Title I school students, who were found to be less likely to implement peer feedback than students from the non-Title I school. The observed differences might be related to differences in home support, prior experiences with peer feedback, and level of teaching resources (Hughes, 2012; Irvin et al., 2011; Xuan et al., 2019). Although it is difficult to change some of the factors underlying the school-level differences, some efforts could be made to improve students' academic motivation and academic engagement. In writing instruction, for example, student perceptions of peer feedback can be modified to promote their engagement in revision.

6.2. Caveats and conclusion

There are several caveats to be considered. First, the current study was fundamentally correlational in nature. Many controls were included in the models, but it is still possible that other factors were the actual factors underlying changes in feedback perception or implementation. Intervention studies need to be conducted in which reviewers are required to change their feedback contents in order to test the causality of the feedback features on implementation. Correlational studies, like the current one, are useful precursor studies to help identify which features are good focal targets for intervention studies.

In a related issue, the variables were limited to a range of feedback features that had been suggested as being critical to explain students' implementations in previous studies. But other variables such as learner characteristics may also account for implementation. For example, high-achieving and low-achieving learners, or high and low self-efficacious learners may interpret or respond to feedback differently (Hattie & Timperley, 2007; Shute, 2008). Future research can include learner characteristics.

Third, the current study focused on many high-level aspects of writing, but it did not track whether the feedback-to-implementation process varied by more fine-grained distinctions. Different high-level problems may need feedback with different amount of information. For example, feedback comments that focus on argument or rhetorical strategies may be different from thesis feedback because problems on argument and strategies may require more explanations than do problems with the thesis. Initial analyses were conducted separately for each of the specific dimensions and found roughly similar results, a larger dataset is needed to verify homogeneity of effects and thus provide specific support to instruction. Similarly, the studies should also verify the role of feedback features for low level aspects of writing, which were too infrequently commented upon in the current student for analysis.

Additionally, the results can only be cautiously generalized to contexts using other methods/parameters for conducting peer assessment. For example, peer assessment was conducted anonymously in the present study, in which students did not know who their peer reviewers were. Their perceptions of peer feedback and the feedback-to-implementation process might be different if peer assessment was conducted in a traditional, non-anonymous context. Other features of the peer assessment procedure in the current study could also matter, especially the use of multi-peer review, back-evaluations, or detailed rubrics, all of which were included to increase the overall quantity and quality of peer feedback. Future research could compare students' perceptions of feedback and the influence of feedback features on implementation across variations in peer feedback procedures. In a similar way, future research should also examine whether different feedback features matter for peer versus teacher feedback, AP course students and non-AP course students. For example, there may be less of a need to persuade students of the expertise of the reviewer in the case of teacher comments. Students taking AP courses may benefit more from peer review because AP courses involve a homogeneous group of students with relatively stronger academic literacy skills, while students in non-AP courses are diverse and with mixed academic literacy skills.

Funding

The original data collection was funded by a grant from the US Institute of Education Sciences.

Declaration of Competing Interest

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

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 nosalgia, 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. Sample of coding feedback features

Reviewer comment	Mitigating praise	Identification	Explanation	Solution	Suggestion	Hedges
I can't grasp the purpose of your first two quotes in the first body paragraph. They don't seem to relate to the thesis or how Louv formulates his argument in any way, and seem a little random and unhelpful.	0	1	1	0	0	1
The organization does not match the order the author provided in the thesis statement. If the author switches around the order of the devices in the thesis statement, the organization will be much better and easier for the audience to follow.	0	1	0	1	0	0
The author did not provide explicit references to the text aside from restating where these devices were incorporated. In the third paragraph, he states "The salesman's jaw dropped". This is the only specific evidence that the author quotes in his analysis of hyperbole. In this way, the author needs to incorporate explicit references to the given text in order to strengthen his essay.	0	1	1	0	1	0
All of the rhetorical strategies are relevant and relate to aiding in Louv's purpose, but the analysis was very simplistic and lacked depth. Maybe tying back to Louv's purpose throughout the body paragraphs to ensure clarity.	1	1	0	0	1	1
The format of the essay is well organized; however, they are missing a paragraph for the analysis of sarcasm.	1	1	0	0	0	0
The author's thesis includes imagery and repetition; however, imagery is not mentioned nor analyzed. Positively, he or she thoroughly analyzes the latter rhetorical strategy!	1	1	0	0	0	0
The organization was clear. The paragraph separation was good. But it seems a bit abrupt of a change between paragraphs. It would further your essay if you included transition sentences in between the last body paragraph and the conclusion.	1	1	0	0	1	1
You also had no analysis of the devices. You just stated some evidence of the device in context, but you didn't explain the importance of each device and how it contributes to Louv's argument. You need to remember to analyze all of the devices, not just state them and throw in a quote of them. You need to make sure you analyze the purpose of each device and how it helps convey Louv's argument.	0	1	1	0	1	0
The evidence, while well inserted, lacks the "so what" in the first instance, but better explains in the second.	1	1	0	0	0	0
You have 4 rhetorical devices in your thesis, yet you only have 2 body paragraphs. You need to have 4 body paragraphs, one for each strategy.	0	1	0	0	1	0
Your organization is adequate; you have a strong thesis, two body paragraphs, and a conclusion. However, in your conclusion, you're pretty much restating your thesis. Instead, you should restate your main arguments of what Louv is actually doing. You should also mention the "so what" of the piece.	1	1	0	0	1	0
The quotes you used were just placed randomly into the paragraphs, without any sort of connections within the paragraphs themselves. It would help you to explain the significance of the quotes more. You explain the quote in the first body paragraph, but it is lost in the second and final body paragraphs, which greatly takes away from any possible analysis of the text, you are simply repeating what has already been said in Louv's passage; it needs a little bit further analysis.	0	1	1	0	1	1

Note. The features are coded as 1 if present and 0 if absent.

Appendix C. Sample of coding back-review comments

Back-review comment	Agree with problem	Agree with solution	Understand problem	Understand solution
True, more quotes would've strengthened the essay as a whole.	1	1	1	1
Explanation and detail for my quotes was needed, I agree, thanks for the examples.	1	1	1	1
This review helps me because it states what I did wrong. You made some really good points. But how should I have analyzed it then?	1	0	1	0
Helpful, I could've incorporated more details.	1	1	1	1
Make sense. But I think you misunderstood my essay, because I don't believe I lacked understanding of the prompt. The comment provides no suggestions to improve the writing.	0	0	1	0
Very helpful, this allows me to try and create a better flow and smoother style.	1	1	1	1
I agree. But how should I fix this?	1	0	1	0
I will establish more from the quotes that I included.	1	1	1	1
I have no idea which quote you are talking about.	0	0	0	0
Make sense, thanks. This will improve my essay.	1	1	1	1
I thank you for your review because it allowed me to know what I fell short in. However, you made no constructive criticism on how I could improve.	1	0	1	0
I agree that more analysis was needed in order to make my essay stronger. At the moment I only have examples and quotes with minimal analysis attached. Definitely more analysis is needed to provide additional analysis.	1	1	1	1

Note. 1: the writer agreed with or understood the problem or solution. 0: the writer did not agree with or understand the problem or solution.

References

Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership, 57*(5), 13–18.

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

Baker, W., & Bricker, R. H. (2010). The effects of direct and indirect speech acts on native English and ESL speakers' perception of teacher written feedback. *System, 38*(1), 75–84. https://doi.org/10.1016/j.system.2009.12.007.

Bangert-Drowns, R. L., Kulik, C. L. C., 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.

Bolzer, M., Strijbos, J. W., & Fischer, F. (2014). Inferring mindful cognitive-processing of peer-feedback via eye-tracking: Role of feedback-characteristics, fixation-durations and transitions. *Journal of Computer Assisted Learning, 31*(5), 422–434. https://doi.org/10.1111/jcal.12091.

Brown, G. T., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing, 9*(2), 105–121. https://doi.org/10.1016/j.asw.2004.07.001.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education, 31*(2), 219–233. https://doi.org/10.1080/03075070600572132.

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 and Education, 48*(3), 409–426. https://doi.org/10.1016/j.compedu.2005.02.004.

Cho, K., Schunn, C. D., & Charney, D. (2006). Commenting on writing: Typology and perceived helpfulness of comments from novice peer reviewers and subject matter experts. *Written Communication, 23*(3), 260–294. https://doi.org/10.1177/0741088306289261.

Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology, 98*(4), 891–901. https://doi.org/10.1037/0022-0663.98.4.891.

Chory, R. M., & Westerman, C. Y. K. (2009). Feedback and fairness: The relationship between negative performance feedback and organizational justice. *Western Journal of Communication, 73*(2), 157–181. https://doi.org/10.1080/10570310902856055.

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

Ferris, D. R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly, 31*(2), 315–339. https://doi.org/10.2307/3588049.

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education, 1*, 3–31.

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.

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

Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: The role of the feedback's perceived usefulness. *Educational Psychology, 34*(3), 269–290. https://doi.org/10.1080/01443410.2013.785384.

Higgins, R. (2000). "Be more critical": Rethinking assessment feedback. Paper presented at British Educational Research Association Conference. Cardiff University. Available online at: http://www.leeds.ac.uk/educo/documents/00001548.htm.

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

Horcajo, J., Briñol, P., & Petty, R. E. (2010). Consumer persuasion: Indirect change and implicit balance. *Psychology & Marketing, 27*(10), 938–963. https://doi.org/10.1002/mar.20367.

Hosman, L. A. (1989). The evaluative consequences of hedges, hesitations, and intensifies: Powerful and powerless speech styles. *Human Communication Research, 15*(3), 383–406. https://doi.org/10.1111/j.1468-2958.1989.tb00190.x.

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.

Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing, 10*(3), 185–212. https://doi.org/10.1016/S1060-3743(01)00038-8.

Hyland, K. (1996). Talking to the academy: Forms of hedging in science research articles. *Written Communication, 13*(2), 251–281. https://doi.org/10.1177/0741088396013002004.

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.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology, 64*(4), 349–371. https://doi.org/10.1037/0021-9010.64.4.349.

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

Jönsson, A. (2013). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education, 14*(1), 63–76. https://doi.org/10.1177/1469787412467125.

Jönsson, A., & Panadero, E. (2017). The use and design of rubrics to support assessment for learning. In D. Carless, (Ed.). *Scaling up assessment for learning in higher education* (pp. 99–111). Singapore: Springer.

Kaufman, J. H., & Schunn, C. D. (2010). Students' perceptions about peer assessment for writing: Their origin and impact on revision work. *Instructional Science, 39*(3), 387–406. https://doi.org/10.1007/s11251-010-9133-6.

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.

Lee, K. Y. (2013). Hedging expressions used in academic written feedback: A study on the use of modal verbs. *Research in Corpus Linguistics, 1*, 33–45.

Leung, K., Su, S., & Morris, M. W. (2001). When is criticism not constructive? The roles of fairness perceptions and dispositional attributions in employee acceptance of critical supervisory feedback. *Human Relations, 54*(9), 1155–1187. https://doi.org/10.1177/0018726701549002.

Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279–290. https://doi.org/10.1080/13562510600680582.

Lizzio, A., & Wilson, K. (2008). Feedback on assessment: Students' perceptions of quality and effectiveness. *Assessment & Evaluation in Higher Education, 33*(3), 263–275. https://doi.org/10.1080/02602930701292548.

Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance, 3*(2), 157–189. https://doi.org/10.1016/0030-5073(68)90004-4.

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.

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

Moskal, B. M. (2000). Scoring rubrics: What, when and how? *Practical Assessment, Research & Evaluation, 7*(3), 1–5. http://PAREonline.net/getvn.asp?v=7&n=3.

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* (pp. 125–143). (3rd ed.). Mahwah, NJ: 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.

Paré, D. E., & Joordens, S. (2008). Peering into large lectures: Examining peer and expert mark agreement using peerScholar, an online peer assessment tool. *Journal of Computed Assisted Learning, 24*(16), 526–540. https://doi.org/10.1111/j.1365-2729.2008.00290.x.

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

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

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*, 547–577. https://doi.org/10.1146/annurev.psych.54.101601.145041.

Price, M., Handley, K., & Millar, J. (2011). Feedback: Focusing attention on engagement. *Studies in Higher Education, 36*(8), 879–896. https://doi.org/10.1080/03075079.2010.483513.

Reder, L. M., & Schunn, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In L. M. Reder (Ed.). *Implicit memory and metacognition* (pp. 45–77). Hillsdale, NJ: Erlbaum.

Russell, J., Horne, S. V., Ward, A. S., Bettis, E. A., & Gikonyo, J. (2017). Variability in students' evaluating processes in peer assessment with calibrated peer review. *Journal of Computed Assisted Learning, 33*(2), 178–190. https://doi.org/10.1111/jcal.12176.

Ryan, T., & Henderson, M. (2017). Feeling feedback: Students' emotional responses to educator feedback. *Assessment & Evaluation in Higher Education, 43*(6), 880–892. https://doi.org/10.1080/02602938.2017.1416456.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144. https://doi.org/10.1007/BF00117714.

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

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

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

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4.

Topping, K. (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.

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

Van der Pol, J., Van den Berg, B. A. M., Admiraal, W. F., & Simons, P. R. J. (2008). The nature, reception, and use of online peer feedback in higher education. *Computers & Education, 51*(4), 1804–1817. https://doi.org/10.1016/j.compedu.2008.06.001.

Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education, 31*(3), 379–394. https://doi.org/10.1080/02602930500353061.

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.

Xuan, X., Xue, Y., Zhang, C., Luo, Y., Jiang, W., Qi, M., et al. (2019). Relationship among school socioeconomic status, teacher-student relationship, and middle school students' academic achievement in China: Using the multilevel mediation model. *PLoS ONE, 14*(3), e0213783. https://doi.org/10.1371/journal.pone.0213783.

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

Young, P. (2000). "I might as well give up": Self-esteem and mature students' feelings about feedback on assignments. *Journal of Further and Higher Education, 24*(3), 409–418. https://doi.org/10.1080/030987700750022325.

Yu, S., & Hu, G. (2017). Understanding university students' peer feedback practices in EFL writing: Insights from a case study. *Assessing Writing, 33*, 25–35. https://doi.org/10.1016/j.asw.2017.03.004.